ANALYSIS & PDE

Volume 16 No. 8 2023

msp

Analysis & PDE
msp.org/apde

PRODUCTION
production@msp.org
Silvio Levy, Scientific Editor

See inside back cover or msp.org/apde for submission instructions.

The subscription price for 2023 is US $405/year for the electronic version, and $630/year (+$65, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online.

APDE peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

ANALYSIS AND PDE
Vol. 16 (2023), No. 8, pp. 1745–1798
DOI: 10.2140/apde.2023.16.1745

GROUND STATE PROPERTIES IN THE QUASICLASSICAL REGIME

MICHELE CORREGGI, MARCO FALCONI AND MARCO OLIVIERI

We study the ground state energy and ground states of systems coupling nonrelativistic quantum particles and force-carrying Bose fields, such as radiation, in the quasiclassical approximation. The latter is very useful whenever the force-carrying field has a very large number of excitations and thus behaves in a semiclassical way, while the nonrelativistic particles, on the other hand, retain their microscopic features. We prove that the ground state energy of the fully microscopic model converges to that of a nonlinear quasiclassical functional depending on both the particles' wave function and the classical configuration of the field. Equivalently, this energy can be interpreted as the lowest energy of a Pekar-like functional with an effective nonlinear interaction for the particles only. If the particles are confined, the ground state of the microscopic system converges as well, to a probability measure concentrated on the set of minimizers of the quasiclassical energy.

1. Introduction and main results

The description and rigorous derivation of effective models for complex quantum systems is a flourishing line of research in modern mathematical physics. Typically, in suitable regimes, the fundamental quantum description can be approximated in terms of some simpler model retaining the salient physical features, but also allowing a more manageable computational or numerical treatment. The questions addressed in this work naturally belong to such a wide class of problems.

We consider indeed a quantum system composed of N nonrelativistic particles interacting with a quantized bosonic field in the *quasiclassical regime*. We refer to [Carlone et al. 2021; Correggi and Falconi 2018; Correggi et al. 2019; 2023] for a detailed discussion of such a regime: in extreme synthesis, we plan to study field configurations with a suitable semiclassical behavior. We require indeed that there is a large number of field excitations, although each one of the latter is carrying a very small amount of energy, in such a way that the field's degrees of freedom are almost classical. More precisely, we assume that the average number of force carriers $\langle \mathcal{N} \rangle$ is of order $1/\varepsilon$, for some $0 < \varepsilon \ll 1$, and thus much larger than the commutator between $a^\dagger$ and a, which is of order 1 (we use units in which $\hbar = 1$). Concretely, this can be realized by rescaling the canonical variables $a^\dagger$ and a by $\sqrt{\varepsilon}$, i.e., setting $a_\varepsilon^\sharp := \sqrt{\varepsilon} a^\sharp$, which leads to

$$[a_\varepsilon(\boldsymbol{k}), a_\varepsilon^\dagger(\boldsymbol{k}')] = \varepsilon \delta(\boldsymbol{k} - \boldsymbol{k}'), \qquad \varepsilon \ll 1. \tag{1-1}$$

MSC2020: primary 81Q20, 81T10; secondary 81Q05, 81Q10, 81S30, 81V10.
Keywords: quasiclassical limit, interaction of matter and light, semiclassical analysis.

On the other hand, the degrees of freedom associated with the particles are not affected by the scaling limit $\varepsilon \to 0$ and the particles remain quantum. Our goal is precisely to set up and rigorously derive an effective quantum model for the lowest energy state of the system in the quasiclassical regime $\varepsilon \to 0$, when the field becomes classical.

Let us now describe in more detail the type of microscopic models we plan to address. The space of states of the full system is[1]

$$\mathscr{H}_\varepsilon := L^2(\mathbb{R}^{dN}) \otimes \mathcal{G}_\varepsilon(\mathfrak{h}), \tag{1-2}$$

where $d \in \{1, 2, 3\}$, the single one-excitation space of the field is $\mathfrak{h}$ and $\mathcal{G}_\varepsilon$ stands for the second quantization map, so that $\mathcal{G}_\varepsilon(\mathfrak{h})$ is the bosonic Fock space constructed over $\mathfrak{h}$ with canonical commutation relations

$$[a_\varepsilon(\xi), a_\varepsilon^\dagger(\eta)] = \varepsilon \langle \xi | \eta \rangle_{\mathfrak{h}}, \tag{1-3}$$

for any $\xi, \eta \in \mathfrak{h}$.

The energy of the microscopic system and thus its Hamiltonian is given by the nonrelativistic energy of the particles, the field energy and the interaction between the particles and the field, in such a way that

- the particle and field energies are a priori of the same order $\mathcal{O}(1)$;
- the interaction is weak, i.e., a priori subleading with respect to the unperturbed energies.

This is concretely realized by considering Hamiltonians of the form

$$H_\varepsilon = \mathcal{K}_0 \otimes 1 + 1 \otimes \mathrm{d}\mathcal{G}_\varepsilon(\omega) + H_I, \tag{1-4}$$

where:

- $\mathcal{K}_0$ is the (ε-independent) free particle Hamiltonian
$$\mathcal{K}_0 = \sum_{j=1}^{N} (-\Delta_j) + \mathcal{W}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) \tag{1-5}$$
which is assumed to be self-adjoint and bounded from below;
- $\mathrm{d}\mathcal{G}_\varepsilon(\omega)$ is the free field energy and is the second quantization of the positive operator ω on $\mathfrak{h}$, admitting a possibly unbounded inverse ω^{-1};
- the interaction H_I is the only nonfactorized term of the Hamiltonian, it depends on ε only through the creation and annihilation operators $a_\varepsilon^\sharp$ and it is a polynomial of such operators of order between one and two.

Such requests meet the scaling conditions mentioned above. Indeed, assuming that the average number $\langle \mathcal{N} \rangle$ of bare excitations of the field is $\mathcal{O}(\varepsilon^{-1})$, the field energy is of order $\varepsilon \langle \mathcal{N} \rangle = \mathcal{O}(1)$, due to the rescaling of $a_\varepsilon^\dagger$ and a_ε. For the same reason and since the interaction is at least of order 1 in the creation and annihilation operators, we have that H_I is of order $\mathcal{O}(\sqrt{\varepsilon})$, i.e., a priori subleading with respect to the rest of H_ε.

[1] We do not take into account the spin degrees of freedom nor the symmetry constraints induced by the presence of identical particles, but such features can be included in the discussion without any effort and the results trivially apply to the corresponding models. In fact, we may even allow for a coupling term between the radiation field and the particle spins [Correggi et al. 2019], as the one often included in the Pauli–Fierz model.

The specific models we consider in the following are:

(a) the *Nelson model* [Nelson 1964]: the coupling in H_I is simply linear, i.e.,

$$H_I = \sum_{j=1}^{N} A_\varepsilon(\boldsymbol{x}_j), \tag{1-6}$$

where

$$A_\varepsilon(\boldsymbol{x}) := a_\varepsilon^\dagger(\boldsymbol{\lambda}(\boldsymbol{x})) + a_\varepsilon(\boldsymbol{\lambda}(\boldsymbol{x})) \tag{1-7}$$

is the field operator and

$$\lambda,\ \omega^{-1/2}\lambda \in L^\infty(\mathbb{R}^3; \mathfrak{h}) \tag{1-8}$$

(a typical choice is $\mathfrak{h} = L^2(\mathbb{R}^d)$, ω a multiplication operator such that $\omega(\boldsymbol{k}) \geq 0$ and also $\lambda(\boldsymbol{x}; \boldsymbol{k}) = \lambda_0(\boldsymbol{k}) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}$, with $\lambda_0, \omega^{-1/2}\lambda_0 \in \mathfrak{h}$);

(b) the *Fröhlich polaron* [Fröhlich 1937]: a variant of the Nelson model where $\mathfrak{h} = L^2(\mathbb{R}^d)$, $\omega = 1$ and

$$\lambda(\boldsymbol{x}; \boldsymbol{k}) = \sqrt{\alpha}\, \frac{e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}}{|\boldsymbol{k}|^{(d-1)/2}}, \tag{1-9}$$

for some $\alpha > 0$;

(c) the *Pauli–Fierz model* [Pauli and Fierz 1938]: the most elaborate model and we consider only its three-dimensional realization, namely $d = 3$; the interaction is provided by the minimal coupling

$$H_\varepsilon = \sum_{j=1}^{N} \frac{1}{2m_j} (-i\nabla_j + e\boldsymbol{A}_{\varepsilon,j}(\boldsymbol{x}_j))^2 + \mathcal{W}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) + 1 \otimes \mathrm{d}\mathcal{G}_\varepsilon(\omega), \tag{1-10}$$

where $\omega \geq 0$, $m_j > 0$, $j = 1, \dots, N$, and e are the particles' masses and charge, respectively, and the field operators $\boldsymbol{A}_{\varepsilon,j}$, $j = 1, \dots, N$, have here the same formal expression as in (1-7) but $\boldsymbol{\lambda}_j = (\lambda_{j,1}, \lambda_{j,2}, \lambda_{j,3})$, with

$$\lambda_{j,\ell},\ \omega^{\pm 1/2}\lambda_{j,\ell} \in L^\infty(\mathbb{R}^3; \mathfrak{h}), \tag{1-11}$$

is a vector function to account for the electromagnetic polarizations and the charge distributions of the particles (the standard choice is, indeed, $\mathfrak{h} = L^2(\mathbb{R}^3; \mathbb{C}^2)$) and we fix for convenience the gauge to be Coulomb's gauge, i.e., $\nabla_j \cdot \boldsymbol{\lambda}_j = 0$.

The physical meaning of the three models above is quite different and we refer, e.g., to the monograph [Spohn 2004] for a detailed discussion. The Nelson model is the simplest and can be applied to model nucleons interacting with a meson field or, in first approximation, to model the interaction of particles with radiation fields, although the case of the electromagnetic field is typically described through the Pauli–Fierz model. The polaron, on the other hand, provides an effective description of quantum particles in a phonon field, e.g., generated by the vibrational models of a crystal. Note also that the quasiclassical limit $\varepsilon \to 0$ itself can have different interpretations in each model. For instance, in the framework of the polaron model, it can be reformulated as a *strong coupling limit*, which has recently attracted a lot of attention; see, e.g., [Frank and Gang 2020; Griesemer 2017; Leopold et al. 2021; Lieb and Seiringer 2020; Mitrouskas 2021].

In the Nelson and Pauli–Fierz Hamiltonians, there is an ultraviolet regularization, made apparent in the assumptions on λ; we do not consider here the renormalization procedure to remove such ultraviolet cut-off, even if for the Nelson model it is possible to perform it rigorously. We plan to address such a problem in a future work. We also skip at this stage the discussion of the well-posedness of such models (see Sections 4A–4C for further details), but we point out that, with the assumptions made, the operator (1-4) is self-adjoint and bounded from below in each model.

The main problem we study concerns the behavior of the ground state of the microscopic Hamiltonian H_ε in the quasiclassical limit $\varepsilon \to 0$ and, more precisely, we investigate the convergence in the same limit of the bottom of the spectrum

$$E_\varepsilon := \inf \sigma(H_\varepsilon) = \inf_{\Gamma_\varepsilon \in \mathscr{L}^1(\mathscr{H}_\varepsilon), \|\Gamma_\varepsilon\|_1 = 1} \operatorname{tr}(H_\varepsilon \Gamma_\varepsilon) \tag{1-12}$$

of H_ε as well as the limiting behavior of any corresponding *approximate ground state* or *minimizing sequence* $\Psi_{\varepsilon,\delta} \in \mathscr{D}(H_\varepsilon)$ satisfying

$$\langle \Psi_{\varepsilon,\delta} | H_\varepsilon | \Psi_{\varepsilon,\delta} \rangle_{\mathscr{H}_\varepsilon} < E_\varepsilon + \delta, \tag{1-13}$$

for some small $\delta > 0$.

We state our main results with all details in Section 1C. After a brief outlook on the existing literature in Section 1A, we first introduce and discuss the quasiclassical variational problems in Section 1B. In the rest of the paper, we present the proofs.

1A. ***State of the art.*** Our paper fits within the framework of infinite-dimensional semiclassical analysis, which was introduced in the series of works [Ammari and Nier 2008; 2009; 2011; 2015] and further discussed in [Falconi 2018a; 2018b]. Apart from the aforementioned works on quasiclassical analysis, semiclassical techniques have already been used in the study of variational problems, both for systems with creation and annihilation of particles [Ammari and Falconi 2014] and for systems with many bosons, using a slightly different approach called quantum de Finetti theorem; see [Lewin et al. 2014; 2015; 2016]. We also point out that partially classical regimes have already been explored in [Amour and Nourrigat 2015; Amour et al. 2017; 2019; Ginibre et al. 2006], although in other contexts and with different purposes.

The question of the ground state energy convergence in the quasiclassical regime has partially been addressed in [Correggi and Falconi 2018] and [Correggi et al. 2019] for the Nelson and polaron models and the Pauli–Fierz model, respectively. In fact, Theorem 1.3 below completes and extends the corresponding results proven in [Correggi and Falconi 2018, Theorem 2.4] and [Correggi et al. 2019, Theorem 1.9]. More precisely, we develop a more general and self-contained proof strategy, based on the new mathematical structure of *quasiclassical Wigner measures* first introduced in [Correggi et al. 2023], allowing us to relax the assumptions on the microscopic models and taking into account more general settings.

On the other hand, the convergence of microscopic ground states and minimizing sequences in the quasiclassical regime is studied here for the first time; see Theorems 1.7 and 1.15 below. Let us point out that our results *do not require* the existence of a microscopic ground state (and imply the existence of quasiclassical minimizers), although in the presence of the latter they become more transparent. In fact, the problem of the ground state existence in quantum field theory is tricky and has been extensively studied in the past. We refer to [Abdesselam and Hasler 2012; Arai 2001; Arai et al. 1999; Betz et al. 2002; Dereziński 2003; Georgescu et al. 2004; Gérard 2000; Gérard et al. 2011; Griesemer et al. 2001; Hirokawa 2006; Hiroshima 2001; Hiroshima and Matte 2022; Møller 2005; Pizzo 2003] for a detailed discussion of the problem.

1B. ***Quasiclassical variational problems.*** As discussed in detail in [Correggi and Falconi 2018; Correggi et al. 2019; 2023], each of the microscopic models introduced so far admits a quasiclassical counterpart in the limit $\varepsilon \to 0$. More precisely, both their stationary [Correggi and Falconi 2018; Correggi et al. 2019] and dynamical [Correggi et al. 2023] properties can be approximated in such a regime in terms of effective models, where the quantum particle system is driven by a classical field, which in turn is the classical counterpart of the quantized field. In extreme synthesis, the quantum field operator gets replaced by a classical field, which is just a function on $\mathbb{R}^d$, and the interaction term H_I in H_ε gives rise to a potential $\mathcal{V}_z$ depending on the classical field configuration $z \in \mathfrak{h}$. Concretely, the quasiclassical effective Hamiltonian reads

$$\mathcal{H}_z = \mathcal{K}_0 + \sum_{j=1}^{N} \mathcal{V}_z(\boldsymbol{x}_j) + \langle z|\omega|z\rangle_{\mathfrak{h}}, \tag{1-14}$$

and it is self-adjoint on some dense $\mathscr{D} \subset L^2(\mathbb{R}^{dN})$ for any $z \in \mathfrak{h}$; see [Correggi and Falconi 2018, Theorems 2.1–2.3] and [Correggi et al. 2019, Theorem 1.1]. In each model the explicit expression of such an effective potential can be identified explicitly:

(a) In the Nelson model, each particle feels a potential of the form

$$\mathcal{V}_z(\boldsymbol{x}) = 2\mathrm{Re}\langle z|\lambda(\boldsymbol{x})\rangle_{\mathfrak{h}} \in \mathscr{B}(L^2(\mathbb{R}^d)); \tag{1-15}$$

(b) For the polaron, the formal expression of the potential $\mathcal{V}_z$ is the same as in (1-15) above, although, since (1-9) does not belong to $L^\infty(\mathbb{R}^d; \mathfrak{h})$, the expression on the right-hand side must be interpreted in the proper way (see Section 4B); in addition, the obtained potential is no longer bounded but it is infinitesimally form-bounded with respect to $-\Delta$;

(c) In the Pauli–Fierz model, the effective operator is obtained via the replacement of the field $\boldsymbol{A}_\varepsilon$ by its classical counterpart $\boldsymbol{a}_z(\boldsymbol{x}) = 2\mathrm{Re}\langle z|\boldsymbol{\lambda}(\boldsymbol{x})\rangle_{\mathfrak{h}}$, which is continuous and vanishing at ∞ (see [Correggi et al. 2019, Remark 1.5]), and thus, in order to recover the expression (1-14), $\mathcal{V}_z$ must be the operator

$$\mathcal{V}_z(\boldsymbol{x}) = 2\sum_{j=1}^{N} \frac{1}{m_j}[-ie\mathrm{Re}\langle z|\boldsymbol{\lambda}_j(\boldsymbol{x})\rangle_{\mathfrak{h}} \cdot \nabla_j + e^2(\mathrm{Re}\langle z|\boldsymbol{\lambda}_j(\boldsymbol{x})\rangle_{\mathfrak{h}})^2]. \tag{1-16}$$

Note that in case (c) the effective operator can in fact be simply rewritten as[2]

$$\mathcal{H}_z = \sum_{j=1}^{N} \frac{1}{2m_j}(-i\nabla_j + e\boldsymbol{a}_z(\boldsymbol{x}_j))^2 + \mathcal{W}(\boldsymbol{X}) + \langle z|\omega|z\rangle_{\mathfrak{h}}. \tag{1-17}$$

[2] We use the compact notation $\boldsymbol{X} := (\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) \in \mathbb{R}^{dN}$.

We can now define the effective quasiclassical ground state energy in terms of the energy functional

$$\mathcal{E}_{\mathrm{qc}}[\psi, z] := \langle \psi | \mathcal{H}_z | \psi \rangle_{L^2(\mathbb{R}^{dN})}, \quad (\psi, z) \in L^2(\mathbb{R}^{dN}) \oplus \mathfrak{h}_\omega, \tag{1-18}$$

as

$$E_{\mathrm{qc}} := \inf_{(\psi,z)\in\mathscr{D}_{\mathrm{qc}}} \mathcal{E}[\psi, z], \tag{1-19}$$

where

$$\mathscr{D}_{\mathrm{qc}} := \{(\psi, z) \in L^2(\mathbb{R}^{dN}) \oplus \mathfrak{h}_\omega : \|\psi\|_2 = 1,\ |\mathcal{E}_{\mathrm{qc}}[\psi, z]| < +\infty\}. \tag{1-20}$$

Here, $\mathfrak{h}_\omega$ is the Hilbert completion of $\bigcap_{k\in\mathbb{N}} \mathscr{D}(\omega^k)$ with respect to the scalar product $\langle \cdot | \cdot \rangle_{\mathfrak{h}_\omega} := \langle \cdot | \omega | \cdot \rangle_{\mathfrak{h}}$, i.e.,

$$\mathfrak{h}_\omega := \overline{\bigcap_{k\in\mathbb{N}} \mathscr{D}(\omega^k)}^{\langle \cdot | \cdot \rangle_{\mathfrak{h}_\omega}}. \tag{1-21}$$

We denote by $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}}) \in \mathscr{D}_{\mathrm{qc}}$ a corresponding minimizing configuration (if any), i.e., such that

$$E_{\mathrm{qc}} = \mathcal{E}_{\mathrm{qc}}[\psi_{\mathrm{qc}}, z_{\mathrm{qc}}]. \tag{1-22}$$

Concretely, the functional $\mathcal{E}_{\mathrm{qc}}$ plays the role of the *quasiclassical energy* of the system under consideration. However, the reader should be careful and be aware that $\mathcal{H}_z$ *is not* the Hamiltonian energy of the whole system: the complete environment-small system evolution is indeed not of Hamiltonian type. For each fixed $z \in \mathfrak{h}_\omega$, the Hamilton–Jacobi equations of $\mathcal{E}_{\mathrm{qc}}[\psi, z]$, with respect to the (complex) ψ variable, yield the dynamics of the small system; the environment on the other hand is stationary in the problems under consideration in this paper; see [Correggi et al. 2023] for a detailed analysis of quasiclassical dynamical systems.

The preliminary questions to address towards the derivation of the above quasiclassical effective models are whether such models are stable and, if this is the case, whether a minimizing configuration does exist: explicitly,

$$\text{“Is } E_{\mathrm{qc}} \text{ greater than } -\infty\text{?” (stability),} \tag{VP1}$$

$$\text{“Does there exist } (\psi_{\mathrm{qc}}, z_{\mathrm{qc}}) \in \mathscr{D}_{\mathrm{qc}} \text{ such that } \mathcal{E}_{\mathrm{qc}}(\psi_{\mathrm{qc}}, z_{\mathrm{qc}}) = E_{\mathrm{qc}}\text{?” (existence of a ground state).} \tag{VP2}$$

Note that any critical point $(\psi, z) \in \mathscr{D}_{\mathrm{qc}}$ of the functional $\mathcal{E}_{\mathrm{qc}}[\psi, z]$ must satisfy the condition

$$\delta_{(\psi,z)}\big[\mathcal{E}_{\mathrm{qc}}[\psi, z] - \epsilon \|\psi\|_2^2\big] = 0,$$

which yields the Euler–Lagrange equations

$$\begin{cases} \mathcal{H}_z \psi = \epsilon \psi, \\ \omega z + \big\langle \psi \big| \partial_{\bar{z}} \sum_j \mathcal{V}_z(\boldsymbol{x}_j) \big| \psi \big\rangle_{L^2(\mathbb{R}^{dN})} = 0, \end{cases} \tag{1-23}$$

where the Lagrange multiplier $\epsilon = \langle \psi | H_z | \psi \rangle \in \mathbb{R}$ takes into account the normalization constraint on ψ. We anticipate that a consequence of the convergence of the microscopic ground state, stated in Corollary 1.10, is that, under suitable assumptions on $\mathcal{K}_0$ (for instance if $\mathcal{W}$ is trapping), the answer to both questions in (VP1) and (VP2) is positive and, in particular, the set of minimizers is not empty.

The variational problem above is strictly related to the more general issue of rigorous derivation of effective theories, since, at least for the polaron model, it is known that the minimization of the microscopic energy can be approximated in the limit $\varepsilon \to 0$ in terms of a nonlinear problem on ψ alone. Indeed, focusing on the particle system, one can naturally approach (1-19) in a different and a priori inequivalent way, i.e., *first* one gets rid of the classical field by minimizing over $z \in \mathfrak{h}_\omega$ and *then* investigates the minimization of the remaining functional on ψ, which is obviously nonlinear, since the minimizing z depends on ψ itself. As anticipated, this strategy has been already followed in the literature in the case of the polaron in the strong coupling regime, leading to the *Pekar functional* and the corresponding variational problem [Donsker and Varadhan 1983; Lieb and Thomas 1997; Pekar 1954]. Such a feature is however not exclusive of the polaron and can be observed in all the models mentioned above: we present below a formal derivation of a Pekar-like functional $\mathcal{E}_{\mathrm{Pekar}}[\psi]$ for both the Nelson and polaron model. The Pauli–Fierz case is also discussed below; let us remark however that in this case such a procedure does not yield an explicit nonlinear functional of ψ (see (1-34) below), because it is in general not possible to solve explicitly the variational equation expressing the minimizing z in terms of ψ.

The formal procedure goes as follows: solving the critical point condition $\delta_z \mathcal{E}_{\mathrm{qc}} = 0$ with respect to the variable z for fixed ψ, we find some z_ψ, that we can plug into $\mathcal{E}_{\mathrm{qc}}$, thus obtaining the Pekar energy

$$\mathcal{E}_{\mathrm{Pekar}}[\psi] := \mathcal{E}_{\mathrm{qc}}[\psi, z_\psi].$$

Such a scheme can be made to work rigorously for the polaron (case (b)) with some care, but the variable z is not the right one to consider in cases (a) and (c). Under the assumptions we have made — recall in particular (1-8) and (1-11) — it is indeed more natural to set, since $z \in \mathfrak{h}_\omega$,

$$\eta := \omega^{1/2} z, \tag{1-24}$$

(note however that in case (b) $\eta = z$) and consider the functional $\mathcal{F}_{\mathrm{qc}}[\psi, \eta] := \mathcal{E}_{\mathrm{qc}}[\psi, \omega^{-1/2}\eta]$, which in case (a) reads

$$\begin{aligned}\mathcal{F}_{\mathrm{qc}}[\psi, \eta] &= \Big\langle \psi \Big| \mathcal{K}_0 + 2\mathrm{Re} \sum_j \langle \eta | \omega^{-1/2} \lambda(\boldsymbol{x}_j) \rangle_{\mathfrak{h}} \Big| \psi \Big\rangle_{L^2(\mathbb{R}^{dN})} + \|\eta\|_{\mathfrak{h}}^2 \\ &= \langle \psi | \mathcal{K}_0 | \psi \rangle_{L^2(\mathbb{R}^{dN})} + 2\mathrm{Re} \big\langle \eta \big| \langle \psi | \Lambda | \psi \rangle_{L^2(\mathbb{R}^{dN})} \big\rangle_{\mathfrak{h}} + \|\eta\|_{\mathfrak{h}}^2,\end{aligned} \tag{1-25}$$

where $\Lambda \in L^\infty(\mathbb{R}^{dN}; \mathfrak{H})$ is given by

$$\Lambda(\boldsymbol{X}) := \sum_{j=1}^N (\omega^{-1/2}\lambda)(\boldsymbol{x}_j)$$

(recall the assumption (1-8) on λ) and we have exploited the linearity of the scalar product. Taking the functional derivative with respect to η, we get the Euler–Lagrange equation for the minimization of the above energy with respect to $\eta \in \mathfrak{h}$, i.e.,

$$\eta + \langle \psi | \Lambda(\,\cdot\,) | \psi \rangle_{L^2(\mathbb{R}^{dN})} = 0, \tag{1-26}$$

yielding the minimizing η_{Pekar} written as

$$\eta_{\mathrm{Pekar}}[\psi] = -\sum_{j=1}^{N} \int_{\mathbb{R}^{dN}} \mathrm{d}\boldsymbol{x}_1 \cdots \boldsymbol{x}_N (\omega^{-1/2}\lambda)(\boldsymbol{x}_j) |\psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)|^2, \tag{1-27}$$

which can be easily seen to belong to $\mathfrak{h}$ under the assumptions made. Plugging η_{Pekar} back into (1-25), we get

$$\mathcal{E}_{\mathrm{Pekar}}[\psi] := \inf_{\eta \in \mathfrak{h}} \mathcal{F}_{\mathrm{qc}}[\psi, \eta] = \mathcal{F}_{\mathrm{qc}}[\psi, \eta_{\mathrm{Pekar}}[\psi]] = \big\langle \psi \big| \mathcal{K}_0 + \mathcal{V}_{\mathrm{Pekar}} \star |\psi|^2 \big| \psi \big\rangle. \tag{1-28}$$

Here we have denoted by $\star$ the action of the integral kernel $\mathcal{V}_{\mathrm{Pekar}}(\boldsymbol{X}, \boldsymbol{Y})$ on $|\psi|^2$, i.e.,

$$(\mathcal{V}_{\mathrm{Pekar}} \star |\psi|^2)(\boldsymbol{X}) := \int_{\mathbb{R}^{dN}} \mathrm{d}\boldsymbol{Y} \mathcal{V}_{\mathrm{Pekar}}(\boldsymbol{X}, \boldsymbol{Y}) |\psi(\boldsymbol{Y})|^2, \tag{1-29}$$

and

$$\mathcal{V}_{\mathrm{Pekar}}(\boldsymbol{X}, \boldsymbol{Y}) = -\mathrm{Re} \sum_{i,j=1}^{N} \langle \lambda(\boldsymbol{x}_i) | \omega^{-1} | \lambda(\boldsymbol{y}_j) \rangle_{\mathfrak{h}} \in L^\infty(\mathbb{R}^{2dN}). \tag{1-30}$$

Note that in the case of identical particles — either fermionic or bosonic — the above expressions may be conveniently rewritten using the one-particle density $\rho_\psi \in L^1(\mathbb{R}^d)$ associated with ψ, i.e.,

$$\rho_\psi(\boldsymbol{x}) := N \int_{\mathbb{R}^{d(N-1)}} \mathrm{d}\boldsymbol{x}_2 \cdots \mathrm{d}\boldsymbol{x}_N |\Psi(\boldsymbol{x}, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)|^2. \tag{1-31}$$

Indeed, in this case, (1-27) reads

$$\eta_{\mathrm{Pekar}}[\psi] = -\langle \rho_\psi | (\omega^{-1/2}\lambda)(\,\cdot\,) \rangle_{L^2(\mathbb{R}^d)},$$

and the Pekar energy becomes

$$\mathcal{E}_{\mathrm{Pekar}}[\psi] = \langle \psi | \mathcal{K}_0 | \psi \rangle_{L^2(\mathbb{R}^{dN})} + \langle \rho_\psi | \mathcal{U} | \rho_\psi \rangle_{L^2(\mathbb{R}^d)}, \tag{1-32}$$

where

$$\mathcal{U} = U(\boldsymbol{x}, \boldsymbol{y}) := \langle \lambda(\boldsymbol{x}) | \omega^{-1} | \lambda(\boldsymbol{y}) \rangle_{\mathfrak{h}}, \tag{1-33}$$

which is its typical form in the literature. For instance, in the polaron case, one recovers the self-interacting potential generated by the kernel $\mathcal{U}(\boldsymbol{x} - \boldsymbol{y}) = -\alpha |\boldsymbol{x} - \boldsymbol{y}|^{-1}$.

The above derivation is easily seen to be correct under the assumptions made in case (a). In case (b), however, one cannot apply such a derivation straightforwardly because $\lambda \notin L^\infty(\mathbb{R}^{dN}; \mathfrak{h})$, but a simple well-known trick (see Section 4B) allows us to split it into two terms, which can be handled separately as above. In case (c) on the other hand the Pekar functional takes the implicit form

$$\begin{cases} \eta_{\mathrm{Pekar}} + \sum_j \frac{1}{m_j} \big\langle \psi \big| e\omega^{-1/2} \boldsymbol{\lambda}_j \cdot (-i\nabla_j) + 2e^2 \omega^{-1/2} \boldsymbol{\lambda}_j \cdot \mathrm{Re} \langle \eta_{\mathrm{Pekar}} | \omega^{-1/2} \boldsymbol{\lambda}_j \rangle_{\mathfrak{h}} \big| \psi \big\rangle_{L^2(\mathbb{R}^{3N})} = 0, \\ \mathcal{E}_{\mathrm{Pekar}}[\psi] = \langle \psi | \mathcal{H}_{z_\psi} | \psi \rangle_{L^2(\mathbb{R}^{3N})}, \end{cases} \tag{1-34}$$

where $\mathcal{H}_z$ is given by (1-17), and we set $z_\psi := \omega^{-1/2} \eta_{\mathrm{Pekar}}[\psi]$ for short. As before, all the terms in the first equation belong to $\mathfrak{h}$, thanks to the assumptions on $\boldsymbol{\lambda}_j$ and the fact that any $(\psi, z) \in \mathscr{D}_{\mathrm{qc}}$ is such that

$\psi \in H^1(\mathbb{R}^{3N})$. Furthermore, the last term can be thought of as the action on η_{Pekar} of a linear operator T on $\mathfrak{h}$ whose norm is bounded by

$$2e^2 \sum_{j=1}^{N} \frac{1}{m_j} \|\omega^{-1/2} \boldsymbol{\lambda}_j\|_{\mathfrak{h}}^2,$$

which is smaller than one if e is small enough. In this case, $1+T$ is invertible and there exists a unique solution $\eta_{\text{Pekar}}[\psi] \in \mathfrak{h}$ of the first equation. More generally, existence and uniqueness of $\eta_{\text{Pekar}}[\psi]$ for any value of e follows from the strict convexity of the energy in η; see Remark 1.2 and Lemma 2.4. Note however that unfortunately it is not possible to write explicitly $\mathcal{E}_{\text{Pekar}}$ as a functional of ψ alone, since, due to the presence of an operator — the gradient — one cannot exchange the scalar product in $L^2(\mathbb{R}^{3N})$ with the one in $\mathfrak{h}$, as was done in (1-25). In particular, even for identical particles, the second term in the first equation in (1-34) depends on the reduced density matrix, while the last term is a function of the density alone.

We now define

$$E_{\text{Pekar}} := \inf_{\psi \in \mathscr{D}_{\text{Pekar}}} \mathcal{E}_{\text{Pekar}}[\psi] \tag{1-35}$$

with

$$\mathscr{D}_{\text{Pekar}} := \{\psi \in L^2(\mathbb{R}^{dN}) : \|\psi\|_2 = 1,\ |\mathcal{E}_{\text{Pekar}}[\psi]| < +\infty\}$$

as the ground state energy of the Pekar functionals (1-28) and (1-34), and denote by $\psi_{\text{Pekar}} \in \mathscr{D}_{\text{Pekar}}$ any corresponding minimizer. It is then natural to wonder whether there is any connection between the questions (VP1) and (VP2) and the analogous stability and ground state existence questions for $\mathcal{E}_{\text{Pekar}}$, i.e.,

$$\text{“Is } E_{\text{Pekar}} \text{ greater than } -\infty\text{?”} \tag{VP$'$1}$$

$$\text{“Does there exist } \psi_{\text{Pekar}} \in L^2(\mathbb{R}^{dN}) \text{ such that } \mathcal{E}_{\text{Pekar}}(\psi_{\text{Pekar}}) = E_{\text{Pekar}}\text{?”} \tag{VP$'$2}$$

This is of particular interest for physical applications, since the minimization of the nonlinear functional $\mathcal{E}_{\text{Pekar}}$ may be easier to address also in numerical experiments. A priori however it is not at all obvious that such a relation exists, but in Proposition 1.1 (see Section 2A for the proof) we are going to state that the two variational problems are actually equivalent, which is particularly interesting in case (c) since the explicit form of $\mathcal{E}_{\text{Pekar}}$ is not available.

Proposition 1.1 (equivalence of variational problems). *Under the assumptions made above*,

$$E_{\text{Pekar}} = E_{\text{qc}} > -\infty. \tag{1-36}$$

Furthermore, if $(\psi_{\text{qc}}, z_{\text{qc}}) \in \mathscr{D}_{\text{qc}}$ *is a minimizer of* $\mathcal{E}_{\text{qc}}[\psi, z]$, *then*

$$\mathcal{E}_{\text{Pekar}}[\psi_{\text{qc}}] = E_{\text{Pekar}}. \tag{1-37}$$

Conversely, if ψ_{Pekar} *is a minimizer of* $\mathcal{E}_{\text{Pekar}}[\psi]$, *then* $\eta_{\text{Pekar}}[\psi_{\text{Pekar}}] \in \mathfrak{h}$ *(given by* (1-26) *and* (1-34) *with* $\psi = \psi_{\text{Pekar}}$, *respectively) and*

$$\mathcal{E}[\psi_{\text{Pekar}}, \eta_{\text{Pekar}}] = E_{\text{qc}}. \tag{1-38}$$

Remark 1.2 (uniqueness of η_{Pekar}). We prove in Lemma 2.4 that the quasiclassical functional $\mathcal{F}_{\mathrm{qc}}[\psi, \eta]$ (or, equivalently, $\mathcal{E}_{\mathrm{qc}}[\psi, z]$) is strictly convex in $\eta \in \mathfrak{h}$ for given $\psi \in L^2(\mathbb{R}^{dN})$. Hence, $\eta_{\mathrm{Pekar}}[\psi]$ is unique (for fixed ψ). Note however that the functional $\mathcal{F}_{\mathrm{qc}}$ is not jointly convex in $(|\psi|^2, \eta)$.

1C. *Ground state in the quasiclassical regime.*

We can now state in detail our main results. We work with a minimal set of assumptions on the microscopic models, which are the weakest ones guaranteeing the self-adjointness and boundedness from below of the microscopic Hamiltonians.

Assumptions. The following conditions are satisfied:

(A1) The external potential $\mathcal{W}$ is such that[3]

$$\mathcal{W} \in L^1_{\mathrm{loc}}(\mathbb{R}^{dN}; \mathbb{R}^+); \tag{1-39}$$

(A2) the operator ω is positive and admits a possibly unbounded inverse ω^{-1};

(A3) the form factor λ of the microscopic model must satisfy condition (1-8), (1-9) or (1-11) for the Nelson, polaron or Pauli–Fierz models, respectively.

Observe in particular that the quantum potential $\mathcal{W}$ may not be trapping, so that there might not be a ground state for both the microscopic and the macroscopic problems. In some of the results stated below however we are going to assume this explicitly by requiring an additional property of the unperturbed particle operator:

(A4) The operator $\mathcal{K}_0$ has compact resolvent.

We now consider the microscopic ground state energy E_ε defined in (1-12) and its quasiclassical limit. Recall the definition of the quasiclassical energy E_{qc} in (1-19).

Theorem 1.3 (ground state energy). *Under assumptions* (A1), (A2) *and* (A3), *there exists $C < +\infty$ such that $E_\varepsilon > -C$ and*

$$E_\varepsilon \xrightarrow[\varepsilon \to 0]{} E_{\mathrm{qc}}, \tag{1-40}$$

which in particular implies that (VP1) *holds true.*

The proof of the result above is given in Section 3A. Once the energy convergence has been stated, it is natural to ask whether, in the presence of a microscopic approximate ground state $\Psi_{\varepsilon,\delta}$ or ground state $\Psi_{\varepsilon,\mathrm{gs}}$, one can prove a suitable convergence to quasiclassical minimizing sequences or configurations $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}}) \in \mathscr{D}_{\mathrm{qc}}$, respectively. Let us stress that the question of existence of a ground state of the microscopic energy has been widely studied in the literature and there are more restrictive conditions on the models guaranteeing that $E_\varepsilon \in \sigma_{\mathrm{pp}}(H_\varepsilon)$ (see Sections 4A–4C); our results about approximate ground states apply even if the microscopic ground state does not exist, and whenever it exists we are able to provide its quasiclassical characterization.

[3] As anticipated above, it is sufficient to have an unperturbed particle operator which is self-adjoint and bounded from below. For instance, one could extend the results to potentials with a negative part which is Kato-small with respect to the Laplacian. We stick however to (A1) for the sake of concreteness.

In order to properly formulate the convergence, we first need to introduce a key structure in quasiclassical analysis: the *quasiclassical Wigner measures* and their relative topologies. We preliminarily recall the definition of the space $\mathscr{P}(\mathfrak{h}_\omega; L^2(\mathbb{R}^{dN}))$ of *state-valued probability measures* (see [Correggi et al. 2023, Definition 2.1]), given by measures $\mathfrak{m}$ on $\mathfrak{h}_\omega$ taking values in $\mathscr{L}^1_+(L^2(\mathbb{R}^{dN}))$ — the space of positive trace class operators on $L^2(\mathbb{R}^{dN})$ — such that $\mathfrak{m}(\varnothing) = 0$, the measure is unconditionally σ-additive in the trace class norm and $\|\mathfrak{m}(\mathfrak{h}_\omega)\|_{L^2} = 1$. Starting from such a notion, it is possible to construct a theory of integration of functions with values in the space of bounded operators on $L^2(\mathbb{R}^{dN})$ with respect to state-valued measures, so that, for any measurable $\mathcal{B}(z) \in \mathscr{B}(L^2(\mathbb{R}^{dN}))$,

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{m}(z)\mathcal{B}(z) \in \mathscr{L}^1(L^2(\mathbb{R}^{dN})). \tag{1-41}$$

We refer to the Appendix, or to the existing literature (e.g., [Balazard-Konlein 1985; Fermanian-Kammerer and Gérard 2002; Gérard 1991; Gérard et al. 1991; Teufel 2003]) for further details. In particular, we point out that any such state-valued measure $\mathfrak{m}$ admits a Radon–Nikodým decomposition, i.e., there exists a scalar Borel measure $\mu_\mathfrak{m}$ and a $\mu_\mathfrak{m}$-integrable function $\gamma_\mathfrak{m}(z) \in \mathscr{L}^1_{+,1}(L^2(\mathbb{R}^{dN}))$ defined a.e. and with values in normalized density matrices, such that

$$\mathrm{d}\mathfrak{m}(z) = \gamma_\mathfrak{m}(z)\mathrm{d}\mu_\mathfrak{m}(z). \tag{1-42}$$

Hence, (1-41) can be rewritten as

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{m}(z)\mathcal{B}(z) = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_\mathfrak{m}(z)\gamma_\mathfrak{m}(z)\mathcal{B}(z). \tag{1-43}$$

Finally, let us denote by $W_\varepsilon(z)$, $z \in \mathfrak{h}$, the Weyl operator constructed over the creation and annihilation operators $a^\sharp_\varepsilon$, i.e.,

$$W_\varepsilon(z) := e^{i(a^\dagger_\varepsilon(z)+a_\varepsilon(z))}. \tag{1-44}$$

Definition 1.4 (quasiclassical Wigner measures). For any family of normalized microscopic states $\{\Psi_\varepsilon\}_{\varepsilon\in(0,1)} \subset \mathscr{H}_\varepsilon$, the associated set of quasiclassical Wigner measures

$$\mathscr{W}(\Psi_\varepsilon, \varepsilon \in (0, 1)) \subset \mathscr{P}(\mathfrak{h}_\omega; \mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))$$

is the subset of all probability measures $\mathfrak{m}$ such that there exists $\{\varepsilon_n\}_{n\in\mathbb{N}}$, $\varepsilon_n \xrightarrow[n\to\infty]{} 0$, so that

$$\Psi_{\varepsilon_n} \xrightarrow[n\to\infty]{\mathrm{qc}} \mathfrak{m}, \tag{1-45}$$

where the above convergence yields, for all $\eta \in \mathscr{D}(\omega^{-1/2})$ and all compact operators $\mathcal{K} \in \mathscr{L}^\infty(L^2(\mathbb{R}^{dN}))$,

$$\begin{aligned}\lim_{n\to\infty} \langle \Psi_{\varepsilon_n} | \mathcal{K} \otimes W_{\varepsilon_n}(\eta)\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} &= \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_\mathfrak{m}(z) e^{2i\mathrm{Re}\langle\eta|z\rangle_\mathfrak{h}} \operatorname{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_\mathfrak{m}(z)\mathcal{K}] \\ &= \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_\mathfrak{m}(z) e^{2i\mathrm{Re}\langle\omega^{-1/2}\eta|\omega^{1/2}z\rangle_\mathfrak{h}} \operatorname{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_\mathfrak{m}(z)\mathcal{K}].\end{aligned} \tag{1-46}$$

Remark 1.5 (measures on $\mathfrak{h}_\omega$ and test functions). A reader familiar with infinite dimensional semiclassical analysis or quasiclassical analysis will find the definition of Wigner measures given here differs slightly from the usual definition [Ammari and Nier 2008; Correggi et al. 2023]. Typically, one considers microscopic states that satisfy a number operator estimate, namely for which the expectation of $\mathrm{d}\mathcal{G}_\varepsilon(1)^c$ is ε-uniformly bounded for some $c > 0$. The corresponding Wigner measures are concentrated on $\mathfrak{h}$ [Ammari and Nier 2008], and it is natural to test the convergence with Weyl operators having arguments $\eta \in \mathfrak{h}$. However, in studying variational problems the number operator estimate may not always be available, in particular whenever the field is massless, such as in electromagnetism (Pauli–Fierz model). In that case, only energy estimates, i.e., involving $\mathrm{d}\mathcal{G}_\varepsilon(\omega)$, are available. The Wigner measures of states satisfying such an energy estimate are concentrated in $\mathfrak{h}_\omega$, and it is natural to test convergence with Weyl operators having arguments $\eta \in \mathscr{D}(\omega^{-1/2})$ belonging to a dense subset of the continuous dual space [Falconi 2018a]. If both the number estimate and the free energy estimate are available, then the measure is concentrated in $\mathfrak{h} \cap \mathfrak{h}_\omega$; this happens for massive fields, where in addition $\mathfrak{h} \cap \mathfrak{h}_\omega = \mathfrak{h}_\omega$. Finally, let us remark as well that in all concrete applications $\mathfrak{h}_\omega$ is in fact the natural domain of definition of the quasiclassical energy $\mathcal{E}_{\mathrm{qc}}$.

The above notion of quasiclassical convergence, defined in (1-46), is however not the only meaningful topology one can consider for sequences of microscopic states. More precisely, the test in (1-46) may be extended to bounded operators, which means that one is considering the weak* topology on $\mathscr{B}(L^2(\mathbb{R}^{dN}))'$ instead of $\mathscr{L}^1(L^2(\mathbb{R}^{dN})) = \mathscr{L}^\infty(L^2(\mathbb{R}^{dN}))'$. In this case, the cluster points belong to a larger space than $\mathscr{P}(\mathfrak{h}_\omega; \mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))$, namely the space of *generalized state-valued measures*; see [Falconi 2018b] for a detailed and more general discussion. We thus introduce the set of positive states $\overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN}))$ in the closure with respect to the weak* topology of the space of trace class operators on $L^2(\mathbb{R}^{dN})$: we denote the action of a functional $F \in \overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN}))$ on a bounded operator $\mathcal{B} \in \mathscr{B}(L^2(\mathbb{R}^{dN}))$ by $F[\mathcal{B}] \in \mathbb{C}$ and its norm by

$$\|F\|_{\mathscr{B}'} := \sup_{\mathcal{B} \in \mathscr{B}(L^2(\mathbb{R}^{dN})),\, \|\mathcal{B}\|=1} |F[\mathcal{B}]|. \tag{1-47}$$

Definition 1.6 (generalized quasiclassical Wigner measures). For any family of normalized microscopic states $\{\Psi_\varepsilon\}_{\varepsilon\in(0,1)} \subset L^2(\mathbb{R}^{dN})_\varepsilon$, the associated set of generalized quasiclassical Wigner measures

$$\mathscr{GW}(\Psi_\varepsilon, \varepsilon \in (0,1)) \subset \mathscr{P}(\mathfrak{h}_\omega; \overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN})))$$

is the subset of all probability measures $\mathfrak{n}$ such that there exists $\{\varepsilon_n\}_{n\in\mathbb{N}}$, $\varepsilon_n \xrightarrow[n\to\infty]{} 0$, so that

$$\Psi_{\varepsilon_n} \xrightarrow[n\to\infty]{\mathrm{gqc}} \mathfrak{n}, \tag{1-48}$$

where the above convergence means that, for all $\eta \in \mathscr{D}(\omega^{-1/2})$ and all bounded operators $\mathcal{B} \in \mathscr{B}(L^2(\mathbb{R}^{dN}))$,

$$\lim_{n\to\infty} \langle \Psi_{\varepsilon_n} | \mathcal{B} \otimes W_{\varepsilon_n}(\eta) \Psi_{\varepsilon_n} \rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z)[\mathcal{B}] e^{2i\mathrm{Re}\langle \omega^{-1/2}\eta | \omega^{1/2} z\rangle_{\mathfrak{h}}}. \tag{1-49}$$

We can now formulate the results about the convergence of microscopic minimizing sequences $\Psi_{\varepsilon,\delta}$ and microscopic minimizers $\Psi_{\varepsilon,\mathrm{gs}}$ (for the proofs see Section 3B). We start by stating a stronger result with some additional assumptions on the microscopic models. Without such assumptions we are still able to prove a weaker convergence, but it requires the introduction of a generalized variational problem.

Theorem 1.7 (convergence of approximate ground states (I)). *Under assumptions* (A1), (A2), (A3) *and* (A4), *for any $\delta > 0$ and for any family of approximate ground states $\Psi_{\varepsilon,\delta}$ satisfying* (1-13), *we have*

$$\mathscr{W}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0,1)) \neq \varnothing.$$

Moreover, any family of quasiclassical Wigner measures $\{\mathfrak{m}_\delta\}_{\delta>0}$, with $\mathfrak{m}_\delta \in \mathscr{W}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0,1))$ for any δ, is such that, for all $\delta > 0$, we have that $\operatorname{tr}_{L^2(\mathbb{R}^{dN})} \mathfrak{m}_\delta(\mathfrak{h}_\omega) = 1$ and $\mathfrak{m}_\delta$ is an approximate ground state of $\mathcal{E}_{\mathrm{qc}}[\psi, z]$, i.e.,

$$\mathcal{E}_{\mathrm{svm}}[\mathfrak{m}_\delta] := \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}_\delta}(z) \operatorname{tr}_{L^2(\mathbb{R}^{dN})}(\gamma_{\mathfrak{m}_\delta}(z)\mathcal{H}_z) < E_{\mathrm{qc}} + \delta. \tag{1-50}$$

Remark 1.8 (concentration in probability). The result stated in Theorem 1.7 does not imply that the energy $E(z) := \operatorname{tr}_{L^2(\mathbb{R}^{dN})}(\gamma_{\mathfrak{m}_\delta}(z)\mathcal{H}_z)$ is smaller than $E_{\mathrm{qc}} + \delta$ for any $z \in \mathfrak{h}_\omega$. Roughly speaking, there might be a nonzero probability that $\mathfrak{m}_\delta$ is concentrated on pairs $(\psi(z), z)$ with large $\mathcal{E}_{\mathrm{qc}}$ energy. However, $E(z)$ can be much larger than $E_{\mathrm{qc}} + \delta$ only with small $\mu_{\mathfrak{m}_\delta}$-probability. More precisely, for any $k \geq 1$,

$$\mathbb{P}_{\mu_{\mathfrak{m}_\delta}}\{E(z) \geq E_{\mathrm{qc}} + k\delta\} < \frac{1}{k}. \tag{1-51}$$

Corollary 1.9 (convergence to ground states (I)). *If* (A4) *holds, then any quasiclassical Wigner measure $\mathfrak{m} \in \mathscr{W}(\Psi_{\varepsilon,o_\varepsilon(1)}, \varepsilon \in (0,1))$, corresponding to approximate ground states $\Psi_{\varepsilon,o_\varepsilon(1)}$ satisfying* (1-13) *with $\delta = o_\varepsilon(1)$, is such that $\operatorname{tr}_{L^2(\mathbb{R}^{dN})} \mathfrak{m}(\mathfrak{h}_\omega) = 1$ and $\mathfrak{m}$ is concentrated on the set of ground states $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}}) \in \mathscr{D}_{\mathrm{qc}}$ of $\mathcal{E}_{\mathrm{qc}}[\psi, z]$. Consequently, $\mathcal{E}_{\mathrm{qc}}[\psi, z]$ has at least one ground state and both* (VP2) *and* (VP′2) *hold true.*

Corollary 1.10 (convergence of ground states (I)). *If* (A4) *holds and H_ε has a ground state $\Psi_{\varepsilon,\mathrm{gs}}$, then any corresponding quasiclassical Wigner measure $\mathfrak{m} \in \mathscr{W}(\Psi_{\varepsilon,\mathrm{gs}}, \varepsilon \in (0,1))$ is such that $\operatorname{tr}_{L^2(\mathbb{R}^{dN})} \mathfrak{m}(\mathfrak{h}_\omega) = 1$ and $\mathfrak{m}$ is concentrated on the set of ground states $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}}) \in \mathscr{D}_{\mathrm{qc}}$ of $\mathcal{E}_{\mathrm{qc}}[\psi, z]$.*

Remark 1.11 (uniqueness and gauge invariance). Concerning uniqueness, we point out that both the microscopic and the quasiclassical variational problems are gauge invariant, namely the multiplication by a constant phase factor of Ψ or ψ does not change the energy. Hence, even if one could prove uniqueness of the quasiclassical minimizer $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}})$ up to gauge transformations, one could not conclude that the set of limit points $\mathscr{W}(\Psi_{\varepsilon,o_\varepsilon(1)}, \varepsilon \in (0,1))$ or $\mathscr{W}(\Psi_{\varepsilon,\mathrm{gs}}, \varepsilon \in (0,1))$ are just given by a Dirac delta measure centered at $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}})$. Indeed, because of gauge invariance, the quasiclassical Wigner measures would be supported over the unit one-dimensional sphere generated by the configurations $(e^{i\vartheta}\psi_{\mathrm{qc}}, z_{\mathrm{qc}})$ with $\vartheta \in \mathbb{R}$.

Remark 1.12 (condition on $\mathcal{K}_0$). The assumption that $\mathcal{K}_0$ has compact resolvent is reasonable, since that is typically the case in which one can also prove the existence of a microscopic minimizer at least for massive systems (see Remark 1.13 below), e.g., in the presence of a trapping potential. However, it is also needed in a technical step in the proof to ensure that there is no loss of mass along the convergence (1-46), i.e., $\operatorname{tr}_{L^2(\mathbb{R}^{dN})} \mathfrak{m}(\mathfrak{h}_\omega) = 1$. Similar assumptions are present also in [Correggi et al. 2023]; see in particular the discussion in [Correggi et al. 2023, Remarks 1.9–1.10 and Section 1.6].

Remark 1.13 (existence of $\Psi_{\varepsilon,\mathrm{gs}}$). In all three cases (a)–(c), if the Bose field is *massive*, i.e., there exists $m > 0$ such that $\omega \geq m > 0$ (which is always the case for the polaron), then it is known [Dereziński and Gérard 1999, Theorem 4.1] that the microscopic Hamiltonian H_ε admits a ground state $\Psi_{\varepsilon,\mathrm{gs}} \in \mathscr{H}_\varepsilon$, if $\mathcal{K}_0$

has compact resolvent. Hence, in the massive case, one can remove the assumption on the existence of $\Psi_{\varepsilon,\mathrm{gs}}$. When the field is *massless*, on the other hand, it is also known that microscopic ground states might not exist or belong to a non-Fock representation of the algebra of observables [Pizzo 2003]. This second case is not covered by Corollary 1.10, but it may be treated with our techniques. We plan to come back to such a question in a future work.

Remark 1.14 (existence of quasiclassical minimizers). Our analysis shows that the quasiclassical energy functionals $\mathcal{E}_{\mathrm{qc}}[\psi, z]$ *always* have at least one minimizer, provided that $\mathcal{K}_0$ has compact resolvent, i.e., provided that the quantum subsystem is trapped. This gives additional evidence that the behavior of the ground state in quantum field theories can differ quite dramatically (nonexistence or non-Fock-representability, see Remark 1.13) from that of their classical and quantum finite-dimensional counterparts.

As anticipated, if we drop the assumption on the operator $\mathcal{K}_0$ there is still convergence, but the variational problem (1-19) has to be generalized: we thus set, for any pure state $\rho \in \overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN}))$ and any $z \in \mathfrak{h}_\omega$,

$$\mathcal{E}_{\mathrm{gqc}}[\rho, z] := \rho[\mathcal{H}_z]. \tag{1-52}$$

We consider the corresponding variational problem: setting (recall the definition (1-47))

$$\mathscr{D}_{\mathrm{gqc}} := \{(\rho, z) \in \overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN})) \oplus \mathfrak{h}_\omega : \|\rho\|_{\mathscr{B}'} = 1,\ |\rho[\mathcal{H}_z]| < +\infty\}, \tag{1-53}$$

we define

$$E_{\mathrm{gqc}} := \inf_{(\rho,z)\in\mathscr{D}_{\mathrm{gqc}}} \mathcal{E}_{\mathrm{gqc}}[\rho, z], \tag{1-54}$$

and we denote by $(\rho_\delta, z_\delta) \in \mathscr{D}_{\mathrm{gqc}}$ a minimizing sequence satisfying

$$\mathcal{E}_{\mathrm{gqc}}[\rho_\delta, z_\delta] < E_{\mathrm{gqc}} + \delta$$

and by $(\rho_{\mathrm{gqc}}, z_{\mathrm{gqc}}) \in \mathscr{D}_{\mathrm{gqc}}$ any corresponding minimizing configuration.

Theorem 1.15 (convergence of approximate ground states (II)). *Under assumptions* (A1), (A2) *and* (A3), *for any* $\delta > 0$ *and for any family of approximate ground states* $\Psi_{\varepsilon,\delta}$ *satisfying* (1-13), *we have that* $\mathscr{GW}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0, 1)) \neq \varnothing$. *Moreover, any family of generalized quasiclassical Wigner measures* $\{\mathfrak{n}_\delta\}_{\delta>0}$, *with* $\mathfrak{n}_\delta \in \bigcup_{\delta>0} \mathscr{GW}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0, 1))$ *for any* δ, *is such that, for all* $\delta > 0$, *we have that* $\|\mathfrak{n}_\delta(\mathfrak{h}_\omega)\|_{\mathscr{B}'} = 1$ *and* $\mathfrak{n}_\delta$ *is an approximate ground state of* $\mathcal{E}_{\mathrm{gqc}}[\rho, z]$, *i.e.*,

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}_\delta(z)[\mathcal{H}_z] < E_{\mathrm{gqc}} + \delta. \tag{1-55}$$

Corollary 1.16 (convergence to ground states (II)). *Any generalized quasiclassical Wigner measure* $\mathfrak{n} \in \mathscr{GW}(\Psi_{\varepsilon,o_\varepsilon(1)}, \varepsilon \in (0, 1))$, *corresponding to approximate ground states* $\Psi_{\varepsilon,o_\varepsilon(1)}$ *satisfying* (1-13) *with* $\delta = o_\varepsilon(1)$, *is such that* $\|\mathfrak{n}(\mathfrak{h}_\omega)\|_{\mathscr{B}'} = 1$ *and* $\mathfrak{n}$ *is concentrated on the set of ground states* $(\varrho_{\mathrm{gqc}}, z_{\mathrm{gqc}}) \in \mathscr{D}_{\mathrm{gqc}}$ *of* $\mathcal{E}_{\mathrm{gqc}}[\varrho, z]$. *Consequently, the functional* $\mathcal{E}_{\mathrm{gqc}}[\rho, z]$ *admits at least one ground state in* $\mathscr{D}_{\mathrm{gqc}}$.

Corollary 1.17 (convergence of ground states (II)). *If* H_ε *has a ground state* $\Psi_{\varepsilon,\mathrm{gs}}$, *then any generalized Wigner measure* $\mathfrak{n} \in \mathscr{GW}(\Psi_{\varepsilon,\mathrm{gs}}, \varepsilon \in (0, 1))$ *is such that* $\|\mathfrak{n}(\mathfrak{h}_\omega)\|_{\mathscr{B}'} = 1$ *and* $\mathfrak{n}$ *is concentrated on the set of ground states* $(\rho_{\mathrm{gqc}}, z_{\mathrm{gqc}}) \in \mathscr{D}_{\mathrm{gqc}}$ *of* $\mathcal{E}_{\mathrm{gqc}}[\rho, z]$.

Remark 1.18 (quasiclassical energy and generalized quasiclassical energy). As proved in Section 2 (see Proposition 2.8),

$$E_{\mathrm{qc}} = E_{\mathrm{gqc}},$$

which is in fact crucial to prove convergence of the ground state energy for systems without trapping on the quantum particles.

2. Quasiclassical minimization problems

Here we consider minimization problems in the quasiclassical setting: we study the functionals introduced in Section 1B and the relative minimizations, but also define and investigate more general problems.

2A. ***Quasiclassical functionals, states and related minimization problems.*** A quasiclassical system behaves like an open system in which a *classical environment* (of infinite dimension) drives a quantum *small system*, described by a Hilbert space $L^2(\mathbb{R}^{dN})$. The classical environment is described by a space of configurations $\mathfrak{h}_\omega$, usually a complex Hilbert space identifiable with the complex phase space of the environment's degrees of freedom. A probability distribution μ on $\mathfrak{h}_\omega$ tells how probable each environment's configuration is, while a state-valued function $\mathfrak{h}_\omega \ni z \mapsto \gamma(z) \in \mathscr{L}^1_+(L^2(\mathbb{R}^{dN}))$ tells how each environment's configuration drives the small system's quantum state. Analogously, both the value of observables $\mathcal{F}(z)$ and the small system's dynamics $\mathcal{U}_t(z)$ are driven by the environment.

A quasiclassical minimization problem consists of finding the lowest energy and possibly the ground states of a suitable functional $\mathcal{E}[\psi, z] : L^2(\mathbb{R}^{dN}) \oplus \mathfrak{h}_\omega \to \mathbb{R}$ depending on the configuration of both the small system and the environment. The first energy functional to consider is $\mathcal{E}_{\mathrm{qc}}[\psi, z]$, defined in (1-18):

$$\mathcal{E}_{\mathrm{qc}}[\psi, z] := \langle \psi | \mathcal{H}_z | \psi \rangle_{L^2(\mathbb{R}^{dN})}, \qquad (\psi, z) \in \mathscr{D}_{\mathrm{qc}},$$

where $\mathcal{H}_z$ and $\mathscr{D}_{\mathrm{qc}}$ are given in (1-14) and (1-20), respectively. We also recall that the ground state energy and minimizer of $\mathcal{E}_{\mathrm{qc}}$ are denoted by E_{qc} and $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}})$, respectively.

Although the above is the foremost functional coming to mind in this context, another minimization problem emerges naturally in studying the quasiclassical limit. To this purpose, we recall the notion of a state-valued measure [Correggi et al. 2023; Falconi 2018b], already mentioned in Section 1C: a state-valued *probability* measure $\mathfrak{m} \in \mathscr{P}(\mathfrak{h}_\omega; \mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))$ is a vector Borel Radon measure on $\mathfrak{h}_\omega$, taking values in the density matrices $\mathscr{L}^1_+(L^2(\mathbb{R}^{dN}))$ of the small system, such that

$$\|\mathfrak{m}(\mathfrak{h}_\omega)\|_{\mathscr{L}^1} = 1. \tag{2-1}$$

Thanks to the Radon–Nikodým property enjoyed by the separable dual space $\mathscr{L}^1(L^2(\mathbb{R}^{dN}))$, it is possible to decompose $\mathfrak{m}$ in a scalar Borel Radon probability measure $\mu_{\mathfrak{m}} \in \mathscr{P}(\mathfrak{h}_\omega)$ such that $\mu_{\mathfrak{m}}(\mathfrak{h}) = 1$, and in an a.e.-defined function (the Radon–Nikodým derivative)

$$\mathfrak{h}_\omega \ni z \mapsto \gamma_{\mathfrak{m}}(z) \in \mathscr{L}^1_{+,1}(L^2(\mathbb{R}^{dN}))$$

taking values in the normalized density matrices of the small system:

$$\mathrm{d}\mathfrak{m}(z) = \gamma_{\mathfrak{m}}(z)\mathrm{d}\mu_{\mathfrak{m}}(z).$$

The quasiclassical energy $\mathcal{E}_{\mathrm{qc}}$, constrained to $\|\psi\|_{L^2(\mathbb{R}^{dN})} = 1$, is the expectation of the quasiclassical Hamiltonian $\mathcal{H}_z$. Therefore, its generalization to state-valued measures obviously reads

$$\mathcal{E}_{\mathrm{svm}}[\mathfrak{m}] := \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\, \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}}(z)\mathcal{H}_z]. \tag{2-2}$$

This leads to the following minimization problem: setting

$$\mathscr{D}_{\mathrm{svm}} := \{\mathfrak{m} \in \mathscr{P}(\mathfrak{h}_\omega; \mathscr{L}^1_+(L^2(\mathbb{R}^{dN}))) : \mathrm{tr}_{L^2(\mathbb{R}^{dN})}\, \mathfrak{m}(\mathfrak{h}_\omega) = 1,\ |\mathcal{E}_{\mathrm{svm}}[\mathfrak{m}]| < +\infty\}, \tag{2-3}$$

we ask (stability)

$$\text{“Is } E_{\mathrm{svm}} := \inf_{\mathfrak{m}\in\mathscr{D}_{\mathrm{svm}}} \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}] \text{ greater than } -\infty\text{?”} \tag{vp1}$$

and (existence of a ground state)

$$\text{“Does there exist } \mathfrak{m}_{\mathrm{svm}} \in \mathscr{D}_{\mathrm{svm}} \text{ such that } \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}_{\mathrm{svm}}] = E_{\mathrm{svm}}\text{?”} \tag{vp2}$$

A variant of the above problem is obtained by assuming that $\gamma_{\mathfrak{m}}(z) = |\psi\rangle\langle\psi|$ for some $\psi \in L^2(\mathbb{R}^{dN})$ independent of z, in which case the functional depends only on a wave function ψ and a probability measure μ over $\mathfrak{h}_\omega$. We thus set

$$\mathcal{E}_{\mathrm{pm}}[\psi, \mu] := \int_{\mathfrak{h}_\omega} \mathrm{d}\mu(z)\langle\psi|\mathcal{H}_z|\psi\rangle_{L^2(\mathbb{R}^{dN})}. \tag{2-4}$$

The variational problem (stability) reads

$$\text{“Is } E_{\mathrm{pm}} := \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}}} \mathcal{E}_{\mathrm{pm}}[\psi, \mu] \text{ greater than } -\infty\text{?”} \tag{vp$'$1}$$

where

$$\mathscr{D}_{\mathrm{pm}} := \{(\psi, \mu) \in L^2(\mathbb{R}^{dN}) \oplus \mathscr{P}(\mathfrak{h}_\omega) : \|\psi\|_2 = 1,\ \mu(\mathfrak{h}_\omega) = 1,\ |\mathcal{E}_{\mathrm{pm}}[\psi, \mu]| < +\infty\}, \tag{2-5}$$

and (existence of a ground state)

$$\text{“Does there exist } (\psi_{\mathrm{pm}}, \mu_{\mathrm{pm}}) \in \mathscr{D}_{\mathrm{pm}} \text{ such that } \mathcal{E}_{\mathrm{pm}}[(\psi_{\mathrm{pm}}, \mu_{\mathrm{pm}})] = E_{\mathrm{pm}}\text{?”} \tag{vp$'$2}$$

Note that the functional $\mathcal{E}_{\mathrm{qc}}$ and the corresponding variational problems (VP1) and (VP2) are recovered by simply imposing in $\mathcal{E}_{\mathrm{pm}}$ above that μ is a Dirac delta, i.e., there exists $z_0 \in \mathfrak{h}_\omega$ such that $\mu = \delta_{z_0}$. Yet another minimization problem can be formulated by substituting the minimization over $\mathscr{P}(\mathfrak{h}_\omega; \mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))$ and $\mathscr{P}(\mathfrak{h}_\omega)$ in (2-2) and (2-4) with the one over atomic measures $\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega; \mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))$ and $\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)$, respectively.

Finally, in the spirit of derivation of effective functionals of ψ or z alone, as the Pekar-like functionals defined in (1-28) and (1-34), we can also define the effective energy

$$\mathcal{I}[z] := \inf_{\psi\in L^2(\mathbb{R}^{dN}),\, \|\psi\|_2=1} \mathcal{E}_{\mathrm{qc}}[\psi, z]. \tag{2-6}$$

The rest of this section is devoted to proving equivalences between the minimization problems defined above. In fact, the natural variational problem emerging in the quasiclassical limit is the one involving state-valued measures (see (vp1) and (vp2)), however the most relevant from the physical and practical point of view is the one formulated in terms of wave functions and classical fields (see (VP1) and (VP2)). Therefore, the fact that all the infima turn out to be equal (Propositions 2.1 and 2.8) and that the existence of the various minimizers are related (Propositions 2.3 and 2.9) allow us to derive a more concrete physical statement.

Proposition 2.1 (quasiclassical energies). *Under assumptions* (A1), (A2) *and* (A3),

$$\begin{aligned} E_{\mathrm{qc}} = E_{\mathrm{svm}} &= \inf_{\mathfrak{m}\in\mathscr{D}_{\mathrm{svm}}\cap\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega;\mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))} \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}] = E_{\mathrm{pm}} \\ &= \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu] = E_{\mathrm{Pekar}} = \inf_{z\in\mathfrak{h}_\omega} \mathcal{I}[z]. \end{aligned} \tag{2-7}$$

Proof. We use the weak density of atomic scalar measures, supported on a finite number of points, in the space of all finite measures, that holds for $\mathfrak{h}_\omega$ separable [Parthasarathy 1967]. Thanks to that it is possible to prove the following (see [Correggi and Falconi 2018, Lemma 3.20] for a detailed proof):

$$E_{\mathrm{svm}} = \inf_{\mathfrak{m}\in\mathscr{D}_{\mathrm{svm}}} \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}] = \inf_{\mathfrak{m}\in\mathscr{D}_{\mathrm{svm}}\cap\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega;\mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))} \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}],$$

$$E_{\mathrm{pm}} = \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}}} \mathcal{E}_{\mathrm{pm}}[\psi,\mu] = \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu].$$

Now, let us prove that

$$\inf_{\mathfrak{m}\in\mathscr{D}_{\mathrm{svm}}\cap\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega;\mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))} \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}] = \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu]. \tag{2-8}$$

Let $\delta > 0$, and let $\mathfrak{m}_\delta = \sum_{k=1}^K \lambda_k\gamma_k\delta_{z_k}$ — with $\lambda_k \geq 0$ (recall that $\mathfrak{m}_\delta$ takes values in positive operators), $\sum_{k=1}^K \lambda_k = 1$ and $\gamma_k \in \mathscr{L}^1_{+,1}(L^2(\mathbb{R}^{dN}))$ — be an atomic state-valued measure such that

$$\mathcal{E}_{\mathrm{svm}}[\mathfrak{m}_\delta] = \sum_{k=1}^K \lambda_k \operatorname{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_k\mathcal{H}_{z_k}] < \inf_{\mathfrak{m}\in\mathscr{D}_{\mathrm{svm}}\cap\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega;\mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))} \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}] + \delta.$$

For fixed k, since γ_k is a normalized density matrix,

$$\inf_{\psi\in L^2(\mathbb{R}^{dN}),\, \|\psi\|_2=1} \langle\psi|\mathcal{H}_{z_k}|\psi\rangle_{L^2(\mathbb{R}^{dN})} \leq \operatorname{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_k\mathcal{H}_{z_k}].$$

Therefore,

$$\begin{aligned} \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu] &= \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \int_{\mathfrak{h}_\omega} \mathrm{d}\mu\langle\psi|\mathcal{H}_z|\psi\rangle_{L^2(\mathbb{R}^{dN})} \\ &\leq \sum_{k=1}^K \lambda_k \inf_{\psi\in L^2(\mathbb{R}^{dN}),\, \|\psi\|_2=1} \langle\psi|\mathcal{H}_{z_k}|\psi\rangle_{L^2(\mathbb{R}^{dN})} \leq \sum_{k=1}^K \lambda_k \operatorname{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_k\mathcal{H}_{z_k}] \\ &< \inf_{\mathfrak{m}\in\mathscr{D}_{\mathrm{svm}}\cap\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega;\mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))} \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}] + \delta. \end{aligned} \tag{2-9}$$

Since $\delta > 0$ is arbitrary, we conclude that

$$\inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu] \leq \inf_{\mathfrak{m}\in\mathscr{D}_{\mathrm{svm}}\cap\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega;\mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))} \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}]. \tag{2-10}$$

To prove the opposite inequality, we follow a similar reasoning. Let $\delta > 0$ and $\mu_\delta = \sum_{k=1}^K \lambda_k\delta_{z_k}$ be a scalar atomic measure and $\psi_{\delta,z_k} \in L^2(\mathbb{R}^{dN})$ a family of normalized wave functions such that $\mu_\delta(\mathfrak{h}_\omega) = 1$

and

$$\sum_{k=1}^{K} \lambda_k \langle \psi_{\delta,z_k} | \mathcal{H}_{z_k} | \psi_{\delta,z_k} \rangle_{L^2(\mathbb{R}^{dN})} < \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu] + \delta.$$

Now, $\mathfrak{m}_\delta := \sum_{k=1}^{K} \lambda_k |\psi_{\delta,z_k}\rangle\langle\psi_{\delta,z_k}| \delta_{z_k}$ is an atomic state-valued measure belonging to $\mathscr{D}_{\mathrm{svm}}$. Therefore,

$$\begin{aligned}\inf_{\mathfrak{m}\in\mathscr{D}_{\mathrm{svm}}\cap\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega;\mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))} \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}] \le \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}_\delta] &= \sum_{k=1}^{K} \lambda_k \langle \psi_{\delta,z_k} | \mathcal{H}_{z_k} | \psi_{\delta,z_k} \rangle_{L^2(\mathbb{R}^{dN})} \\ &< \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu] + \delta, \end{aligned} \tag{2-11}$$

which yields the desired inequality.

To complete the proof, we show that

$$\inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu] = \inf_{z\in\mathfrak{h}_\omega} \mathcal{I}[z] = E_{\mathrm{Pekar}} = E_{\mathrm{qc}}. \tag{2-12}$$

Let us prove the first equality beforehand. Let $\mu_\delta = \sum_{k=1}^{K} \lambda_k \delta_{z_k}$ be the atomic minimizing family of measures defined before and ψ_{δ,z_k} the corresponding minimizing vectors. Then

$$\begin{aligned}\sum_{k=1}^{K} \lambda_k \inf_{\psi\in L^2(\mathbb{R}^{dN}),\, \|\psi\|_2=1} \langle \psi | \mathcal{H}_{z_k} | \psi \rangle_{L^2(\mathbb{R}^{dN})} &\le \sum_{k=1}^{K} \lambda_k \langle \psi_{\delta,z_k} | \mathcal{H}_{z_k} | \psi_{\delta,z_k} \rangle_{L^2(\mathbb{R}^{dN})} \\ &< \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu] + \delta. \end{aligned} \tag{2-13}$$

Since the left-hand side is a convex combination and δ is arbitrary, we immediately deduce that

$$\inf_{z\in\mathfrak{h}_\omega} \mathcal{I}[z] \le \inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu]. \tag{2-14}$$

On the other hand, since a measure concentrated in a single point is atomic,

$$\inf_{(\psi,\mu)\in\mathscr{D}_{\mathrm{pm}},\, \mu\in\mathscr{P}_{\mathrm{atom}}(\mathfrak{h}_\omega)} \mathcal{E}_{\mathrm{pm}}[\psi,\mu] \le \inf_{z\in\mathfrak{h}_\omega} \inf_{\psi\in L^2(\mathbb{R}^{dN}),\, \|\psi\|_2=1} \mathcal{E}_{\mathrm{qc}}[\psi,z] = \inf_{z\in\mathfrak{h}_\omega} \mathcal{I}[z],$$

which implies the first identity in (2-12).

Now, let us prove the second equality above, namely

$$\inf_{z\in\mathfrak{h}_\omega} \mathcal{I}[z] = E_{\mathrm{Pekar}}. \tag{2-15}$$

Let again $\delta > 0$ and let z_δ be a minimizing family of vectors for $\mathcal{I}$, i.e., such that $\mathcal{I}[z_\delta] < \inf_{z\in\mathfrak{h}_\omega} \mathcal{I}[z] + \delta$. For each z_δ, let ψ_{δ,z_δ} be a minimizing vector for $\mathcal{E}_{\mathrm{qc}}[\,\cdot\,, z_\delta]$, i.e., such that

$$\mathcal{E}_{\mathrm{qc}}[\psi_{\delta,z_\delta}, z_\delta] < \mathcal{I}[z_\delta] + \delta.$$

Now,

$$E_{\mathrm{Pekar}} \le \mathcal{E}_{\mathrm{Pekar}}[\psi_{\delta,z_\delta}] \le \mathcal{E}_{\mathrm{qc}}[\psi_{\delta,z_\delta}, z_\delta],$$

therefore

$$E_{\mathrm{Pekar}} \le \inf_{z\in\mathfrak{h}_\omega} \mathcal{I}[z]. \tag{2-16}$$

On the other hand, let ψ_δ be a minimizing family of states for E_{Pekar}, and, after fixing ψ_δ, let z_{δ,ψ_δ} be a minimizing family for $\mathcal{E}_{\mathrm{qc}}[\psi_\delta, \cdot\,]$:

$$\mathcal{E}_{\mathrm{qc}}[\psi_\delta, z_{\delta,\psi_\delta}] < E_{\mathrm{Pekar}} + \delta. \tag{2-17}$$

As above, we then get

$$\inf_{z\in\mathfrak{h}_\omega} \mathcal{I}[z] \le \inf_{\psi\in L^2(\mathbb{R}^{dN}),\|\psi\|_2=1} \mathcal{E}_{\mathrm{qc}}[\psi, z_{\delta,\psi_\delta}] \le \mathcal{E}_{\mathrm{qc}}[\psi_\delta, z_{\delta,\psi_\delta}] < E_{\mathrm{Pekar}} + \delta,$$

which yields

$$\inf_{z\in\mathfrak{h}_\omega} \mathcal{I}[z] \le E_{\mathrm{Pekar}}. \tag{2-18}$$

Finally, we prove that

$$E_{\mathrm{Pekar}} = E_{\mathrm{qc}}. \tag{2-19}$$

Now, let $(\psi_\delta, z_{\delta,\psi_\delta})$ be as above, i.e., such that (2-17) holds true. Hence,

$$E_{\mathrm{qc}} \le \mathcal{E}_{\mathrm{qc}}[\psi_\delta, z_{\delta,\psi_\delta}] < E_{\mathrm{Pekar}} + \delta,$$

and thus $E_{\mathrm{qc}} \le E_{\mathrm{Pekar}}$. On the other hand, let (ψ_δ, z_δ) be a minimizing family of configurations for $\mathcal{E}_{\mathrm{qc}}$:

$$\mathcal{E}_{\mathrm{qc}}[\psi_\delta, z_\delta] < E_{\mathrm{qc}} + \delta.$$

Clearly, now one has

$$E_{\mathrm{Pekar}} \le \mathcal{E}_{\mathrm{Pekar}}[\psi_\delta] \le \mathcal{E}_{\mathrm{qc}}[\psi_\delta, z_\delta] < E_{\mathrm{qc}} + \delta,$$

yielding the opposite inequality, i.e., $E_{\mathrm{Pekar}} \le E_{\mathrm{qc}}$. □

Remark 2.2 (stability). In the above proof we have implicitly assumed that the energies under considerations are bounded from below, but in fact it is easy to see that, if one of the functionals is unbounded from below, then all the others must be unstable as well. We do not provide any detail of such an argument, because our main result (Theorem 1.3) implies that (VP1) holds true, so that (VP′1), (vp1) and (vp′1) immediately follow.

The other important result concerns equivalences for the existence of minimizers in the variational problems above.

Proposition 2.3 (quasiclassical minimizers). *Under assumptions* (A1), (A2) *and* (A3),

$$\text{(VP2)} \iff \text{(VP}'\text{2)} \iff \text{(vp2)} \iff \text{(vp}'\text{2)}. \tag{2-20}$$

Furthermore, any minimizer $\mathfrak{m}_{\mathrm{svm}}$ *of* (vp2) *is concentrated on the set of minimizers* $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}})$ *of* (VP2).

Proof. Some implications are easy to prove. Let us first prove that (VP2) $\Longrightarrow$ (vp′2). Let $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}})$ be a minimizer of $\mathcal{E}_{\mathrm{qc}}$ in $\mathscr{D}_{\mathrm{qc}}$. Then, evaluating the energy $\mathcal{E}_{\mathrm{pm}}$ on the configuration $(\psi_{\mathrm{qc}}, \mu_0)$, with $\mu_0 = \delta_{z_{\mathrm{qc}}}$, we get

$$\mathcal{E}_{\mathrm{pm}}[\psi_{\mathrm{qc}}, \mu_0] = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_0(z)\mathcal{E}_{\mathrm{qc}}[\psi_{\mathrm{qc}}, z] = \mathcal{E}_{\mathrm{qc}}[\psi_{\mathrm{qc}}, z_{\mathrm{qc}}] = E_{\mathrm{qc}}.$$

By Proposition 2.1, $(\psi_{\mathrm{qc}}, \mu_0)$ thus solves (vp′2). Analogously, let us prove (vp′2) $\Longrightarrow$ (vp2): let $(\psi_{\mathrm{pm}}, \mu_{\mathrm{pm}})$ be a minimizer for (vp′2); then, the state-valued measure $\mathfrak{m}_0$, with $\mu_{\mathfrak{m}_0} = \mu_{\mathrm{pm}}$ and $\gamma_{\mathfrak{m}_0}(z) = |\psi_{\mathrm{pm}}\rangle\langle\psi_{\mathrm{pm}}|$, solves (VP2) by Proposition 2.1.

We prove now that (vp2) $\Longrightarrow$ (VP2). Given a minimizer $\mathfrak{m}_{\mathrm{svm}}$ of $\mathcal{E}_{\mathrm{svm}}$, for $\mu_{\mathfrak{m}_{\mathrm{svm}}}$-a.e. $z \in \mathfrak{h}_\omega$ there exists $\{\lambda_k(z)\}_{k\in\mathbb{N}}$, with $\lambda_k(z) \geq 0$ and $\sum_{k\in\mathbb{N}} \lambda_k(z) = 1$, and $\{\psi_k(z)\}_{k\in\mathbb{N}}$, with $\|\psi_k(z)\|_{L^2(\mathbb{R}^{dN})} = 1$, such that

$$E_{\mathrm{svm}} = \mathcal{E}_{\mathrm{svm}}[\mathfrak{m}_{\mathrm{svm}}] = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}_{\mathrm{svm}}}(z) \sum_{k\in\mathbb{N}} \lambda_k(z) \mathcal{E}_{\mathrm{qc}}[\psi_k(z), z].$$

The above is due to the fact that $\gamma_{\mathfrak{m}_{\mathrm{svm}}}(z)$ is a density matrix on $L^2(\mathbb{R}^{dN})$ for $\mu_{\mathfrak{m}_{\mathrm{svm}}}$-a.e. z. The measure $\mu_{\mathfrak{m}_{\mathrm{svm}}} \in \mathscr{P}(\mathfrak{h}_\omega)$ is a probability measure, hence the right-hand side of the above equation is a (double) convex combination of numerical values of the real-valued function $\mathcal{E}_{\mathrm{qc}}$. However, a convex combination of values of a function equals its infimum, if and only if the infimum is a minimum, and all variables appearing in the convex combination are minimizers. Therefore, $\mathcal{E}_{\mathrm{qc}}$ admits at least one minimizer. Actually, the measure $\mathfrak{m}_{\mathrm{svm}}$ is concentrated on the set of minimizers $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}})$, in the above sense.

Finally, we consider the Pekar-like variational problem (VP′2) and its equivalence with (VP2). Let us first prove that (VP′2) $\Longrightarrow$ (VP2): given a Pekar minimizer $\psi_{\mathrm{Pekar}} \in L^2(\mathbb{R}^{dN})$, we immediately deduce that $\psi_{\mathrm{Pekar}} \in H^1(\mathbb{R}^{dN})$ by boundedness from above of the energy and regularity of the classical field $\boldsymbol{a}(\boldsymbol{x})$, which is continuous and vanishing at infinity [Correggi et al. 2019, Remark 1.5]. Furthermore, Lemma 2.4 guarantees the existence (and uniqueness) of $\eta_{\mathrm{Pekar}}[\psi_{\mathrm{Pekar}}] \in \mathfrak{h}$ minimizing $\mathcal{E}_{\mathrm{qc}}[\psi_{\mathrm{Pekar}}, z]$ with respect to z. Therefore, the configuration $(\psi_{\mathrm{Pekar}}, \eta_{\mathrm{Pekar}}[\psi_{\mathrm{Pekar}}])$ is admissible for $\mathcal{E}_{\mathrm{qc}}$ and we deduce from Proposition 2.1 that $\mathcal{E}_{\mathrm{qc}}[\psi_{\mathrm{Pekar}}, \eta_{\mathrm{Pekar}}[\psi_{\mathrm{Pekar}}]] = E_{\mathrm{qc}}$.

Conversely, given a minimizer $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}}) \in \mathscr{D}$ of $\mathcal{E}_{\mathrm{qc}}$, we know that the configuration must satisfy the Euler–Lagrange equations (1-23) at least in the weak sense. However, the second equation in (1-23) is easily seen to coincide with (1-27) or the first equation in (1-34), when the change of variable $\eta = \omega^{1/2} z$ has been performed. Furthermore, any weak solution η of such equations is in fact a strong solution, i.e., $\eta \in \mathfrak{h}$, under the assumptions made. Hence, by strict convexity of $\mathcal{F}_{\mathrm{qc}}[\psi, \eta]$ in η proven in Lemma 2.4 and then uniqueness of η_{Pekar}, we deduce that $\eta_{\mathrm{Pekar}}[\psi_{\mathrm{qc}}] = \omega^{1/2} z_{\mathrm{qc}}$, and the equivalence (VP2) $\Longrightarrow$ (VP′2) is readily proven via Proposition 2.1. □

The next result about the quasiclassical functional defined in (1-18) or, more precisely, about its variant $\mathcal{F}_{\mathrm{qc}}$ introduced in (1-25) is important to explore the connection with the Pekar-like functionals (1-28) and (1-34).

Lemma 2.4. *For any fixed ψ, the functional $\mathcal{F}_{\mathrm{qc}}[\psi, \eta]$ is strictly convex in $\eta \in \mathfrak{h}_\omega$.*

Proof. In cases (a) and (b) the proof is trivial, since $\mathcal{F}_{\mathrm{qc}}$ contains only two terms depending on η: one is quadratic in η (the free field energy) and therefore strictly convex, while the other (the interaction) is linear and thus convex.

So we have to investigate in detail only case (c), namely the Pauli–Fierz quasiclassical energy, and, specifically, only the kinetic part of the energy involving the interaction, which reads

$$\sum_{j=1}^{N} \frac{1}{2m_j} (-i\nabla_j + 2\mathrm{Re}\langle \eta | (\omega^{-1/2} \boldsymbol{\lambda}_j)(\boldsymbol{x}_j)\rangle_{\mathfrak{h}})^2.$$

Let us then set $\eta = \beta\eta_1 + (1-\beta)\eta_2$ for some $\eta_1, \eta_2 \in \mathfrak{h}$ and $\beta \in (0,1)$. Expanding the square and setting $\boldsymbol{\xi}_j(\boldsymbol{x}) := \omega^{-1/2}\lambda_j(\boldsymbol{x})$ for short, we get (for any nonzero ψ)

$$\begin{aligned}\big\langle\psi\big|(-i\nabla_j + 2\mathrm{Re}\langle\eta|\boldsymbol{\xi}_j(\boldsymbol{x}_j)\rangle_{\mathfrak{h}})^2\big|\psi\big\rangle_{L^2(\mathbb{R}^{3N})}&\\ < \langle\psi| - \Delta_j|\psi\rangle_{L^2(\mathbb{R}^{3N})} - 2\big\langle\psi\big|i\beta\mathrm{Re}\langle\eta_1|\boldsymbol{\xi}_j(\boldsymbol{x}_j)\rangle_{\mathfrak{h}}\cdot\nabla_j + i(1-\beta)\mathrm{Re}\langle\eta_2|\boldsymbol{\xi}_j(\boldsymbol{x}_j)\rangle_{\mathfrak{h}}\cdot\nabla_j\big|\psi\big\rangle_{L^2(\mathbb{R}^{3N})}&\\ + 4\big\langle\psi\big|\beta(\mathrm{Re}\langle\eta_1|\boldsymbol{\xi}_j(\boldsymbol{x}_j)\rangle_{\mathfrak{h}})^2 + (1-\beta)(\mathrm{Re}\langle\eta_2|\boldsymbol{\xi}_j(\boldsymbol{x}_j)\rangle_{\mathfrak{h}})^2\big|\psi\big\rangle_{L^2(\mathbb{R}^{3N})},& \quad (2\text{-}21)\end{aligned}$$

again by the strict convexity of the square, i.e., the bound $(\beta a + (1-\beta)b)^2 < \beta a^2 + (1-\beta)b^2$, valid for any $a, b \in \mathbb{R}$ and $\beta \in (0,1)$. The result easily follows, since the remaining term in the functional depending on η is the free field energy, which is quadratic in η and thus strictly convex as well. □

Remark 2.5 (minimizers for (vp′2)). The existence of a solution for (vp′2) obtained here is trivial, i.e., it involves a measure concentrated in a single point $z_{\mathrm{qc}} \in \mathfrak{h}_\omega$ and a $\psi_{z_{\mathrm{qc}}}$ dependent on such a point. It would be interesting, but outside the scope of this paper, to know whether there are nontrivial minimizers in which μ_0 is not concentrated at a single point. This is obviously related to the question of uniqueness of the minimizing configuration $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}})$. Note that this would not be in contradiction with Lemma 2.4, since we prove there strict convexity of $\mathcal{F}_{\mathrm{qc}}[\psi, \eta]$ only in η, while the full functional $\mathcal{E}_{\mathrm{qc}}[\psi, z]$ is in general not jointly convex in ψ and z nor in $|\psi|^2$ and z (see also Remark 1.2).

Proof of Proposition 1.1. Combining Proposition 2.1 with Proposition 2.3 one obtains the equivalence of the variational problems. □

2B. ***Minimization problem for generalized state-valued measures.*** We discuss now the generalization of the concepts introduced above needed to deal with the minimization (1-52), which is particularly useful to treat small systems consisting of unconfined particles. Taking the double dual, it is well known that $\mathscr{L}^1(L^2(\mathbb{R}^{dN}))$ can be continuously embedded in $\mathscr{B}(L^2(\mathbb{R}^{dN}))'$, the dual of bounded operators, in a positivity preserving way. By an abuse of notation, we will write $\mathscr{L}^1(L^2(\mathbb{R}^{dN})) \subset \mathscr{B}(L^2(\mathbb{R}^{dN}))'$. We recall that we denoted by $\overline{\mathscr{L}^1}(L^2(\mathbb{R}^{dN}))$ the closure of $\mathscr{L}^1(L^2(\mathbb{R}^{dN}))$ with respect to the weak* topology $\sigma(\mathscr{B}(L^2(\mathbb{R}^{dN}))', \mathscr{B}(L^2(\mathbb{R}^{dN})))$ on $\mathscr{B}(L^2(\mathbb{R}^{dN}))'$. Also, $\overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN}))$ and $\overline{\mathscr{L}^1}_{+,1}(L^2(\mathbb{R}^{dN}))$ stand for the subsets of positive and normalized positive elements, respectively. A generalized state-valued measure is then a measure on $\mathfrak{h}_\omega$ with values in the space of generalized states $\overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN}))$. Properties of generalized state-valued measures are discussed in the Appendix. Since the dual space $\mathscr{B}(L^2(\mathbb{R}^{dN}))'$ is not separable, it does not have the Radon–Nikodým property, therefore integration of functions $\mathcal{F} : \mathfrak{h}_\omega \to \mathscr{B}(L^2(\mathbb{R}^{dN}))$ is restricted only to those with separable range.

Such integration can be extended to functions valued in unbounded operators in the following sense.

Definition 2.6 (domains of generalized Wigner measures). Let $\mathcal{T}$ be a strictly positive unbounded operator on $L^2(\mathbb{R}^{dN})$. A generalized state-valued measure $\mathfrak{n}$ is *in the domain of* $\mathcal{T}$ if and only if there exists a measure $\mathfrak{n}_{\mathcal{T}} \in \mathscr{P}(\mathfrak{h}_\omega, \overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN})))$ such that for all $\mathcal{B} \in \mathscr{B}(L^2(\mathbb{R}^{dN}))$ and all Borel sets $S \subseteq \mathfrak{h}_\omega$,

$$\mathfrak{n}_{\mathcal{T}}(S)[\mathcal{T}^{-1/2}\mathcal{B}\mathcal{T}^{-1/2}] = \mathfrak{n}(S)[\mathcal{B}]. \quad (2\text{-}22)$$

Therefore, if $\mathfrak{n}$ is in the domain of $\mathcal{T}$, with a little abuse of notation we may write

$$\mathfrak{n}(S)[\mathcal{T}^{1/2}\cdot\mathcal{T}^{1/2}]=\mathfrak{n}_{\mathcal{T}}(S)[\,\cdot\,] \tag{2-23}$$

as a state-valued measure "absorbing a singularity" of order $\mathcal{T}$. Now, let $\mathcal{F}(z)$ be a function with values in unbounded operators such that for all $z\in\mathfrak{h}_\omega$,

- $\mathcal{T}^{-1/2}\mathcal{F}(z)\mathcal{T}^{-1/2}\in\mathscr{B}(L^2(\mathbb{R}^{dN}))$,
- the range of $z\mapsto\mathcal{T}^{-1/2}\mathcal{F}(z)\mathcal{T}^{-1/2}$ is separable,
- $\mathcal{T}^{-1/2}\mathcal{F}(z)\mathcal{T}^{-1/2}$ is $\mathfrak{n}_{\mathcal{T}}$-absolutely integrable.

Then, it follows that we can define the integral of $\mathcal{F}$ with respect to $\mathfrak{n}$ as

$$\int_{\mathfrak{h}_\omega}\mathrm{d}\mathfrak{n}(z)[\mathcal{F}(z)]:=\int_{\mathfrak{h}_\omega}\mathrm{d}\mathfrak{n}_{\mathcal{T}}(z)[\mathcal{T}^{-1/2}\mathcal{F}(z)\mathcal{T}^{-1/2}]. \tag{2-24}$$

A simple but useful example of such $\mathcal{F}(z)$ is the following: let $\mathcal{S}$ be a self-adjoint operator, and let $\mathfrak{n}$ be in the domain of $\mathcal{T}=|\mathcal{S}|+1$; then the function $\mathcal{F}(z)=\mathcal{S}$ satisfies all above hypotheses and thus it makes sense to write, for all Borel set $S\subseteq\mathfrak{h}_\omega$,

$$\int_S\mathrm{d}\mathfrak{n}(z)[\mathcal{S}]=\mathfrak{n}(S)[\mathcal{S}]:=\mathfrak{n}_{\mathcal{T}}(S)[\mathcal{T}^{-1/2}\mathcal{S}\mathcal{T}^{-1/2}]\in\mathbb{R}. \tag{2-25}$$

The other cases useful for our analysis are discussed in Section 3.

We are now in a position to define another quasiclassical minimization problem. Recall the definition of the domain $\mathscr{D}_{\mathrm{gqc}}$ (1-53), the ground state energy E_{gqc} given by (1-54) and any corresponding minimizing configuration $(\rho_{\mathrm{gqc}},z_{\mathrm{gqc}})\in\mathscr{D}_{\mathrm{gqc}}$; then the analogues of (VP1) and (VP2) are (stability)

$$\text{“Is } E_{\mathrm{gqc}} \text{ greater than } -\infty\text{?”} \tag{GVP1}$$

and (existence of a ground state)

$$\text{“Does there exist } (\rho_{\mathrm{gqc}},z_{\mathrm{gqc}})\in\mathscr{D}_{\mathrm{gqc}} \text{ such that } \mathcal{E}_{\mathrm{gqc}}(\rho_{\mathrm{gqc}},z_{\mathrm{gqc}})=E_{\mathrm{gqc}}\text{?”} \tag{GVP2}$$

The functional $\mathcal{E}_{\mathrm{gqc}}$ can indeed be seen as the generalized quasiclassical energy: let H_z be the abstract realization of $\mathcal{H}_z$ as an operator affiliated to the abstract C^*-algebra $\mathscr{B}(L^2(\mathbb{R}^{dN}))$. Then, given a normalized pure state $\rho\in\overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN}))$, we define the corresponding irreducible GNS representation by $(\mathscr{K}_\rho,\pi_\rho,\psi_\rho)$, where $\mathscr{K}_\rho$ is a suitable Hilbert space, $\pi_\varrho:\mathscr{B}(L^2(\mathbb{R}^{dN}))\to\mathscr{B}(\mathscr{K}_\rho)$ is a C^*-homomorphism (that can be extended to operators affiliated to the algebra) and $\psi_\rho\in\mathscr{K}_\varrho$ is the normalized cyclic vector associated to ρ. Therefore, it follows that

$$\mathcal{E}_{\mathrm{gqc}}[\rho,z]=\langle\psi_\rho|\pi_\rho(H_z)|\psi_\rho\rangle_{\mathscr{K}_\rho}.$$

This expression is analogous to the one for $\mathcal{E}_{\mathrm{qc}}$ (see (1-18)) and it reduces exactly to the latter whenever ρ is a pure state belonging to $\mathscr{L}^1(L^2(\mathbb{R}^{dN}))$ (see Remark 2.7).

The generalization of the variational problems for state-valued measures (vp1) and (vp2) is obtained as follows: setting

$$\mathscr{D}_{\mathrm{gsvm}}:=\left\{\mathfrak{n}\in\overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN})):\|\mathfrak{n}(\mathfrak{h}_\omega)\|_{\mathscr{B}'}=1,\ \left|\int_{\mathfrak{h}_\omega}\mathrm{d}\mathfrak{n}(z)[\mathcal{H}_z]\right|<+\infty\right\}, \tag{2-26}$$

we consider the questions (stability)

$$\text{“Is } E_{\mathrm{gsvm}} := \inf_{\mathfrak{n}\in\mathscr{D}_{\mathrm{gsvm}}} \int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z)[\mathcal{H}_z] \text{ greater than } -\infty\text{?”} \tag{gvp1}$$

and (existence of a ground state)

$$\text{“Does there exists } \mathfrak{n}_{\mathrm{gsvm}} \in \mathscr{D}_{\mathrm{gsvm}} \text{ such that } \int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}_{\mathrm{gsvm}}(z)[\mathcal{H}_z] = E_{\mathrm{gsvm}}\text{?”} \tag{gvp2}$$

Remark 2.7 (state-valued and generalized state-valued measures). We point out that, if a generalized state-valued measure $\mathfrak{n} \in \mathscr{D}_{\mathrm{gsvm}}$ is actually a state-valued measure, i.e., such that, for all Borel sets $S \subseteq \mathfrak{h}_\omega$,

$$\mathfrak{n}(S) \in \mathscr{L}^1_+(L^2(\mathbb{R}^{dN})),$$

then $\mathfrak{n} \in \mathscr{D}_{\mathrm{svm}}$ and

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z)[\mathcal{H}_z] = \mathcal{E}_{\mathrm{svm}}[\mathfrak{n}].$$

Proposition 2.8 (generalized quasiclassical ground state energy). *Under assumptions* (A1), (A2) *and* (A3),

$$E_{\mathrm{qc}} = E_{\mathrm{gqc}} = E_{\mathrm{gsvm}}. \tag{2-27}$$

Proof. Firstly, let us prove that

$$E_{\mathrm{qc}} = E_{\mathrm{gqc}}.$$

Since ρ belongs to the weak* closure of $\mathscr{L}^1_{+,1}(L^2(\mathbb{R}^{dN}))$, there exists a filter base $\mathfrak{S} \subset 2^{\mathscr{L}^1_{+,1}(L^2(\mathbb{R}^{dN}))}$ such that $\mathfrak{S} \to \rho$ in the weak* topology. Hence, for any fixed $z \in \mathfrak{h}_\omega$,[4]

$$\lim_{\mathfrak{S}\to\rho} \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\mathfrak{S}(\mathcal{H}_z)] = \rho[\mathcal{H}_z].$$

Now, on one hand, each $|\psi\rangle\langle\psi|$, $\psi \in L^2(\mathbb{R}^{dN})$, is also a pure generalized state and therefore

$$E_{\mathrm{gqc}} \le \inf_{(\psi,z)\in\mathscr{D}_{\mathrm{qc}}} \mathcal{E}_{\mathrm{qc}}[\psi, z] = E_{\mathrm{qc}}. \tag{2-28}$$

On the other hand, let $(\rho_\delta, z_\delta) \in \mathscr{D}_{\mathrm{gqc}}$ be a minimizing sequence:

$$\mathcal{E}_{\mathrm{gqc}}[\rho_\delta, z_\delta] = \rho_\delta[\mathcal{H}_{z_\delta}] < E_{\mathrm{gqc}} + \delta,$$

for some $\delta > 0$ and let $\mathfrak{S}_\delta \subset 2^{\mathscr{L}^1_{+,1}(L^2(\mathbb{R}^{dN}))}$ be the corresponding approximating filter base for ρ_δ. Then,

$$\begin{aligned} E_{\mathrm{qc}} &= \inf_{(\psi,z)\in\mathscr{D}_{\mathrm{qc}}} \mathcal{E}_{\mathrm{qc}}[\psi, z] = \inf_{(\gamma,z)\in\mathscr{L}^1_{+,1}(L^2(\mathbb{R}^{dN}))\oplus\mathfrak{h}_\omega} \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma\mathcal{H}_z] \le \sup_{X\in\mathfrak{S}_\delta} \inf_{\gamma\in X} \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma\mathcal{H}_{z_\delta}] \\ &= \liminf_{\mathfrak{S}_\delta} \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\mathfrak{S}_\delta(\mathcal{H}_{z_\delta})] = \lim_{\mathfrak{S}_\delta\to\rho_\delta} \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\mathfrak{S}_\delta(\mathcal{H}_{z_\delta})] = \rho_\delta[\mathcal{H}_{z_\delta}] < E_{\mathrm{gqc}} + \delta. \end{aligned} \tag{2-29}$$

[4] The notation $\mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\mathfrak{S}(\mathcal{H}_z)]$ stands for the filter base that is the image of $\mathfrak{S}$ on $\mathbb{R}$ via the map $\gamma \mapsto \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma\mathcal{H}_z]$; given any $X \in \mathfrak{S}$, we have that $\{\mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma\mathcal{H}_z], \gamma \in X\} \in \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\mathfrak{S}(\mathcal{H}_z)]$.

Since the above chain of inequalities is valid for all $\delta > 0$, it follows that the opposite inequality of (2-28) holds true, i.e.,

$$E_{\mathrm{qc}} \leq E_{\mathrm{gqc}}, \tag{2-30}$$

which implies the claim.

The proof of the identity $E_{\mathrm{gsvm}} = E_{\mathrm{qc}}$ is perfectly analogous, where we remind the reader that it is possible to approximate any measure $\mathfrak{n} \in \mathscr{P}(\mathfrak{h}_\omega, \overline{\mathscr{L}^1}_+(L^2(\mathbb{R}^{dN})))$ with a filter base $\mathfrak{T} \subset 2^{\mathscr{P}(\mathfrak{h}_\omega, \mathscr{L}^1_+(L^2(\mathbb{R}^{dN})))}$ with respect to the product of weak* topologies

$$\prod_{S \subset \mathfrak{h}_\omega \text{ Borel}} \sigma(\mathcal{B}(L^2(\mathbb{R}^{dN}))', \mathcal{B}(L^2(\mathbb{R}^{dN}))),$$

which implies the convergence of integrals[5]

$$\lim_{\mathfrak{T} \to \mathfrak{n}} \mathrm{tr}_{L^2(\mathbb{R}^{dN})}\left[\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{T}(z)\mathcal{H}_z\right] = \int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z)[\mathcal{H}_z]. \qquad \square$$

Finally, also for the generalized minimization problems, it is possible to prove equivalence of existence of minimizers.

Proposition 2.9 (generalized quasiclassical minimizers). *Under assumptions* (A1), (A2) *and* (A3),

$$\text{(GVP2)} \iff \text{(gvp2)}. \tag{2-31}$$

Furthermore, any minimizer $\mathfrak{n}_{\mathrm{gsvm}}$ *of* (gvp2) *is concentrated on the set of minimizers* $(\rho_{\mathrm{gqc}}, z_{\mathrm{gqc}})$ *of* (GVP2).

Proof. The forward implication is trivial: let $(\rho_{\mathrm{gqc}}, z_{\mathrm{gqc}})$ be a minimizer for (GVP2). Then, evaluating the energy of the generalized state-valued measure $\mathfrak{n}_0 = \delta_{z_{\mathrm{gqc}}} \rho_{\mathrm{gqc}}$, we get

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}_0(z)[\mathcal{H}_z] = \rho_{\mathrm{gqc}}[\mathcal{H}_{z_{\mathrm{gqc}}}] = E_{\mathrm{gqc}}. \tag{2-32}$$

By Proposition 2.8, $\mathfrak{n}_0$ is thus a minimizer for (gvp2).

To prove the reverse implication, note that the integral with respect to a generalized state-valued probability measure is a convex combination of expectations over possibly mixed generalized states. Since the mixed states are themselves convex combinations of pure states, it follows that the measure $\mathfrak{n}_{\mathrm{gsvm}}$ must be concentrated on the set of minimizers for (gvp2), and thus the latter is not empty. □

3. Ground state energies and ground states in the quasiclassical regime

In this section we study the quasiclassical limit of ground state energies and ground states of the microscopic models introduced in Section 1.

The microscopic interaction is described by a fully quantum system, in which both the small system and the environment are quantum. The Hilbert space is thus (see (1-2)) given by $\mathscr{H}_\varepsilon = L^2(\mathbb{R}^{dN}) \otimes \mathcal{G}_\epsilon(\mathfrak{h})$, where $\mathcal{G}_\epsilon(\mathfrak{h}) = \bigoplus_{n \in \mathbb{N}} \mathfrak{h}^{\otimes_s n}$ is the symmetric Fock space over $\mathfrak{h}$ and ε is the quasiclassical parameter whose dependence is given by a semiclassical choice of canonical commutation relations (1-3), i.e.,

[5] As before, the integral with respect to $\mathrm{d}\mathfrak{T}$ is just a short-hand notation to denote the integral over elements belonging to the filter $\mathfrak{T}$.

$[a_\varepsilon(z), a_\varepsilon^\dagger(w)] = \varepsilon\langle z|w\rangle_{\mathfrak{h}}$, with $a_\varepsilon^\sharp$ the annihilation and creation operators on the Fock space. A state of the whole system is given by a density matrix

$$\Gamma_\varepsilon \in \mathscr{L}^1_{+,1}(L^2(\mathbb{R}^{dN}) \otimes \mathcal{G}_\varepsilon(\mathfrak{h})),$$

the positive trace-class operators with unit trace.

The dynamics of the system is described by a self-adjoint Hamiltonian operator H_ε whose general form is given in (1-4). Such an operator is the partial Wick quantization of the quasiclassical Schrödinger energy operator $\mathcal{H}_z$ provided in (1-14). Wick quantization consists in substituting each z appearing in $\mathcal{H}$ with a_ε and each $\bar{z}$ with $a_\varepsilon^\dagger$, and of ordering all the $a_\varepsilon^\dagger$ to the left of all the a_ε. Such a quantization procedure is well-defined for symbols $\mathcal{F}_z$ that are polynomial in z and z^*, as is the case in the concrete models we are considering; see Section 4 for additional details and [Ammari and Nier 2008] for the rigorous procedure. Hence, we write

$$H_\varepsilon = \mathrm{Op}_\varepsilon^{\mathrm{Wick}}(\mathcal{H}_z), \tag{3-1}$$

and, more precisely, $\mathcal{H}_z$ can be split into three terms, at least in the sense of quadratic forms, i.e.,

$$\mathcal{H}_z = \mathcal{K}_0 + \sum_{i=1}^{N} \mathcal{V}_z(\boldsymbol{x}_i) + \langle z|\omega|z\rangle_{\mathfrak{h}} \tag{3-2}$$

with $\mathcal{K}_0$ self-adjoint and bounded from below, yielding

$$H_\varepsilon = \mathcal{K}_0 \otimes 1 + 1 \otimes \mathrm{Op}_\varepsilon^{\mathrm{Wick}}(\langle z|\omega|z\rangle_{\mathfrak{h}}) + \sum_{i=1}^{N} \mathrm{Op}_\varepsilon^{\mathrm{Wick}}(\mathcal{V}_z(\boldsymbol{x}_i)) \tag{3-3}$$

as a quadratic form. The first and second terms on the right-hand side are the free energies of the small system and environment, respectively, and the third term is the *small system-environment* interaction.

The minimization problem for the quantum system described by H_ε is defined in (1-12): the microscopic ground state energy is $E_\varepsilon := \inf \sigma(H_\varepsilon)$, while $\Psi_{\varepsilon,\mathrm{gs}}$ stands for any corresponding minimizer. Such a minimization problem has been thoroughly studied for the concrete models under consideration in this paper; see Section 1A. A crucial ingredient of our proof is the uniform boundedness from below of the spectrum of H_ε. Note again that we do not need the existence of a microscopic ground state $\Psi_{\varepsilon,\mathrm{gs}}$.

Proposition 3.1 (stability and existence of the ground state). *Under assumptions* (A1), (A2) *and* (A3), *there exist finite constants* $c, C > 0$ *independent of* ε *such that*

$$-c \le E_\varepsilon \le C. \tag{3-4}$$

The proof of the above result is model-dependent and therefore it is postponed to Section 4.

We now investigate the link between the microscopic ground state problem and the quasiclassical minimization problems described in Section 2, starting from the proof of Theorem 1.3. The strategy of proof can be outlined as follows:

- Derive an energy upper bound (Proposition 3.2) by means of a suitable trial state.
- Prove a matching lower bound (Proposition 3.3) by showing the convergence of the expectation of each term in the energy over a suitable minimizing sequence.

Although both cases could be treated at once, we provide a separate discussion of the main results for trapped and nontrapped particle systems, whose difference is apparent in the statements of Corollary 1.10 and Corollary 1.17. The convergence of minimizing sequences and ground states (Theorem 1.7), if present, is then obtained as a direct consequence of the above arguments.

3A. ***Proof of Theorem 1.3.*** The proof of Theorem 1.3 is obtained by putting together the energy upper bound (Proposition 3.2) and lower bound (Proposition 3.3).

In the following, we denote by $\Psi_{\varepsilon,\delta} \in \mathscr{D}(H_\varepsilon)$, $\delta > 0$, a minimizing sequence for H_ε:

$$\langle \Psi_{\varepsilon,\delta} | H_\varepsilon | \Psi_{\varepsilon,\delta} \rangle_{\mathscr{H}_\varepsilon} < E_\varepsilon + \delta. \tag{3-5}$$

Proposition 3.2 (energy upper bound). *Under assumptions* (A1), (A2) *and* (A3),

$$\limsup_{\varepsilon \to 0} E_\varepsilon \leq E_{\mathrm{qc}}. \tag{3-6}$$

Proof. In order to prove the upper bound we use a coherent trial state: let us denote by $\Omega_\varepsilon \in \mathcal{G}_\varepsilon(\mathfrak{h})$ the Fock vacuum and let

$$\Xi_\varepsilon[\psi, z] := \psi \otimes W_\varepsilon\Big(\frac{z}{i\varepsilon}\Big)\Omega_\varepsilon \tag{3-7}$$

be a coherent product state constructed over the particle state ψ and the classical configuration $z \in \mathfrak{h}$. We shall restrict to $\psi \in \mathscr{Q}(\mathcal{K}_0)$, where $\mathscr{Q}(\mathcal{K}_0)$ is the form domain of $\mathcal{K}_0$, and $z \in \mathfrak{h}$ such that $\omega^{1/2} z \in \mathfrak{h}$. As discussed in Section 4, this is sufficient to make $\Xi_\varepsilon[\psi, z] \in \mathscr{Q}(H_\varepsilon)$ and $(\psi, z) \in \mathscr{D}_{\mathrm{qc}}$. The energy of the above trial state is

$$\langle \Xi_\varepsilon[\psi, z] | H_\varepsilon | \Xi_\varepsilon[\psi, z] \rangle_{\mathscr{H}_\varepsilon} = \mathcal{E}_{\mathrm{qc}}[\psi, z] + o_\varepsilon(1). \tag{3-8}$$

The proof of the above estimate depends on the microscopic model involved. The computation of the expectation over the trial states (3-7) can be found in [Correggi and Falconi 2018, Proposition 3.11 and Section 3.6] for the Nelson and polaron models, and in [Correggi et al. 2019, Proof of Theorem 1.9] for the Pauli–Fierz model. Hence, we have that

$$E_\varepsilon \leq \inf_{(\psi,z)\in\mathscr{D}_{\mathrm{qc}}} \langle \Xi_\varepsilon[\psi, z] | H_\varepsilon | \Xi_\varepsilon[\psi, z] \rangle_{\mathscr{H}_\varepsilon} = \inf_{(\psi,z)\in\mathscr{D}_{\mathrm{qc}}} \mathcal{E}_{\mathrm{qc}}[\psi, z] + o_\varepsilon(1) = E_{\mathrm{qc}} + o_\varepsilon(1). \tag{3-9}$$

The result is then obtained by taking the $\limsup_{\varepsilon\to 0}$ on both sides. □

The symmetric result of Proposition 3.2 is stated in the following proposition.

Proposition 3.3 (energy lower bound). *Under assumptions* (A1), (A2) *and* (A3),

$$\liminf_{\varepsilon \to 0} E_\varepsilon \geq E_{\mathrm{qc}}. \tag{3-10}$$

Although not necessary in principle, we find it convenient to present two different proofs of (3-10), one valid only when $\mathcal{K}_0$ has compact resolvent, e.g., when the small system is trapped, and one valid for nontrapped small systems as well. The main reason is that the former does not require the use of generalized Wigner measures, since conventional state-valued measures are sufficient, resulting in a more accessible proof.

3A1. *Energy lower bound*: *trapped particle systems.* If $\mathcal{K}_0$ has compact resolvent, the set of quasiclassical Wigner measures (as in Definition 1.4) associated with minimizing sequences for H_ε is not empty. In addition, the expectation of $\mathrm{Op}_\varepsilon^{\mathrm{Wick}}(\mathcal{V}_z)$ converges to the quasiclassical integral of $\mathcal{V}_z$. Let us formulate some preliminary results about the convergence of the expectation values of the operators involved. Such results rely on suitable a priori bounds on the family of states $\Psi_\varepsilon \in \mathscr{H}_\varepsilon$, as ε varies in $(0, 1)$. Lemma 3.7 guarantees that there exists a minimizing sequence $\Psi_{\varepsilon,\delta}$ in the sense of (3-5) satisfying such bounds.

Lemma 3.4. *If* (A4) *holds and there exists $C < +\infty$ such that*, *uniformly with respect to $\varepsilon \in (0, 1)$*,

$$\left|\langle \Psi_\varepsilon | (\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega) + 1) | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon}\right| \le C, \tag{3-11}$$

then $\mathscr{W}(\Psi_\varepsilon, \varepsilon \in (0, 1)) \neq \varnothing$. Furthermore, if $\Psi_{\varepsilon_n} \xrightarrow[\varepsilon_n \to 0]{\mathrm{qc}} \mathfrak{m}$, then $\mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}}(z)\mathcal{K}_0]$ is $\mu_{\mathfrak{m}}$-a.e. finite and $\mu_{\mathfrak{m}}$-absolutely integrable, *and*

$$\lim_{n\to\infty} \langle \Psi_{\varepsilon_n} | \mathcal{K}_0 | \Psi_{\varepsilon_n} \rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\, \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}}(z)\mathcal{K}_0]. \tag{3-12}$$

Proof. For $\omega = 1$ this proposition is proved in [Correggi et al. 2023, Propositions 2.3 and 2.6]. For a generic $\omega \ge 0$, the proof (in the presence of semiclassical degrees of freedom only) can be found in [Falconi 2018a, Theorem 3.3]; the extension to the quasiclassical setting is straightforward, testing with compact observables of the small system, as in the aforementioned [Correggi et al. 2023, Propositions 2.3 and 2.6]. Let us stress that the fact that all Wigner measures are probability measures, i.e., there is no loss of mass and $\mathfrak{m}(\mathfrak{h}_\omega) = 1$, is due to the fact that $\mathcal{K}_0$ has compact resolvent. Otherwise, there may be a loss of probability mass due to the interplay between the particle system and the environment; see [Correggi et al. 2023, Corollary 1.7 and Remark 1.9] for additional details. □

In order to control the convergence of the free field energy, we first have to regularize it: we pick a sequence of positive self-adjoint compact operators $\{\mathbb{1}_r\}_{r\in\mathbb{N}} \subset \mathscr{B}(\mathfrak{h})$ approximating the identity: for all $r \in \mathbb{N}$, $\mathbb{1}_r \le \mathbb{1}$, and for all $z \in \mathfrak{h}_\omega$,

$$\lim_{r\to\infty} \langle z | \omega_r | z \rangle_{\mathfrak{h}} = \lim_{r\to\infty} \langle z | \mathbb{1}_r | z \rangle_{\mathfrak{h}_\omega} = \|z\|_{\mathfrak{h}_\omega}^2 = \langle z | \omega | z \rangle_{\mathfrak{h}}, \tag{3-13}$$

where we have written $\omega_r := \omega^{1/2}\mathbb{1}_r\omega^{1/2}$. Recall also that $\mathrm{Op}_\varepsilon^{\mathrm{Wick}}(\langle z|\omega|z\rangle_{\mathfrak{h}}) = 1 \otimes \mathrm{d}\mathcal{G}_\varepsilon(\omega)$, where $\mathrm{d}\mathcal{G}_\varepsilon(\omega)$ stands for the second quantization of ω as above.

Lemma 3.5. *If* (A4) *holds and there exist $C < +\infty$ and $\delta > 1$ such that*, *uniformly with respect to $\varepsilon \in (0, 1)$*,

$$\left|\langle \Psi_\varepsilon | (\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega)^\delta + 1) | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon}\right| \le C, \tag{3-14}$$

then, *if $\Psi_{\varepsilon_n} \xrightarrow[\varepsilon_n \to 0]{\mathrm{qc}} \mathfrak{m} \in \mathscr{W}(\Psi_\varepsilon, \varepsilon \in (0, 1))$, it follows that for any $\eta \le \delta$*,

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z) \langle z | \omega | z \rangle_{\mathfrak{h}}^{\eta} \le C, \tag{3-15}$$

and, *for all $r \in \mathbb{N}$*,

$$\lim_{n\to\infty} \langle \Psi_{\varepsilon_n} | 1 \otimes \mathrm{d}\mathcal{G}_{\varepsilon_n}(\omega_r) | \Psi_{\varepsilon_n} \rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z) \langle z | \omega_r | z \rangle_{\mathfrak{h}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z) \langle z | \mathbb{1}_r | z \rangle_{\mathfrak{h}_\omega}. \tag{3-16}$$

Proof. The proof of $\mu_{\mathfrak{m}}$-integrability of $\langle z|\omega|z\rangle_{\mathfrak{h}}^{\eta}$ (and the relative bound) is a consequence of the corresponding result for semiclassical (scalar) Wigner measures proved in [Ammari and Nier 2008; Falconi 2018a]. Analogously, the convergence holds because $\langle z|\mathbb{1}_r|z\rangle_{\mathfrak{h}_\omega}$ is a compact scalar symbol; see [Falconi 2018a] for the convergence of compact symbols in $\mathfrak{h}_\omega$, and [Correggi et al. 2023, Propositions 2.3 and 2.6] for additional details on the generalization of results in semiclassical analysis to the quasiclassical case. □

Lemma 3.6. *If* (A4) *holds and there exists* $C < +\infty$ *such that, uniformly with respect to* $\varepsilon \in (0, 1)$,

$$\big|\langle \Psi_\varepsilon|(\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega)^2 + 1)|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}\big| \le C, \tag{3-17}$$

then, if $\Psi_{\varepsilon_n} \xrightarrow[\varepsilon_n\to 0]{\mathrm{qc}} \mathfrak{m}$, *for any* $i = 1, \dots, N$,

$$\lim_{n\to\infty} \langle \Psi_{\varepsilon_n}|\mathrm{Op}_{\varepsilon_n}^{\mathrm{Wick}}(\mathcal{V}_z(\boldsymbol{x}_i))|\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\, \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}}(z)\mathcal{V}_z(\boldsymbol{x}_i)]. \tag{3-18}$$

Lemma 3.7. *Under assumptions* (A1), (A2) *and* (A3), *there exists a minimizing sequence* $\{\Psi_{\varepsilon,\delta}\}_{\varepsilon,\delta\in(0,1)}$ *such that, for all fixed* $\delta \in (0, 1)$, (3-5) *holds true and there exists* $C_\delta < +\infty$ *such that*

$$\big|\langle \Psi_{\varepsilon,\delta}|(\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega)^2 + 1)|\Psi_{\varepsilon,\delta}\rangle_{\mathscr{H}_\varepsilon}\big| \le C_\delta. \tag{3-19}$$

The proofs of Lemma 3.6 and Lemma 3.7, like the form of the quasiclassical potential $\mathcal{V}_z$, depend on the model considered. We thus provide them in Section 4.

Remark 3.8. Observe that if Lemma 3.7 holds, then the assumptions of Lemmas 3.4–3.6 are verified for the minimizing sequence $\Psi_{\varepsilon,\delta}$.

We are now in a position to prove the lower bound in the trapped case.

Proof of Proposition 3.3. Let $\Psi_{\varepsilon,\delta}$ be the minimizing sequence for H_ε of Lemma 3.7. Since for any $r \in \mathbb{N}$, $\omega_r \le \omega$, it follows that $\mathrm{d}\mathcal{G}_\varepsilon(\omega_r) \le \mathrm{d}\mathcal{G}_\varepsilon(\omega)$. Hence,

$$\left\langle \Psi_{\varepsilon,\delta}\middle|\left(\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega_r) + \mathrm{Op}_\varepsilon^{\mathrm{Wick}}\left(\sum_i \mathcal{V}_z(\boldsymbol{x}_i)\right)\right)\middle|\Psi_{\varepsilon,\delta}\right\rangle_{\mathscr{H}_\varepsilon} \le \langle \Psi_{\varepsilon,\delta}|H_\varepsilon|\Psi_{\varepsilon,\delta}\rangle_{\mathscr{H}_\varepsilon} < E_\varepsilon + \delta. \tag{3-20}$$

Now, let us recall that, by Lemmas 3.4–3.7,

- for any $\delta > 0$, $\mathscr{W}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0, 1)) \neq \varnothing$;
- the expectation value of each term in the Hamiltonian converges as $\varepsilon \to 0$ or, more precisely, there exists $\mathfrak{m} \in \mathscr{W}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0, 1))$ such that

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\, \mathrm{tr}_{L^2(\mathbb{R}^{dN})}\left[\gamma_{\mathfrak{m}}\left(\mathcal{K}_0 + \langle z|\omega_r|z\rangle_{\mathfrak{h}} + \sum_i \mathcal{V}_z(\boldsymbol{x}_i)\right)\right] \le \liminf_{\varepsilon\to 0}\left\langle \Psi_{\varepsilon,\delta}\middle|\left(\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega_r) + \mathrm{Op}_\varepsilon^{\mathrm{Wick}}\left(\sum_i \mathcal{V}_z(\boldsymbol{x}_i)\right)\right)\middle|\Psi_{\varepsilon,\delta}\right\rangle_{\mathscr{H}_\varepsilon}. \tag{3-21}$$

Hence, we deduce that

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\,\mathrm{tr}_{L^2(\mathbb{R}^{dN})}\Big[\gamma_{\mathfrak{m}}\Big(\mathcal{K}_0 + \langle z|\omega_r|z\rangle_{\mathfrak{h}} + \sum_i \mathcal{V}_z(\boldsymbol{x}_i)\Big)\Big] < \liminf_{\varepsilon\to 0} E_\varepsilon + \delta. \tag{3-22}$$

Now, $\langle z|\omega_r|z\rangle_{\mathfrak{h}} \xrightarrow[r\to\infty]{} \langle z|\omega|z\rangle_{\mathfrak{h}}$ for any $z \in \mathfrak{h}_\omega$ by construction, and any $\mathfrak{m} \in \mathscr{W}(\Psi_\varepsilon, \varepsilon \in (0,1))$ is concentrated on $\mathfrak{h}_\omega$ by Lemma 3.4. Furthermore,

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\langle z|\omega_r|z\rangle_{\mathfrak{h}} \le \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\langle z|\omega|z\rangle_{\mathfrak{h}} \le C < +\infty.$$

Hence, by dominated convergence,

$$\lim_{r\to\infty}\int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\langle z|\omega_r|z\rangle_{\mathfrak{h}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\langle z|\omega|z\rangle_{\mathfrak{h}}. \tag{3-23}$$

Thus, one gets

$$\inf_{\mathfrak{m}\in\mathscr{W}(\Psi_{\varepsilon,\delta},\varepsilon\in(0,1))}\int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\,\mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}}\mathcal{H}_z] < \liminf_{\varepsilon\to 0} E_\varepsilon + \delta, \tag{3-24}$$

which, via Proposition 2.1, implies that

$$E_{\mathrm{qc}} \le \inf_{\mathfrak{m}\in\mathscr{W}(\Psi_{\varepsilon,\delta},\varepsilon\in(0,1))}\int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\,\mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}}\mathcal{H}_z] < \liminf_{\varepsilon\to 0} E_\varepsilon + \delta.$$

Since $\delta > 0$ is arbitrary, the claim follows. □

3A2. *Energy lower bound*: *nontrapped particle systems.* In the nontrapped case, the strategy of proof is very similar, however it is not ensured that the set of quasiclassical Wigner measures for the minimizing sequence is not empty. It is then necessary to use generalized Wigner measures (recall Definition 1.6).

We first generalize the preparatory lemmas that we needed in the trapped case to the general situation. Note that for Lemma 3.7 it is not necessary that $\mathcal{K}_0$ has compact resolvent and therefore we can use it directly also in the nontrapped case. We also use the same notation as in the trapped case; in particular, we make use of the same compact approximation ω_r of ω we introduced in (3-13).

Lemma 3.9. *If there exists $C < +\infty$ such that*, *uniformly with respect to $\varepsilon \in (0,1)$*,

$$\big|\langle\Psi_\varepsilon|(\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega) + 1)|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}\big| \le C, \tag{3-25}$$

then $\mathscr{G}\mathscr{W}(\Psi_\varepsilon, \varepsilon \in (0,1)) \neq \varnothing$. Furthermore, *if $\Psi_{\varepsilon_n} \xrightarrow[\varepsilon_n\to 0]{\mathrm{gqc}} \mathfrak{n}$*, *then $\mathfrak{n}$ is in the domain of $\mathcal{K}_0 + 1$ in the sense of Definition 2.6 and*

$$\lim_{n\to\infty}\langle\Psi_{\varepsilon_n}|\mathcal{K}_0|\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z)[\mathcal{K}_0]. \tag{3-26}$$

In addition, *it follows that*

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z)[1]\langle z|\omega|z\rangle_{\mathfrak{h}} \le C, \tag{3-27}$$

and, *for all $r \in \mathbb{N}$*,

$$\lim_{n\to\infty}\langle\Psi_{\varepsilon_n}|1\otimes\mathrm{d}\mathcal{G}_{\varepsilon_n}(\omega_r)|\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z)[1]\langle z|\omega_r|z\rangle_{\mathfrak{h}}. \tag{3-28}$$

Proof. These lemmas extend to generalized Wigner measures Lemmas 3.4 and 3.5, respectively. The proof is, *mutatis mutandis*, completely analogous to those of the latter. Contrary to Lemma 3.4, since now $\mathcal{K}_0$ has a noncompact resolvent, the set of Wigner measures of Ψ_ε may be empty and there might be a loss of mass along the quasiclassical convergence. The set of generalized Wigner measures is, however, always nonempty: no mass is lost due to the fact that

$$\|\Psi_\varepsilon\|^2_{\mathscr{H}_\varepsilon} = \langle \Psi_\varepsilon | 1 \otimes W_\varepsilon(0) | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon} = 1,$$

and the identity operator belongs to $\mathscr{B}(L^2(\mathbb{R}^{dN}))$ but is not compact. More precisely, the above quantity can be immediately identified, in the limit $\varepsilon \to 0$, with the total mass of all generalized Wigner measures associated to Ψ_ε, as defined in Definition 1.6, whereas it is a priori only bigger than or equal to the total mass of measures defined by the convergence in Definition 1.4 (if all cluster points for the aforementioned convergence have total mass strictly less than one, the set of Wigner measures associated to Ψ_ε, which are required by Definition 1.4 to have total mass 1, is thus empty). □

Lemma 3.10. *If there exists $C < +\infty$ such that, uniformly with respect to $\varepsilon \in (0, 1)$,*

$$\left| \langle \Psi_\varepsilon | (\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega)^2 + 1) | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon} \right| \le C, \tag{3-29}$$

then, if $\Psi_{\varepsilon_n} \xrightarrow[\varepsilon_n \to 0]{\mathrm{gqc}} \mathfrak{n}$, for any $i = 1, \dots, N$,

$$\lim_{n\to\infty} \langle \Psi_{\varepsilon_n} | \mathrm{Op}^{\mathrm{Wick}}_{\varepsilon_n}(\mathcal{V}_z(\boldsymbol{x}_i)) | \Psi_{\varepsilon_n} \rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z) [\mathcal{V}_z(\boldsymbol{x}_i)]. \tag{3-30}$$

Like its analogue Lemma 3.6, the proof of Lemma 3.10 is model-dependent and thus given in Section 4.

The proof of the lower bound for the nontrapped case is now equivalent to the one in the trapped case, using generalized Wigner measures.

Proof of Proposition 3.3. Let $\Psi_{\varepsilon,\delta}$ be the minimizing sequence for H_ε of Lemma 3.7 satisfying (3-20). By Lemmas 3.7–3.10,

- for any $\delta > 0$, we have that $\mathscr{GW}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0, 1)) \neq \varnothing$;
- for Wigner measures, there exists $\mathfrak{n} \in \mathscr{GW}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0, 1))$ such that

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z) \Big[\mathcal{K}_0 + \langle z | \omega_r | z \rangle_{\mathfrak{h}} + \sum_i \mathcal{V}_z(\boldsymbol{x}_i) \Big] \le \liminf_{\varepsilon \to 0} \Big\langle \Psi_{\varepsilon,\delta} \Big| \Big(\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega_r) + \mathrm{Op}^{\mathrm{Wick}}_\varepsilon \Big(\sum_i \mathcal{V}_z(\boldsymbol{x}_i) \Big) \Big) \Big| \Psi_{\varepsilon,\delta} \Big\rangle_{\mathscr{H}_\varepsilon},$$

and therefore

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z) \Big[\mathcal{K}_0 + \langle z | \omega_r | z \rangle_{\mathfrak{h}} + \sum_i \mathcal{V}_z(\boldsymbol{x}_i) \Big] < \liminf_{\varepsilon \to 0} E_\varepsilon + \delta. \tag{3-31}$$

However, by dominated convergence, see Theorem A.18 in the Appendix,

$$\lim_{r\to\infty} \int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z)[1] \langle z | \omega_r | z \rangle_{\mathfrak{h}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z)[1] \langle z | \omega | z \rangle_{\mathfrak{h}}.$$

Hence,

$$E_{\mathrm{gqc}} \le \inf_{\mathfrak{n} \in \mathscr{GW}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0,1))} \int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}(z) [\mathcal{H}_z] < \liminf_{\varepsilon \to 0} E_\varepsilon + \delta,$$

and the result follows from the arbitrariness of $\delta > 0$, via Proposition 2.8. □

3B. ***Convergence of minimizing sequences and minimizers.*** Once the energy convergence is proven, we investigate the behavior of minimizing sequences and minimizers, if any.

Proof of Theorem 1.7. Let $\Psi_{\varepsilon,\delta} \in \mathscr{D}(H_\varepsilon)$ be a minimizing sequence. Then by Lemmas 3.4–3.7, any $\mathfrak{m}_\delta \in \mathscr{W}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0,1))$, corresponding to a sequence $\{\Psi_{\varepsilon_n,\delta}\}_{n\in\mathbb{N}}$, $\varepsilon_n \to 0$, satisfies

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}_\delta}(z)\,\mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}_\delta}(z)\mathcal{H}_z] = \lim_{n\to\infty} \langle \Psi_{\varepsilon_n,\delta} | H_{\varepsilon_n} | \Psi_{\varepsilon_n,\delta}\rangle_{\mathscr{H}_{\varepsilon_n}} < \lim_{n\to\infty} E_{\varepsilon_n} + \delta = E_{\mathrm{qc}} + \delta,$$

as proven in Theorem 1.3. □

Proof of Corollary 1.9. If $\delta = o_\varepsilon(1)$, then considering $\mathfrak{m}_0 \in \mathscr{W}(\Psi_{\varepsilon,o_\varepsilon(1)}, \varepsilon \in (0,1))$, corresponding to a sequence $\{\Psi_{\varepsilon_n,\delta}\}_{n\in\mathbb{N}}$, $\varepsilon_n \to 0$, we have

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}_0}(z)\,\mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}_0}(z)\mathcal{H}_z] \le \lim_{n\to\infty} (E_{\varepsilon_n} + o_{\varepsilon_n}(1)) = E_{\mathrm{qc}}.$$

By Proposition 2.1 it follows that $\mathfrak{m}_0$ is a minimizer of (vp2) and, by Proposition 2.3, is concentrated on the set $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}})$ of minimizers of (VP2). □

Proof of Corollary 1.10. Let $\Psi_{\varepsilon,\mathrm{gs}}$ be a ground state of H_ε. Then it is also an (exact) minimizing sequence with $\delta = 0$, and thus as above $\mathfrak{m}_0$ is a minimizer of (vp2) and is concentrated on the set $(\psi_{\mathrm{qc}}, z_{\mathrm{qc}})$ of minimizers of (VP2). □

The proof of Theorem 1.15 is also completely analogous to the proof of Theorem 1.7 for trapped systems.

Proof of Theorem 1.15. If $\mathcal{K}_0$ does not have compact resolvent, then by Lemmas 3.7, 3.9 and 3.10, any $\mathfrak{n}_\delta \in \mathscr{GW}(\Psi_{\varepsilon,\delta}, \varepsilon \in (0,1))$ satisfies

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}_\delta(z)[\mathcal{H}_z] < \lim_{\varepsilon_n\to 0} E_{\varepsilon_n} + \delta = E_{\mathrm{qc}} + \delta = E_{\mathrm{gqc}} + \delta, \tag{3-32}$$

by Theorem 1.3 and Proposition 2.8. □

Proof of Corollary 1.16. If $\delta = o_\varepsilon(1)$, it follows that $\mathfrak{n}_0 \in \mathscr{GW}(\Psi_{\varepsilon,o_\varepsilon(1)}, \varepsilon \in (0,1))$ satisfies

$$\int_{\mathfrak{h}_\omega} \mathrm{d}\mathfrak{n}_0(z)[\mathcal{H}_z] = \lim_{\varepsilon_n\to 0} E_{\varepsilon_n} = E_{\mathrm{gqc}}.$$

Therefore $\mathfrak{n}_0$ solves (gvp2), and thus it is concentrated on minimizers solving (GVP2). □

Proof of Corollary 1.17. This proof is completely analogous to that of Corollary 1.10. □

4. Concrete models

In this section we discuss the concrete models introduced in Section 1, and in particular we provide the proofs of results used in Section 3 that require a model-dependent treatment.

4A. ***The Nelson model.*** The simplest model under consideration is the so-called Nelson model [1964]. It consists of a small system of N nonrelativistic particles coupled with a scalar bosonic field, both moving in d spatial dimensions.

We recall the explicit expression of the quasiclassical energy (1-14) in the Nelson model:

$$\mathcal{H}_z = \sum_{j=1}^{N} \{-\Delta_j + \mathcal{V}_z(\boldsymbol{x}_j)\} + \mathcal{W}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) + \langle z|\omega|z\rangle_{\mathfrak{h}},$$

acting on $L^2(\mathbb{R}^{dN})$ and dependent on $z \in \mathfrak{h}$, where $\mathcal{V}_z$ is the potential (1-15), i.e., $\mathcal{V}_z(\boldsymbol{x}) = 2\mathrm{Re}\langle z|\lambda(\boldsymbol{x})\rangle_{\mathfrak{h}}$, $\mathcal{W} \in L^1_{\mathrm{loc}}(\mathbb{R}^{dN}; \mathbb{R}_+)$ is a field-independent potential,[6] e.g., a trap or an interaction between the particles, $\omega \geq 0$ is a self-adjoint operator on $\mathfrak{h}$ with an inverse that is possibly unbounded and λ, $\omega^{-1/2}\lambda \in L^\infty(\mathbb{R}^d, \mathfrak{h})$. Both $\mathcal{W}$ and $\mathcal{V}_z$ are multiplication operators and $\mathcal{H}_z$ is self-adjoint on $\mathscr{D}(-\Delta + \mathcal{W})$ and bounded from below for all $z \in \mathfrak{h}_\omega$. The associated quasiclassical energy of the system is the quadratic form $\mathcal{E}_{\mathrm{qc}}$, whose form domain is thus contained in $\mathscr{Q}(-\Delta + \mathcal{W}) \oplus \mathscr{Q}(\omega)$, where we recall that $\mathscr{Q}(A)$ stands for the quadratic form domain associated with the self-adjoint operator A.

The quasiclassical Wick quantization of $\mathcal{H}_z$ yields the quantum field Hamiltonian

$$H_\varepsilon = \sum_{j=1}^{N} \{-\Delta_j \otimes 1 + a_\varepsilon(\lambda(\boldsymbol{x}_j)) + a_\varepsilon^\dagger(\lambda(\boldsymbol{x}_j))\} + \mathcal{W}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) \otimes 1 + 1 \otimes \mathrm{d}\mathcal{G}_\varepsilon(\omega)$$

acting on $\mathscr{H}_\varepsilon = L^2(\mathbb{R}^{dN}) \otimes \mathcal{G}_\varepsilon(\mathfrak{h})$, where we have explicitly highlighted the trivial action of some terms of H_ε on either the particle's or the field's degrees of freedom. Whenever $\lambda \in L^\infty(\mathbb{R}^d; \mathfrak{h})$, the operator H_ε is self-adjoint, with domain of essential self-adjointness

$$\mathscr{D}(-\Delta + \mathcal{W} + \mathrm{d}\mathcal{G}_\varepsilon(\omega)) \cap \mathscr{C}_0^\infty(\mathrm{d}\mathcal{G}_\varepsilon(1)),$$

where the latter is the set of vectors with a finite number of field excitations [Falconi 2015], but it may be unbounded from below if $0 \in \sigma(\omega)$. It is however well known that, if for a.e. $\boldsymbol{x} \in \mathbb{R}^d$ we have $\lambda(\boldsymbol{x}) \in \mathscr{D}(\omega^{-1/2})$, that we assume in (1-8), then H_ε is bounded from below by Kato–Rellich's theorem. Nonetheless, it may still not have a ground state if $0 \in \sigma(\omega)$ or if $\mathcal{W}$ is not regular enough. We simply remark here that the ground state exists if $0 \notin \sigma(\omega)$ and $-\Delta + \mathcal{W}$ has compact resolvent (trapped particle system), or if $0 \in \sigma(\omega)$ and λ and $\mathcal{W}$ satisfy suitable conditions, irrespective of compactness of the resolvent of $-\Delta + \mathcal{W}$.

Proof of Proposition 3.1. The upper and lower bounds in (3-4) are well known; see, e.g., [Ammari and Falconi 2014; Correggi and Falconi 2018; Ginibre et al. 2006]. The lower bound is a direct consequence of Kato–Rellich's inequality, while the upper bound is proved using coherent states for the field. We provide some details for the sake of completeness.

[6] Of course we may allow for a negative part of the potential $\mathcal{W}$, provided it is bounded, but we choose a positive potential for the sake of simplicity.

Setting[7]

$$H_{\mathrm{free}} := \mathcal{K}_0 \otimes 1 + 1 \otimes \mathrm{d}\mathcal{G}_\varepsilon(\omega), \tag{4-1}$$

we get, for all $\alpha > 0$ and all $\Psi_\varepsilon \in \mathscr{D}(H_{\mathrm{free}})$,

$$\begin{aligned}\Big\| \sum_{j=1}^{N} (a_\varepsilon(\lambda(\boldsymbol{x}_j)) + a_\varepsilon^\dagger(\lambda(\boldsymbol{x}_j)))\Psi_\varepsilon \Big\|_{\mathscr{H}_\varepsilon} & \\ &\le 2N \|\omega^{-1/2}\lambda\|_{L^\infty(\mathbb{R}^d;\mathfrak{h})} \|\mathrm{d}\mathcal{G}_\varepsilon(\omega)^{1/2}\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon} + \sqrt{\varepsilon}\|\lambda\|_{L^\infty(\mathbb{R}^d;\mathfrak{h})}\|\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon} \\ &\le \alpha \langle \Psi_\varepsilon | \mathrm{d}\mathcal{G}_\varepsilon(\omega) | \Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} + \Big[\frac{N^2}{\alpha}\|\omega^{-1/2}\lambda\|^2_{L^\infty(\mathbb{R}^d;\mathfrak{h})} + \sqrt{\varepsilon}\|\lambda\|_{L^\infty(\mathbb{R}^d;\mathfrak{h})}\Big]\|\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon}. \end{aligned} \tag{4-2}$$

Therefore, choosing $\alpha = 1$, we deduce that (recall that $\varepsilon \in (0, 1)$)

$$E_\varepsilon \ge -N^2 \|\omega^{-1/2}\lambda\|^2_{L^\infty(\mathbb{R}^d;\mathfrak{h})} - \|\lambda\|_{L^\infty(\mathbb{R}^d;\mathfrak{h})}. \tag{4-3}$$

The upper bound is trivial to show by exploiting (4-2) and evaluating the energy on any state such that $\langle \Psi_\varepsilon | \mathrm{d}\mathcal{G}_\varepsilon(\omega) | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon} \le C < +\infty$, e.g., a product state $\Psi_\varepsilon = \psi \otimes \Omega_\varepsilon$, with $\psi \in \mathscr{D}(\mathcal{K}_0)$ and Ω_ε the field vacuum. Note that the uniform boundedness of E_ε from above could as well be deduced by the boundedness of E_0, which in turn follows from the evaluation of $\mathcal{E}_{\mathrm{qc}}$ on, e.g., a configuration $(\psi, 0)$, with $\psi \in \mathscr{D}(\mathcal{K}_0)$. □

We now prove Lemmas 3.6 and 3.7 for the Nelson model. We have however to state first a technical result, which generalizes the convergence of expectation values proven in [Correggi et al. 2023]: indeed, in [Correggi et al. 2023, Proposition 2.6] it is shown that,[8] if

$$\langle \Psi_\varepsilon, (\mathrm{d}\mathcal{G}_\varepsilon(\omega) + 1)^\delta \Psi_\varepsilon \rangle_{L^2(\mathbb{R}^{dN})\otimes\mathscr{H}_\varepsilon} \le C,$$

for any $\delta > \frac{1}{2}$, and $\Psi_{\varepsilon_n} \xrightarrow[n\to\infty]{\mathrm{qc}} \mathfrak{m}$, then, for all $\mathcal{K} \in \mathscr{L}^\infty(L^2(\mathbb{R}^{dN}))$,

$$\lim_{n\to\infty} \langle \Psi_{\varepsilon_n} | \mathrm{Op}^{\mathrm{Wick}}_{\varepsilon_n}(\mathcal{V}_z)\mathcal{K}\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\, \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}}(z)\mathcal{V}_z\mathcal{K}], \tag{4-4}$$

but our goal is to apply the above convergence to the identity, which is not compact. We have then to approximate it with compact operators.

Lemma 4.1. *If* (A4) *holds and there exist $C < +\infty$ and $\delta \ge 1$ such that, uniformly with respect to $\varepsilon \in (0, 1)$,*

$$\big|\langle \Psi_\varepsilon | (\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega)^\delta + 1) | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon}\big| \le C \tag{4-5}$$

and $\Psi_{\varepsilon_n} \xrightarrow[\varepsilon_n\to 0]{\mathrm{qc}} \mathfrak{m}$, then, for all $\mathcal{B} \in \mathscr{B}(L^2(\mathbb{R}^{dN}))$ and any $j = 1, \dots, N$,

$$\lim_{n\to\infty} \langle \Psi_{\varepsilon_n} | \mathrm{Op}^{\mathrm{Wick}}_{\varepsilon_n}(\mathcal{V}_z(\boldsymbol{x}_j))\mathcal{B}\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\, \mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}}(z)\mathcal{V}_z(\boldsymbol{x}_j)\mathcal{B}]. \tag{4-6}$$

[7] Even if not stated explicitly, we use the notation H_{free} also in Sections 4B and 4C with the same meaning.

[8] In [Correggi et al. 2023, Proposition 2.6] the result is proved for $\omega = 1$. The extension to a generic ω is done straightforwardly by combining the proof of Proposition 2.6 with the techniques introduced in [Falconi 2018a].

Proof. Let us introduce compact approximate identities $\{1_m\}_{m\in\mathbb{N}} \subset \mathscr{L}^\infty(L^2(\mathbb{R}^{dN}))$ as follows:

$$1_m := \mathbb{1}_{[-m,m]}(\mathcal{K}_0),$$

where $\mathbb{1}_{[-m,m]} : \mathbb{R} \to \{0, 1\}$ is the characteristic function of the interval $[-m, m]$, so that the right-hand side of the above expression is the usual spectral projector of $\mathcal{K}_0$ constructed via spectral theory. For later convenience, let us also define $\mathcal{B}_m := \mathcal{B}1_m$. Therefore, we have

$$\begin{aligned}\langle \Psi_{\varepsilon_n} | \mathrm{Op}_{\varepsilon_n}^{\mathrm{Wick}}(\mathcal{V}_z(\boldsymbol{x}_j))\mathcal{B}\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} \\ = \langle \Psi_{\varepsilon_n} | \mathrm{Op}_{\varepsilon_n}^{\mathrm{Wick}}(\mathcal{V}_z(\boldsymbol{x}_j))\mathcal{B}_m\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} + \langle \Psi_{\varepsilon_n} | \mathrm{Op}_{\varepsilon_n}^{\mathrm{Wick}}(\mathcal{V}_z(\boldsymbol{x}_j))(\mathcal{B}-\mathcal{B}_m)\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}}.\end{aligned} \tag{4-7}$$

The first term on the right-hand side converges when $n \to \infty$ for any fixed $m \in \mathbb{N}$, since we have that $\mathcal{B}_m \in \mathscr{L}^\infty(L^2(\mathbb{R}^{dN}))$ (see (4-4)), i.e.,

$$\lim_{n\to\infty} \langle \Psi_{\varepsilon_n} | \mathrm{Op}_{\varepsilon_n}^{\mathrm{Wick}}(\mathcal{V}_z(\boldsymbol{x}_j))\mathcal{B}_m\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\,\mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}}(z)\mathcal{V}_z(\boldsymbol{x}_j)\mathcal{B}_m].$$

By dominated convergence, we can then take the limit $m \to \infty$, to obtain

$$\lim_{m\to\infty}\lim_{n\to\infty} \langle \Psi_{\varepsilon_n} | \mathrm{Op}_{\varepsilon_n}^{\mathrm{Wick}}(\mathcal{V}_z(\boldsymbol{x}_j))\mathcal{B}_m\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} = \int_{\mathfrak{h}_\omega} \mathrm{d}\mu_{\mathfrak{m}}(z)\,\mathrm{tr}_{L^2(\mathbb{R}^{dN})}[\gamma_{\mathfrak{m}}(z)\mathcal{V}_z(\boldsymbol{x}_j)\mathcal{B}]. \tag{4-8}$$

It remains to prove that

$$\lim_{m\to\infty}\sup_{\varepsilon\in(0,1)} \left|\langle \Psi_\varepsilon | \mathrm{Op}_\varepsilon^{\mathrm{Wick}}(\mathcal{V}_z(\boldsymbol{x}_j))(\mathcal{B}-\mathcal{B}_m)\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}\right| = 0. \tag{4-9}$$

For any $0 < s \leq \frac{1}{2}$ and for any $c_0 > |\inf\sigma(\mathcal{K}_0)|$,

$$\begin{aligned}&\left|\langle \Psi_\varepsilon | \mathrm{Op}_\varepsilon^{\mathrm{Wick}}(\mathcal{V}_z(\boldsymbol{x}_j))(\mathcal{B}-\mathcal{B}_m)\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}\right| \\ &\quad\leq 2\|(\mathcal{B}-\mathcal{B}_m)(\mathcal{K}_0+c_0)^{-s/2}\|_{\mathscr{B}(L^2(\mathbb{R}^{dN}))}\|(\mathrm{d}\mathcal{G}_\varepsilon(\omega)^{1/2}+1)^{-1/2}a_\varepsilon(\lambda(\boldsymbol{x}_j))(\mathrm{d}\mathcal{G}_\varepsilon(\omega)^{1/2}+1)^{-1/2}\|_{\mathscr{B}(\mathscr{H}_\varepsilon)} \\ &\qquad\qquad \times \|(\mathrm{d}\mathcal{G}_\varepsilon(\omega)^{1/2}+1)^{1/2}(\mathcal{K}_0+c_0)^{s/2}\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon}^2 \\ &\quad\leq C\|\mathcal{B}\|_{\mathscr{B}}\|(1-1_m)(\mathcal{K}_0+c_0)^{-s/2}\|_{\mathscr{B}}\|\omega^{-1/2}\lambda\|_{L^\infty(\mathbb{R}^d,\mathfrak{h})}\langle\Psi_\varepsilon|\mathcal{K}_0^{2s}+\mathrm{d}\mathcal{G}_\varepsilon(1)+1|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} \\ &\quad\leq C\sup_{\eta\in[(-\infty,-m)\cup(m,+\infty)]\cap\sigma(\mathcal{K}_0)}\frac{1}{(\eta+c_0)^{s/2}} \leq Cm^{-s/2}\end{aligned}$$

for m large enough, e.g., larger than $|\inf\sigma(\mathcal{K}_0)|$. Therefore, since the above quantity vanishes as $m \to \infty$ uniformly with respect to $\varepsilon \in (0, 1)$, we conclude that (4-9) holds true and the result follows. □

Proof of Lemma 3.6. The result follows by taking $\mathcal{B} = 1$ in Lemma 4.1. Again, this makes crucial use of the fact that $\mathcal{K}_0 = -\Delta + \mathcal{W}$ has compact resolvent, and that Ψ_ε is regular enough with respect to $\mathcal{K}_0$. □

Proof of Lemma 3.7. The proof of Lemma 3.7 stems from a known result that allows us to compare the expectation of the square of the free energy H_{free}^2 with the expectation of the square of the full Hamiltonian H_ε^2. This is a consequence of Kato–Rellich's inequality: there exists $C > 0$ (independent of ε) such that

$$\langle\Psi_\varepsilon|H_{\mathrm{free}}^2|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} \leq C\langle\Psi_\varepsilon|H_\varepsilon^2+1|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}. \tag{4-10}$$

The idea of the proof of this standard inequality goes as follows: From the triangular inequality we get

$$\langle\Psi_\varepsilon|H_{\mathrm{free}}^2|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} \leq 2\langle\Psi_\varepsilon|H_\varepsilon^2|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} + 2\langle\Psi_\varepsilon|(H_\varepsilon-H_{\mathrm{free}})^2|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}.$$

Now, using inequality (4-2), we get that for any $\alpha < 1/\sqrt{2}$,

$$(1-2\alpha^2)\langle\Psi_\varepsilon|H_{\text{free}}^2|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} \leq 2\langle\Psi_\varepsilon|H_\varepsilon^2|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} + C_\alpha\|\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon}^2,$$

with C_α independent of ε. The result then easily follows.

It remains to prove that there exists a minimizing sequence $\{\Psi_{\varepsilon,\delta}\}_{\varepsilon,\delta\in(0,1)} \subset \mathscr{D}(H_\varepsilon)$ for H_ε such that

$$\langle\Psi_\varepsilon|H_\varepsilon^2|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} \leq \max\{E_\varepsilon^2, (E_\varepsilon+\delta)^2\} \leq C, \tag{4-11}$$

with the last inequality given by Proposition 3.1. Indeed, combining the above estimate with (4-10), we immediately deduce that (3-17) holds true. Let us denote by $\mathbb{1}_{(a,b)}(H_\varepsilon)$ the spectral projections of H_ε, and by $\mathscr{P}_{(a,b)} := \mathbb{1}_{(\alpha,\beta)}(H_\varepsilon)\mathscr{H}_\varepsilon$ the associated spectral subspaces. Let us now choose, for any $\delta > 0$,

$$\Psi_{\varepsilon,\delta} \in \{\Psi \in \mathscr{P}_{(E_\varepsilon-\delta,E_\varepsilon+\delta)} : \|\Psi\|_{\mathscr{H}_\varepsilon} = 1\}.$$

Each spectral subspace above is not empty by definition of $E_\varepsilon = \inf\sigma(H_\varepsilon)$. Therefore, on one hand,

$$\langle\Psi_{\varepsilon,\delta}|H_\varepsilon|\Psi_{\varepsilon,\delta}\rangle_{\mathscr{H}_\varepsilon} \leq E_\varepsilon + \delta,$$

and, on the other,

$$\|H_\varepsilon\Psi_{\varepsilon,\delta}\|_{\mathscr{H}_\varepsilon}^2 \leq \max\{E_\varepsilon^2, (E_\varepsilon+\delta)^2\}.$$ □

It remains only to prove Lemma 3.10, used in the nontrapped case.

Proof of Lemma 3.10. To prove the result, it is sufficient to show that, if Ψ_ε is such that

$$\left|\langle\Psi_\varepsilon|(\mathrm{d}\mathcal{G}_\varepsilon(\omega)+1)^\delta|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}\right| \leq C,$$

for some $\delta > \frac{1}{2}$ and some finite constant C, and if $\Psi_{\varepsilon_n} \xrightarrow[n\to\infty]{\text{gqc}} \mathfrak{n}$, then (3-30) holds true, i.e., for all $\mathcal{B} \in \mathscr{B}(L^2(\mathbb{R}^{dN}))$,

$$\lim_{n\to\infty}\langle\Psi_{\varepsilon_n}|\mathrm{Op}_{\varepsilon_n}^{\text{Wick}}(\mathcal{V}_z))\mathcal{B}\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_\varepsilon} = \int_{\mathfrak{h}_\omega}\mathrm{d}\mathfrak{n}[\mathcal{V}_z\mathcal{B}].$$

Such a result is however a special case of [Correggi et al. 2023, Proposition 2.6], if in that statement Wigner measures are substituted by generalized Wigner measures, the test with compact operators of the small system is replaced with the test with bounded operators, and $\mathrm{d}\mathcal{G}_\varepsilon(1)$ is replaced by $\mathrm{d}\mathcal{G}_\varepsilon(\omega)$. The proof given there is generalized to this setting straightforwardly, recalling the properties of generalized Wigner measures outlined in the Appendix. There is only one thing we need to mention explicitly: the integration of operator-valued functions with respect to generalized Wigner measures makes sense only if $\mathrm{Ran}(z \mapsto \mathcal{V}_z) \subset \mathscr{B}(L^2(\mathbb{R}^{dN}))$ is separable in the norm topology of $\mathscr{B}(L^2(\mathbb{R}^{dN}))$. Let us check explicitly that $\mathrm{Ran}(z \mapsto \mathcal{V}_z)$ is indeed separable: Since $\mathfrak{h}_\omega$ is separable, let us denote by $\mathfrak{k} \subset \mathfrak{h}_\omega$ a countable dense subset and denote by

$$\mathcal{V}_{\mathfrak{k}} := \{\mathcal{V}_\zeta(\boldsymbol{x}) \in \mathscr{B}(L^2(\mathbb{R}^{dN})) : \zeta \in \mathfrak{k}\}$$

the image of $\mathfrak{k}$ by means of $z \mapsto \mathcal{V}_z$. Now, for any $z \in \mathfrak{h}_\omega$, $\zeta \in \mathfrak{k}$, we have that

$$\|\mathcal{V}_z - \mathcal{V}_\zeta\|_{\mathscr{B}(L^2(\mathbb{R}^{dN}))} \leq 2\|\omega^{-1/2}\lambda\|_{L^\infty(\mathbb{R}^d;\mathfrak{h})}\|z-\zeta\|_{\mathfrak{h}_\omega},$$

which implies that $\mathcal{V}_{\mathfrak{k}}$ is dense in $\mathrm{Ran}(z \mapsto \mathcal{V}_z)$ with respect to the $\mathscr{B}(L^2(\mathbb{R}^{dN}))$-norm topology. □

4B. ***The polaron model.*** The polaron model, introduced in [Fröhlich 1937], describes N electrons (spinless for simplicity) subjected to the vibrational (phonon) field of a lattice. This model is similar to Nelson's, however the coupling is slightly more singular. The one-excitation space is $\mathfrak{h} = L^2(\mathbb{R}^d)$, while the form factor is given by (1-9): the quasiclassical energy has the same form as in the Nelson model, as well as the effective potential $\mathcal{V}_z$ (see (1-15)), although now

$$\lambda(\boldsymbol{x}; \boldsymbol{k}) = \sqrt{\alpha}\frac{e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}}{|\boldsymbol{k}|^{(d-1)/2}}, \quad \omega = 1,$$

where $\alpha > 0$ is a constant measuring the coupling's strength. The assumptions on $\mathcal{K}_0 = -\Delta + \mathcal{W}$ are the same as in the Nelson model. Let us remark that in this case since $\omega = 1$, we have that $\mathfrak{h}_\omega = \mathfrak{h}$.

The key difference with the aforementioned Nelson model is thus that there exists $z \in \mathfrak{h}$ such that

$$\mathcal{V}_z(\,\cdot\,) \notin L^\infty(\mathbb{R}^d),$$

due to the fact that $\lambda \notin L^\infty(\mathbb{R}^d; \mathfrak{h})$. However, it is possible to write $\mathcal{V}_z$ as the sum of an L^∞ function and the commutator between an L^∞ vector function and the momentum operator $-i\nabla_{\boldsymbol{x}}$:

$$\mathcal{V}_z(\boldsymbol{x}) = \sqrt{\alpha}(\mathcal{V}_{<,z}(\boldsymbol{x}) + [-i\nabla_{\boldsymbol{x}}, \boldsymbol{\mathcal{V}}_{>,z}(\boldsymbol{x})]), \tag{4-12}$$

where

$$\mathcal{V}_{<,z}(\boldsymbol{x}) = 2\mathrm{Re}\mathscr{F}^{-1}[\lambda_< z](\boldsymbol{x}), \quad \lambda_<(\boldsymbol{k}) := \mathbb{1}_{|\boldsymbol{k}|\le\varrho}|\boldsymbol{k}|^{-(d-1)/2},$$
$$\boldsymbol{\mathcal{V}}_{>,z}(\boldsymbol{x}) = 2\mathrm{Re}\mathscr{F}^{-1}[\boldsymbol{\lambda}_> z](\boldsymbol{x}), \quad \boldsymbol{\lambda}_>(\boldsymbol{k}) := \mathbb{1}_{|\boldsymbol{k}|>\varrho}|\boldsymbol{k}|^{-(d+1)/2}\hat{\boldsymbol{k}},$$

where $\hat{\boldsymbol{k}} := \boldsymbol{k}/|\boldsymbol{k}|$ and $\mathscr{F}$ stands for the Fourier transform in $\mathbb{R}^d$. Note that, for any $\varrho > 0$, we have that $\lambda_< \in \mathfrak{h}$ and $\boldsymbol{\lambda}_> \in \mathfrak{h} \otimes \mathbb{C}^d$. By the KLMN theorem, it then follows that $\mathcal{H}_z$ is self-adjoint and bounded from below for all $z \in \mathfrak{h}$, with z-independent form domain $\mathscr{Q}(\mathcal{H}_z) = \mathscr{Q}(\mathcal{K}_0)$. Let us remark that, choosing ρ suitably large (independent of z) in the above decomposition, it is possible to make the operator $\mathcal{H}_z$ bounded from below uniformly with respect to $z \in \mathfrak{h}$; see, e.g., [Correggi and Falconi 2018, Proposition 3.21].

The quasiclassical Wick quantization of $\mathcal{H}_z$ formally yields the same expression as in the Nelson model (with $\omega = 1$ and λ as above). Such a formal operator gives rise to a closed and bounded from below quadratic form, via the decomposition (4-12) (this can also be proved by the KLMN theorem, choosing ϱ sufficiently large; see, e.g., [Frank and Schlein 2014; Lieb and Thomas 1997]). We still denote the corresponding self-adjoint operator by H_ε with a little abuse of notation. The polaron Hamiltonian H_ε has a ground state, if $-\Delta + \mathcal{W}$ has compact resolvent by an application of the HVZ theorem analogous to the one for the Nelson model (see the aforementioned result in [Dereziński and Gérard 1999]). It is known that ground states exist also for nonconfining but suitably regular external potentials $\mathcal{W}$.

Proof of Proposition 3.1. These lower and upper bounds are well known; see, e.g., [Correggi and Falconi 2018; Lieb and Thomas 1997]. The lower bound is a direct consequence of the KLMN theorem, while the upper bound is proved using coherent states for the field in a fashion that is completely analogous to the one discussed for the Nelson model. Thus here we focus on the lower bound.

Let us introduce the unperturbed operator $H_{\text{free}} = \mathcal{K}_0 \otimes 1 + 1 \otimes \mathrm{d}\mathcal{G}_\varepsilon(1)$, as in the Nelson model. Then, for any $\Psi_\varepsilon \in \mathscr{Q}(H_{\text{free}})$, for all $\varrho > 0$ and for all $\beta > 0$, we can bound the interaction term in the polaron quadratic form via

$$\begin{aligned}\bigl|\langle\Psi_\varepsilon|\mathrm{Op}_\varepsilon^{\mathrm{Wick}}(\mathcal{V}_{<,z}(\boldsymbol{x})) - i\nabla_{\boldsymbol{x}}\cdot\mathrm{Op}_\varepsilon^{\mathrm{Wick}}(\boldsymbol{\mathcal{V}}_{>,z}(\boldsymbol{x})) + i\mathrm{Op}_\varepsilon^{\mathrm{Wick}}(\boldsymbol{\mathcal{V}}_{>,z}(\boldsymbol{x}))\cdot\nabla_{\boldsymbol{x}}|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}\bigr| \\ \leq 2\|\lambda_<\|_{\mathfrak{h}}\langle\Psi_\varepsilon|H_{\text{free}}^{1/2}|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} + 4\|\boldsymbol{\lambda}_>\|_{\mathfrak{h}}\langle\Psi_\varepsilon|H_{\text{free}}|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} \\ \leq \frac{1}{\beta}\|\lambda_<\|_{\mathfrak{h}}^2\|\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon}^2 + (\beta + 4\|\boldsymbol{\lambda}_>\|_{\mathfrak{h}})\langle\Psi_\varepsilon|H_{\text{free}}|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}.\end{aligned} \tag{4-13}$$

Obviously, the norms of $\lambda_<$ and $\boldsymbol{\lambda}_>$ depend on ϱ. However, since the norm of $\boldsymbol{\lambda}_>$ diverges as $\varrho \to 0$ and vanishes as $\varrho \to +\infty$, we can always choose $\varrho = \varrho(\beta)$ such that

$$4\|\boldsymbol{\lambda}_>\|_{\mathfrak{h}} = \beta. \tag{4-14}$$

Hence, we can bound

$$\bigl|\langle\Psi_\varepsilon|H_I|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}\bigr| \leq \sqrt{\alpha}N\Bigl[2\beta\langle\Psi_\varepsilon|H_{\text{free}}|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon} + \frac{1}{\beta}\|\lambda_<\|_{\mathfrak{h}}^2\|\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon}^2\Bigr],$$

and therefore, taking $\beta = (2\sqrt{\alpha}N)^{-1}$, we conclude that

$$E_\varepsilon \geq -2\alpha N^2\|\lambda_<\|_{\mathfrak{h}}^2, \tag{4-15}$$

where the last norm is evaluated at $\varrho((2\sqrt{\alpha}N)^{-1})$. □

Let us now prove Lemmas 3.6 and 3.7. The assumption in the former takes the following simplified form for the polaron model: assuming that there exists a finite constant C such that

$$\bigl|\langle\Psi_\varepsilon|(\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(1)^2 + 1)|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}\bigr| \leq C, \tag{4-16}$$

the convergence (3-18) holds true for any limit point in $\mathscr{W}(\Psi_\varepsilon, \varepsilon \in (0,1))$.

Proof of Lemma 3.6. Using again the splitting (4-12), we see that the term involving the quantization of $\mathcal{V}_{<,z}$ converges by Lemma 4.1. Let us consider then the other term. Analogously to the proof of Lemma 4.1, we define compact approximate identities $\{1_m\}_{m\in\mathbb{N}} \subset \mathscr{L}^\infty(L^2(\mathbb{R}^{dN}))$ as

$$1_m := \mathbb{1}_{[-m,m]}(\mathcal{K}_0).$$

We can now rewrite explicitly the term involving the quantization of $\boldsymbol{\mathcal{V}}_{>,z}$, by introducing $\boldsymbol{\xi} \in L^\infty(\mathbb{R}^d;\mathfrak{h})$ given by

$$\boldsymbol{\xi}(\boldsymbol{x};\boldsymbol{k}) := \boldsymbol{\lambda}_> e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{4-17}$$

as

$$\begin{aligned}\sqrt{\alpha}\sum_{j=1}^{N}\langle\Psi_{\varepsilon_n}|[-i\nabla_j, \mathrm{Op}_{\varepsilon_n}^{(\mathrm{Wick})}(\boldsymbol{\mathcal{V}}_>(\boldsymbol{x}_j))]|\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}} \\ = 2\sqrt{\alpha}\sum_{j=1}^{N}\mathrm{Re}\langle -i\nabla_j\Psi_{\varepsilon_n}|[a_{\varepsilon_n}^\dagger(\boldsymbol{\xi}(\boldsymbol{x}_j)) + a_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))]\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}}.\end{aligned} \tag{4-18}$$

In order to prove its convergence, we estimate

$$\begin{aligned}\left|\langle -i\nabla_j\Psi_{\varepsilon_n}|[a^{\dagger}_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))+a_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))]\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}}\right| \\ \le \left|\langle -i\nabla_j\Psi_{\varepsilon_n}|[a^{\dagger}_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))+a_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))]1_m\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}}\right| \\ + \left|\langle -i\nabla_j\Psi_{\varepsilon_n}|[a^{\dagger}_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))+a_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))](1-1_m)\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}}\right|. \qquad (4\text{-}19)\end{aligned}$$

The first term on the right-hand side converges when $n\to\infty$ and $m\in\mathbb{N}$ is fixed, thanks to [Correggi et al. 2023, Proposition 7.1]; then, a dominated convergence argument allows us to take the limit $m\to\infty$, yielding the desired result. It remains therefore to prove that the second term on the right-hand side converges to zero as $m\to\infty$, uniformly with respect to $\varepsilon\in(0,1)$. This is done as follows:

$$\begin{aligned}\left|\langle -i\nabla_j\Psi_{\varepsilon_n}|[a^{\dagger}_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))+a_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))](1-1_m)\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}}\right| \\ \le \|\nabla_j\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon}\|[a^{\dagger}_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))+a_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))](1-1_m)\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon} \\ \le 2(\varepsilon+\|\boldsymbol{\xi}\|_{L^\infty(\mathbb{R}^d;\mathfrak{h})})\|\nabla_j\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon}\|(\mathrm{d}\mathcal{G}_\varepsilon(1)+1)^{1/2}(1-1_m)\Psi_\varepsilon\|_{\mathscr{H}_\varepsilon}. \qquad (4\text{-}20)\end{aligned}$$

Thus, for all $\beta>0$, $\varepsilon\in(0,1)$ and $s>0$ and for any $c_0>|\inf\sigma(\mathcal{K}_0)|$,

$$\begin{aligned}&\left|\langle -i\nabla_j\Psi_{\varepsilon_n}|[a^{\dagger}_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))+a_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))](1-1_m)\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}}\right| \\ &\le (1+\|\boldsymbol{\xi}\|_{L^\infty})\Big[\beta\|\mathcal{K}_0^{1/2}\Psi_\varepsilon\|^2_{\mathscr{H}_\varepsilon} \\ &\qquad + \frac{1}{\beta}\|(1-1_m)(\mathcal{K}_0+c_0)^{-s/2}\|^2_{\mathscr{B}(L^2(\mathbb{R}^{dN}))}\|(\mathrm{d}\mathcal{G}_\varepsilon(1)+1)^{1/2}(\mathcal{K}_0+c_0)^{s/2}\Psi_\varepsilon\|^2_{\mathscr{H}_\varepsilon}\Big] \\ &\le 2(1+\|\boldsymbol{\xi}\|_{L^\infty})\Big(\beta+\frac{1}{\beta}\|(1-1_m)(\mathcal{K}_0+c_0)^{-s/2}\|^2_{\mathscr{B}}\Big)\langle\Psi_\varepsilon|\mathcal{K}_0+\mathcal{K}_0^{2s}+\mathrm{d}\mathcal{G}_\varepsilon(1)^2+1|\Psi_\varepsilon\rangle_{\mathscr{H}_\varepsilon}. \qquad (4\text{-}21)\end{aligned}$$

Hence, using (4-16), for any $s\le\frac{1}{2}$, we get

$$\left|\langle -i\nabla_j\Psi_{\varepsilon_n}|[a^{\dagger}_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))+a_{\varepsilon_n}(\boldsymbol{\xi}(\boldsymbol{x}_j))](1-1_m)\Psi_{\varepsilon_n}\rangle_{\mathscr{H}_{\varepsilon_n}}\right| \le C\beta_m, \qquad (4\text{-}22)$$

where we have chosen

$$\beta=\beta_m:=\|(1-1_m)(\mathcal{K}_0+c_0)^{-s/2}\|_{\mathscr{B}}=\sup_{\eta\in[(-\infty,-m)\cup(m,+\infty)]\cap\sigma(\mathcal{K}_0)}\frac{1}{(\eta+c_0)^{s/2}}\xrightarrow[m\to\infty]{}0.$$

Since the right-hand side of (4-22) is independent of ε and converges to zero as $m\to\infty$, the result is proven. □

Proof of Lemma 3.7. The proof is analogous to the one for the Nelson model. The expectation of the number operator squared is bounded via the *pull-through formula* by means of the expectation of H_ε^2. As discussed in [Correggi et al. 2023], the pull-through formula was originally proved for the renormalized Nelson Hamiltonian with a bound that is ε-dependent in [Ammari 2000]; the uniformity of such bound with respect to $\varepsilon\in(0,1)$ has been proved in [Ammari and Falconi 2017]. Since the renormalized Nelson model "contains" all type of terms appearing in the polaron model, the proof of the formula extends to the polaron model immediately, see [Olivieri 2020] for additional details.

The pull-through formula reads as follows: there exists a finite constant C (independent of ε) such that

$$\langle \Psi_\varepsilon | \mathrm{d}\mathcal{G}_\varepsilon(1)^2 | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon} \leq C \langle \Psi_\varepsilon | (H_\varepsilon + 1)^2 | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon}. \tag{4-23}$$

The expectation of H_{free} is bounded by means of the expectation of H_ε, using the KLMN inequality, already discussed in the proof of Proposition 3.1, in the very same way we used the Kato–Rellich inequality for the Nelson model. The fact that there exists a minimizing sequence such that the expectation of H_ε^2 is bounded uniformly with respect to $\varepsilon \in (0, 1)$ is also discussed in the proof for the Nelson model and it does not depend on the model at hand. We omit further details for the sake of brevity. □

It remains only to prove Lemma 3.10 for nontrapping potentials.

Proof of Lemma 3.10. The proof uses the following fact: if Ψ_ε is such that there exists $\delta \geq 1$ and a finite constant C such that

$$\left| \langle \Psi_\varepsilon | (\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(1)^\delta + 1) | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon} \right| \leq C, \tag{4-24}$$

then, if $\mathfrak{n} \in \mathscr{GW}(\Psi_\varepsilon, \varepsilon \in (0, 1))$ and $\Psi_{\varepsilon_n} \xrightarrow[n\to\infty]{\mathrm{gqc}} \mathfrak{n}$, one has that (3-30) holds true.

Such a result is proved by a combination of [Correggi et al. 2023, Propositions 2.6 and 7.1], if in these propositions Wigner measures are substituted by generalized Wigner measures and the test with compact operators of the small system is substituted by the test with the identity operator. The proof given there is generalized to this setting straightforwardly, recalling the properties of generalized Wigner measures outlined in the Appendix.

As in the proof for the Nelson model, let us check explicitly that $\mathrm{Ran}(z \mapsto \mathcal{V}_z)$ is separable in the norm operator topology.[9] By using the decomposition (4-12), we see that the term containing $\mathcal{V}_{<,z}$ has separable range, since it is equivalent to the one appearing in the Nelson model. Let us focus then on the remaining term containing the expectation of the operator $[-i\nabla_x, \boldsymbol{\mathcal{V}}_{>,z}]$. Such an operator is not bounded. Nonetheless, it is $\mathfrak{n}_{\mathcal{T}}$-integrable with $\mathcal{T} = \mathcal{K}_0 + 1$ by (4-24), provided that

$$\mathfrak{h} \ni z \mapsto \sum_{j=1}^{N} \mathcal{T}^{-1/2} [-i\nabla_j, \boldsymbol{\mathcal{V}}_{>,z}(\boldsymbol{x}_j)] \mathcal{T}^{-1/2} \in \mathscr{B}(L^2(\mathbb{R}^{dN})) \tag{4-25}$$

has separable range. Since $\mathfrak{h}$ is separable, let us denote by $\mathfrak{k} \subset \mathfrak{h}$ a countable dense subset and denote by

$$\mathcal{T}^{-1/2} \widetilde{\mathcal{V}}_{\mathfrak{k}} \mathcal{T}^{-1/2} := \left\{ \sum_j \mathcal{T}^{-1/2} [-i\nabla_j, \boldsymbol{\mathcal{V}}_{>,\zeta(\boldsymbol{x}_j)}] \mathcal{T}^{-1/2} \in \mathscr{B}(L^2(\mathbb{R}^{dN})) : \zeta \in \mathfrak{k} \right\}$$

the image of $\mathfrak{k}$ through $\mathcal{T}^{-1/2} \sum_j [-i\nabla_j, \boldsymbol{\mathcal{V}}_{>,\,\cdot}\,(\boldsymbol{x}_j)] \mathcal{T}^{-1/2}$. Now, for any $z \in \mathfrak{h}$, $\zeta \in \mathfrak{k}$ and $j = 1, \ldots, N$, we have that (recall (4-17))

$$\| \mathcal{T}^{-1/2} [-i\nabla_j, \boldsymbol{\mathcal{V}}_{>,z}(\boldsymbol{x}_j)] \mathcal{T}^{-1/2} - \mathcal{T}^{-1/2} [-i\nabla_j, \boldsymbol{\mathcal{V}}_{>,\zeta}(\boldsymbol{x}_j)] \mathcal{T}^{-1/2} \|_{\mathscr{B}(L^2(\mathbb{R}^{dN}))} \leq 4 \| \boldsymbol{\xi} \|_{L^\infty} \| z - \zeta \|_{\mathfrak{h}},$$

which implies that $\mathcal{T}^{-1/2} \widetilde{\mathcal{V}}_{\mathfrak{k}} \mathcal{T}^{-1/2}$ is dense in the image of the map (4-25) with respect to the norm topology in $\mathscr{B}(L^2(\mathbb{R}^{dN}))$. □

[9] More precisely, we prove that $\mathrm{Ran}(z \mapsto (\mathcal{K}_0 + 1)^{-1/2} \mathcal{V}_z (\mathcal{K}_0 + 1)^{-1/2})$ has separable range. This is sufficient to prove that $\mathcal{V}_{(\,\cdot\,)}$ is integrable with respect to $\mathfrak{n}$, since the latter is in the domain of $\mathcal{K}_0 + 1$.

4C. ***The Pauli–Fierz model.*** The Pauli–Fierz model describes N spinless charges (with an extended and sufficiently smooth charge distribution) interacting with the electromagnetic field in the Coulomb gauge, in three dimensions. Generalizations to other gauges, to particles with spin or to two dimensions are possible without much effort. The one-excitation Hilbert space is thus $\mathfrak{h} = L^2(\mathbb{R}^3; \mathbb{C}^2)$. Let the charge density of each particle be given by $\lambda_j(\boldsymbol{x})$, with $\lambda_j \in L^\infty(\mathbb{R}^3; L^2(\mathbb{R}^3))$, $j = 1, \dots, N$, such that $-i\nabla_j \lambda_j(\boldsymbol{x}; \boldsymbol{k}) = \boldsymbol{k}\lambda_j(\boldsymbol{x}; \boldsymbol{k})$ and let the polarization vectors be denoted $\boldsymbol{e}_p \in L^\infty(\mathbb{R}^3; \mathbb{R}^3)$, $p = 1, 2$, such that for a.e. $\boldsymbol{k} \in \mathbb{R}^3$ and $\boldsymbol{e}_p(\boldsymbol{k}) \cdot \boldsymbol{e}_{p'}(\boldsymbol{k}) = \delta_{pp'}$, $\boldsymbol{k} \cdot \boldsymbol{e}_p(\boldsymbol{k}) = 0$ (Coulomb gauge). The quasiclassical energy functional is then given by (1-17), i.e.,[10]

$$\mathcal{H}_z = \sum_{j=1}^{N} \frac{1}{2m_j}(-i\nabla_j + \boldsymbol{a}_{z,j}(\boldsymbol{x}_j))^2 + \mathcal{W}(\boldsymbol{X}) + \langle z|\omega|z\rangle_{\mathfrak{h}},$$

where the classical field is

$$\boldsymbol{a}_{z,j}(\boldsymbol{x}) = 2\mathrm{Re}\langle z|\boldsymbol{\lambda}_j(\boldsymbol{x})\rangle_{\mathfrak{h}} = 2\mathrm{Re}\sum_{p=1}^{2}\langle z_p|\lambda_j(\boldsymbol{x})\boldsymbol{e}_p\rangle_{L^2(\mathbb{R}^3)} \in \mathbb{C}^3$$

and, as usual, $\mathcal{W}$ is an external positive potential acting on the particles. Note that the field free energy is

$$\langle z|\omega|z\rangle_{\mathfrak{h}} = \sum_{p=1}^{2}\langle z_p|\omega|z_p\rangle_{L^2(\mathbb{R}^3)}.$$

The operator $\mathcal{H}_z$ is self-adjoint for all $z \in \mathfrak{h}_\omega$, with domain of self-adjointness $\mathscr{D}(\mathcal{K}_0)$, where we recall that $\mathcal{K}_0 = -\Delta + \mathcal{W}$, where in this case we adopt the notation

$$-\Delta = \sum_{j=1}^{N} -\frac{\Delta_j}{2m_j}.$$

The quasiclassical Wick quantization of $\mathcal{H}_z$ yields the Pauli–Fierz Hamiltonian in (1-10):

$$H_\varepsilon = \sum_{j=1}^{N} \frac{1}{2m_j}(-i\nabla_j + \boldsymbol{A}_{\varepsilon,j}(\boldsymbol{x}_j))^2 + \mathcal{W}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) + 1 \otimes \mathrm{d}\mathcal{G}_\varepsilon(\omega),$$

where

$$\boldsymbol{A}_{\varepsilon,j}(\boldsymbol{x}) = a_\varepsilon^\dagger(\boldsymbol{\lambda}_j(\boldsymbol{x})) + a_\varepsilon(\boldsymbol{\lambda}_j(\boldsymbol{x})) = \sum_{p=1}^{2}(a_{\varepsilon,p}^\dagger(\lambda_j(\boldsymbol{x})\boldsymbol{e}_p) + a_{\varepsilon,p}(\lambda_j(\boldsymbol{x})\boldsymbol{e}_p))$$

is the quantized magnetic potential. The Pauli–Fierz Hamiltonian is self-adjoint on $\mathscr{D}(\mathcal{K}_0 + \mathrm{d}\mathcal{G}_\varepsilon(\omega))$, provided that for almost all $\boldsymbol{x} \in \mathbb{R}^3$ and for all $j = 1, \dots, N$, we have $\lambda_j(\boldsymbol{x}) \in \mathscr{Q}(\omega + \omega^{-1})$ (see [Falconi 2015; Hasler and Herbst 2008; Hiroshima 2000; 2002; Matte 2017]), that we assumed in (1-11). The Pauli–Fierz Hamiltonian has a ground state for suitable choices of the potential $\mathcal{W}$, e.g., if it is the sum of single particle and pair potentials with suitable properties (clustering, binding, etc.); see, e.g., [Arai et al. 1999; Gérard 2000; Griesemer et al. 2001; Hiroshima 2001]. In particular, this holds true when the field is massive [Gérard 2000], i.e., for $\omega > 0$. As for the other models, we refrain from giving a detailed

[10] Without loss of generality, we fix the charge $e = 1$ since it does not play any relevant role in these arguments.

description of the conditions allowing them to have a ground state, since for our purposes it is sufficient that a ground state does exist in some cases.

Proof of Proposition 3.1. The lower bound follows from the diamagnetic inequality [Matte 2017]:

$$\langle \Psi_\varepsilon | -\Delta_j | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon} \leq \langle \Psi_\varepsilon | (-i\nabla_j + \boldsymbol{A}_{\varepsilon,j}(\boldsymbol{x}_j))^2 | \Psi_\varepsilon \rangle_{\mathscr{H}_\varepsilon}, \tag{4-26}$$

which in particular implies that H_ε is positive. The upper bound is proved using coherent states for the field, analogously to the Nelson model and the polaron. □

Let us now prove Lemmas 3.6 and 3.7 for the Pauli–Fierz model. The former takes the following form.

Proof of Lemma 3.6. The "potential" (1-16) is composed of two parts:

$$\mathcal{V}_z(\boldsymbol{x}) = 2\sum_{j=1}^{N} \frac{1}{m_j} [-i \mathrm{Re}\langle z | \boldsymbol{\lambda}_j(\boldsymbol{x}) \rangle_{\mathfrak{h}} \cdot \nabla_j + (\mathrm{Re}\langle z | \boldsymbol{\lambda}_j(\boldsymbol{x}) \rangle_{\mathfrak{h}})^2]$$

as well as its Wick quantization. The convergence of the quantization of the second term is perfectly analogous to the one given for the Nelson model in Lemma 4.1. The proof of convergence for the quantization of the term involving the gradient is given in the proof of Lemma 3.6 for the polaron. □

Proof of Lemma 3.7. The proof follows from the next estimate, due to F. Hiroshima, and whose detailed proof will be given in [Ammari et al. 2022]. There exists a finite constant $C > 0$ such that, for all $\Psi_\varepsilon \in \mathscr{D}(H_{\text{free}})$,

$$\| H_{\text{free}} \Psi_\varepsilon \|_{\mathscr{H}_\varepsilon} \leq C \| H_\varepsilon \Psi_\varepsilon \|_{\mathscr{H}_\varepsilon}. \tag{4-27}$$

Let us remark that the expectation of

$$\mathcal{K}_0 = -\sum_j (1/(2m_j))\Delta_j + \mathcal{W}$$

could also be bounded by means of the expectation of H_ε using the diamagnetic inequality (4-26). Hence if $\omega > 0$, (3-19) could be proved combining the diamagnetic inequality and the pull-through formula (4-23).

Finally, the fact that there exists a minimizing sequence such that the expectation of H_ε^2 is bounded uniformly with respect to $\varepsilon \in (0, 1)$ is also discussed in the proof of Lemma 3.7 for the Nelson model. □

It remains only to prove Lemma 3.10 for nontrapped systems.

Proof of Lemma 3.10. The proof here is obtained combining the proofs given for the Nelson and polaron models. In fact, the quadratic terms can be treated exactly as the linear terms in the Nelson model, and the gradient terms are equivalent to those appearing in the polaron. □

Appendix: Algebraic state-valued measures

The quasiclassical Wigner measures are state-valued by construction [Correggi et al. 2023; Falconi 2018b]. In other words, quasiclassical measures are countably additive (in a sense to be clarified below) measures on the measurable phase space of classical fields, taking values in quantum states, or, more generally, in the Banach cone $\mathfrak{A}'_+$ of positive elements in the dual of a C*-algebra $\mathfrak{A}$. In addition, the quasiclassical symbols are measurable functions from the phase space to a W*-algebra $\mathfrak{B} \supseteq \mathfrak{A}$ of observables (operators), where $\mathfrak{A}$ is supposed to be an ideal of $\mathfrak{B}$. It is therefore necessary to properly define integration of operator-valued symbols with respect to a state-valued measure. In this appendix we collect some technical properties of state-valued measures and integration, from a general algebraic standpoint that includes both state-valued and generalized state-valued measures, as used throughout the paper. The ideas developed here in great generality are particularly suited for what we called generalized state-valued measures, and they are mostly taken from [Bartle 1956; Neeb 1998]. In fact, if state-valued measures have been already studied in semiclassical analysis and adiabatic theories (see [Balazard-Konlein 1985; Fermanian-Kammerer and Gérard 2002; Gérard 1991; Gérard et al. 1991; Teufel 2003]), the reader might not be so familiar with generalized state-valued measures. Since for the latter there is no Radon–Nikodým property, their description is more abstract, and there are some limitations, especially concerning integration of operator-valued functions. This justifies the abstract approach followed in this appendix.

A1. ***Algebraic state-valued measures.*** Let $\mathfrak{A}$ be a C*-algebra and denote by $\mathfrak{A}'_+$ the cone of positive elements in the dual of $\mathfrak{A}$. In addition, let (X, Σ) be a measurable space. There are two *equivalent* ways of defining an $\mathfrak{A}'_+$-valued measure on (X, Σ).

Definition A.1 (state-valued measure [Neeb 1998]). A family of real-valued measures $(\mu_A)_{A\in\mathfrak{A}_+}$ defines a weak* σ-additive measure $\mathfrak{m} : \Sigma \to \mathfrak{A}'_+$ as

$$[\mathfrak{m}(S)](A_1 - A_2 + iA_3 - iA_4) = \mu_{A_1}(S) - \mu_{A_2}(S) + i\mu_{A_3}(S) - i\mu_{A_4}(S),$$

for any $S \in \Sigma$ and $A_1, A_2, A_3, A_4 \in \mathfrak{A}_+$, if and only if for any $A, B \in \mathfrak{A}_+$ and $\lambda \in \mathbb{R}_+$, we have $\mu_{\lambda A+B} = \lambda\mu_A + \mu_B$.

Definition A.2 (algebraic state-valued measure [Bartle 1956]). An application $\mathfrak{m} : \Sigma \to \mathfrak{A}'_+$ is a measure if and only if $\mathfrak{m}(\varnothing) = 0$, and for any family $(S_n)_{n\in\mathbb{N}} \subset \Sigma$ of mutually disjoint measurable sets,

$$\mathfrak{m}\Bigl(\bigcup_{n\in\mathbb{N}} S_n\Bigr) = \sum_{n\in\mathbb{N}} \mathfrak{m}(S_n),$$

where the right-hand side converges unconditionally in the norm of $\mathfrak{A}'$.

It is clear that any $\mathfrak{m}$ satisfying Definition A.2 satisfies also Definition A.1, since σ-additivity in norm implies weak* σ-additivity. The converse, i.e., that an $\mathfrak{m}$ satisfying Definition A.1 also satisfies Definition A.2 is nontrivial, and follows from properties of uniform boundedness in Banach spaces, as proved in [Dunford 1938, Chapter II]. We use these two definitions interchangeably, depending on the context. Let us remark that with the definitions above, any state-valued measure is automatically finite, since $\mathfrak{m}(X) \in \mathfrak{A}'_+$. Actually, in the main body of the paper, we consider probability measures, i.e., $\|\mathfrak{m}(X)\|_{\mathfrak{A}'} = 1$.

Remark A.3 (state-valued and generalized state-valued measures). The state-valued measures used in the paper correspond to choosing $\mathfrak{A} = \mathscr{L}^\infty(\mathscr{H})$; generalized state-valued measures are in a subset of the measures obtained by picking $\mathfrak{A} = \mathscr{L}^1(\mathscr{H})$.

For algebraic state-valued (cylindrical) measures on vector spaces, Bochner's theorem holds, and the Fourier transforms are completely positive maps that are weak* continuous when restricted to any finite-dimensional subspace; see [Falconi 2018b] for additional details. An algebraic state-valued measure is also *monotone*:

Lemma A.4. *For any* $S_1 \subseteq S_2 \in \Sigma$,

$$\mathfrak{m}(S_1) \leq \mathfrak{m}(S_2),$$

i.e., $\mathfrak{m}(S_2) - \mathfrak{m}(S_1) \in \mathfrak{A}'_+$.

Proof. The scalar measures $\mu_A,\ A \in \mathfrak{A}_+$, are monotonic. Therefore, for all $A \in \mathfrak{A}_+$,

$$[\mathfrak{m}(S_2)](A) := \mu_A(S_2) \geq \mu_A(S_1) =: [\mathfrak{m}(S_1)](A). \tag{A-1}$$

Hence, for all $A \in \mathfrak{A}_+$,

$$[\mathfrak{m}(S_2) - \mathfrak{m}(S_1)](A) \geq 0. \qquad \square$$

We can now introduce the scalar norm measure m, satisfying $\mu_A(S) \leq \|A\|_{\mathfrak{A}} m(S)$, for any $S \in \Sigma$, that proves to be a very useful tool to compare vector integrals with scalar integrals.

Definition A.5 (norm measure). Let $\mathfrak{m}$ be an algebraic state-valued measure. Then, its norm measure $m : \Sigma \to \mathbb{R}_+$ is defined as

$$m(S) := \|\mathfrak{m}(S)\|_{\mathfrak{A}'}, \tag{A-2}$$

for any measurable set S.

Using the cone properties of positive states in a C*-algebra, it is possible to prove that m is a finite measure. Let us recall that the C*-algebra $\mathfrak{A}$ may not be unital, so from now on we assume that there exists a W*-algebra $\mathfrak{B} \supseteq \mathfrak{A}$. If $\mathfrak{A} = \mathscr{L}^\infty(\mathscr{K})$ — the compact operators on a separable Hilbert space $\mathscr{K}$ — and $\mathfrak{B} = \mathscr{B}(\mathscr{K})$, it is well known that the aforementioned property is satisfied: $\mathfrak{A}$ is actually in this case a two-sided ideal of $\mathfrak{B}$. Let us denote by $e \in \mathfrak{B}$ the identity element.

Proposition A.6 (properties of the norm measure). *Let* $\mathfrak{m}$ *be an algebraic state-valued measure. Then its norm measure* m *is a finite measure on* (X, Σ) *and* $\mathfrak{m} \ll m$.

Proof. The proof that $m(\varnothing) = 0$ and $m(X) < +\infty$ follows immediately from the definition, while σ-additivity is proved as follows: Let $(S_n)_{n\in\mathbb{N}} \subset \Sigma$ be a family of mutually disjoint measurable sets. We are going to prove that, for any $N \in \mathbb{N}$,

$$m\biggl(\bigcup_{n=1}^{N} S_n\biggr) = \sum_{n=1}^{N} m(S_n). \tag{A-3}$$

Indeed, let $(e_\alpha)_{\alpha\in I} \subset \mathfrak{A}_+$ be an approximate identity of $e \in \mathfrak{B}$. It is well known that for any $\omega \in \mathfrak{A}'_+$ we have $\|\omega\|_{\mathfrak{A}'} = \lim_{\alpha\in I} \omega(w_\alpha)$. Hence, by Definition A.1 and Definition A.5,

$$m\biggl(\bigcup_{n=1}^{N} S_n\biggr) = \lim_{\alpha\in I} \mathfrak{m}\biggl(\bigcup_{n=1}^{N} S_n\biggr)(e_\alpha) = \lim_{\alpha\in I} \mu_{e_\alpha}\biggl(\bigcup_{n=1}^{N} S_n\biggr) = \lim_{\alpha\in I} \sum_{n=1}^{N} \mu_{e_\alpha}(S_n) = \sum_{n=1}^{N} m(S_n).$$

Next, we show

$$\lim_{N\to\infty} m\Big(\bigcup_{n\in\mathbb{N}} S_n\Big) - \sum_{n=1}^{N} m(S_n) = 0, \tag{A-4}$$

which directly implies σ-additivity: Using again the approximate identity on the left-hand side, we obtain

$$\lim_{N\to\infty}\lim_{\alpha\in I} m\Big(\bigcup_{n\in\mathbb{N}} S_n\Big) - \sum_{n=1}^{N} \mu_{e_\alpha}(S_n).$$

We know that every μ_{e_α}, $\alpha\in I$, is σ-additive, and therefore that $\lim_{N\to\infty}\sum_{n=1}^{N}\mu_{e_\alpha}(S_n) = \mu_{e_\alpha}\big(\bigcup_{n\in\mathbb{N}} S_n\big)$ and $\lim_{\alpha\in I}\mu_{e_\alpha}\big(\bigcup_{n\in\mathbb{N}} S_n\big) = m\big(\bigcup_{n\in\mathbb{N}} S_n\big)$. Hence, it remains to show that the limits in N and α can be exchanged. In order to do that, it suffices to show that the limit in α exists uniformly with respect to N:

$$\begin{aligned}
\lim_{\alpha\in I}\sup_{N\in\mathbb{N}}\left| m\Big(\bigcup_{n=1}^{N} S_n\Big) - \sum_{n=1}^{N}\mu_{e_\alpha}(S_n)\right| &= \lim_{\alpha\in I}\sup_{N\in\mathbb{N}}\left| m\Big(\bigcup_{n=1}^{N} S_n\Big) - \mu_{e_\alpha}\Big(\bigcup_{n=1}^{N} S_n\Big)\right| \\
&= \lim_{\alpha\in I}\sup_{N\in\mathbb{N}}(m-\mu_{e_\alpha})\Big(\bigcup_{n=1}^{N} S_n\Big) \\
&\le \lim_{\alpha\in I}(m-\mu_{e_\alpha})(X) = 0,
\end{aligned} \tag{A-5}$$

where we have used finite additivity of m and $m-\mu_{e_\alpha}$ and the fact that for any $S\in\Sigma$, $\mu_{e_\alpha}(S)\le m(S)$.

It remains to prove that $\mathfrak{m}$ is absolutely continuous with respect to m. For absolute continuity of a vector measure with respect to a scalar measure, we adopt the definition of [Diestel and Uhl 1977, Section I.2, Definition 3]. Since both $\mathfrak{m}$ and m are countably additive, it suffices to prove that, for any $S\in\Sigma$, $m(S)=0$ implies $\mathfrak{m}(S)=0$. However, since $m(S)=\|\mathfrak{m}(S)\|_{\mathfrak{A}'}$ and $\|\cdot\|_{\mathfrak{A}'}$ is a norm, then the aforementioned implication follows directly by the properties of norms. □

A2. ***Integration of scalar functions.*** The theory of integration for algebraic state-valued measures could be done in a unified way for scalar- and operator-valued functions. However, it is instructive to deal with scalar functions first. Let us recall that a function $g : X\to\mathbb{R}^+$ is simple if there exist a number $N\in\mathbb{N}$, mutually disjoint measurable sets $S_1,\dots,S_N\in\Sigma$ and nonnegative numbers $c_1,\dots,c_N\in\mathbb{R}^+$, such that, for all $x\in X$,

$$g(x) = \sum_{j=1}^{N} c_j \mathbb{1}_{S_j}(x), \tag{A-6}$$

where $\mathbb{1}_{S_j}$ is the characteristic function of the set S_j. Integration of simple functions with respect to an algebraic state-valued measure μ is straightforwardly defined as

$$\int_X \mathrm{d}\mathfrak{m}(x) g(x) = \sum_{j=1}^{N} c_j \mathfrak{m}(S_j) \in \mathfrak{A}'_+. \tag{A-7}$$

The integral of a nonsimple function can be defined again in two equivalent ways.

Definition A.7 (integrability I [Neeb 1998, Lemma I.12]). We say that a measurable function $f : X \to \mathbb{R}^+$ is $\mathfrak{m}$-integrable if and only if f is μ_A-integrable for any $A \in \mathfrak{A}_+$. Furthermore, its integral belongs to $\mathfrak{A}'_+$ and is uniquely defined by the integral with respect to μ_A, i.e.,

$$\begin{aligned}\Big(\int_S \mathrm{d}\mathfrak{m}(x) f(x)\Big)&(A_1 - A_2 + iA_3 - iA_4) \\ &= \int_S \mathrm{d}\mu_{A_1}(x) f(x) - \int_S \mathrm{d}\mu_{A_2}(x) f(x) + i \int_S \mathrm{d}\mu_{A_3}(x) f(x) - i \int_S \mathrm{d}\mu_{A_4}(x) f(x), \quad \text{(A-8)}\end{aligned}$$

for any $A_1, A_2, A_3, A_4 \in \mathfrak{A}_+$.

Definition A.8 (integrability II [Bartle 1956, Definition 1]). We say that a measurable function $f : X \to \mathbb{R}^+$ is $\mathfrak{m}$-integrable if and only if for any $S \in \Sigma$ the sequence of simple integrals

$$\Big\{\int_X \mathrm{d}\mathfrak{m}(x) f_n(x) \mathbb{1}_S(x)\Big\}_{n\in\mathbb{N}} \in \mathfrak{A}'$$

is Cauchy, where $(f_n)_{n\in\mathbb{N}}$ is any approximation of f in terms of simple functions. The integral is then defined as

$$\int_S \mathrm{d}\mathfrak{m}(x) f(x) = \lim_{n\to\infty} \int_X \mathrm{d}\mathfrak{m}(x) f_n(x) \mathbb{1}_S(x), \quad \text{(A-9)}$$

and it is independent of the chosen approximation.

In both cases one says that a complex function $f : X \to \mathbb{C}$ is μ-integrable if and only if $|f|$ is $\mathfrak{m}$-integrable and, in this case, its integral is given by the complex combination of the integrals of its real positive, real negative, imaginary positive and imaginary negative parts.

Since the weak* and strong limits coincide if they both exist, it follows that the integrals of a function that is $\mathfrak{m}$-integrable with respect to Definitions A.7 and A.8 coincide. In addition, if f is $\mathfrak{m}$-integrable in the "strong" sense of Definition A.8, then it is also $\mathfrak{m}$-integrable in the weak* sense of Definition A.7. It remains to show that if f is $\mathfrak{m}$-integrable in the sense of Definition A.7, then it is $\mathfrak{m}$-integrable in the sense of Definition A.8, but this can be done by exploiting the norm measure m.

Lemma A.9. *If a measurable function $f : X \to \mathbb{R}^+$ is $\mathfrak{m}$-integrable in the sense of Definition A.7, then it is m-integrable as well.*

Proof. If f is $\mathfrak{m}$-integrable, then for any $S \in \Sigma$, $\int_S \mathrm{d}\mu_A(x) f(x)$ is finite and nonnegative for any $A \in \mathfrak{A}_+$. Applying [Neeb 1998, Lemma I.5], we deduce that there exists a finite constant C, depending only on S, $\mathfrak{m}$, and f, such that

$$\int_S \mathrm{d}\mu_A(x) f(x) \le C \|A\|_{\mathfrak{A}}. \quad \text{(A-10)}$$

Now, let $(f_n)_{n\in\mathbb{N}}$ be a simple pointwise nondecreasing approximation of f from below. Then, by the monotone convergence theorem,

$$\int_S \mathrm{d}m(x) f(x) = \lim_{n\to\infty} \int_X \mathrm{d}m(x) f_n(x) \mathbb{1}_S(x).$$

Hence, by Definition A.5 and μ_{e_α}-integrability of f,

$$\int_X \mathrm{d}m(x) f_n(x) \mathbb{1}_S(x) = \lim_{\alpha\in I} \int_X \mathrm{d}\mu_{e_\alpha}(x) f_n(x) \mathbb{1}_S(x) \le \lim_{\alpha\in I} \int_S \mathrm{d}\mu_{e_\alpha}(x) f(x) \le C \lim_{\alpha\in I} \|e_\alpha\|_{\mathfrak{A}} \le C, \quad \text{(A-11)}$$

and taking the limit $n \to \infty$, we get the result. □

Proposition A.10 (equivalence of Definitions A.7 and A.8). *If a measurable function $f : X \to \mathbb{R}^+$ is $\mathfrak{m}$-integrable in the sense of Definition A.7, then it is $\mathfrak{m}$-integrable in the sense of Definition A.8. In addition, for any $S \in \Sigma$,*

$$\left\| \int_S \mathrm{d}\mathfrak{m}(x) f(x) \right\|_{\mathfrak{A}'} \le \int_S \mathrm{d}m(x) f(x). \quad \text{(A-12)}$$

Proof. We prove that

$$\left\{ \int_S \mathrm{d}\mathfrak{m}(x) f_n(x) \right\}_{n\in\mathbb{N}} \in \mathfrak{A}'_+$$

is a Cauchy sequence, where $(f_n)_{n\in\mathbb{N}}$ is a nondecreasing simple approximation of f. Observe that for any $n \ge m \in \mathbb{N}$, $f_n - f_m$ is a simple positive function, which can be written as

$$f_n - f_m = \sum_{j=1}^{N(n,m)} c_j^{(n,m)} \mathbb{1}_{S_j^{(n,m)}}. \quad \text{(A-13)}$$

Hence,

$$\left\| \int_S \mathrm{d}\mathfrak{m}(x)(f_n(x) - f_m(x)) \right\|_{\mathfrak{A}'} \le \sum_{j=1}^{N(n,m)} c_j^{(n,m)} m(S_j^{(n,m)} \cap S) = \int_S \mathrm{d}m(x)(f_n - f_m)(x) \xrightarrow[n,m\to\infty]{} 0, \quad \text{(A-14)}$$

where in the last limit we have used the dominated convergence theorem, since $f_n - f_m \le 2f$, and f is m-integrable by Lemma A.9. This proves both $\mathfrak{m}$-integrability of f in the sense of Definition A.8, and the bound (A-12). □

Therefore, the two definitions are indeed equivalent: Definition A.8 has the advantage of identifying constructively the integral as the limit of the integrals of simple approximations of the integrand, while Definition A.7 is useful to prove properties of the integral. The integral defined above is indeed linear in the integrand and monotonic.

Lemma A.11. *Let $f, g : X \to \mathbb{R}$ be two $\mathfrak{m}$-integrable functions. If for $\mathfrak{m}$-a.e. $x \in X$ we have that $g(x) \le f(x)$, then*

$$\int_X \mathrm{d}\mathfrak{m}(x)(f(x) - g(x)) \in \mathfrak{A}'_+. \quad \text{(A-15)}$$

Proof. The result follows from Definition A.7 and monotonicity of the usual integral. □

The dominated convergence theorem holds in a general form (see Theorems A.17 and A.18 below), which in particular implies that it applies to scalar functions.

A3. ***Integration of operator-valued functions.*** The integration of operator-valued functions is defined similarly to Definition A.8. Let us discuss first the integration of simple operator-valued functions and the approximation with simple functions in this context. An operator-valued function $g : X \to \mathfrak{B}$ is simple if there exist $N \in \mathbb{N}$, mutually disjoint measurable sets $S_1, \dots, S_N \in \Sigma$ and $c_1, \dots, c_N \in \mathfrak{B}$ such that for all $x \in X$,

$$g(x) = \sum_{j=1}^{N} c_j \mathbb{1}_{S_j}(x). \tag{A-16}$$

Let us recall that since $\mathfrak{A} \subset \mathfrak{B}$, for any $\omega \in \mathfrak{A}'$ and $B \in \mathfrak{B}$, we can define $\omega \circ B \in \mathfrak{A}'$ as

$$(\omega \circ B)(\,\cdot\,) := \omega(\,\cdot\, B) \quad \text{or} \quad (\omega \circ B)(\,\cdot\,) := \omega(B\,\cdot\,), \tag{A-17}$$

depending on which side $\mathfrak{A}$ is an ideal of $\mathfrak{B}$. If it is a two-sided ideal, both definitions are equivalent. Keeping this definition in mind, we can define the integral of simple functions as

$$\int_X \mathrm{d}\mathfrak{m}(x) g(x) = \sum_{j=1}^{N} \mathfrak{m}(S_j) \circ c_j \in \mathfrak{A}'. \tag{A-18}$$

Next, we recall hypotheses under which an operator-valued function admits a simple approximation.

Proposition A.12 (simple approximation [Cohn 2013, Proposition E.2]). *Let $f : X \to \mathfrak{B}$ be a measurable function. If $f(X)$ is separable, then f admits a simple approximation, i.e., there exists a sequence $\{f_n\}_{n\in\mathbb{N}}$ of simple functions such that for all $x \in X$ and $n \in \mathbb{N}$,*

$$\|f_n(x)\|_{\mathfrak{B}} \le \|f(x)\|_{\mathfrak{B}} \quad \textit{and} \quad \lim_{n\to\infty} \|f(x) - f_n(x)\|_{\mathfrak{B}} = 0. \tag{A-19}$$

Due to this result, in the following we only consider operator-valued functions with *separable range*, even if not stated explicitly.

Definition A.13 (integrability III). A measurable function with separable range $f : X \to \mathfrak{B}$ is $\mathfrak{m}$-integrable if and only if, for any $S \in \Sigma$, the sequence of simple integrals

$$\left\{ \int_X \mathrm{d}\mathfrak{m}(x) f_n(x) \mathbb{1}_S(x) \right\}_{n\in\mathbb{N}} \in \mathfrak{A}' \tag{A-20}$$

is Cauchy, where $\{f_n\}_{n\in\mathbb{N}}$ is any approximation of f in terms of simple functions. The integral is then defined as

$$\int_S \mathrm{d}\mathfrak{m}(x) f(x) = \lim_{n\to\infty} \int_X \mathrm{d}\mathfrak{m}(x) f_n(x) \mathbb{1}_S(x), \tag{A-21}$$

and it is independent of the chosen approximation.

Definition A.14 (absolute integrability). A measurable function with separable range $f : X \to \mathfrak{B}$ is $\mathfrak{m}$-absolutely integrable if and only if $\|f(\,\cdot\,)\|_{\mathfrak{B}}$ is m-integrable.

In fact, any $\mathfrak{m}$-absolutely integrable function is also $\mathfrak{m}$-integrable.

Proposition A.15 (integrability and absolute integrability). *Let $f : X \to \mathfrak{B}$ be an $\mathfrak{m}$-absolutely integrable function. Then f is also $\mathfrak{m}$-integrable and, for all $S \in \Sigma$,*

$$\left\| \int_S \mathrm{d}\mathfrak{m}(x) f(x) \right\|_{\mathfrak{A}'} \leq \int_S \mathrm{d}m(x) \| f(x) \|_{\mathfrak{B}}. \tag{A-22}$$

Proof. The proof is completely analogous to that of Proposition A.10. We omit it for the sake of brevity. □

Corollary A.16 (integrability of bounded functions). *Any function $f : X \to \mathfrak{B}$ with separable range such that $\| f(\,\cdot\,) \|_{\mathfrak{B}}$ is m-a.e. uniformly bounded is $\mathfrak{m}$-integrable.*

We are now in a position to state two versions of the dominated convergence theorem for operator-valued functions. The second, that makes crucial use of absolute integrability, is the most convenient in our concrete applications. Note that both results easily apply to the special case of scalar functions discussed in the previous section.

Theorem A.17 (dominated convergence I [Bartle 1956, Theorem 6]). *Let $\{f_n\}_{n\in\mathbb{N}}$, $f_n : X \to \mathfrak{B}$ for all $n \in \mathbb{N}$, be a sequence of $\mathfrak{m}$-integrable operator-valued functions strongly converging $\mathfrak{m}$-a.e. to $f : X \to \mathfrak{B}$. If there exists an $\mathfrak{m}$-integrable operator-valued function g such that for all $n \in \mathbb{N}$ and $S \in \Sigma$*

$$\left\| \int_S \mathrm{d}\mathfrak{m}(x) f_n(x) \right\| \leq \left\| \int_S \mathrm{d}\mathfrak{m}(x) g(x) \right\|, \tag{A-23}$$

then f is $\mathfrak{m}$-integrable and for any $S \in \Sigma$

$$\int_S \mathrm{d}\mathfrak{m}(x) f(x) = \lim_{n\to\infty} \int_S \mathrm{d}\mathfrak{m}(x) f_n(x). \tag{A-24}$$

Theorem A.18 (dominated convergence II). *Let $\{f_n\}_{n\in\mathbb{N}}$, $f_n : X \to \mathfrak{B}$ for all $n \in \mathbb{N}$, be a sequence of operator-valued functions strongly converging μ-a.e. to $f : X \to \mathfrak{B}$. If there exists an m-integrable function $G : X \to \mathbb{R}^+$ such that $\mathfrak{m}$-a.e.*

$$\| f_n(x) \|_{\mathfrak{B}} \leq G(x), \tag{A-25}$$

then, for any $n \in \mathbb{N}$, f_n and f are $\mathfrak{m}$-absolutely integrable, and

$$\int_S \mathrm{d}\mathfrak{m}(x) f(x) = \lim_{n\to\infty} \int_S \mathrm{d}\mathfrak{m}(x) f_n(x). \tag{A-26}$$

Proof. By the dominated convergence theorem for scalar measures and functions applied to m and $\{\| f_n(\,\cdot\,) \|_{\mathfrak{B}}\}_{n\in\mathbb{N}}$, respectively, we get that $\| f_n(\,\cdot\,) \|_{\mathfrak{B}}$ and $\| f(\,\cdot\,) \|_{\mathfrak{B}}$ are both m-integrable and therefore, by Proposition A.15, it follows that f_n and f are also $\mathfrak{m}$-integrable. Now, for any $S \in \Sigma$, again by Proposition A.15,

$$\left\| \int_S \mathrm{d}\mathfrak{m}(x) (f - f_n)(x) \right\|_{\mathfrak{A}'} \leq \int_S \mathrm{d}m(x) \| (f - f_n)(x) \|_{\mathfrak{B}}.$$

Therefore by the dominated convergence theorem for m applied to the sequence of scalar functions $\{\| (f - f_n)(x) \|_{\mathfrak{B}}\}_{n\in\mathbb{N}}$, it follows that in the strong topology of $\mathfrak{A}'$,

$$\int_S \mathrm{d}\mathfrak{m}(x) f(x) = \lim_{n\to\infty} \int_S \mathrm{d}\mathfrak{m}(x) f_n(x). \qquad \square$$

A4. ***Integration of functions with values in unbounded operators.*** Let us restrict attention, for this section, to the concrete case $\mathfrak{A} = \mathscr{B}(L^2(\mathbb{R}^{dN}))$. In the applications described above, it is sometimes necessary to integrate functions from some measurable space X to the unbounded operators on $L^2(\mathbb{R}^{dN})$ (albeit with a rather explicit form). It is possible to define the integration of such functions with respect to suitable generalized state-valued measures, as already outlined in Section 2B. Let us repeat here the argument for the sake of completeness.

Let $\mathcal{T} > 0$ be an operator on $L^2(\mathbb{R}^{dN})$, possibly unbounded. A generalized state-valued measure is in the domain of $\mathcal{T}$ if and only if there exists a generalized state-valued measure $\mathfrak{n}_{\mathcal{T}}$ such that for all $\mathcal{B} \in \mathscr{B}(L^2(\mathbb{R}^{dN}))$ and for any $S \in \Sigma$,

$$\mathfrak{n}_{\mathcal{T}}(S)[\mathcal{T}^{-1/2}\mathcal{B}\mathcal{T}^{-1/2}] = \mathfrak{n}(S)[\mathcal{B}].$$

Given a measure in the domain of $\mathcal{T}$, we can integrate functions singular “at most as $\mathcal{T}$”. Let $\mathcal{F}$ be a function from X to the (closed and densely defined) operators on $L^2(\mathbb{R}^{dN})$. Then $\mathcal{F}$ is $\mathfrak{n}$-absolutely integrable, with $\mathfrak{n}$ in the domain of $\mathcal{T}$, if and only if for $\mathfrak{n}$-a.e. $x \in X$,

- $\mathcal{T}^{-1/2}\mathcal{F}(x)\mathcal{T}^{-1/2} \in \mathscr{B}(L^2(\mathbb{R}^{dN}))$;
- $\mathcal{T}^{-1/2}\mathcal{F}(x)\mathcal{T}^{-1/2}$ is $\mathfrak{n}_{\mathcal{T}}$-absolutely integrable.

Given an absolutely integrable function, we can define the integral as follows: for any $S \in \Sigma$,

$$\int_S d\mathfrak{n}(x)[\mathcal{F}(x)] = \int_S d\mathfrak{n}_{\mathcal{T}}(x)[\mathcal{T}^{-1/2}\mathcal{F}(x)\mathcal{T}^{-1/2}].$$

A5. ***Two-sided integration.*** If $\mathfrak{A}$ is a two-sided ideal of $\mathfrak{B}$, we can give a slight generalization of the operator-valued integration, to accommodate integration of one function to the left and one function to the right of the measure. We use the notations and definitions of Section A3. Let $g, h : X \to \mathfrak{B}$ be two simple functions,

$$g(x) = \sum_{j=1}^{N} c_j \mathbb{1}_{S_j}(x), \quad h(x) = \sum_{j=1}^{M} d_j \mathbb{1}_{T_j}(x).$$

In addition, for any $B, C \in \mathfrak{B}$ and for any $\omega \in \mathfrak{A}'$, let us define $B \circ \omega \circ C \in \mathfrak{A}'$ by

$$(B \circ \omega \circ C)(\,\cdot\,) := \omega(B \cdot C). \tag{A-27}$$

Hence, it is possible to define two-sided simple integration as

$$\int_X g(x) d\mathfrak{m}(x) h(x) = \sum_{j=1}^{N} \sum_{k=1}^{M} c_j \circ \mu(S_j \cap T_k) \circ d_k. \tag{A-28}$$

Moreover, if $f_1, f_2 : X \to \mathfrak{B}$ have separable range, it is straightforward to extend Definition A.13 to define the two-sided integral

$$\int_S f_1(x) d\mathfrak{m}(x) f_2(x) \in \mathfrak{A}'. \tag{A-29}$$

If the above integral exists, we say that the pair f_1, f_2 is $\mathfrak{m}$-two-sided-integrable (the order is relevant). This notion also preserves positivity: for all f such that f^*, f is $\mathfrak{m}$-two-sided-integrable, then

$$\int_S f^*(x)\mathrm{d}\mathfrak{m}(x)f(x) \in \mathfrak{A}'_+. \tag{A-30}$$

A pair of functions with separable range $f_1, f_2 : X \to \mathfrak{B}$ are $\mathfrak{m}$-two-sided-absolutely integrable if and only if $\|f_1(\,\cdot\,)\|_{\mathfrak{B}}\|f_2(\,\cdot\,)\|_{\mathfrak{B}}$ is m-integrable. The analogue of Proposition A.15 is the following.

Proposition A.19 (integrability and absolute integrability). *Let $f_1, f_2 : X \to \mathfrak{B}$ be $\mathfrak{m}$-two-sided-absolutely integrable. Then, f_1, f_2 and f_2, f_1 are both $\mathfrak{m}$-two-sided-integrable and, for all $S \in \Sigma$,*

$$\left\| \int_S f_1(x)\mathrm{d}\mathfrak{m}(x)f_2(x) \right\|_{\mathfrak{A}'} \leq \int_S \mathrm{d}m(x)\|f_1(x)\|_{\mathfrak{B}}\|f_2(x)\|_{\mathfrak{B}}, \tag{A-31}$$

with analogous bound when f_1 and f_2 are exchanged on the left-hand side.

Finally, dominated convergence applies to two-sided integration too.

Theorem A.20 (dominated convergence III). *Let $\{f_n\}_{n\in\mathbb{N}}$ and $\{g_n\}_{n\in\mathbb{N}}$, $f_n, g_n : X \to \mathfrak{B}$ for all $n \in \mathbb{N}$, be two sequences of operator-valued functions strongly converging $\mathfrak{m}$-a.e. to $f, g : X \to \mathfrak{B}$, respectively. If there exists an m-square-integrable function $G : X \to \mathbb{R}^+$ such that $\mathfrak{m}$-a.e.*

$$\|f_n(x)\|_{\mathfrak{B}} \leq G(x), \quad \|g_n(x)\|_{\mathfrak{B}} \leq G(x), \tag{A-32}$$

then, for any $n \in \mathbb{N}$, f_n, g_n and f, g are $\mathfrak{m}$-two-sided-absolutely integrable, and

$$\int_S f(x)\mathrm{d}\mathfrak{m}(x)g(x) = \lim_{n\to\infty}\int_S f_n(x)\mathrm{d}\mathfrak{m}(x)g_n(x), \tag{A-33}$$

$$\int_S g(x)\mathrm{d}\mathfrak{m}(x)f(x) = \lim_{n\to\infty}\int_S g_n(x)\mathrm{d}\mathfrak{m}(x)f_n(x). \tag{A-34}$$

A6. ***Radon–Nikodým property and push-forward.*** If an operator-valued function does not have a separable range, it may fail to have an approximation with simple functions. It is possible to give an alternative definition of integration if $\mathfrak{A}'$ is a *separable* space, as it is the case for the trace class operators on a separable Hilbert space $\mathscr{L}^1(\mathscr{K})$, thanks to the following property.

Theorem A.21 (Radon–Nikodým property [Dunford and Pettis 1940, Theorem 2.1.0]). *If $\mathfrak{A}'$ is separable, then it has the* ***Radon–Nikodým property****: for every algebraic state-valued measure $\mathfrak{m}$, there exists a function $\varrho : X \to \mathfrak{A}'_+$, which is m-Bochner-integrable and such that, for all $S \in \Sigma$,*

$$\mathfrak{m}(S) = \int_S \mathrm{d}m(x)\varrho(x). \tag{A-35}$$

The function ϱ is the ***Radon–Nikodým derivative*** *of $\mathfrak{m}$ with respect to m, denoted by $\varrho = \mathrm{d}\mathfrak{m}/\mathrm{d}m$.*

Therefore, it is natural to give the following alternative definition of integrability. Recall that for any $\Gamma \in \mathfrak{A}'$ and $B \in \mathfrak{B}$ we define $(\Gamma \circ B)(\,\cdot\,) = \Gamma(B\,\cdot\,)$ if $\mathfrak{A}$ is a left ideal of $\mathfrak{B}$, and $(\Gamma \circ B)(\,\cdot\,) = \Gamma(\,\cdot\,B)$ if $\mathfrak{A}$ is a right ideal of $\mathfrak{B}$. If $\mathfrak{A}$ is a two-sided ideal, the notation ΓB denotes indifferently either of the two. In this case, for any $B, C \in \mathfrak{B}$, we can define $(B \circ \Gamma \circ C)(\,\cdot\,) = \Gamma(B \cdot C)$.

Definition A.22 (integrability IV). Suppose that $\mathfrak{A}'$ is separable, and let $f, g : X \to \mathfrak{B}$ be measurable functions (possibly with nonseparable range) and $\mathfrak{m}$ be an algebraic state-valued measure with Radon–Nikodým derivative $\varrho = \mathrm{d}\mathfrak{m}/\mathrm{d}m$. Then f is $\mathfrak{m}$-integrable if and only if $\varrho \circ f \in \mathfrak{A}'$ is m-Bochner-integrable and, for any $S \in \Sigma$,

$$\int_S \mathrm{d}\mathfrak{m}(x) f(x) := \int_S \mathrm{d}m(x) \varrho(x) \circ f(x) \in \mathfrak{A}'. \tag{A-36}$$

If in addition $\mathfrak{A}$ is a two-sided ideal of $\mathfrak{B}$, then f, g is $\mathfrak{m}$-two-sided-integrable if and only if $f \varrho g \in \mathfrak{A}'$ is m-Bochner-integrable and, for any $S \in \Sigma$,

$$\int_S f(x) \mathrm{d}\mathfrak{m}(x) g(x) := \int_S \mathrm{d}m(x) f(x) \circ \varrho(x) \circ g(x) \in \mathfrak{A}'. \tag{A-37}$$

It is straightforward to see that Definition A.22 is equivalent to Definition A.13 and the analogous one for the two-sided integral for any f, g with separable range, and therefore Definition A.22 extends Definition A.13 to any separable $\mathfrak{A}'$. In addition, since m-Bochner-integrability is equivalent to $\mathfrak{m}$-absolute integrability, it follows that, if $\mathfrak{A}'$ is separable, then $\mathfrak{m}$-integrability is equivalent to $\mathfrak{m}$-absolute-integrability. Hence, all the results of Sections A2, A3 and A5 extend, if $\mathfrak{A}'$ is separable, to functions with nonseparable range.

Suppose now that X is a topological vector space and Σ is the corresponding Borel σ-algebra. In this context, Bochner's theorem holds for algebraic state-valued measures [Falconi 2018b]: the Fourier transform

$$\hat{\mathfrak{m}}(\xi) := \int_X \mathrm{d}\mathfrak{m}(x) e^{2i\xi(x)} \in \mathfrak{A}', \quad \text{with } \xi \in X', \tag{A-38}$$

identifies uniquely a measure. Therefore, the push-forward of an algebraic state-valued measure $\mathfrak{m}$ by means of a linear continuous map $\Phi : X \to Y$, where Y is again a topological vector space with the Borel σ-algebra, is conveniently defined using the Fourier transform, and this definition suffices for the purposes of this paper: more precisely, the push-forward measure $\Phi \sharp \mathfrak{m}$ is the measure on Y whose Fourier transform is defined by

$$(\widehat{\Phi \sharp \mathfrak{m}})(\eta) := \int_X \mathrm{d}\mathfrak{m}(x) e^{2i\eta(\Phi(x))} \in \mathfrak{A}', \quad \text{with } \eta \in Y'. \tag{A-39}$$

Acknowledgements

The authors would like to thank Z. Ammari, N. Rougerie, and F. Hiroshima for many stimulating discussions and helpful insights during the redaction of this paper. Falconi has been supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC CoG UniCoSM, grant agreement no. 724939). Olivieri has been supported by the Deutsche Forschungsgemeinschaft (DFG, project number 258734477 - SFB 1173) and by GNFM of INdAM through the Progetto Giovani 2019 "Derivation of effective theories for large quantum systems".

References

[Abdesselam and Hasler 2012] A. Abdesselam and D. Hasler, "Analyticity of the ground state energy for massless Nelson models", *Comm. Math. Phys.* **310**:2 (2012), 511–536. MR Zbl

[Ammari 2000] Z. Ammari, "Asymptotic completeness for a renormalized nonrelativistic Hamiltonian in quantum field theory: the Nelson model", *Math. Phys. Anal. Geom.* **3**:3 (2000), 217–285. MR Zbl

[Ammari and Falconi 2014] Z. Ammari and M. Falconi, "Wigner measures approach to the classical limit of the Nelson model: convergence of dynamics and ground state energy", *J. Stat. Phys.* **157**:2 (2014), 330–362. MR Zbl

[Ammari and Falconi 2017] Z. Ammari and M. Falconi, "Bohr's correspondence principle for the renormalized Nelson model", *SIAM J. Math. Anal.* **49**:6 (2017), 5031–5095. MR Zbl

[Ammari and Nier 2008] Z. Ammari and F. Nier, "Mean field limit for bosons and infinite dimensional phase-space analysis", *Ann. Henri Poincaré* **9**:8 (2008), 1503–1574. MR Zbl

[Ammari and Nier 2009] Z. Ammari and F. Nier, "Mean field limit for bosons and propagation of Wigner measures", *J. Math. Phys.* **50**:4 (2009), art. id. 042107. MR Zbl

[Ammari and Nier 2011] Z. Ammari and F. Nier, "Mean field propagation of Wigner measures and BBGKY hierarchies for general bosonic states", *J. Math. Pures Appl.* (9) **95**:6 (2011), 585–626. MR Zbl

[Ammari and Nier 2015] Z. Ammari and F. Nier, "Mean field propagation of infinite-dimensional Wigner measures with a singular two-body interaction potential", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **14**:1 (2015), 155–220. MR Zbl

[Ammari et al. 2022] Z. Ammari, M. Falconi, and F. Hiroshima, "Towards a derivation of classical electrodynamics of charges and fields from QED", preprint, 2022. arXiv 2202.05015

[Amour and Nourrigat 2015] L. Amour and J. Nourrigat, "Hamiltonian systems and semiclassical dynamics for interacting spins in QED", preprint, 2015. arXiv 1512.08429

[Amour et al. 2017] L. Amour, R. Lascar, and J. Nourrigat, "Weyl calculus in QED, I: The unitary group", *J. Math. Phys.* **58**:1 (2017), art. id. 013501. MR Zbl

[Amour et al. 2019] L. Amour, L. Jager, and J. Nourrigat, "Infinite dimensional semiclassical analysis and applications to a model in nuclear magnetic resonance", *J. Math. Phys.* **60**:7 (2019), art. id. 071503. MR Zbl

[Arai 2001] A. Arai, "Ground state of the massless Nelson model without infrared cutoff in a non-Fock representation", *Rev. Math. Phys.* **13**:9 (2001), 1075–1094. MR Zbl

[Arai et al. 1999] A. Arai, M. Hirokawa, and F. Hiroshima, "On the absence of eigenvectors of Hamiltonians in a class of massless quantum field models without infrared cutoff", *J. Funct. Anal.* **168**:2 (1999), 470–497. MR Zbl

[Balazard-Konlein 1985] A. Balazard-Konlein, "Asymptotique semi-classique du spectre pour des opérateurs à symbole opératoriel", *C. R. Acad. Sci. Paris Sér. I Math.* **301**:20 (1985), 903–906. MR Zbl

[Bartle 1956] R. G. Bartle, "A general bilinear vector integral", *Studia Math.* **15** (1956), 337–352. MR Zbl

[Betz et al. 2002] V. Betz, F. Hiroshima, J. Lőrinczi, R. A. Minlos, and H. Spohn, "Ground state properties of the Nelson Hamiltonian: a Gibbs measure-based approach", *Rev. Math. Phys.* **14**:2 (2002), 173–198. MR Zbl

[Carlone et al. 2021] R. Carlone, M. Correggi, M. Falconi, and M. Olivieri, "Emergence of time-dependent point interactions in polaron models", *SIAM J. Math. Anal.* **53**:4 (2021), 4657–4691. MR Zbl

[Cohn 2013] D. L. Cohn, *Measure theory*, 2nd ed., Birkhäuser, New York, 2013. MR Zbl

[Correggi and Falconi 2018] M. Correggi and M. Falconi, "Effective potentials generated by field interaction in the quasi-classical limit", *Ann. Henri Poincaré* **19**:1 (2018), 189–235. MR Zbl

[Correggi et al. 2019] M. Correggi, M. Falconi, and M. Olivieri, "Magnetic Schrödinger operators as the quasi-classical limit of Pauli–Fierz-type models", *J. Spectr. Theory* **9**:4 (2019), 1287–1325. MR Zbl

[Correggi et al. 2023] M. Correggi, M. Falconi, and M. Olivieri, "Quasi-classical dynamics", *J. Eur. Math. Soc.* **25**:2 (2023), 731–783. MR Zbl

[Dereziński 2003] J. Dereziński, "Van Hove Hamiltonians: exactly solvable models of the infrared and ultraviolet problem", *Ann. Henri Poincaré* **4**:4 (2003), 713–738. MR Zbl

[Dereziński and Gérard 1999] J. Dereziński and C. Gérard, “Asymptotic completeness in quantum field theory: massive Pauli–Fierz Hamiltonians”, *Rev. Math. Phys.* **11**:4 (1999), 383–450. MR Zbl

[Diestel and Uhl 1977] J. Diestel and J. J. Uhl, Jr., *Vector measures*, Math. Surv. **15**, Amer. Math. Soc., Providence, RI, 1977. MR Zbl

[Donsker and Varadhan 1983] M. D. Donsker and S. R. S. Varadhan, “Asymptotics for the polaron”, *Comm. Pure Appl. Math.* **36**:4 (1983), 505–528. MR Zbl

[Dunford 1938] N. Dunford, “Uniformity in linear spaces”, *Trans. Amer. Math. Soc.* **44**:2 (1938), 305–356. MR Zbl

[Dunford and Pettis 1940] N. Dunford and B. J. Pettis, “Linear operations on summable functions”, *Trans. Amer. Math. Soc.* **47** (1940), 323–392. MR Zbl

[Falconi 2015] M. Falconi, “Self-adjointness criterion for operators in Fock spaces”, *Math. Phys. Anal. Geom.* **18**:1 (2015), art. id. 2. MR Zbl

[Falconi 2018a] M. Falconi, “Concentration of cylindrical Wigner measures”, *Comm. Cont. Math.* **20**:5 (2018), art. id. 1750055. MR Zbl

[Falconi 2018b] M. Falconi, “Cylindrical Wigner measures”, *Doc. Math.* **23** (2018), 1677–1756. MR Zbl

[Fermanian-Kammerer and Gérard 2002] C. Fermanian-Kammerer and P. Gérard, “Mesures semi-classiques et croisement de modes”, *Bull. Soc. Math. France* **130**:1 (2002), 123–168. MR Zbl

[Frank and Gang 2020] R. L. Frank and Z. Gang, “A non-linear adiabatic theorem for the one-dimensional Landau–Pekar equations”, *J. Funct. Anal.* **279**:7 (2020), art. id. 108631. MR Zbl

[Frank and Schlein 2014] R. L. Frank and B. Schlein, “Dynamics of a strongly coupled polaron”, *Lett. Math. Phys.* **104**:8 (2014), 911–929. MR Zbl

[Fröhlich 1937] H. Fröhlich, “Theory of electrical breakdown in ionic crystals”, *Proc. A* **160**:901 (1937), 230–241.

[Georgescu et al. 2004] V. Georgescu, C. Gérard, and J. S. Møller, “Spectral theory of massless Pauli–Fierz models”, *Comm. Math. Phys.* **249**:1 (2004), 29–78. MR Zbl

[Gérard 1991] P. Gérard, “Microlocal defect measures”, *Comm. Partial Differential Equations* **16**:11 (1991), 1761–1794. MR Zbl

[Gérard 2000] C. Gérard, “On the existence of ground states for massless Pauli–Fierz Hamiltonians”, *Ann. Henri Poincaré* **1**:3 (2000), 443–459. MR Zbl

[Gérard et al. 1991] C. Gérard, A. Martinez, and J. Sjöstrand, “A mathematical approach to the effective Hamiltonian in perturbed periodic problems”, *Comm. Math. Phys.* **142**:2 (1991), 217–244. MR Zbl

[Gérard et al. 2011] C. Gérard, F. Hiroshima, A. Panati, and A. Suzuki, “Infrared problem for the Nelson model on static space-times”, *Comm. Math. Phys.* **308**:2 (2011), 543–566. MR Zbl

[Ginibre et al. 2006] J. Ginibre, F. Nironi, and G. Velo, “Partially classical limit of the Nelson model”, *Ann. Henri Poincaré* **7**:1 (2006), 21–43. MR Zbl

[Griesemer 2017] M. Griesemer, “On the dynamics of polarons in the strong-coupling limit”, *Rev. Math. Phys.* **29**:10 (2017), art. id. 173003. MR Zbl

[Griesemer et al. 2001] M. Griesemer, E. H. Lieb, and M. Loss, “Ground states in non-relativistic quantum electrodynamics”, *Invent. Math.* **145**:3 (2001), 557–595. MR Zbl

[Hasler and Herbst 2008] D. Hasler and I. Herbst, “On the self-adjointness and domain of Pauli–Fierz type Hamiltonians”, *Rev. Math. Phys.* **20**:7 (2008), 787–800. MR Zbl

[Hirokawa 2006] M. Hirokawa, “Infrared catastrophe for Nelson’s model: non-existence of ground state and soft-boson divergence”, *Publ. Res. Inst. Math. Sci.* **42**:4 (2006), 897–922. MR Zbl

[Hiroshima 2000] F. Hiroshima, “Essential self-adjointness of translation-invariant quantum field models for arbitrary coupling constants”, *Comm. Math. Phys.* **211**:3 (2000), 585–613. MR Zbl

[Hiroshima 2001] F. Hiroshima, “Ground states and spectrum of quantum electrodynamics of nonrelativistic particles”, *Trans. Amer. Math. Soc.* **353**:11 (2001), 4497–4528. MR Zbl

[Hiroshima 2002] F. Hiroshima, "Self-adjointness of the Pauli–Fierz Hamiltonian for arbitrary values of coupling constants", *Ann. Henri Poincaré* **3**:1 (2002), 171–201. MR Zbl

[Hiroshima and Matte 2022] F. Hiroshima and O. Matte, "Ground states and associated path measures in the renormalized Nelson model", *Rev. Math. Phys.* **34**:2 (2022), art. id. 2250002. MR Zbl

[Leopold et al. 2021] N. Leopold, D. Mitrouskas, and R. Seiringer, "Derivation of the Landau–Pekar equations in a many-body mean-field limit", *Arch. Ration. Mech. Anal.* **240**:1 (2021), 383–417. MR Zbl

[Lewin et al. 2014] M. Lewin, P. T. Nam, and N. Rougerie, "Derivation of Hartree's theory for generic mean-field Bose systems", *Adv. Math.* **254** (2014), 570–621. MR Zbl

[Lewin et al. 2015] M. Lewin, P. T. Nam, and N. Rougerie, "Remarks on the quantum de Finetti theorem for bosonic systems", *Appl. Math. Res. Express* **2015**:1 (2015), 48–63. MR Zbl

[Lewin et al. 2016] M. Lewin, P. T. Nam, and N. Rougerie, "The mean-field approximation and the non-linear Schrödinger functional for trapped Bose gases", *Trans. Amer. Math. Soc.* **368**:9 (2016), 6131–6157. MR Zbl

[Lieb and Seiringer 2020] E. H. Lieb and R. Seiringer, "Divergence of the effective mass of a polaron in the strong coupling limit", *J. Stat. Phys.* **180**:1-6 (2020), 23–33. MR Zbl

[Lieb and Thomas 1997] E. H. Lieb and L. E. Thomas, "Exact ground state energy of the strong-coupling polaron", *Comm. Math. Phys.* **183**:3 (1997), 511–519. MR Zbl

[Matte 2017] O. Matte, "Pauli–Fierz type operators with singular electromagnetic potentials on general domains", *Math. Phys. Anal. Geom.* **20**:2 (2017), art. id. 18. MR Zbl

[Mitrouskas 2021] D. Mitrouskas, "A note on the Fröhlich dynamics in the strong coupling limit", *Lett. Math. Phys.* **111**:2 (2021), art. id. 45. MR Zbl

[Møller 2005] J. S. Møller, "The translation invariant massive Nelson model, I: The bottom of the spectrum", *Ann. Henri Poincaré* **6**:6 (2005), 1091–1135. MR Zbl

[Neeb 1998] K.-H. Neeb, "Operator-valued positive definite kernels on tubes", *Monatsh. Math.* **126**:2 (1998), 125–160. MR Zbl

[Nelson 1964] E. Nelson, "Interaction of nonrelativistic particles with a quantized scalar field", *J. Math. Phys.* **5** (1964), 1190–1197. MR

[Olivieri 2020] M. Olivieri, *Quasi-classical dynamics of quantum particles interacting with radiation*, Ph.D. thesis, Sapienza Università di Roma, 2020, available at https://tinyurl.com/tesiolivieri.

[Parthasarathy 1967] K. R. Parthasarathy, *Probability measures on metric spaces*, Probab. Math. Statist. **3**, Academic Press, New York, 1967. MR Zbl

[Pauli and Fierz 1938] W. Pauli and M. Fierz, "Zur Theorie der Emission langwelliger Lichtquanten", *Il Nuovo Cimento* **15** (1938), 167–188. Zbl

[Pekar 1954] S. I. Pekar, *Untersuchungen über die Elektronentheorie der Kristalle*, Akad. Verlag, Berlin, 1954. Zbl

[Pizzo 2003] A. Pizzo, "One-particle (improper) states in Nelson's massless model", *Ann. Henri Poincaré* **4**:3 (2003), 439–486. MR Zbl

[Spohn 2004] H. Spohn, *Dynamics of charged particles and their radiation field*, Cambridge Univ. Press, 2004. MR Zbl

[Teufel 2003] S. Teufel, *Adiabatic perturbation theory in quantum dynamics*, Lecture Notes in Math. **1821**, Springer, 2003. MR Zbl

Received 13 Nov 2020. Revised 2 Nov 2021. Accepted 14 Feb 2022.

MICHELE CORREGGI: michele.correggi@polimi.it
Dipartimento di Matematica, Politecnico di Milano, Milan, Italy

MARCO FALCONI: marco.falconi@polimi.it
Dipartimento di Matematica, Politecnico di Milano, Milan, Italy

MARCO OLIVIERI: olivieri.math@gmail.com
Fakultät für Mathematik, Karlsruhe Institut für Technologie, Karlsruhe, Germany
Current address: *Department of Mathematics, Aarhus Universitet, Aarhus, Denmark*

ANALYSIS AND PDE
Vol. 16 (2023), No. 8, pp. 1799–1824
DOI: 10.2140/apde.2023.16.1799

A CHARACTERIZATION OF THE RAZAK–JACELON ALGEBRA

NORIO NAWATA

Combining Elliott, Gong, Lin and Niu's result and Castillejos and Evington's result, we see that if A is a simple separable nuclear monotracial C*-algebra, then $A \otimes \mathcal{W}$ is isomorphic to $\mathcal{W}$, where $\mathcal{W}$ is the Razak–Jacelon algebra. In this paper, we give another proof of this. In particular, we show that if $\mathcal{D}$ is a simple separable nuclear monotracial M_{2^∞}-stable C*-algebra which is KK-equivalent to $\{0\}$, then $\mathcal{D}$ is isomorphic to $\mathcal{W}$ without considering tracial approximations of C*-algebras with finite nuclear dimension. Our proof is based on Matui and Sato's technique, Schafhauser's idea in his proof of the Tikuisis–White–Winter theorem and properties of Kirchberg's central sequence C*-algebra $F(\mathcal{D})$ of $\mathcal{D}$. Note that some results for $F(\mathcal{D})$ are based on Elliott, Gong, Lin and Niu's stable uniqueness theorem. Also, we characterize $\mathcal{W}$ by using properties of $F(\mathcal{W})$. Indeed, we show that a simple separable nuclear monotracial C*-algebra D is isomorphic to $\mathcal{W}$ if and only if D satisfies the following properties:

(i) For any $\theta \in [0, 1]$, there exists a projection p in $F(D)$ such that $\tau_{D,\omega}(p) = \theta$.

(ii) If p and q are projections in $F(D)$ such that $0 < \tau_{D,\omega}(p) = \tau_{D,\omega}(q)$, then p is Murray–von Neumann equivalent to q.

(iii) There exists an injective homomorphism from D to $\mathcal{W}$.

1. Introduction

The Razak–Jacelon algebra $\mathcal{W}$ is a certain simple separable nuclear monotracial C*-algebra which is KK-equivalent to $\{0\}$. Note that such a C*-algebra must be stably projectionless; that is, $\mathcal{W} \otimes M_n(\mathbb{C})$ has no nonzero projections for any $n \in \mathbb{N}$. In particular, every stably projectionless C*-algebra is nonunital. Jacelon [2013] constructed $\mathcal{W}$ as an inductive limit C*-algebra of Razak's building blocks [2002]. We can regard $\mathcal{W}$ as a stably finite analogue of the Cuntz algebra $\mathcal{O}_2$. In particular, $\mathcal{W}$ is expected to play a central role in the classification theory of simple separable nuclear stably projectionless C*-algebras as $\mathcal{O}_2$ played in the classification theory of Kirchberg algebras; see, for example, [Rørdam 2002; Gabe 2020]. We refer the reader to [Elliott et al. 2020a; 2020b; Gong and Lin 2020] for recent progress in the classification of simple separable nuclear stably projectionless C*-algebras. Note that there exist many interesting examples of simple stably projectionless C*-algebras. See, for example, [Connes 1982; Elliott 1996; Kishimoto 1999; Kishimoto and Kumjian 1996; 1997; Robert 2012].

Combining Elliott, Gong, Lin and Niu's result [Elliott et al. 2020a] and Castillejos and Evington's result [2020] (see also [Castillejos et al. 2021]), we see that if A is a simple separable nuclear monotracial

This work was supported by JSPS KAKENHI Grant Number 20K03630.
MSC2020: primary 46L35; secondary 46L40, 46L80.
Keywords: stably projectionless C*-algebra, Kirchberg's central sequence C*-algebra, KK-contractible C*-algebra, Razak–Jacelon algebra.

C*-algebra, then $A \otimes \mathcal{W}$ is isomorphic to $\mathcal{W}$. This can be considered as a Kirchberg–Phillips-type absorption theorem [2000] for $\mathcal{W}$. In this paper, we give another proof of this. In our proof, we do not consider tracial approximations of C*-algebras with finite nuclear dimension. Also, we mainly consider abstract settings and do not use any classification theorem based on inductive limit structures of $\mathcal{W}$ other than Razak's classification theorem [2002]. (Actually, we need Razak's classification theorem only for $\mathcal{W} \otimes M_{2^\infty} \cong \mathcal{W}$.) We obtain a Kirchberg–Phillips-type absorption theorem for $\mathcal{W}$ as a corollary of the following theorem.

Theorem 6.1. *Let $\mathcal{D}$ be a simple separable nuclear monotracial M_{2^∞}-stable* C**-algebra which is KK-equivalent to $\{0\}$. Then $\mathcal{D}$ is isomorphic to $\mathcal{W}$.*

Our proof of the theorem above is based on Matui and Sato's technique [2012; 2014a; 2014b], Schafhauser's idea [2020a] (see also [Schafhauser 2020b]) in his proof of the Tikuisis–White–Winter theorem [Tikuisis et al. 2017] and properties of Kirchberg's central sequence C*-algebra $F(\mathcal{D})$ of $\mathcal{D}$.

Matui and Sato's technique enables us to show that certain (relative) central sequence C*-algebras have strict comparison. Note that a key concept in their technique is property (SI). This concept was introduced in [Sato 2009; 2010].

Borrowing Schafhauser's idea, we show that if $\mathcal{D}$ is a simple separable nuclear monotracial (M_{2^∞}-stable) C*-algebra which is KK-equivalent to $\{0\}$, then there exist "trace-preserving" homomorphisms from $\mathcal{D}$ to ultrapowers B^ω of certain C*-algebras B. Combining this and a uniqueness result for approximate homomorphisms from $\mathcal{D}$, we obtain an existence result, that is, existence of homomorphisms from $\mathcal{D}$ to certain C*-algebras. Schafhauser's arguments are based on extension theory (or KK-theory) and Elliott and Kucerovsky's result [2001] with a correction by Gabe [2016]. Hence Schafhauser's arguments are suitable for our purpose, that is, a study of C*-algebras which are KK-equivalent to $\{0\}$.

We studied properties of $F(\mathcal{W})$ in [Nawata 2019; 2021] by using the stable uniqueness theorem in [Elliott et al. 2020a]. In particular, we showed that $F(\mathcal{W})$ has many projections and satisfies a certain comparison theory for projections. By these properties and Connes' 2×2 matrix trick, we can show that every trace-preserving endomorphism of $\mathcal{W}$ is approximately inner. (Note that Jacelon [2013, Corollary 4.6] showed this result as an application of Razak's results [2002].) This argument is a traditional argument in the theory of operator algebras; see [Connes 1976]. In this paper, we remark that arguments in [Nawata 2019; 2021] work for a simple separable nuclear monotracial M_{2^∞}-stable C*-algebra $\mathcal{D}$ which is KK-equivalent to $\{0\}$. Also, we characterize $\mathcal{W}$ by using these properties of $F(\mathcal{W})$. Indeed, we show the following theorem.

Theorem 6.4. *Let D be a simple separable nuclear monotracial* C**-algebra. Then D is isomorphic to $\mathcal{W}$ if and only if D satisfies the following properties*:

(i) *For any $\theta \in [0, 1]$, there exists a projection p in $F(D)$ such that $\tau_{D,\omega}(p) = \theta$.*

(ii) *If p and q are projections in $F(D)$ such that $0 < \tau_{D,\omega}(p) = \tau_{D,\omega}(q)$, then p is Murray–von Neumann equivalent to q.*

(iii) *There exists an injective homomorphism from D to $\mathcal{W}$.*

This paper is organized as follows. In Section 2, we collect notation, definitions and some results. In particular, we recall Matui and Sato's technique. In Section 3, we introduce the property W, which is a key property for uniqueness results. Also, we remark that arguments in [Nawata 2019; 2021] work for more general settings. In Section 4, we show uniqueness results. First, we show that if D has property W, then every trace-preserving endomorphism of D is approximately inner. Secondly, we consider a uniqueness theorem for approximate homomorphisms from a simple separable nuclear monotracial M_{2^∞}-stable C*-algebra $\mathcal{D}$ which is KK-equivalent to $\{0\}$ for an existence result in Section 5. In Section 5, we show an existence result by borrowing Schafhauser's idea. In Section 6, we show the main results in this paper.

2. Preliminaries

In this section we shall collect notation, definitions and some results. We refer the reader to [Blackadar 2006; Pedersen 1979] for basics of operator algebras.

For a C*-algebra A, we denote by A_+ the sets of positive elements of A and by $A^\sim$ the unitization algebra of A. Note that if A is unital, then $A = A^\sim$. For $a, b \in A_+$, we say that a is *Murray–von Neumann equivalent to* b, written $a \sim b$, if there exists an element z in A such that $z^*z = a$ and $zz^* = b$. Note that $\sim$ is an equivalence relation by [Pedersen 1998, Theorem 3.5]. For $a, b \in A$, we denote by $[a, b]$ the commutator $ab - ba$. For a subset F of A and $\varepsilon > 0$, we say that a completely positive (c.p.) map $\varphi : A \to B$ is (F, ε)-*multiplicative* if

$$\|\varphi(ab) - \varphi(a)\varphi(b)\| < \varepsilon$$

for any $a, b \in F$. Let $\mathcal{Z}$ and M_{2^∞} denote the Jiang–Su algebra and the CAR algebra, respectively. We say a C*-algebra A is *monotracial* if A has a unique tracial state and no unbounded traces. In the case where A is monotracial, we denote by τ_A the unique tracial state on A unless otherwise specified.

2A. ***Razak–Jacelon algebra $\mathcal{W}$.*** The *Razak–Jacelon algebra* $\mathcal{W}$ is a certain simple separable nuclear monotracial C*-algebra which is KK-equivalent to $\{0\}$. In [Jacelon 2013], $\mathcal{W}$ is constructed as an inductive limit C*-algebra of Razak's building blocks. By Razak's classification theorem [2002], $\mathcal{W}$ is M_{2^∞}-stable, and hence $\mathcal{W}$ is $\mathcal{Z}$-stable. In this paper, we do not assume any classification theorem for $\mathcal{W}$ other than Razak's classification theorem.

2B. ***Kirchberg's central sequence* C**-algebras.*** We shall recall the definition of Kirchberg's central sequence C*-algebras [2006]. Fix a free ultrafilter ω on $\mathbb{N}$. For a C*-algebra B, put

$$c_\omega(B) := \bigl\{\{x_n\}_{n\in\mathbb{N}} \in \ell^\infty(\mathbb{N}, B) \mid \lim_{n\to\omega} \|x_n\| = 0\bigr\}, \quad B^\omega := \ell^\infty(\mathbb{N}, B)/c_\omega(B).$$

We denote by $(x_n)_n$ a representative of an element in B^ω. Let A be a C*-subalgebra of B^ω. Set

$$\mathrm{Ann}(A, B^\omega) := \{(x_n)_n \in B^\omega \cap A' \mid (x_n)_n a = 0 \text{ for any } a \in A\}.$$

Then $\mathrm{Ann}(A, B^\omega)$ is a closed ideal of $B^\omega \cap A'$. Define a (*relative*) *central sequence* C**-algebra* $F(A, B)$ of $A \subseteq B^\omega$ by

$$F(A, B) := B^\omega \cap A' / \mathrm{Ann}(A, B^\omega).$$

We identify B with the C*-subalgebra of B^ω consisting of equivalence classes of constant sequences. In the case $A = B$, we denote $F(B, B)$ by $F(B)$ and call it the *central sequence* C**-algebra of* B. If A is σ-unital, then $F(A, B)$ is unital by [Kirchberg 2006, Proposition 1.9]. Indeed, let $s = (s_n)_n$ be a strictly positive element in $A \subseteq B^\omega$. Since we have $\lim_{k\to\infty} s^{1/k}s = s$, taking a suitable sequence $\{k(n)\}_{n\in\mathbb{N}} \subset \mathbb{N}$, we obtain $s' = (s_n^{1/k(n)})_n \in B^\omega$ such that $s's = s$. Then it is easy to see that $s' \in B^\omega \cap A'$ and $[s'] = 1$ in $F(A, B)$. Note that the inclusion $B \subset B^\sim$ induces an isomorphism from $F(A, B)$ onto $F(A, B^\sim)$ because we have $[xs'] = [x]$ in $F(A, B^\sim)$ for any $x \in (B^\sim)^\omega \cap A'$.

Let τ_B be a tracial state on B. Define $\tau_{B,\omega} : B^\omega \to \mathbb{C}$ by $\tau_{B,\omega}((x_n)_n) = \lim_{n\to\omega} \tau_B(x_n)$ for any $(x_n)_n \in B^\omega$. Since ω is an ultrafilter, it is easy to see that $\tau_{B,\omega}$ is a well-defined tracial state on B^ω. The following proposition is a relative version of [Nawata 2019, Proposition 2.1].

Proposition 2.1. *Let B be a* C**-algebra with a faithful tracial state τ_B, and let A be a* C**-subalgebra of B^ω. Assume that $\tau_{B,\omega}|_A$ is a state. Then $\tau_{B,\omega}((x_n)_n) = 0$ for any* $(x_n)_n \in \mathrm{Ann}(A, B^\omega)$.

Proof. Let $\{h_\lambda\}_{\lambda\in\Lambda}$ be an approximate unit for A. Since $\tau_{B,\omega}|_A$ is a state, we have $\lim \tau_{B,\omega}(h_\lambda) = 1$. The rest of proof is same as the proof of [Nawata 2019, Proposition 2.1]. □

By the proposition above, if $\tau_{B,\omega}|_A$ is a state, then $\tau_{B,\omega}$ induces a tracial state on $F(A, B)$. We denote it by the same symbol $\tau_{B,\omega}$ for simplicity.

2C. *Invertible elements in unitization algebras.* Let $\mathrm{GL}(A^\sim)$ denote the set of invertible elements in $A^\sim$. The following proposition is trivial if $1_{A^\sim} = 1_{B^\sim}$.

Proposition 2.2. *Let $A \subseteq B$ be an inclusion of* C**-algebras. Then* $\mathrm{GL}(A^\sim) \subset \overline{\mathrm{GL}(B^\sim)}$.

Proof. Let $x \in \mathrm{GL}(A^\sim)$. There exists $\varepsilon_0 > 0$ such that for any $0 \le \varepsilon < \varepsilon_0$ we have $x + \varepsilon 1_{A^\sim} \in \mathrm{GL}(A^\sim)$ because $\mathrm{GL}(A^\sim)$ is open. Since $\mathrm{Sp}_A(x) \cup \{0\} = \mathrm{Sp}_B(x) \cup \{0\}$, we have $x + \varepsilon 1_{B^\sim} \in \mathrm{GL}(B^\sim)$ for any $0 < \varepsilon < \varepsilon_0$. Therefore $x \in \overline{\mathrm{GL}(B^\sim)}$. □

The following corollary is an immediate consequence of the proposition above.

Corollary 2.3. *Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of* C**-algebras with $A_n \subseteq A_{n+1}$, and let $A = \overline{\bigcup_{n=1}^\infty A_n}$. If $A_n \subseteq \overline{\mathrm{GL}(A_n^\sim)}$ for any $n \in \mathbb{N}$, then $A \subseteq \overline{\mathrm{GL}(A^\sim)}$.*

The following proposition is well known if B is unital. See, for example, the proof of [Schafhauser 2020a, Proposition 3.2].

Proposition 2.4. *Let B be a* C**-algebra with $B \subseteq \overline{\mathrm{GL}(B^\sim)}$. Then $B^\omega \subseteq \overline{\mathrm{GL}((B^\omega)^\sim)}$.*

Proof. We shall show only the case where B is nonunital. Let $(x_n)_n \in B^\omega$. Because of $B \subseteq \overline{\mathrm{GL}(B^\sim)}$, there exists $(z_n)_n \in (B^\sim)^\omega$ such that $z_n \in \mathrm{GL}(B^\sim)$ for any $n \in \mathbb{N}$ and $(x_n)_n = (z_n)_n$ in $(B^\sim)^\omega$. For any $n \in \mathbb{N}$, put $u_n := z_n(z_n^* z_n)^{-1/2}$. Then u_n is a unitary element and $z_n = u_n(z_n^* z_n)^{1/2}$. Note that we have $(x_n)_n = (u_n)_n(x_n^* x_n)_n^{1/2}$. For any $n \in \mathbb{N}$, there exist $y_n \in B$ and $\lambda_n \in \mathbb{C}$ such that $u_n = y_n + \lambda_n 1_{B^\sim}$ and $|\lambda_n| = 1$ because u_n is a unitary element in $B^\sim$. Since ω is an ultrafilter, there exists $\lambda_0 \in \mathbb{C}$ such that $\lim_{n\to\omega} \lambda_n = \lambda_0$. Hence

$$(u_n)_n = (y_n)_n + \lambda_0 1_{(B^\omega)^\sim} \in (B^\omega)^\sim.$$

Since

$$((y_n)_n + \lambda_0 1_{(B^\omega)^\sim})((x_n^* x_n)_n^{1/2} + \varepsilon 1_{(B^\omega)^\sim}) \to (x_n)_n$$

as $\varepsilon \to 0$, we have $(x_n)_n \in \overline{\mathrm{GL}((B^\omega)^\sim)}$. □

Note that if B has almost stable rank 1 (see [Robert 2016] for the definition), then $B \subseteq \overline{\mathrm{GL}(B^\sim)}$. Also, if B is unital, then $B \otimes \mathbb{K} \subseteq \overline{\mathrm{GL}((B \otimes \mathbb{K})^\sim)}$, where $\mathbb{K}$ is the C*-algebra of compact operators on an infinite-dimensional separable Hilbert space.

2D. *Matui and Sato's technique.* We shall review Matui and Sato's technique [2012; 2014a; 2014b]. Let B be a monotracial C*-algebra, and let A be a simple separable nuclear monotracial C*-subalgebra of B^ω. Assume that τ_B is faithful and $\tau_{B,\omega}|_A$ is a state. Consider the Gelfand–Naimark–Segal (GNS) representation π_{τ_B} of B associated with τ_B, and put

$$M := \ell^\infty(\mathbb{N}, \pi_{\tau_B}(B)'')/\{\{x_n\}_{n\in\mathbb{N}} \mid \tilde{\tau}_{B,\omega}((x_n^* x_n)_n) := \lim_{n\to\omega} \tilde{\tau}_B(x_n^* x_n) = 0\},$$

where $\tilde{\tau}_B$ is the unique normal extension of τ_B on $\pi_{\tau_B}(B)''$. Note that M is a von Neumann algebraic ultrapower of $\pi_{\tau_B}(B)''$ and $\tilde{\tau}_{B,\omega}$ is a faithful normal tracial state on M. Since B is monotracial, $\pi_{\tau_B}(B)''$ is a finite factor, and hence M is also a finite factor. Define a homomorphism ϱ from B^ω to M by $\varrho((x_n)_n) = (\pi_{\tau_B}(x_n))_n$. Kaplansky's density theorem implies that ϱ is surjective. Moreover, [Matui and Sato 2014a, Theorem 3.1] (see also [Kirchberg and Rørdam 2014, Theorem 3.3]) implies that the restriction ϱ on $B^\omega \cap A'$ is a surjective homomorphism onto $M \cap \varrho(A)'$.

Proposition 2.5. *With notation as above, $M \cap \varrho(A)'$ is a finite factor.*

Proof. Note that $\tilde{\tau}_{B,\omega}$ is the unique tracial state on M since M is a finite factor. It is enough to show that $M \cap \varrho(A)'$ is monotracial. Let τ be a tracial state on $M \cap \varrho(A)'$. Since we assume that $\tau_{B,\omega}|_A$ is a state, we see that if A is unital, then $\varrho(1_A) = 1_M$. Hence ϱ can be extended to a unital homomorphism $\varrho^\sim$ from $A^\sim$ to M, and $M \cap \varrho(A)' = M \cap \varrho^\sim(A^\sim)'$. By [Bosa et al. 2019, Lemma 3.21], there exists a positive element a in $A^\sim$ such that $\tilde{\tau}_{B,\omega}(\varrho^\sim(a)) = 1$ and $\tau(x) = \tilde{\tau}_{B,\omega}(\varrho^\sim(a)x)$ for any $x \in M \cap \varrho(A)'$. Since A is monotracial,

$$\tau(x) = \tilde{\tau}_{B,\omega}(\varrho^\sim(a)x) = \tilde{\tau}_{B,\omega}(\varrho^\sim(a))\tilde{\tau}_{B,\omega}(x) = \tilde{\tau}_{B,\omega}(x).$$

Indeed, let x_0 be a positive contraction in $M \cap \varrho(A)'$. For any $a \in A$, define $\tau'(a) := \tilde{\tau}_{B,\omega}(\varrho(a)x_0)$. Then τ' is a tracial positive linear functional on A. Since A is monotracial and $\tau_{B,\omega}|_A$ is a tracial state on A, there exists a positive number t such that $\tau'(a) = t\,\tau_{B,\omega}(a)$ for any $a \in A$. Note that if $\{h_n\}_{n\in\mathbb{N}}$ is an approximate unit for A, then $t = \lim_{n\to\infty} \tau'(h_n)$. On the other hand, we have

$$\begin{aligned}|\tilde{\tau}_{B,\omega}(x_0) - \tau'(h_n)| &= |\tilde{\tau}_{B,\omega}((1-\varrho(h_n))x_0)| = |\tilde{\tau}_{B,\omega}((1-\varrho(h_n))^{1/2}x_0(1-\varrho(h_n))^{1/2})| \\ &\le |\tilde{\tau}_{B,\omega}(1-\varrho(h_n))| = |1-\tau_{B,\omega}(h_n)| \to 0\end{aligned}$$

as $n \to \infty$. Hence $t = \tilde{\tau}_{B,\omega}(x_0)$, and $\tilde{\tau}_{B,\omega}(\varrho(a)x_0) = \tilde{\tau}_{B,\omega}(\varrho(a))\tilde{\tau}_{B,\omega}(x_0)$ for any $a \in A$. It is easy to see that this implies $\tilde{\tau}_{B,\omega}(\varrho^\sim(a)x) = \tilde{\tau}_{B,\omega}(\varrho^\sim(a))\tilde{\tau}_{B,\omega}(x)$ for any $a \in A^\sim$ and $x \in M \cap \varrho(A)'$. Therefore we have $\tau(x) = \tilde{\tau}_{B,\omega}(x)$ for any $x \in M \cap \varrho(A)'$. Consequently, $M \cap \varrho(A)'$ is monotracial. □

For $a, b \in A_+$, we say that a is *Cuntz smaller than* b, written $a \precsim b$, if there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ of A such that $\|x_n^* b x_n - a\| \to 0$. A monotracial C*-algebra B is said to have *strict comparison* if, for any $k \in \mathbb{N}$, $a, b \in M_k(B)_+$ with $d_{\tau_B \otimes \mathrm{Tr}_k}(a) < d_{\tau_B \otimes \mathrm{Tr}_k}(b)$ implies $a \precsim b$, where Tr_k is the unnormalized trace on $M_k(\mathbb{C})$ and $d_{\tau_B \otimes \mathrm{Tr}_k}(a) = \lim_{n\to\infty} \tau_B \otimes \mathrm{Tr}_k(a^{1/n})$. Using [Nawata 2013, Lemma 5.7], essentially the same proofs as [Matui and Sato 2012, Theorem 1.1; 2014a, Lemma 3.2] show the following proposition. See also the proof of [Nawata 2021, Lemma 3.6].

Proposition 2.6. *Let B be a monotracial* C**-algebra, and let A be a simple separable non-type-I nuclear monotracial* C**-subalgebra of B^ω. Assume that τ_B is faithful, $\tau_{B,\omega}|_A$ is a state and B has strict comparison. Then B has property* (*SI*) *relative to A*; *that is*, *for any positive contractions a and b in $B^\omega \cap A'$ satisfying*

$$\tau_{B,\omega}(a) = 0 \quad \textit{and} \quad \inf_{m\in\mathbb{N}} \tau_{B,\omega}(b^m) > 0,$$

there exists an element s in $B^\omega \cap A'$ such that $s^*s = a$ and $bs = s$.

By Proposition 2.1, ϱ induces a surjective homomorphism from $F(A, B)$ to $M \cap \varrho(A)'$. We denote it by the same symbol ϱ for simplicity. Using Propositions 2.5 and 2.6, essentially the same proofs as [Matui and Sato 2014a, Proposition 3.3; 2014b, Proposition 4.8] show the following proposition. See also the proof of [Nawata 2021, Proposition 3.8].

Proposition 2.7. *Let B be a monotracial* C**-algebra, and let A be a simple separable non-type-I nuclear monotracial* C**-subalgebra of B^ω. Assume that τ_B is faithful, $\tau_{B,\omega}|_A$ is a state and B has strict comparison. Then $F(A, B)$ is monotracial and has strict comparison. Furthermore*, *if a and b are positive elements in $F(A, B)$ satisfying $d_{\tau_{B,\omega}}(a) < d_{\tau_{B,\omega}}(b)$, then there exists an element r in $F(A, B)$ such that $r^*br = a$.*

3. Property W

In this section we shall introduce the property W, which is a key property in Section 4.

Definition 3.1. Let D be a simple separable nuclear monotracial C*-algebra. We say that D has *property W* if $F(D)$ satisfies the following properties:

(i) For any $\theta \in [0, 1]$, there exists a projection p in $F(D)$ such that $\tau_{D,\omega}(p) = \theta$.

(ii) If p and q are projections in $F(D)$ such that $0 < \tau_{D,\omega}(p) = \tau_{D,\omega}(q)$, then p is Murray–von Neumann equivalent to q.

By arguments in [Nawata 2019; 2021], we see that if $\mathcal{D}$ is a simple separable nuclear monotracial M_{2^∞}-stable C*-algebra which is KK-equivalent to $\{0\}$, then $\mathcal{D}$ has property W. We shall give a sketch of a proof for reader's convenience and show a slight generalization (or a relative version).

In this section, we assume that $\mathcal{D}$ is a simple separable nuclear monotracial M_{2^∞}-stable C*-algebra which is KK-equivalent to $\{0\}$ and B is a simple monotracial C*-algebra with strict comparison and $B \subseteq \overline{\mathrm{GL}(B^\sim)}$. Let Φ be a homomorphism from $\mathcal{D}$ to B^ω such that $\tau_{\mathcal{D}} = \tau_{B,\omega} \circ \Phi$. By the Choi–Effros lifting theorem, there exists a sequence $\{\Phi_n\}_{n\in\mathbb{N}}$ of contractive c.p. maps from $\mathcal{D}$ to B such that

$\Phi(x) = (\Phi_n(x))_n$ for any $x \in \mathcal{D}$. Since we assume $\tau_{\mathcal{D}} = \tau_{B,\omega} \circ \Phi$, we have $\tau_{B,\omega}|_{\Phi(\mathcal{D})}$ is a state. Hence $\tau_{B,\omega}$ is the unique tracial state on $F(\Phi(\mathcal{D}), B)$ by Proposition 2.7. The following proposition is analogous to [Nawata 2019, Proposition 4.2; 2021, Proposition 2.6].

Proposition 3.2. (i) *For any $N \in \mathbb{N}$, there exists a unital homomorphism from $M_{2^N}(\mathbb{C})$ to $F(\Phi(\mathcal{D}), B)$.*

(ii) *For any $\theta \in [0, 1]$, there exists a projection p in $F(\Phi(\mathcal{D}), B)$ such that $\tau_{B,\omega}(p) = \theta$.*

(iii) *Let h be a positive element in $F(\Phi(\mathcal{D}), B)$ such that $d_{\tau_{B,\omega}}(h) > 0$. For any $\theta \in [0, d_{\tau_{B,\omega}}(h))$, there exists a nonzero projection p in $\overline{hF(\Phi(\mathcal{D}), B)h}$ such that $\tau_{B,\omega}(p) = \theta$.*

Proof. (i) Since $\mathcal{D}$ is isomorphic to $\mathcal{D} \otimes M_{2^\infty} = \mathcal{D} \otimes \bigotimes_{n\in\mathbb{N}} M_{2^N}(\mathbb{C})$, an argument similar to that in the proof of Proposition 4.2 in [Nawata 2019], henceforth abbreviated [N19], shows that there exists a family $\{(e_{ij,m})_m\}_{i,j=1}^{2^N}$ of contractions in $\mathcal{D}^\omega \cap \mathcal{D}'$ such that

$$\left(\sum_{\ell=1}^{2^N} e_{\ell\ell,m} x\right)_m = x \quad \text{and} \quad (e_{ij,m} e_{kl,m} x)_m = (\delta_{jk} e_{il,m} x)_m$$

for any $1 \le i, j, k, l \le 2^N$ and $x \in \mathcal{D}$. Note that we have

$$\lim_{m\to\omega} \|([\Phi_n(e_{ij,m}), \Phi_n(x)])_n\| = 0, \quad \lim_{m\to\omega} \left\| \left(\sum_{\ell=1}^{2^N} \Phi_n(e_{\ell\ell,m}) \Phi_n(x) - \Phi_n(x) \right)_n \right\| = 0$$

and

$$\lim_{m\to\omega} \|((\Phi_n(e_{ij,m}) \Phi_n(e_{kl,m}) - \delta_{jk} \Phi_n(e_{il,m})) \Phi_n(x))_n\| = 0$$

for any $1 \le i, j, k, l \le 2^N$ and $x \in \mathcal{D}$. Hence, for any finite subset $F \subset \mathcal{D}$ and $\varepsilon > 0$, there exists a family of $\{(\Phi_n(e_{ij,(F,\varepsilon)}))_n\}_{i,j=1}^{2^N}$ of contractions in B^ω such that

$$\lim_{n\to\omega} \|[\Phi_n(e_{ij,(F,\varepsilon)}), \Phi_n(x)]\| < \varepsilon, \quad \lim_{n\to\omega} \left\| \sum_{\ell=1}^{2^N} \Phi_n(e_{\ell\ell,(F,\varepsilon)}) \Phi_n(x) - \Phi_n(x) \right\| < \varepsilon$$

and

$$\lim_{n\to\omega} \|(\Phi_n(e_{ij,(F,\varepsilon)}) \Phi_n(e_{kl,(F,\varepsilon)}) - \delta_{jk} \Phi_n(e_{il,(F,\varepsilon)})) \Phi_n(x)\| < \varepsilon$$

for any $1 \le i, j, k, l \le 2^N$ and $x \in F$. Let $\{F_m\}_{m\in\mathbb{N}}$ be an increasing sequence of finite subsets in $\mathcal{D}$ such that $\mathcal{D} = \overline{\bigcup_{m\in\mathbb{N}} F_m}$. We can find a sequence $\{X_m\}_{m\in\mathbb{N}}$ of elements in ω such that $X_{m+1} \subset X_m$, $\bigcap_{m\in\mathbb{N}} X_m = \varnothing$, and, for any $n \in X_m$,

$$\|[\Phi_n(e_{ij,(F_m,1/m)}), \Phi_n(x)]\| < \frac{1}{m}, \quad \left\| \sum_{\ell=1}^{2^N} \Phi_n(e_{\ell\ell,(F_m,1/m)}) \Phi_n(x) - \Phi_n(x) \right\| < \frac{1}{m}$$

and

$$\|(\Phi_n(e_{ij,(F_m,1/m)}) \Phi_n(e_{kl,(F_m,1/m)}) - \delta_{jk} \Phi_n(e_{il,(F_m,1/m)})) \Phi_n(x)\| < \frac{1}{m}$$

for any $1 \le i, j, k, l \le 2^N$ and $x \in F_m$. For any $1 \le i, j \le 2^N$, put

$$E_{ij,n} := \begin{cases} 0 & \text{if } n \notin X_1, \\ \Phi_n(e_{ij,(F_m,1/m)}) & \text{if } n \in X_m \setminus X_{m+1} \quad (m \in \mathbb{N}). \end{cases}$$

Then we have $(E_{ij,n})_n \in B^{\omega} \cap \Phi(\mathcal{D})'$,

$$\sum_{\ell=1}^{2^N} [(E_{\ell\ell,n})_n] = 1 \quad \text{and} \quad [(E_{ij,n})_n][(E_{kl,n})_n] = \delta_{jk}[(E_{il,n})_n]$$

in $F(\Phi(\mathcal{D}), B)$ for any $1 \leq i, j, k, l \leq 2^N$. Therefore there exists a unital homomorphism from $M_{2^N}(\mathbb{C})$ to $F(\Phi(\mathcal{D}), B)$.

(ii) Since $\mathcal{D}$ is isomorphic to $\mathcal{D} \otimes M_{2^\infty} = \mathcal{D} \otimes \bigotimes_{n\in\mathbb{N}} M_{2^\infty}$, an argument similar to that in the proof of [N19, Proposition 4.2] shows that there exists a positive contraction $(p_m)_m$ in $\mathcal{D}^{\omega} \cap \mathcal{D}$ such that $((p_m^2 - p_m)x)_m = 0$ for any $x \in \mathcal{D}$ and $\tau_{\mathcal{D},\omega}((p_m)_m) = \theta$. By an argument similar to that above, we obtain a projection p in $F(\Phi(\mathcal{D}), B)$ such that $\tau_{B,\omega}(p) = \theta$.

(iii) Using Proposition 2.7 instead of [N19, Proposition 4.1], we obtain the conclusion by the same argument as in the proof of [N19, Proposition 4.2]. □

The proposition above and the same arguments as in [N19, Section 4] show the following corollary.

Corollary 3.3 ((cf. [N19, Proposition 4.8])). *Let p and q be projections in $F(\Phi(\mathcal{D}), B)$ such that $\tau_{B,\omega}(p) < 1$. Then p and q are Murray–von Neumann equivalent if and only if p and q are unitarily equivalent.*

Since we assume $B \subseteq \overline{\mathrm{GL}(B^{\sim})}$, we obtain the following proposition by the same argument as in the proof of [N19, Proposition 4.9].

Proposition 3.4. *Let u be a unitary element in $F(\Phi(\mathcal{D}), B)$. Then there exists a unitary element w in $(B^{\sim})^{\omega} \cap \Phi(\mathcal{D})'$ such that $u = [w]$.*

There exists a homomorphism ρ from $F(\Phi(\mathcal{D}), B) \otimes \mathcal{D}$ to B^{ω} such that

$$\rho([(x_n)_n] \otimes a) = (x_n \Phi_n(a))_n$$

for any $[(x_n)_n] \in F(\Phi(\mathcal{D}), B)$ and $a \in \mathcal{D}$. For a projection p in $F(\Phi(\mathcal{D}), B)$, put

$$B_p^{\omega} := \overline{\rho(p \otimes s) B^{\omega} \rho(p \otimes s)},$$

where s is a strictly positive element in $\mathcal{D}$. Define a homomorphism σ_p from $\mathcal{D}$ to B_p^{ω} by $\sigma_p(a) := \rho(p \otimes a)$ for any $a \in \mathcal{D}$. Since B has strict comparison, we see that if p is a projection in $F(\Phi(\mathcal{D}), B)$ such that $\tau_{B,\omega}(p) > 0$, then σ_p is (L, N)-full for some maps L and N by [N19, Lemma 3.5 and Proposition 3.7]. (We refer the reader to [N19, Section 3] for details of the (L, N)-fullness.) Therefore [N19, Proposition 3.3] implies the following theorem. We may regard this theorem as a variant of Elliott, Gong, Lin and Niu's stable uniqueness theorem [Elliott et al. 2020a, Corollary 3.15]; see also [Elliott and Niu 2016, Corollary 8.16]. Note that [N19, Proposition 3.3] is also based on the results in [Elliott and Kucerovsky 2001; Gabe 2016; Dadarlat and Eilers 2001; 2002].

Theorem 3.5. *Let Ω be a compact metrizable space. For any finite subsets $F_1 \subset C(\Omega)$, $F_2 \subset \mathcal{D}$ and $\varepsilon > 0$, there exist finite subsets $G_1 \subset C(\Omega)$, $G_2 \subset \mathcal{D}$, $m \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Let p be a projection in $F(\Phi(\mathcal{D}), B)$ such that $\tau_{B,\omega}(p) > 0$. For any contractive $(G_1 \odot G_2, \delta)$-multiplicative* *maps $\psi_1, \psi_2 : C(\Omega)\otimes\mathcal{D}\to B_p^{\omega}$, there exist a unitary element u in $M_{m^2+1}(B_p^{\omega})^{\sim}$ and $z_1, z_2, \dots, z_m\in\Omega$ such that*

$$\left\| u\Big(\psi_1(f\otimes b)\oplus \overbrace{\bigoplus_{k=1}^{m} f(z_k)\rho(p\otimes b)\oplus\cdots\oplus\bigoplus_{k=1}^{m} f(z_k)\rho(p\otimes b)}^{m}\Big)u^* - \psi_2(f\otimes b)\oplus \overbrace{\bigoplus_{k=1}^{m} f(z_k)\rho(p\otimes b)\oplus\cdots\oplus\bigoplus_{k=1}^{m} f(z_k)\rho(p\otimes b)}^{m}\right\| < \varepsilon$$

for any $f\in F_1$ and $b\in F_2$.

Using Proposition 2.7, Proposition 3.2 and Corollary 3.3 instead of Propositions 4.1, 4.2, and 4.8 of [N19], the same proof as [N19, Lemma 5.1] shows the following lemma.

Lemma 3.6. *Let Ω be a compact metrizable space, and let F be a finite subset of $C(\Omega)$ and $\varepsilon > 0$. Suppose that ψ_1 and ψ_2 are unital homomorphisms from $C(\Omega)$ to $F(\Phi(\mathcal{D}), B)$ such that $\tau_{B,\omega}\circ\psi_1 = \tau_{B,\omega}\circ\psi_2$. Then there exist a projection $p\in F(\Phi(\mathcal{D}), B)$, (F,ε)-multiplicative unital c.p. maps ψ_1' and ψ_2' from $C(\Omega)$ to $pF(\Phi(\mathcal{D}), B)p$, a unital homomorphism σ from $C(\Omega)$ to $(1-p)F(\Phi(\mathcal{D}), B)(1-p)$ with finite-dimensional range and a unitary element $u\in F(\Phi(\mathcal{D}), B)$ such that*

$$0 < \tau_{B,\omega}(p) < \varepsilon, \quad \|\psi_1(f) - (\psi_1'(f)+\sigma(f))\| < \varepsilon, \quad \|\psi_2(f) - u(\psi_2'(f)+\sigma(f))u^*\| < \varepsilon$$

for any $f\in F$.

The following lemma is essentially the same as [N19, Theorem 5.2] and [Nawata 2021, Theorem 5.2].

Lemma 3.7. *Let Ω be a compact metrizable space, and let F_1 be a finite subset of $C(\Omega)$ and F_2 a finite subset of $\mathcal{D}$, and let $\varepsilon > 0$. Then there exist mutually orthogonal positive elements $h_1, h_2, \dots, h_l$ in $C(\Omega)$ of norm 1 such that the following holds. If ψ_1 and ψ_2 are unital homomorphisms from $C(\Omega)$ to $F(\Phi(\mathcal{D}), B)$ such that*

$$\tau_{B,\omega}(\psi_1(h_i)) > 0, \quad 1\le\forall i\le l, \qquad \textit{and} \qquad \tau_{B,\omega}\circ\psi_1 = \tau_{B,\omega}\circ\psi_2,$$

then there exists a unitary element u in $(B^{\omega})^{\sim}$ such that

$$\|u\rho(\psi_1(f)\otimes a)u^* - \rho(\psi_2(f)\otimes a)\| < \varepsilon$$

for any $f\in F_1$, $a\in F_2$.

Proof. Take positive elements $h_1, h_2, \dots, h_l$ in $C(\Omega)$ in the same way as in the proof of [N19, Theorem 5.2]. Let ψ_1 and ψ_2 be unital homomorphisms from $C(\Omega)$ to $F(\Phi(\mathcal{D}), B)$ such that $\tau_{B,\omega}(\psi_1(h_i)) > 0$ for any $1\le i\le l$ and $\tau_{B,\omega}\circ\psi_1 = \tau_{B,\omega}\circ\psi_2$. Define homomorphisms Ψ_1 and Ψ_2 from $C(\Omega)\otimes\mathcal{D}$ to B^{ω} by

$$\Psi_1 := \rho\circ(\psi_1\otimes\mathrm{id}_{\mathcal{D}}) \quad \text{and} \quad \Psi_2 := \rho\circ(\psi_2\otimes\mathrm{id}_{\mathcal{D}}).$$

Note that there exists $\nu > 0$ such that $\tau_{B,\omega}(\psi_1(h_i)) \ge \nu$ for any $1\le i\le l$. Using Proposition 3.4, Theorem 3.5 and Lemma 3.6 instead of Corollaries 4.10, 3.8 and Lemma 5.1 in [N19], the same argument as in the proof of [N19, Theorem 5.2] shows that there exists a unitary element u in $(B^{\omega})^{\sim}$ such that

$$\|u\Psi_1(f\otimes a)u^* - \Psi_2(f\otimes a)\| < \varepsilon$$

for any $f\in F_1$, $a\in F_2$. Therefore we obtain the conclusion. □

The following theorem is a generalization of [N19, Theorem 5.3]. See also [N19, Theorem 5.3].

Theorem 3.8. *Let N_1 and N_2 be normal elements in $F(\Phi(\mathcal{D}),B)$ such that $\mathrm{Sp}(N_1)=\mathrm{Sp}(N_2)$ and $\tau_{B,\omega}(f(N_1))>0$ for any $f\in C(\mathrm{Sp}(N_1))_+\setminus\{0\}$. Then there exists a unitary element u in $F(\Phi(\mathcal{D}),B)$ such that $uN_1u^*=N_2$ if and only if $\tau_{B,\omega}(f(N_1))=\tau_{B,\omega}(f(N_2))$ for any $f\in C(\mathrm{Sp}(N_1))$.*

Proof. It is enough to show the "if" part because the "only if" part is obvious. Let $\Omega:=\mathrm{Sp}(N_1)=\mathrm{Sp}(N_2)$, and define unital homomorphisms ψ_1 and ψ_2 from $C(\Omega)$ to $F(\Phi(\mathcal{D}),B)$ by $\psi_1(f):=f(N_1)$ and $\psi_2(f):=f(N_2)$ for any $f\in C(\Omega)$. By the Choi–Effros lifting theorem, there exist sequences of unital c.p. maps $\{\psi_{1,n}\}_{n\in\mathbb{N}}$ and $\{\psi_{2,n}\}_{n\in\mathbb{N}}$ from $C(\Omega)$ to $B^{\sim}$ such that $\psi_1(f)=[(\psi_{1,n}(f))_n]$ and $\psi_2(f)=[(\psi_{2,n}(f))_n]$ for any $f\in C(\Omega)$. Let $F_1:=\{1,\iota\}\subset C(\Omega)$, where ι is the identity function on Ω, that is, $\iota(z)=z$ for any $z\in\Omega$, and let $\{F_{2,m}\}_{m\in\mathbb{N}}$ be an increasing sequence of finite subsets in $\mathcal{D}$ such that $\mathcal{D}=\overline{\bigcup_{m\in\mathbb{N}}F_{2,m}}$. For any $m\in\mathbb{N}$, applying Lemma 3.7 to F_1, $F_{2,m}$ and $1/m$, we obtain mutually orthogonal positive elements $h_{1,m}, h_{2,m},\dots,h_{l(m),m}$ in $C(\Omega)$ of norm 1. Since we have

$$\tau_{B,\omega}(\psi_1(h_{i,m}))>0,\quad 1\le\forall i\le l(m),\qquad\text{and}\qquad \tau_{B,\omega}\circ\psi_1=\tau_{B,\omega}\circ\psi_2$$

by the assumption, Lemma 3.7 implies that there exists a unitary element $(u_{m,n})_n$ in $(B^{\omega})^{\sim}$ such that

$$\|(u_{m,n})_n\rho(\psi_1(f)\otimes a)(u_{m,n}^*)_n-\rho(\psi_2(f)\otimes a)\|<\frac{1}{m}$$

for any $f\in F_1$, $a\in F_{2,m}$. By the definition of ρ, we have

$$\lim_{n\to\omega}\|u_{m,n}\psi_{1,n}(f)\Phi_n(a)u_{m,n}^*-\psi_{2,n}(f)\Phi_n(a)\|<\frac{1}{m}$$

for any $f\in F_1$, $a\in F_{2,m}$. Therefore we inductively obtain a decreasing sequence $\{X_m\}_{m\in\mathbb{N}}$ of elements in ω such that $\bigcap_{m\in\mathbb{N}}X_m=\varnothing$, and, for any $n\in X_m$,

$$\|u_{m,n}\psi_{1,n}(f)\Phi_n(a)u_{m,n}^*-\psi_{2,n}(f)\Phi_n(a)\|<\frac{1}{m}$$

for any $f\in F_1$, $a\in F_{2,m}$. Set

$$u_n:=\begin{cases}1 & \text{if } n\notin X_1,\\ u_{m,n} & \text{if } n\in X_m\setminus X_{m+1}\quad (m\in\mathbb{N}).\end{cases}$$

Then we have

$$\lim_{n\to\omega}\|u_n\Phi_n(a)u_n^*-\Phi_n(a)\|=0,\quad \lim_{n\to\omega}\|u_n\psi_{1,n}(\iota)\Phi_n(a)u_n^*-\psi_{2,n}(\iota)\Phi_n(a)\|=0$$

for any $a\in\mathcal{D}$. Therefore, $(u_n)_n\in(B^{\sim})^{\omega}\cap\Phi(\mathcal{D})'$ and $[(u_n)_n]N_1[(u_n)_n]^*=N_2$ in $F(\Phi(\mathcal{D}),B)$. Since $[(u_n)_n]$ is a unitary element in $F(\Phi(\mathcal{D}),B)$, we obtain the conclusion. □

The following corollary is an immediate consequence of the theorem above.

Corollary 3.9 (cf. [Nawata 2021, Corollary 5.4]). *Let p and q be projections in $F(\Phi(\mathcal{D}), B)$ such that $0 < \tau_{B\omega}(p) < 1$. Then p and q are unitarily equivalent if and only if $\tau_{B,\omega}(p) = \tau_{B,\omega}(q)$.*

The corollary above and the same argument as in the proof of [Nawata 2021, Corollary 5.5] show the following theorem.

Theorem 3.10. *Let p and q be projections in $F(\Phi(\mathcal{D}), B)$ such that $0 < \tau_{B,\omega}(p) \leq 1$. Then p and q are Murray–von Neumann equivalent if and only if $\tau_{B,\omega}(p) = \tau_{B,\omega}(q)$.*

By Proposition 3.2 and applying the theorem above to $B = \mathcal{D}$ and $\Phi = \mathrm{id}_{\mathcal{D}}$, we obtain the following corollary.

Corollary 3.11. *Let $\mathcal{D}$ be a simple separable nuclear monotracial M_{2^∞}-stable* C**-algebra which is KK-equivalent to $\{0\}$. Then $\mathcal{D}$ has property W.*

4. Uniqueness theorem

In this section we shall show that if D has property W, then every trace-preserving endomorphism of D is approximately inner. Furthermore, we shall consider a uniqueness theorem for approximate homomorphisms from a simple separable nuclear monotracial M_{2^∞}-stable C*-algebra $\mathcal{D}$ which is KK-equivalent to $\{0\}$ for an existence theorem in Section 5.

Let D be a simple separable nuclear monotracial C*-algebra, and let φ be a trace-preserving endomorphism of D. Define a homomorphism Φ from D to $M_2(D)$ by

$$\Phi(a) := \begin{pmatrix} a & 0 \\ 0 & \varphi(a) \end{pmatrix}$$

for any $a \in D$. Since φ is trace-preserving, we see that $\tau_{M_2(D),\omega}|_{\Phi(D)}$ is a state. Hence $\tau_{M_2(D),\omega}$ is a tracial state on $F(\Phi(D), M_2(D))$. (See Proposition 2.1.) Define homomorphisms ι_{11} and ι_{22} from $F(D)$ to $F(\Phi(D), M_2(D))$ by

$$\iota_{11}([(x_n)_n]) := \left[\left(\begin{pmatrix} x_n & 0 \\ 0 & 0 \end{pmatrix}\right)_n\right] \quad \text{and} \quad \iota_{22}([(x_n)_n]) := \left[\left(\begin{pmatrix} 0 & 0 \\ 0 & \varphi(x_n) \end{pmatrix}\right)_n\right]$$

for any $[(x_n)_n]$ in $F(D)$. It is easy to see that ι_{11} and ι_{22} are well-defined. Put $p := \iota_{11}(1)$ and $q := \iota_{22}(1)$. Note that p and q are projections in $F(\Phi(D), M_2(D))$ and if $\{h_n\}_{n\in\mathbb{N}}$ is an approximate unit for D, then

$$p = \left[\left(\begin{pmatrix} h_n & 0 \\ 0 & 0 \end{pmatrix}\right)_n\right] \quad \text{and} \quad q = \left[\left(\begin{pmatrix} 0 & 0 \\ 0 & \varphi(h_n) \end{pmatrix}\right)_n\right].$$

It can be easily checked that ι_{11} is an isomorphism from $F(D)$ onto $pF(\Phi(D), M_2(D))p$.

Lemma 4.1. *Let D be a simple separable nuclear monotracial* C**-algebra with property W. Then D is M_{2^∞}-stable, and hence D is $\mathcal{Z}$-stable.*

Proof. Since D has property W, there exists a projection p in $F(D)$ such that $\tau_{D,\omega}(p) = \frac{1}{2}$. Moreover, p is Murray–von Neumann equivalent to $1 - p$. Hence there exists a unital homomorphism from $M_2(\mathbb{C})$ to $F(D)$. By Corollary 1.13 and Proposition 4.11 in [Kirchberg 2006] (see [Blackadar et al. 1992, Proposition 2.12] for the pioneering work), D is M_{2^∞}-stable. □

The lemma above implies that if D has property W, then D has strict comparison and $D \subseteq \overline{\mathrm{GL}(D^{\sim})}$ by [Rørdam 2004a; Robert 2016]. Furthermore, $F(\Phi(D), M_2(D))$ is monotracial and has strict comparison by Proposition 2.7. The following lemma is related to [Nawata 2021, Lemma 6.2].

Lemma 4.2. *With notation as above, if D has property W, then p is Murray–von Neumann equivalent to q in $F(\Phi(D), M_2(D))$.*

Proof. For any $m \in \mathbb{N}$, there exists a projection q_m in $F(D)$ such that $\tau_{D,\omega}(q_m) = 1 - 1/m$ because D has property W. Proposition 2.7 implies that there exists a contraction r_m in $F(\Phi(D), M_2(D))$ such that $r_m^* p r_m = \iota_{22}(q_m)$. By a diagonal argument, we see that there exist a projection q' in $F(D)$ and a contraction r in $F(\Phi(D), M_2(D))$ such that $\tau_{D,\omega}(q') = 1$ and $r^* p r = \iota_{22}(q')$. Note that $\iota_{22}(q')$ is Murray–von Neumann equivalent to prr^*p. There exists a projection p' in $F(D)$ such that $\iota_{11}(p') = prr^*p$ and $\tau_{D,\omega}(p') = 1$ because ι_{11} is an isomorphism from $F(D)$ onto $pF(\Phi(D), M_2(D))p$. Since D has property W, there exist v_1 and v_2 in $F(D)$ such that $v_1^* v_1 = 1$, $v_1 v_1^* = p'$, $v_2^* v_2 = 1$ and $v_2 v_2^* = q'$. Therefore we have

$$p = \iota_{11}(1) \sim \iota_{11}(p') = prr^*p \sim r^*pr = \iota_{22}(q') \sim \iota_{22}(1) = q. \qquad \square$$

The following theorem is one of the main theorems in this section.

Theorem 4.3. *Let D be a simple separable nuclear monotracial* C**-algebra with property W, and let φ be a trace-preserving endomorphism of D. Then φ is approximately inner.*

Proof. By Lemma 4.2, there exists a contraction V in $F(\Phi(D), M_2(D))$ such that

$$V^*V = \left[\left(\begin{pmatrix} h_n & 0 \\ 0 & 0 \end{pmatrix}\right)_n\right] \quad \text{and} \quad VV^* = \left[\left(\begin{pmatrix} 0 & 0 \\ 0 & \varphi(h_n) \end{pmatrix}\right)_n\right],$$

where $\{h_n\}_{n\in\mathbb{N}}$ is an approximate unit for D. It can be easily checked that there exists an element $(v_n)_n$ in D^{ω} such that

$$V = \left[\left(\begin{pmatrix} 0 & 0 \\ v_n & 0 \end{pmatrix}\right)_n\right],$$

and we have

$$(v_n x)_n = (\varphi(x) v_n)_n, \quad (v_n^* v_n x)_n = x \quad \text{and} \quad (v_n v_n^* \varphi(x))_n = \varphi(x)$$

for any $x \in D$. Since $(v_n x)_n = (\varphi(x) v_n)_n$ and $(\varphi(x) v_n v_n^*)_n = \varphi(x)$, we have $(v_n x v_n^*)_n = \varphi(x)$ for any $x \in D$. Because of $D \subseteq \overline{\mathrm{GL}(D^{\sim})}$, we may assume that v_n is an invertible element in $D^{\sim}$ for any $n \in \mathbb{N}$. (See the proof of Proposition 2.4.) For any $n \in \mathbb{N}$, let $u_n := v_n(v_n^* v_n)^{-1/2}$. Then u_n is a unitary element in $D^{\sim}$. Since $(v_n^* v_n x)_n = x$, we have $(u_n x)_n = (v_n (v_n^* v_n)^{-1/2} x)_n = (v_n x)_n$ for any $x \in D$. Therefore

$$\varphi(x) = (v_n x v_n^*)_n = (u_n x v_n^*)_n = (u_n (v_n x^*)^*)_n = (u_n (u_n x^*)^*)_n = (u_n x u_n^*)_n$$

for any $x \in D$. Consequently, φ is approximately inner. $\square$

Let $\mathcal{D}$ be a simple separable nuclear monotracial M_{2^∞}-stable C*-algebra which is KK-equivalent to $\{0\}$. In the rest of this section, we shall consider a uniqueness theorem for approximate homomorphisms from $\mathcal{D}$ to certain C*-algebras. Let B be a simple monotracial C*-algebra with strict comparison, $B \subseteq \overline{\mathrm{GL}(B^{\sim})}$ and $M_2(B) \subseteq \overline{\mathrm{GL}(M_2(B)^{\sim})}$, and let φ and ψ be homomorphisms from $\mathcal{D}$ to B^{ω} such that

$\tau_{\mathcal{D}} = \tau_{B,\omega} \circ \varphi = \tau_{B,\omega} \circ \psi$. By the Choi–Effros lifting theorem, there exist sequences of contractive c.p. maps φ_n and ψ_n from $\mathcal{D}$ to B such that $\varphi(a) = (\varphi_n(a))_n$ and $\psi(a) = (\psi_n(a))_n$ for any $a \in \mathcal{D}$. Define a homomorphism Φ from $\mathcal{D}$ to $M_2(B)^\omega$ by

$$\Phi(a) := \left(\begin{pmatrix} \varphi_n(a) & 0 \\ 0 & \psi_n(a) \end{pmatrix}\right)_n$$

for any $a \in \mathcal{D}$. Since $\tau_{\mathcal{D}} = \tau_{B,\omega} \circ \varphi = \tau_{B,\omega} \circ \psi$, we know $\tau_{M_2(B),\omega}|_{\Phi(\mathcal{D})}$ is a state. Hence $\tau_{M_2(B),\omega}$ is a tracial state on $F(\Phi(\mathcal{D}), M_2(B))$ as above. Since $\mathcal{D}$ is separable, there exist elements $(s_n)_n$ and $(t_n)_n$ in B^ω such that $[(s_n)_n] = 1$ in $F(\varphi(\mathcal{D}), B)$ and $[(t_n)_n] = 1$ in $F(\psi(\mathcal{D}), B)$ by arguments in Section 2B. Put

$$p := \left[\left(\begin{pmatrix} s_n & 0 \\ 0 & 0 \end{pmatrix}\right)_n\right] \quad \text{and} \quad q := \left[\left(\begin{pmatrix} 0 & 0 \\ 0 & t_n \end{pmatrix}\right)_n\right]$$

in $F(\Phi(\mathcal{D}), M_2(B))$. It is easy to see that p and q are projections in $F(\Phi(\mathcal{D}), M_2(B))$ such that $\tau_{M_2(B),\omega}(p) = \tau_{M_2(B),\omega}(q) = \frac{1}{2}$. Theorem 3.10 implies that p is Murray–von Neumann equivalent to q. Therefore we obtain the following theorem by an argument similar to that in the proof of Theorem 4.3.

Theorem 4.4. *Let $\mathcal{D}$ be a simple separable nuclear monotracial M_{2^∞}-stable* C**-algebra which is KK-equivalent to $\{0\}$ and B a simple monotracial* C**-algebra with strict comparison, $B \subseteq \overline{\mathrm{GL}(B^\sim)}$ and $M_2(B) \subseteq \overline{\mathrm{GL}(M_2(B)^\sim)}$. If φ and ψ are homomorphisms from $\mathcal{D}$ to B^ω such that $\tau_{\mathcal{D}} = \tau_{B,\omega} \circ \varphi = \tau_{B,\omega} \circ \psi$, then there exists a unitary element u in $(B^\sim)^\omega$ such that $\varphi(a) = u\psi(a)u^*$ for any $a \in \mathcal{D}$.*

The following corollary is an immediate consequence of the theorem above.

Corollary 4.5. *Let $\mathcal{D}$ be a simple separable nuclear monotracial M_{2^∞}-stable* C**-algebra which is KK-equivalent to $\{0\}$ and B a simple monotracial* C**-algebra with strict comparison, $B \subseteq \overline{\mathrm{GL}(B^\sim)}$ and $M_2(B) \subseteq \overline{\mathrm{GL}(M_2(B)^\sim)}$. If φ and ψ are trace-preserving homomorphisms from $\mathcal{D}$ to B, then φ is approximately unitarily equivalent to ψ.*

Remark 4.6. If B is a simple separable exact monotracial $\mathcal{Z}$-stable C*-algebra, then B has strict comparison, $B \subseteq \overline{\mathrm{GL}(B^\sim)}$ and $M_2(B) \subseteq \overline{\mathrm{GL}(M_2(B)^\sim)}$ by [Rørdam 2004a; Robert 2016].

The following corollary is also an immediate consequence of Theorem 4.4.

Corollary 4.7. *Let $\mathcal{D}$ be a simple separable nuclear monotracial M_{2^∞}-stable* C**-algebra which is KK-equivalent to $\{0\}$ and B a simple monotracial* C**-algebra with strict comparison, $B \subseteq \overline{\mathrm{GL}(B^\sim)}$ and $M_2(B) \subseteq \overline{\mathrm{GL}(M_2(B)^\sim)}$. For any finite subset $F \subset \mathcal{D}$ and $\varepsilon > 0$, there exist a finite subset $G \subset \mathcal{D}$ and $\delta > 0$ such that the following holds. If φ and ψ are (G, δ)-multiplicative maps from $\mathcal{D}$ to B such that*

$$|\tau_B(\varphi(a)) - \tau_{\mathcal{D}}(a)| < \delta \quad \textit{and} \quad |\tau_B(\psi(a)) - \tau_{\mathcal{D}}(a)| < \delta$$

for any $a \in G$, then there exists a unitary element u in $B^\sim$ such that

$$\|\varphi(a) - u\psi(a)u^*\| < \varepsilon$$

for any $a \in F$.

5. Existence theorem

In this section, we assume that $\mathcal{D}$ is a simple separable nuclear monotracial M_{2^∞}-stable C*-algebra which is KK-equivalent to $\{0\}$ and B is a simple separable exact monotracial $\mathcal{Z}$-stable C*-algebra. We shall show that there exists a trace-preserving homomorphism from $\mathcal{D}$ to B. Many arguments in this section are motivated by Schafhauser's proof [2020a] (see also [Schafhauser 2020b]) of the Tikuisis–White–Winter theorem [Tikuisis et al. 2017].

The following lemma is related to [Kirchberg and Phillips 2000, Lemma 2.2].

Lemma 5.1. *Let $\mathcal{D}$ be a simple separable nuclear monotracial M_{2^∞}-stable* C**-algebra which is KK-equivalent to $\{0\}$ and B a simple separable exact monotracial $\mathcal{Z}$-stable* C**-algebra. If there exists a homomorphism φ from $\mathcal{D}$ to B^ω such that $\tau_{B,\omega}\circ\varphi=\tau_{\mathcal{D}}$, then there exists a trace-preserving homomorphism from $\mathcal{D}$ to B.*

Proof. By the Choi–Effros lifting theorem, there exists a sequence $\{\varphi_n\}_{n\in\mathbb{N}}$ of contractive c.p. maps from $\mathcal{D}$ to B such that $\varphi(a)=(\varphi_n(a))_n$ for any $a\in\mathcal{D}$. Let $\{F_m\}_{m\in\mathbb{N}}$ be an increasing sequence of finite subsets in $\mathcal{D}$ such that $\mathcal{D}=\overline{\bigcup_{m\in\mathbb{N}}F_m}$. For any $m\in\mathbb{N}$, applying Corollary 4.7 to F_m and $1/2^m$, we obtain a finite subset G_m of $\mathcal{D}$ and $\delta_m>0$. We may assume that $G_m\subset G_{m+1}$, $\delta_m>\delta_{m+1}$ for any $m\in\mathbb{N}$ and $\lim_{m\to\infty}\delta_m=0$. Since we have

$$\lim_{n\to\omega}\|\varphi_n(ab)-\varphi_n(a)\varphi_n(b)\|=0\quad\text{and}\quad\lim_{n\to\omega}|\tau_B(\varphi_n(a))-\tau_{\mathcal{D}}(a)|=0$$

for any $a,b\in\mathcal{D}$, there exists a subsequence $\{\varphi_{n(m)}\}_{m\in\mathbb{N}}$ of $\{\varphi_n\}_{n\in\mathbb{N}}$ such that

$$\|\varphi_{n(m)}(ab)-\varphi_{n(m)}(a)\varphi_{n(m)}(b)\|<\delta_m\quad\text{and}\quad|\tau_B(\varphi_{n(m)}(a))-\tau_{\mathcal{D}}(a)|<\delta_m$$

for any $a,b\in G_m$. Corollary 4.7 implies that for any $m\in\mathbb{N}$, there exists a unitary element u_m in $B^\sim$ such that

$$\|\varphi_{n(m)}(a)-u_m\varphi_{n(m+1)}(a)u_m^*\|<\frac{1}{2^m}$$

for any $a\in F_m$. Therefore it can easily be checked that the limit

$$\lim_{m\to\infty}u_1u_2\cdots u_{m-1}\varphi_{n(m)}(a)u_{m-1}^*\cdots u_2^*u_1^*$$

exists for any $a\in\mathcal{D}$. Define $\psi(a):=\lim_{m\to\infty}u_1u_2\cdots u_{m-1}\varphi_{n(m)}(a)u_{m-1}^*\cdots u_2^*u_1^*$ for any $a\in\mathcal{D}$. Then ψ is a trace-preserving homomorphism from $\mathcal{D}$ to B. □

By the lemma above, it is enough to show that there exists a homomorphism φ from $\mathcal{D}$ to B^ω such that $\tau_{B,\omega}\circ\varphi=\tau_{\mathcal{D}}$. Borrowing Schafhauser's idea [2020a], we shall show this. By arguments in Section 2D, there exists the extension

$$\eta:0\longrightarrow J\longrightarrow B^\omega\xrightarrow{\varrho}M\longrightarrow 0,$$

where M is a von Neumann algebraic ultrapower of $\pi_{\tau_B}(B)''$ and

$$J=\ker\varrho=\{(x_n)_n\in B^\omega\mid\tilde{\tau}_{B,\omega}((x_n^*x_n)_n)=0\}.$$

Note that J is known as the trace kernel ideal. Also, M is a II_1-factor because B is infinite-dimensional (which is implied by $\mathcal{Z}$-stability) and monotracial. Since $\mathcal{D}$ is monotracial and nuclear, $\pi_{\tau_{\mathcal{D}}}(\mathcal{D})''$ is the injective II_1-factor. Hence there exists a unital homomorphism from $\pi_{\tau_{\mathcal{D}}}(\mathcal{D})''$ to M (see, for example, [Takesaki 2003, Chapter XIV, Proposition 2.15]). In particular, there exists a trace-preserving homomorphism Π from $\mathcal{D}$ to M. Consider the pullback extension

$$\begin{tikzcd}
\Pi^*\eta: \, 0 \arrow[r] & J \arrow[r] \arrow[d, equal] & E \arrow[r, "\hat{\varrho}"] \arrow[d, "\hat{\Pi}"] & \mathcal{D} \arrow[r] \arrow[d, "\Pi"] & 0 \\
\eta: \, 0 \arrow[r] & J \arrow[r] & B^{\omega} \arrow[r, "\varrho"] & M \arrow[r] & 0
\end{tikzcd}$$

where $E=\{(a,x)\in\mathcal{D}\oplus B^{\omega}\mid \Pi(a)=\varrho(x)\}$, $\hat{\varrho}((a,x))=a$ and $\widehat{\Pi}((a,x))=x$ for any $(a,x)\in E$. If we could show that $\Pi^*\eta$ is a split extension with a cross section γ, then $\widehat{\Pi}\circ\gamma$ is a homomorphism from $\mathcal{D}$ to B^{ω} such that $\tau_{B,\omega}\circ\widehat{\Pi}\circ\gamma=\tau_{\mathcal{D}}$. But we were unable to show this, immediately. Note that we need to consider a separable extension in order to use KK-theory and some results in [Elliott and Kucerovsky 2001; Gabe 2016]. We shall construct a suitable separable extension η_0 by Blackadar's technique [2006, Section II.8.5].

We shall recall some definitions and some results in [Elliott and Kucerovsky 2001; Gabe 2016]. An extension $0\to I\to C\to A\to 0$ is said to be *purely large* if, for any $x\in C\setminus I$, $\overline{xIx^*}$ contains a stable C^*-subalgebra which is full in I. Note that $\overline{xIx^*}=\overline{xx^*Ixx^*}=I\cap\overline{xCx^*}$. By [Gabe 2016, Theorem 2.1] (see also [Elliott and Kucerovsky 2001, Corollary 16]), if A is nonunital and I is stable, then a separable extension $0\to I\to C\to A\to 0$ is nuclear-absorbing if and only if it is purely large.

Lemma 5.2. *With notation as above, suppose that there exist separable* C^**-subalgebras* $J_0\subset J$, $B_0\subset B^{\omega}$ *and* $M_0\subset M$ *such that* J_0 *is stable,*

$$\eta_0: 0\longrightarrow J_0\longrightarrow B_0\xrightarrow{\varrho|_{B_0}} M_0\longrightarrow 0$$

is a purely large extension and $\Pi(\mathcal{D})\subset M_0$. *Then there exists a homomorphism* φ *from* $\mathcal{D}$ *to* B^{ω} *such that* $\tau_{B,\omega}\circ\varphi=\tau_{\mathcal{D}}$.

Proof. Consider the pullback extension

$$\begin{tikzcd}
\Pi^*\eta_0: \, 0 \arrow[r] & J_0 \arrow[r] \arrow[d, equal] & E_0 \arrow[r, "\hat{\varrho}"] \arrow[d, "\hat{\Pi}"] & \mathcal{D} \arrow[r] \arrow[d, "\Pi"] & 0 \\
\eta_0: \, 0 \arrow[r] & J_0 \arrow[r] & B_0 \arrow[r, "\varrho"] & M_0 \arrow[r] & 0
\end{tikzcd}$$

where $E_0=\{(a,x)\in\mathcal{D}\oplus B_0\mid \Pi(a)=\varrho(x)\}$, $\hat{\varrho}((a,x))=a$ and $\widehat{\Pi}((a,x))=x$ for any $(a,x)\in E_0$. Since η_0 is purely large, it can be easily checked that $\Pi^*\eta_0$ is purely large. Hence $\Pi^*\eta_0$ is nuclear-absorbing by [Gabe 2016, Theorem 2.1]. Because $\mathcal{D}$ is KK-equivalent to $\{0\}$ and nuclear, we have $\mathrm{Ext}(\mathcal{D}, J_0)=\{0\}$, and hence $[\Pi^*\eta_0]=0$ in $\mathrm{Ext}(\mathcal{D}, J_0)$. Therefore there exists a (nuclear) split extension η' such that $\Pi^*\eta_0\oplus\eta'$ is a split extension. Since $\Pi^*\eta_0$ is nuclear-absorbing, $\Pi^*\eta_0$ is strongly unitarily equivalent to $\Pi^*\eta_0\oplus\eta'$, and hence $\Pi^*\eta_0$ is a split extension. Let γ_0 be a cross section of $\Pi^*\eta_0$, and define $\varphi:=\widehat{\Pi}\circ\gamma_0$. Then φ is the desired homomorphism. □

A key result in the proof of the pure largeness is the following characterization of stable C*-algebras.

Theorem 5.3 [Hjelmborg and Rørdam 1998; Rørdam 2004b, Theorem 2.2]. *Let A be a σ-unital* C**-algebra. Then A is stable if and only if, for any $a \in A_+$ and $\varepsilon > 0$, there exist positive elements a' and c in A such that $\|a - a'\| \le \varepsilon$, $a' \sim c$ and $\|ac\| \le \varepsilon$.*

Before we construct a separable extension η_0, we shall consider properties of η.

Proposition 5.4. *With notation as above, let b be a positive element in $B^\omega \setminus J$.*

(i) *For any positive element a in $\overline{bJb}$, there exists a positive element c in $\overline{bJb}$ such that $a \sim c$ and $ac = 0$.*

(ii) *For any positive element a in J and $\varepsilon > 0$, there exist a positive element d in $\overline{bJb}$ and an element r in J such that $\|r^* d r - a\| < \varepsilon$.*

(iii) *For any element x in B^ω and $\varepsilon > 0$, there exists an element y in* $\mathrm{GL}((B^\omega)^\sim)$ *such that $\|x - y\| < \varepsilon$.*

For the proof of the proposition above, we need some lemmas. For a positive element $a \in A$ and $\varepsilon > 0$, we denote by $(a-\varepsilon)_+$ the element $f(a)$ in A, where $f(t) = \max\{0, t - \varepsilon\}$, $t \in \mathrm{Sp}(a)$. The same proof as in [Rørdam 1992, Proposition 2.4] shows the following lemma. See also [Pedersen 1987, Corollary 8].

Lemma 5.5. *Let A be a* C**-algebra with $A \subseteq \overline{\mathrm{GL}(A^\sim)}$, and let a and b be positive elements in A. Then a is Cuntz smaller than b if and only if, for any $\varepsilon > 0$, there exists a unitary element u in $A^\sim$ such that $u(a-\varepsilon)_+u^* \in \overline{bAb}$.*

The following lemma can be regarded as an application of the construction of $\mathcal{Z}$.

Lemma 5.6. *Let A be a monotracial $\mathcal{Z}$-stable* C**-algebra. For any $\theta \in \big(0, \frac{1}{2}\big)$, there exist positive elements d and d' in A such that $dd' = 0$ and $d_{\tau_A}((d-\varepsilon)_+) = d_{\tau_A}((d'-\varepsilon)_+) = (1-\varepsilon)\theta$ for any $0 \le \varepsilon \le 1$.*

Proof. Let μ be the Lebesgue measure on $[0, 1]$, and define a tracial state τ_0 on $C([0, 1])$ by $\tau_0(f) := \int_{[0,1]} f \, d\mu$ for any $f \in C([0, 1])$. By [Rørdam 2004a, Theorem 2.1(i)], there exists a unital homomorphism ψ from $C([0, 1])$ to $\mathcal{Z}$ such that $\tau_0 = \tau_{\mathcal{Z}} \circ \psi$. Define f and g in $C([0, 1])$ by

$$f(t) := \begin{cases} \frac{2}{\theta}t & \text{if } t \in \big[0, \frac{\theta}{2}\big], \\ -\frac{2}{\theta}t + 2 & \text{if } t \in \big(\frac{\theta}{2}, \theta\big], \\ 0 & \text{if } t \in (\theta, 1] \end{cases} \quad \text{and} \quad g(t) := \begin{cases} 0 & \text{if } t \in [0, \theta], \\ \frac{2}{\theta}t - 2 & \text{if } t \in \big(\theta, \frac{3\theta}{2}\big], \\ -\frac{2}{\theta}t + 4 & \text{if } t \in \big(\frac{3\theta}{2}, 2\theta\big], \\ 0 & \text{if } t \in (2\theta, 1]. \end{cases}$$

Note that for any $0 \le \varepsilon \le 1$, we have

$$(f-\varepsilon)_+(t) = \begin{cases} 0 & \text{if } t \in \big[0, \frac{\varepsilon\theta}{2}\big], \\ \frac{2}{\theta}t - \varepsilon & \text{if } t \in \big(\frac{\varepsilon\theta}{2}, \frac{\theta}{2}\big], \\ -\frac{2}{\theta}t + 2 - \varepsilon & \text{if } t \in \big(\frac{\theta}{2}, \theta - \frac{\varepsilon\theta}{2}\big], \\ 0 & \text{if } t \in \big(\theta - \frac{\varepsilon\theta}{2}, 1\big], \end{cases}$$

$$(g-\varepsilon)_+(t) = \begin{cases} 0 & \text{if } t \in \big[0, \theta + \frac{\varepsilon\theta}{2}\big], \\ \frac{2}{\theta}t - 2 - \varepsilon & \text{if } t \in \big(\theta + \frac{\varepsilon\theta}{2}, \frac{3\theta}{2}\big], \\ -\frac{2}{\theta}t + 4 - \varepsilon & \text{if } t \in \big(\frac{3\theta}{2}, 2\theta - \frac{\varepsilon\theta}{2}\big], \\ 0 & \text{if } t \in \big(2\theta - \frac{\varepsilon\theta}{2}, 1\big]. \end{cases}$$

Hence $d_{\tau_0}((f-\varepsilon)_+) = d_{\tau_0}((g-\varepsilon)_+) = (1-\varepsilon)\theta$. Let s be a strictly positive element in A, and put

$$d := s \otimes \psi(f) \quad \text{and} \quad d' := s \otimes \psi(g)$$

in $A \otimes \mathcal{Z} \cong A$. Then d and d' are desired positive elements in A. □

Lemma 5.7. *Let A be a simple separable exact monotracial $\mathcal{Z}$-stable* C*-*algebra, and let b be a* (*nonzero*) *positive element in A. For any $\theta \in (0, d_{\tau_A}(b)/2)$, there exist positive elements e and e' in $\overline{bAb}$ such that $ee' = 0$ and $d_{\tau_A}(e) = d_{\tau_A}(e') > \theta$.*

Proof. By Lemma 5.6, there exist contractions d and d' in A such that $dd' = 0$ and $\theta < d_{\tau_A}(d) = d_{\tau_A}(d') < d_{\tau_A}(b)/2$. Furthermore, we may assume that there exists $\varepsilon > 0$ such that $d_{\tau_A}((d-\varepsilon)_+) = d_{\tau_A}((d'-\varepsilon)_+) > \theta$. Since A has strict comparison and $d_{\tau_A}(d+d') = d_{\tau_A}(d) + d_{\tau_A}(d') < d_{\tau_A}(b)$, Lemma 5.5 implies that there exists a unitary element u in $A^{\sim}$ such that $u(d+d'-\varepsilon)_+u^* \in \overline{bAb}$. Note that $(d+d'-\varepsilon)_+ = (d-\varepsilon)_+ + (d'-\varepsilon)_+$ because of $dd' = 0$. Put

$$e := u(d-\varepsilon)_+u^* \quad \text{and} \quad e' := u(d'-\varepsilon)_+u^*.$$

Then e and e' are desired positive elements. □

Proof of Proposition 5.4. (i) We may assume $\|a\| = 1$ and $\|b\| = 1$. Since $b \notin J$, we have $\tau_{B,\omega}(b) > 0$. Take a representative $(b_n)_n$ of b such that $\|b_n\| = 1$ for any $n \in \mathbb{N}$, and choose $\varepsilon_0 > 0$ such that $\tau_{B,\omega}(b) - \varepsilon_0 > 0$. Since we have

$$\lim_{n\to\omega} d_{\tau_B}(b_n) \geq \lim_{n\to\omega} \tau_B(b_n) = \tau_{B,\omega}(b),$$

there exists an element $X_1 \in \omega$ such that, for any $n \in X_1$,

$$d_{\tau_B}(b_n) > \tau_{B,\omega}(b) - \varepsilon_0.$$

By an argument similar to that in the proof of [Sato 2010, Lemma 3.2], we see that there exists a representative $(a_n)_n$ of a such that $a_n \in \overline{b_nBb_n}$ and $\|a_n\| = 1$ for any $n \in \mathbb{N}$ and $\lim_{n\to\omega} d_{\tau_B}(a_n) = 0$ because of $a \in \overline{(b_n)_nJ(b_n)_n}$. Hence there exists an element $X_2 \in \omega$ such that for any $n \in X_2$,

$$d_{\tau_B}(a_n) < \frac{\tau_{B,\omega}(b) - \varepsilon_0}{2}.$$

Note that we have $d_{\tau_B}(a_n) < d_{\tau_B}(b_n)/2$ for any $n \in X_1 \cap X_2$. Hence Lemma 5.7 implies that for any $n \in X_1 \cap X_2$, there exist positive elements e_n and e'_n in $\overline{b_nBb_n}$ such that $e_ne'_n = 0$ and $d_{\tau_B}(e_n) = d_{\tau_B}(e'_n) > d_{\tau_B}(a_n)$. Since $\overline{b_nBb_n}$ has strict comparison and $\overline{b_nBb_n} \subseteq \overline{\mathrm{GL}(\overline{b_nBb_n}^{\sim})}$ by [Rørdam 2004a; Robert 2016], Lemma 5.5 shows that for any $n \in X_1 \cap X_2$, there exist unitary elements u_n and v_n in $\overline{b_nBb_n}^{\sim}$ such that

$$u_n(a_n - 1/n)_+u_n^* \in \overline{e_nBe_n} \quad \text{and} \quad v_n(a_n - 1/n)_+v_n^* \in \overline{e'_nBe'_n}.$$

Note that $(a_n - 1/n)_+u_n^*v_n(a_n - 1/n)_+ = 0$ for any $n \in X_1 \cap X_2$. Define $z = (z_n)_n$ and $c = (c_n)_n$ in B^{ω} by

$$z_n := \begin{cases} 0 & \text{if } n \notin X_1 \cap X_2, \\ u_n^*v_n(a_n - 1/n)_+^{1/2} & \text{if } n \in X_1 \cap X_2 \end{cases}$$

and

$$c_n := \begin{cases} 0 & \text{if } n \notin X_1 \cap X_2, \\ u_n^* v_n (a_n - 1/n)_+ v_n^* u_n & \text{if } n \in X_1 \cap X_2. \end{cases}$$

It is easy to see that $z, c \in \overline{bB^{\omega}b}$, $z^*z = a$, $zz^* = c$ and $ac = 0$. Since $\overline{bJb}$ is a closed ideal in $\overline{bB^{\omega}b}$ and $a \in \overline{bJb}$, we know z and c are elements in $\overline{bJb}$. Therefore we obtain the conclusion.

(ii) Note that B^{ω} has strict comparison; see, for example, [Bosa et al. 2019, Lemma 1.23]. Since $a \in J$ and $b \notin J$, we have $d_{\tau_{B,\omega}}(a^{1/5}) = 0$ and $d_{\tau_{B,\omega}}(b) > 0$. Hence there exists a sequence $\{s_N\}_{N \in \mathbb{N}}$ in B^{ω} such that $\lim_{N \to \infty} \|s_N^* b s_N - a^{1/5}\| = 0$. Let $d_N := b s_N a^{1/5} s_N^* b$ and $r_N := s_N a^{1/5}$ for any $N \in \mathbb{N}$. Then we have $d_N \in \overline{bJb}$, $r_N \in J$ for any $N \in \mathbb{N}$ and

$$r_N^* d_N r_N = a^{1/5} s_N^* b s_N a^{1/5} s_N^* b s_N a^{1/5} \to a$$

as $N \to \infty$. Therefore we obtain the conclusion.

(iii) Since B is a simple monotracial $\mathcal{Z}$-stable C*-algebra, $B \subseteq \overline{\mathrm{GL}(B^{\sim})}$ by [Rørdam 2004a; Robert 2016]. Therefore we obtain the conclusion by Proposition 2.4. □

If B is unital, then the following lemma is a well-known consequence of Proposition 2.4 and Blackadar's technique [2006, Proposition II.8.5.4].

Lemma 5.8. *With notation as above, let S be a separable subset of B^{ω}. Then there exists a separable* C**-algebra A such that $S \subseteq A \subset B^{\omega}$ and $A \subseteq \overline{\mathrm{GL}(A^{\sim})}$.*

Proof. We shall show only the case where B is nonunital. Let A_1 be the C*-subalgebra of B^{ω} generated by S. Since A_1 is separable, there exists a countable dense subset $\{x_k \mid k \in \mathbb{N}\}$ of A_1. By Proposition 5.4(iii), for any $k, m \in \mathbb{N}$, there exist $y_{k,m} \in B^{\omega}$ and $\lambda_{k,m} \in \mathbb{C} \setminus \{0\}$ such that

$$\|x_k - (y_{k,m} + \lambda_{k,m} 1_{(B^{\omega})^{\sim}})\| < \frac{1}{m}$$

and $y_{k,m} + \lambda_{k,m} 1_{(B^{\omega})^{\sim}} \in \mathrm{GL}((B^{\omega})^{\sim})$. Let A_2 be the C*-subalgebra of B^{ω} generated by A_1 and $\{y_{k,m} \mid k, m \in \mathbb{N}\}$. Then we have $A_1 \subseteq \overline{\mathrm{GL}(A_2^{\sim})}$. Indeed, we have $y_{k,m} + \lambda_{k,m} 1_{A_2^{\sim}} \in \mathrm{GL}(A_2^{\sim})$ for any $k, m \in \mathbb{N}$ because of $\mathrm{Sp}_{A_2}(y_{k,m}) \cup \{0\} = \mathrm{Sp}_{B^{\omega}}(y_{k,m}) \cup \{0\}$ and $\lambda_{k,m} \neq 0$. Since $A_1 = \overline{\{x_k \mid k \in \mathbb{N}\}}$ and

$$\begin{aligned} \|x_k - (y_{k,m} + \lambda_{k,m} 1_{(A_2)^{\sim}})\| &= \|1_{A_2^{\sim}} x_k - 1_{A_2^{\sim}} (y_{k,m} + \lambda_{k,m} 1_{(B^{\omega})^{\sim}})\| \\ &\leq \|x_k - (y_{k,m} + \lambda_{k,m} 1_{(B^{\omega})^{\sim}})\| < \frac{1}{m} \end{aligned}$$

for any $k, m \in \mathbb{N}$, we have $A_1 \subseteq \overline{\mathrm{GL}(A_2^{\sim})}$. Repeating this process, we obtain a sequence $\{A_n\}_{n \in \mathbb{N}}$ of separable C*-subalgebras of B^{ω} such that $A_n \subseteq A_{n+1}$ and $A_n \subseteq \overline{\mathrm{GL}(A_{n+1}^{\sim})}$ for any $n \in \mathbb{N}$. Put $A := \overline{\bigcup_{n=1}^{\infty} A_n}$. Since $A_n \subseteq \overline{\mathrm{GL}(A_{n+1}^{\sim})} \subseteq \overline{\mathrm{GL}(A^{\sim})}$ for any $n \in \mathbb{N}$ by Proposition 2.2, we have $A \subseteq \overline{\mathrm{GL}(A^{\sim})}$. Therefore A is the desired separable C*-algebra. □

The following lemma is also based on Blackadar's technique.

Lemma 5.9. *With notation as above, let $\{b_k \mid k \in \mathbb{N}\}$ be a countable subset of $B^{\omega} \setminus J$ and S a separable subset of B^{ω}. Then there exists a separable* C**-algebra A such that $\{b_k \mid k \in \mathbb{N}\} \cup S \subseteq A \subset B^{\omega}$ and $\overline{b_k (A \cap J) b_k}$ is full in $A \cap J$ for any $k \in \mathbb{N}$.*

Proof. Let A_1 be the C*-subalgebra of B^ω generated by $\{b_k \mid k \in \mathbb{N}\}$ and S. Since A_1 is separable, there exists a countable dense subset $\{a_l \mid l \in \mathbb{N}\}$ of $(A_1 \cap J)_+$. By Proposition 5.4(ii), for any $k, l, m \in \mathbb{N}$, there exist $d_{k,l,m} \in \overline{b_k J b_k}_+$ and $r_{k,l,m} \in J$ such that

$$\|r_{k,l,m}^* d_{k,l,m} r_{k,l,m} - a_l\| < \frac{1}{m}.$$

Let A_2 be the C*-subalgebra of B^ω generated by A_1 and $\{d_{k,l,m}, r_{k,l,m} \mid k, l, m \in \mathbb{N}\}$. Then we have $A_1 \cap J \subseteq \overline{(A_2 \cap J) b_k (A_2 \cap J) b_k (A_2 \cap J)}$ for any $k \in \mathbb{N}$ because $A_1 \cap J$ is generated by $\{a_l \mid l \in \mathbb{N}\}$. Repeating this process, we obtain a sequence $\{A_n\}_{n\in\mathbb{N}}$ of separable C*-subalgebras of B^ω such that $A_n \subseteq A_{n+1}$ and $A_n \cap J \subseteq \overline{(A_{n+1} \cap J) b_k (A_{n+1} \cap J) b_k (A_{n+1} \cap J)}$ for any $k, n \in \mathbb{N}$. Put $A := \overline{\bigcup_{n=1}^\infty A_n}$. Since we have $A \cap J = \overline{\bigcup_{n=1}^\infty (A_n \cap J)}$, we see that A is the desired separable C*-algebra. □

By Lemmas 5.8 and 5.9, [Blackadar 2006, Proposition II.8.5.3] implies the following lemma.

Lemma 5.10. *With notation as above, let $\{b_k \mid k \in \mathbb{N}\}$ be a countable subset of $B^\omega \setminus J$ and S a separable subset of B^ω. Then there exists a separable* C**-algebra A such that $\{b_k \mid k \in \mathbb{N}\} \cup S \subseteq A \subset B^\omega$, $A \subseteq \overline{\mathrm{GL}(A^\sim)}$ and $\overline{b_k (A \cap J) b_k}$ is full in $A \cap J$ for any $k \in \mathbb{N}$.*

We shall construct the separable extension η_0 of Lemma 5.2.

Since ϱ is surjective and $\mathcal{D}$ is separable, there exists a separable subset S_0 of B^ω such that $\overline{\varrho(S_0)} = \Pi(\mathcal{D})$. Applying Lemma 5.8 to S_0, we obtain a separable C*-algebra B_1 such that $S_0 \subseteq B_1 \subset B^\omega$ and $B_1 \subseteq \overline{\mathrm{GL}(B_1^\sim)}$. Since B_1 is separable, there exist a countable subset $\{a_{1,m} \mid m \in \mathbb{N}\}$ of $(B_1 \cap J)_+$ and a countable subset $\{b_{1,k} \mid k \in \mathbb{N}\}$ of B_{1+} such that

$$\overline{\{a_{1,m} \mid m \in \mathbb{N}\}} = (B_1 \cap J)_+ \quad \text{and} \quad \overline{\{b_{1,k} \mid k \in \mathbb{N}\}} = B_{1+}.$$

Put $T_1 := \{(k,l) \in \mathbb{N} \times \mathbb{N} \mid (b_{1,k} - 1/l)_+ \notin J\}$. Applying Proposition 5.4(i) to $(b_{1,k} - 1/l)_+ a_{1,m} (b_{1,k} - 1/l)_+$ for any $(k,l) \in T_1$ and $m \in \mathbb{N}$, there exist a positive element $c_{1,1,(k,l),m}$ and an element $z_{1,1,(k,l),m}$ in $\overline{(b_{1,k} - 1/l)_+ J (b_{1,k} - 1/l)_+}$ such that

$$\begin{aligned} (b_{1,k} - 1/l)_+ a_{1,m} (b_{1,k} - 1/l)_+ c_{1,1,(k,l),m} &= 0, \\ z_{1,1,(k,l),m}^* z_{1,1,(k,l),m} &= (b_{1,k} - 1/l)_+ a_{1,m} (b_{1,k} - 1/l)_+, \\ z_{1,1,(k,l),m} z_{1,1,(k,l),m}^* &= c_{1,1,(k,l),m}. \end{aligned}$$

Let $S_2 := B_1 \cup \{c_{1,1,(k,l),m}, z_{1,1,(k,l),m} \mid (k,l) \in T_1, m \in \mathbb{N}\}$. Applying Lemma 5.10 to $\{(b_{1,k} - 1/l)_+ \mid (k,l) \in T_1\}$ and S_2, we obtain a separable C*-algebra B_2 such that

$$B_1 \cup \{c_{1,1,(k,l),m}, z_{1,1,(k,l),m} \mid (k,l) \in T_1, m \in \mathbb{N}\} \subseteq B_2 \subset B^\omega,$$

$B_2 \subseteq \overline{\mathrm{GL}(B_2^\sim)}$ and $\overline{(b_{1,k} - 1/l)_+ (B_2 \cap J) (b_{1,k} - 1/l)_+}$ is full in $B_2 \cap J$ for any $(k,l) \in T_1$. In the same way as above, there exist a countable subset $\{a_{2,m} \mid m \in \mathbb{N}\}$ of $(B_2 \cap J)_+$ and a countable subset $\{b_{2,k} \mid k \in \mathbb{N}\}$ of B_{2+} such that

$$\overline{\{a_{2,m} \mid m \in \mathbb{N}\}} = (B_2 \cap J)_+ \quad \text{and} \quad \overline{\{b_{2,k} \mid k \in \mathbb{N}\}} = B_{2+},$$

and we put $T_2 := \{(k,l) \in \mathbb{N}\times\mathbb{N} \mid (b_{2,k}-1/l)_+ \notin J\}$. Applying Proposition 5.4(i) to $(b_{i,k}-1/l)_+ a_{2,m} \times (b_{i,k}-1/l)_+$ for any $1 \le i \le 2$, $(k,l) \in T_i$ and $m \in \mathbb{N}$, there exist a positive element $c_{2,i,(k,l),m}$ and an element $z_{2,i,(k,l),m}$ in $\overline{(b_{i,k}-1/l)_+ J (b_{i,k}-1/l)_+}$ such that

$$\begin{aligned}(b_{i,k}-1/l)_+ a_{2,m}(b_{i,k}-1/l)_+ c_{2,i,(k,l),m} &= 0,\\ z^*_{2,i,(k,l),m} z_{2,i,(k,l),m} &= (b_{i,k}-1/l)_+ a_{2,m}(b_{i,k}-1/l)_+,\\ z_{2,i,(k,l),m} z^*_{2,i,(k,l),m} &= c_{2,i,(k,l),m}.\end{aligned}$$

Let $S_3 := B_2 \cup \{c_{2,i,(k,l),m}, z_{2,i,(k,l),m} \mid 1 \le i \le 2, (k,l) \in T_i, m \in \mathbb{N}\}$. Applying Lemma 5.10 to $\{(b_{i,k}-1/l)_+ \mid 1 \le i \le 2, (k,l) \in T_i\}$ and S_3, we obtain a separable C*-algebra B_3 such that

$$B_2 \cup \{c_{2,i,(k,l),m}, z_{2,i,(k,l),m} \mid 1 \le i \le 2, (k,l) \in T_i, m \in \mathbb{N}\} \subseteq B_3 \subset B^{\omega},$$

$B_3 \subseteq \overline{\mathrm{GL}(B_3^{\sim})}$ and $\overline{(b_{i,k}-1/l)_+ (B_3 \cap J)(b_{i,k}-1/l)_+}$ is full in $B_3 \cap J$ for any $1 \le i \le 2$ and $(k,l) \in T_i$. Repeating this process, for any $n \in \mathbb{N}$, we obtain

$$\begin{gathered}B_n \subset B^{\omega}, \quad \{a_{n,m} \mid m \in \mathbb{N}\} \subset (B_n \cap J)_+, \quad \{b_{n,k} \mid k \in \mathbb{N}\} \subset B_{n+},\\ T_n \subset \mathbb{N}\times\mathbb{N}, \quad \{c_{n,i,(k,l),m}, z_{n,i,(k,l),m} \mid 1 \le i \le n, (k,l) \in T_i, m \in \mathbb{N}\}\end{gathered}$$

such that B_n is separable,

$$\begin{gathered}B_n \subseteq B_{n+1}, \quad B_n \subseteq \overline{\mathrm{GL}(B_n^{\sim})}, \quad \overline{\{a_{n,m} \mid m \in \mathbb{N}\}} = (B_n \cap J)_+,\\ \overline{\{b_{n,k} \mid k \in \mathbb{N}\}} = B_{n+}, \quad T_n = \{(k,l) \in \mathbb{N}\times\mathbb{N} \mid (b_{n,k}-1/l)_+ \notin J\},\\ c_{n,i,(k,l),m},\ z_{n,i,(k,l),m} \in \overline{(b_{i,k}-1/l)_+ (B_{n+1} \cap J)(b_{i,k}-1/l)_+},\\ (b_{i,k}-1/l)_+ a_{n,m}(b_{i,k}-1/l)_+ c_{n,i,(k,l),m} = 0,\\ z^*_{n,i,(k,l),m} z_{n,i,(k,l),m} = (b_{i,k}-1/l)_+ a_{n,m}(b_{i,k}-1/l)_+,\\ z_{n,i,(k,l),m} z^*_{n,i,(k,l),m} = c_{n,i,(k,l),m}\end{gathered}$$

and $\overline{(b_{i,k}-1/l)_+ (B_{n+1} \cap J)(b_{i,k}-1/l)_+}$ is full in $B_{n+1} \cap J$ for any $1 \le i \le n$ and $(k,l) \in T_i$. Define

$$B_0 := \overline{\bigcup_{n=1}^{\infty} B_n}, \quad J_0 := B_0 \cap J \quad \text{and} \quad M_0 := \varrho(B_0).$$

Then

$$\eta_0 : 0 \longrightarrow J_0 \longrightarrow B_0 \xrightarrow{\varrho} M_0 \longrightarrow 0$$

is a separable extension and $\Pi(\mathcal{D}) \subseteq M_0$. Corollary 2.3 implies $B_0 \subseteq \overline{\mathrm{GL}(B_0^{\sim})}$ since we have $B_n \subseteq \overline{\mathrm{GL}(B_n^{\sim})}$ for any $n \in \mathbb{N}$. Furthermore, for any $i \in \mathbb{N}$ and $(k,l) \in T_i$, $\overline{(b_{i,k}-1/l)_+ J_0 (b_{i,k}-1/l)_+}$ is full in J_0 by a similar argument as in the proof of Lemma 5.9. Note that, for any $n_0 \in \mathbb{N}$,

$$J_{0+} = \overline{\bigcup_{n=n_0}^{\infty} \{a_{n,m} \mid m \in \mathbb{N}\}} \quad \text{and} \quad B_{0+} = \overline{\bigcup_{n=n_0}^{\infty} \{b_{n,k} \mid k \in \mathbb{N}\}}.$$

We shall show that J_0 is stable and η_0 is purely large.

Proof of the stability of J_0. Let $a \in J_{0+} \setminus \{0\}$ and $\varepsilon > 0$. Set

$$\varepsilon' := \min\left\{ \frac{\varepsilon}{2\|a\|},\ \sqrt{\frac{\varepsilon}{2}},\ \varepsilon \right\}.$$

Since B_0 is separable, there exists an approximate unit $\{h_n\}_{n\in\mathbb{N}}$ for B_0. Note that $h_n \notin J$ for sufficiently large n because of $M_0 \neq \{0\}$. Hence there exists $N \in \mathbb{N}$ such that $h_N \notin J$ and $\|h_N a h_N - a\| < \varepsilon'/2$. Since $B_{0+} = \overline{\bigcup_{n=1}^{\infty}\{b_{n,k} \mid k \in \mathbb{N}\}}$, for any $l \in \mathbb{N}$, there exist $n(l)$ and $k(l)$ in $\mathbb{N}$ such that

$$\|h_N - b_{n(l),k(l)}\| < \frac{1}{l}.$$

Note that $(b_{n(l),k(l)} - 1/l)_+ \to h_N$ as $l \to \infty$ because we have

$$\|h_N - (b_{n(l),k(l)} - 1/l)_+\| \le \|h_N - b_{n(l),k(l)}\| + \|b_{n(l),k(l)} - (b_{n(l),k(l)} - 1/l)_+\| < \frac{2}{l}.$$

Hence there exists $l_0 \in \mathbb{N}$ such that $(b_{n(l_0),k(l_0)} - 1/l_0)_+ \notin J$, that is, $(k(l_0), l_0) \in T_{n(l_0)}$ and

$$\|a - (b_{n(l_0),k(l_0)} - 1/l_0)_+ a (b_{n(l_0),k(l_0)} - 1/l_0)_+\| < \frac{\varepsilon'}{2}.$$

Since $J_{0+} = \overline{\bigcup_{n=n(l_0)}^{\infty}\{a_{n,m} \mid m \in \mathbb{N}\}}$, there exist $n_0 \ge n(l_0)$ and $m_0 \in \mathbb{N}$ such that

$$\|a - a_{n_0,m_0}\| < \frac{\varepsilon'}{2\|b_{n(l_0),k(l_0)}\|^2}.$$

Put $a' := (b_{n(l_0),k(l_0)} - 1/l_0)_+ a_{n_0,m_0} (b_{n(l_0),k(l_0)} - 1/l_0)_+$. Then

$$\|a - a'\| < \varepsilon' \le \varepsilon.$$

By construction of B_0 and J_0, there exist

$$z = z_{n_0,n(l_0),(k(l_0),l_0),m_0},\ c = c_{n_0,n(l_0),(k(l_0),l_0),m_0} \in J_0$$

such that $a'c = 0$, $z^*z = a'$ and $zz^* = c$. Hence $a' \sim c$ and

$$\|ac\| = \|ac - a'c\| \le \|a - a'\|\|c\| = \|a - a'\|\|a'\| < \varepsilon'(\|a\| + \varepsilon') \le \varepsilon.$$

Therefore J_0 is stable by Hjelmborg and Rørdam's characterization (Theorem 5.3). □

Proof of the pure largeness of η_0. Let $x \in B_0 \setminus J_0$. Note that we have $xx^* \notin J$. Since $B_{0+} = \overline{\bigcup_{n=1}^{\infty}\{b_{n,k} \mid k \in \mathbb{N}\}}$, for any $l \in \mathbb{N}$, there exist $n(l)$ and $k(l)$ in $\mathbb{N}$ such that

$$\|xx^* - b_{n(l),k(l)}\| < \frac{1}{2l}.$$

By an argument similar to that in the proof of stability of J_0, there exists $l_0 \in \mathbb{N}$ such that $(b_{n(l_0),k(l_0)} - 1/l_0)_+ \notin J$, that is, $(k(l_0), l_0) \in T_{n(l_0)}$. On the other hand, [Kirchberg and Rørdam 2002, Lemma 2.2] implies that $(b_{n(l_0),k(l_0)} - 1/2l_0)_+$ is Cuntz smaller than xx^*. Since we have $B_0 \subseteq \overline{\mathrm{GL}(B_0^\sim)}$, there exists a unitary element u in $B_0^\sim$ such that

$$u(b_{n(l_0),k(l_0)} - 1/l_0)_+ u^* = u((b_{n(l_0),k(l_0)} - 1/2l_0)_+ - 1/2l_0)_+ u^* \in \overline{xx^* B_0 xx^*} = \overline{x B_0 x^*}$$

by Lemma 5.5. Put

$$C := u\overline{(b_{n(l_0),k(l_0)}-1/l_0)_+J_0(b_{n(l_0),k(l_0)}-1/l_0)_+}u^* \subseteq \overline{xJ_0x^*}.$$

Then C is full in J_0 because $\overline{(b_{n(l_0),k(l_0)}-1/l_0)_+J_0(b_{n(l_0),k(l_0)}-1/l_0)_+}$ is full in J_0. We shall show that C is stable. Let $a \in C_+ \setminus \{0\}$ and $\varepsilon > 0$. Set

$$\varepsilon' := \min\left\{\frac{\varepsilon}{2\|a\|}, \sqrt{\frac{\varepsilon}{2}}, \varepsilon\right\}.$$

By the definition of C and $J_{0+} = \overline{\bigcup_{n=n(l_0)}^{\infty}\{a_{n,m} \mid m \in \mathbb{N}\}}$, there exist $n_0 \geq n(l_0)$ and $m_0 \in \mathbb{N}$ such that

$$\|a - u(b_{n(l_0),k(l_0)}-1/l_0)_+a_{n_0,m_0}(b_{n(l_0),k(l_0)}-1/l_0)_+u^*\| < \varepsilon' \leq \varepsilon.$$

Put $a' = u(b_{n(l_0),k(l_0)}-1/l_0)_+a_{n_0,m_0}(b_{n(l_0),k(l_0)}-1/l_0)_+u^* \in C$, then $\|a-a'\| < \varepsilon' \leq \varepsilon$. By construction of B_0 and J_0, there exist elements

$$z_{n_0,n(l_0),(k(l_0),l_0),m_0},\ c_{n_0,n(l_0),(k(l_0),l_0),m_0}$$

in $\overline{(b_{n(l_0),k(l_0)}-1/l_0)_+J_0(b_{n(l_0),k(l_0)}-1/l_0)_+}$ such that

$$u^*a'uc_{n_0,n(l_0),(k(l_0),l_0),m_0} = 0, \quad z^*_{n_0,n(l_0),(k(l_0),l_0),m_0}z_{n_0,n(l_0),(k(l_0),l_0),m_0} = u^*a'u$$

and

$$z_{n_0,n(l_0),(k(l_0),l_0),m_0}z^*_{n_0,n(l_0),(k(l_0),l_0),m_0} = c_{n_0,n(l_0),(k(l_0),l_0),m_0}.$$

Put $c := uc_{n_0,n(l_0),(k(l_0),l_0),m_0}u^*$. It is easy to see that $c \in C$, $a'c = 0$ and

$$c \sim c_{n_0,n(l_0),(k(l_0),l_0),m_0} \sim u^*a'u \sim a' \quad \text{in } B_0.$$

Since C is a hereditary C*-subalgebra of B_0 and $a', c \in C$, we see that a' is Murray–von Neumann equivalent to c in C. Therefore, the same argument as in the proof of stability of J_0 shows $\|ac\| < \varepsilon$, and C is stable. Consequently, η_0 is a purely large extension. □

Therefore we obtain the following lemma.

Lemma 5.11. *With notation as above, there exist separable* C*-*subalgebras $J_0 \subset J$, $B_0 \subset B^\omega$ and $M_0 \subset M$ such that J_0 is stable,*

$$\eta_0 : 0 \longrightarrow J_0 \longrightarrow B_0 \xrightarrow{\varrho|_{B_0}} M_0 \longrightarrow 0$$

is a purely large extension and $\Pi(\mathcal{D}) \subset M_0$.

Consequently, we obtain the following theorem by Lemma 5.1, Lemma 5.2 and the lemma above.

Theorem 5.12. *Let $\mathcal{D}$ be a simple separable nuclear monotracial M_{2^∞}-stable* C*-*algebra which is KK-equivalent to $\{0\}$ and B a simple separable exact monotracial $\mathcal{Z}$-stable* C*-*algebra. Then there exists a trace-preserving homomorphism from $\mathcal{D}$ to B.*

Remark 5.13. Actually, we need not assume that $\mathcal{D}$ is M_{2^∞}-stable in the theorem above. Indeed, define a homomorphism φ from $\mathcal{D}$ to $\mathcal{D} \otimes M_{2^\infty}$ by $\varphi(a) = a \otimes 1$. Then φ is a trace-preserving homomorphism from $\mathcal{D}$ to $\mathcal{D} \otimes M_{2^\infty}$. By the theorem above, there exists a trace-preserving homomorphism ψ from $\mathcal{D} \otimes M_{2^\infty}$ to B. Then $\psi \circ \varphi$ is a trace-preserving homomorphism from $\mathcal{D}$ to B.

The following corollary is an immediate consequence of the theorem above.

Corollary 5.14. *Let B a simple separable exact monotracial $\mathcal{Z}$-stable* C*-*algebra. Then there exists a trace-preserving homomorphism from $\mathcal{W}$ to B.*

The injective II_1-factor can embed unitally into every II_1-factor. Hence the following question is natural and interesting.

Question 5.15. (1) Let B be a simple monotracial infinite-dimensional C*-algebra. Does there exist a trace-preserving homomorphism from $\mathcal{W}$ to B?

(2) Let B be a simple non-type-I C*-algebra. Does there exist a (nonzero) homomorphism from $\mathcal{W}$ to B?

Note that Dadarlat, Hirshberg, Toms and Winter [Dadarlat et al. 2009] showed that there exists a unital simple separable nuclear infinite-dimensional C*-algebra B such that $\mathcal{Z}$ does not embed unitally into B.

6. Characterization of $\mathcal{W}$

In this section we shall show that if $\mathcal{D}$ is a simple separable nuclear monotracial M_{2^∞}-stable C*-algebra which is KK-equivalent to $\{0\}$, then $\mathcal{D}$ is isomorphic to $\mathcal{W}$. Also, we shall characterize $\mathcal{W}$ by using properties of $F(\mathcal{W})$.

Theorem 6.1. *Let $\mathcal{D}$ be a simple separable nuclear monotracial M_{2^∞}-stable* C*-*algebra which is KK-equivalent to $\{0\}$. Then $\mathcal{D}$ is isomorphic to $\mathcal{W}$.*

Proof. By Theorem 5.12 and Corollary 5.14, there exist trace-preserving homomorphisms φ and ψ from $\mathcal{D}$ to $\mathcal{W}$ and from $\mathcal{W}$ and $\mathcal{D}$, respectively. Since $\mathcal{D}$ and $\mathcal{W}$ have property W by Corollary 3.11, Theorem 4.3 implies that $\psi \circ \varphi$ and $\varphi \circ \psi$ are approximately inner. Therefore $\mathcal{D}$ is isomorphic to $\mathcal{W}$ by Elliott's approximate intertwining argument [Elliott 1993]; see also [Rørdam 2002, Corollary 2.3.4]. □

The following corollary is an immediate consequence of the theorem above.

Corollary 6.2. (i) *If A is a simple separable nuclear monotracial* C*-*algebra, then $A \otimes \mathcal{W}$ is isomorphic to $\mathcal{W}$. In particular, $\mathcal{W} \otimes \mathcal{W}$ is isomorphic to $\mathcal{W}$.*

(ii) *For any nonzero positive element h in $\mathcal{W}$, $\overline{h\mathcal{W}h}$ is isomorphic to $\mathcal{W}$.*

Following the definition in [Lin and Ng 2023], we say that a C*-algebra A is *$\mathcal{W}$-embeddable* if there exists an injective homomorphism from A to $\mathcal{W}$.

Lemma 6.3. *Let A be a monotracial $\mathcal{W}$-embeddable* C*-*algebra. Then there exists a trace-preserving homomorphism from A to $\mathcal{W}$.*

Proof. By the assumption, there exists an injective homomorphism φ from A to $\mathcal{W}$. Let s be a strictly positive element in A. (Note that A is separable because A is $\mathcal{W}$-embeddable.) Since φ is injective, $\varphi(s)$ is a nonzero positive element. Corollary 6.2 implies that there exists an isomorphism Φ from $\overline{\varphi(s)\mathcal{W}\varphi(s)}$ onto $\mathcal{W}$. Note that φ can be regarded as a homomorphism from A to $\overline{\varphi(s)\mathcal{W}\varphi(s)}$. Define $\psi := \Phi \circ \varphi$. Then ψ is a trace-preserving homomorphism from A to $\mathcal{W}$. □

The following theorem is a characterization of $\mathcal{W}$.

Theorem 6.4. *Let D be a simple separable nuclear monotracial* C**-algebra. Then D is isomorphic to $\mathcal{W}$ if and only if D has property W and is $\mathcal{W}$-embeddable, that is, D satisfies the following properties*:

(i) *For any $\theta \in [0, 1]$, there exists a projection p in $F(D)$ such that $\tau_{D,\omega}(p) = \theta$.*

(ii) *If p and q are projections in $F(D)$ such that $0 < \tau_{D,\omega}(p) = \tau_{D,\omega}(q)$, then p is Murray–von Neumann equivalent to q.*

(iii) *There exists an injective homomorphism from D to $\mathcal{W}$.*

Proof. The "only if" part is obvious by Corollary 3.11. We shall show the "if" part. Since D is $\mathcal{W}$-embeddable, there exists a trace-preserving homomorphism φ from D to $\mathcal{W}$ by Lemma 6.3. Lemma 4.1 implies that D is $\mathcal{Z}$-stable because D has property W. Hence there exists a trace-preserving homomorphism ψ from $\mathcal{W}$ to D by Corollary 5.14. The rest of proof is same as the proof of Theorem 6.1. □

We think that every simple separable nuclear monotracial C*-algebra with property W ought to be $\mathcal{W}$-embeddable. Note that every simple separable nuclear monotracial C*-algebra with property W is stably projectionless by [Kirchberg 2006, Remark 2.13] and an argument similar to that in the proof of [Nawata 2019, Corollary 5.9]. Hence an affirmative answer to the following question, which can be regarded as an analogue of Kirchberg's embedding theorem [Kirchberg and Phillips 2000], would imply this.

Question 6.5. Let A be a simple separable exact stably projectionless monotracial C*-algebra. Assume that τ_A is amenable. Is A $\mathcal{W}$-embeddable?

Note that we need to assume that τ_A is amenable because $\pi_{\tau_{\mathcal{W}}}(\mathcal{W})''$ is the injective II_1-factor.

References

[Blackadar 2006] B. Blackadar, *Operator algebras: theory of* C**-algebras and von Neumann algebras*, Encycl. Math. Sci. **122**, Springer, 2006. MR Zbl

[Blackadar et al. 1992] B. Blackadar, A. Kumjian, and M. Rørdam, "Approximately central matrix units and the structure of noncommutative tori", *K-Theory* **6**:3 (1992), 267–284. MR Zbl

[Bosa et al. 2019] J. Bosa, N. P. Brown, Y. Sato, A. Tikuisis, S. White, and W. Winter, *Covering dimension of* C**-algebras and 2-coloured classification*, Mem. Amer. Math. Soc. **1233**, Amer. Math. Sci., Providence, RI, 2019. MR Zbl

[Castillejos and Evington 2020] J. Castillejos and S. Evington, "Nuclear dimension of simple stably projectionless C*-algebras", *Anal. PDE* **13**:7 (2020), 2205–2240. MR Zbl

[Castillejos et al. 2021] J. Castillejos, S. Evington, A. Tikuisis, S. White, and W. Winter, "Nuclear dimension of simple C*-algebras", *Invent. Math.* **224**:1 (2021), 245–290. MR Zbl

[Connes 1976] A. Connes, "Classification of injective factors: cases II_1, II_∞, III_λ, $\lambda \neq 1$", *Ann. of Math.* (2) **104**:1 (1976), 73–115. MR Zbl

[Connes 1982] A. Connes, "A survey of foliations and operator algebras", pp. 521–628 in *Operator algebras and applications*, *I* (Kingston, ON, 1980), edited by R. V. Kadison, Proc. Sympos. Pure Math. **38**, Amer. Math. Soc., Providence, RI, 1982. MR Zbl

[Dadarlat and Eilers 2001] M. Dadarlat and S. Eilers, "Asymptotic unitary equivalence in *KK*-theory", *K-Theory* **23**:4 (2001), 305–322. MR Zbl

[Dadarlat and Eilers 2002] M. Dadarlat and S. Eilers, "On the classification of nuclear C*-algebras", *Proc. Lond. Math. Soc.* (3) **85**:1 (2002), 168–210. MR Zbl

[Dadarlat et al. 2009] M. Dadarlat, I. Hirshberg, A. S. Toms, and W. Winter, "The Jiang–Su algebra does not always embed", *Math. Res. Lett.* **16**:1 (2009), 23–26. MR Zbl

[Elliott 1993] G. A. Elliott, "On the classification of C*-algebras of real rank zero", *J. Reine Angew. Math.* **443** (1993), 179–219. MR Zbl

[Elliott 1996] G. A. Elliott, "An invariant for simple C*-algebras", pp. 61–90 in *Canadian Mathematical Society*, 1945–1995, *III*, edited by J. B. Carrell and R. Murty, Canadian Math. Soc., Ottawa, ON, 1996. MR Zbl

[Elliott and Kucerovsky 2001] G. A. Elliott and D. Kucerovsky, "An abstract Voiculescu–Brown–Douglas–Fillmore absorption theorem", *Pacific J. Math.* **198**:2 (2001), 385–409. MR Zbl

[Elliott and Niu 2016] G. A. Elliott and Z. Niu, "The classification of simple separable *KK*-contractible C*-algebras with finite nuclear dimension", preprint, 2016. arXiv 1611.05159

[Elliott et al. 2020a] G. A. Elliott, G. Gong, H. Lin, and Z. Niu, "The classification of simple separable *KK*-contractible C*-algebras with finite nuclear dimension", *J. Geom. Phys.* **158** (2020), art. id. 103861. MR Zbl

[Elliott et al. 2020b] G. A. Elliott, G. Gong, H. Lin, and Z. Niu, "Simple stably projectionless C*-algebras with generalized tracial rank one", *J. Noncommut. Geom.* **14**:1 (2020), 251–347. MR Zbl

[Gabe 2016] J. Gabe, "A note on nonunital absorbing extensions", *Pacific J. Math.* **284**:2 (2016), 383–393. MR Zbl

[Gabe 2020] J. Gabe, "A new proof of Kirchberg's $\mathcal{O}_2$-stable classification", *J. Reine Angew. Math.* **761** (2020), 247–289. MR Zbl

[Gong and Lin 2020] G. Gong and H. Lin, "On classification of non-unital amenable simple C*-algebras, II", *J. Geom. Phys.* **158** (2020), art. id. 103865. MR Zbl

[Hjelmborg and Rørdam 1998] J. v. B. Hjelmborg and M. Rørdam, "On stability of C*-algebras", *J. Funct. Anal.* **155**:1 (1998), 153–170. MR Zbl

[Jacelon 2013] B. Jacelon, "A simple, monotracial, stably projectionless C*-algebra", *J. Lond. Math. Soc.* (2) **87**:2 (2013), 365–383. MR Zbl

[Kirchberg 2006] E. Kirchberg, "Central sequences in C*-algebras and strongly purely infinite algebras", pp. 175–231 in *Operator algebras* (Oslo, 2004), edited by S. Neshveyev and C. Skau, Abel Symposium **1**, Springer, 2006. MR Zbl

[Kirchberg and Phillips 2000] E. Kirchberg and N. C. Phillips, "Embedding of exact C*-algebras in the Cuntz algebra $\mathcal{O}_2$", *J. Reine Angew. Math.* **525** (2000), 17–53. MR Zbl

[Kirchberg and Rørdam 2002] E. Kirchberg and M. Rørdam, "Infinite non-simple C*-algebras: absorbing the Cuntz algebras $\mathcal{O}_\infty$", *Adv. Math.* **167**:2 (2002), 195–264. MR Zbl

[Kirchberg and Rørdam 2014] E. Kirchberg and M. Rørdam, "Central sequence C*-algebras and tensorial absorption of the Jiang–Su algebra", *J. Reine Angew. Math.* **695** (2014), 175–214. MR Zbl

[Kishimoto 1999] A. Kishimoto, "Pairs of simple dimension groups", *Int. J. Math.* **10**:6 (1999), 739–761. MR Zbl

[Kishimoto and Kumjian 1996] A. Kishimoto and A. Kumjian, "Simple stably projectionless C*-algebras arising as crossed products", *Canad. J. Math.* **48**:5 (1996), 980–996. MR Zbl

[Kishimoto and Kumjian 1997] A. Kishimoto and A. Kumjian, "Crossed products of Cuntz algebras by quasi-free automorphisms", pp. 173–192 in *Operator algebras and their applications* (Waterloo, ON, 1994–1995), edited by P. A. Fillmore and J. A. Mingo, Fields Inst. Commun. **13**, Amer. Math. Soc., Providence, RI, 1997. MR Zbl

[Lin and Ng 2023] H. Lin and P. W. Ng, "Extensions of C*-algebras by a small ideal", *Int. Math. Res. Not. IMRN* **2023**:12 (2023), 10350–10438. MR

[Matui and Sato 2012] H. Matui and Y. Sato, "Strict comparison and $\mathcal{Z}$-absorption of nuclear C*-algebras", *Acta Math.* **209**:1 (2012), 179–196. MR Zbl

[Matui and Sato 2014a] H. Matui and Y. Sato, "Decomposition rank of UHF-absorbing C*-algebras", *Duke Math. J.* **163**:14 (2014), 2687–2708. MR Zbl

[Matui and Sato 2014b] H. Matui and Y. Sato, "$\mathcal{Z}$-stability of crossed products by strongly outer actions, II", *Amer. J. Math.* **136**:6 (2014), 1441–1496. MR Zbl

[Nawata 2013] N. Nawata, "Picard groups of certain stably projectionless C*-algebras", *J. Lond. Math. Soc.* (2) **88**:1 (2013), 161–180. MR Zbl

[Nawata 2019] N. Nawata, "Trace scaling automorphisms of the stabilized Razak–Jacelon algebra", *Proc. Lond. Math. Soc.* (3) **118**:3 (2019), 545–576. MR Zbl

[Nawata 2021] N. Nawata, "Rohlin actions of finite groups on the Razak–Jacelon algebra", *Int. Math. Res. Not.* **2021**:4 (2021), 2991–3020. MR Zbl

[Pedersen 1979] G. K. Pedersen, C**-algebras and their automorphism groups*, Lond. Math. Soc. Monogr. **14**, Academic Press, London, 1979. MR Zbl

[Pedersen 1987] G. K. Pedersen, "Unitary extensions and polar decompositions in a C*-algebra", *J. Operator Theory* **17**:2 (1987), 357–364. MR Zbl

[Pedersen 1998] G. K. Pedersen, "Factorization in C*-algebras", *Expo. Math.* **16**:2 (1998), 145–156. MR Zbl

[Razak 2002] S. Razak, "On the classification of simple stably projectionless C*-algebras", *Canad. J. Math.* **54**:1 (2002), 138–224. MR Zbl

[Robert 2012] L. Robert, "Classification of inductive limits of 1-dimensional NCCW complexes", *Adv. Math.* **231**:5 (2012), 2802–2836. MR Zbl

[Robert 2016] L. Robert, "Remarks on $\mathcal{Z}$-stable projectionless C*-algebras", *Glasg. Math. J.* **58**:2 (2016), 273–277. MR Zbl

[Rørdam 1992] M. Rørdam, "On the structure of simple C*-algebras tensored with a UHF-algebra, II", *J. Funct. Anal.* **107**:2 (1992), 255–269. MR Zbl

[Rørdam 2002] M. Rørdam, "Classification of nuclear, simple C*-algebras", pp. 1–145 in *Classification of nuclear* C**-algebras: entropy in operator algebras*, Encycl. Math. Sci. **126**, Springer, 2002. MR Zbl

[Rørdam 2004a] M. Rørdam, "The stable and the real rank of $\mathcal{Z}$-absorbing C*-algebras", *Int. J. Math.* **15**:10 (2004), 1065–1084. MR Zbl

[Rørdam 2004b] M. Rørdam, "Stable C*-algebras", pp. 177–199 in *Operator algebras and applications* (Fukuoka, Japan, 1999), edited by H. Kosaki, Adv. Stud. Pure Math. **38**, Math. Soc. Japan, Tokyo, 2004. MR Zbl

[Sato 2009] Y. Sato, "Certain aperiodic automorphisms of unital simple projectionless C*-algebras", *Int. J. Math.* **20**:10 (2009), 1233–1261. MR Zbl

[Sato 2010] Y. Sato, "The Rohlin property for automorphisms of the Jiang–Su algebra", *J. Funct. Anal.* **259**:2 (2010), 453–476. MR Zbl

[Schafhauser 2020a] C. Schafhauser, "A new proof of the Tikuisis–White–Winter theorem", *J. Reine Angew. Math.* **759** (2020), 291–304. MR Zbl

[Schafhauser 2020b] C. Schafhauser, "Subalgebras of simple AF-algebras", *Ann. of Math.* (2) **192**:2 (2020), 309–352. MR Zbl

[Takesaki 2003] M. Takesaki, *Theory of operator algebras, III*, Encycl. Math. Sci. **127**, Springer, 2003. MR Zbl

[Tikuisis et al. 2017] A. Tikuisis, S. White, and W. Winter, "Quasidiagonality of nuclear C*-algebras", *Ann. of Math.* (2) **185**:1 (2017), 229–284. MR Zbl

Received 7 Jan 2021. Revised 12 Jan 2022. Accepted 14 Feb 2022.

NORIO NAWATA: nawata@ist.osaka-u.ac.jp
Department of Pure and Applied Mathematics, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

ANALYSIS AND PDE
Vol. 16 (2023), No. 8, pp. 1825–1868
DOI: 10.2140/apde.2023.16.1825

INVERSE PROBLEMS FOR NONLINEAR MAGNETIC SCHRÖDINGER EQUATIONS ON CONFORMALLY TRANSVERSALLY ANISOTROPIC MANIFOLDS

KATYA KRUPCHYK AND GUNTHER UHLMANN

We study the inverse boundary problem for a nonlinear magnetic Schrödinger operator on a conformally transversally anisotropic Riemannian manifold of dimension $n \geq 3$. Under suitable assumptions on the nonlinearity, we show that the knowledge of the Dirichlet-to-Neumann map on the boundary of the manifold determines the nonlinear magnetic and electric potentials uniquely. No assumptions on the transversal manifold are made in this result, whereas the corresponding inverse boundary problem for the linear magnetic Schrödinger operator is still open in this generality.

1. Introduction and statement of results

Let (M, g) be a smooth compact oriented Riemannian manifold of dimension $n \geq 3$ with smooth boundary. Let $A \in C^\infty(M, T^*M)$ be a 1-form with complex-valued C^∞ coefficients, and let

$$d_A = d + iA : C^\infty(M) \to C^\infty(M, T^*M),$$

where $d : C^\infty(M) \to C^\infty(M, T^*M)$ is the de Rham differential. We define the formal L^2-adjoint of d_A, $d_A^* : C^\infty(M, T^*M) \to C^\infty(M)$, as

$$(d_A u, v)_{L^2(M,T^*M)} = (u, d_A^* v)_{L^2(M)}, \quad u \in C_0^\infty(M^{\text{int}}), \quad v \in C_0^\infty(M^{\text{int}}, T^*M^{\text{int}}),$$

where $M^{\text{int}} = M \setminus \partial M$ stands for the interior of M. Here and in what follows, when $u, v \in C^\infty(M)$, we write

$$(u, v)_{L^2(M)} = \int_M u\bar{v}\, dV_g$$

for the natural L^2-scalar product, where dV_g is the Riemannian volume element on M. Similarly, when $\alpha, \beta \in C^\infty(M, T^*M)$ are 1-forms, we define the L^2-scalar product

$$(\alpha, \beta)_{L^2(M,T^*M)} = \int_M \langle \alpha, \bar{\beta} \rangle_g \, dV_g(x),$$

where $\langle \cdot, \cdot \rangle_g$ is the pointwise scalar product in the space of 1-forms induced by the Riemannian metric g. In the local coordinates $(x_1, \dots, x_n)$, in which $\alpha = \sum_{j=1}^n \alpha_j \, dx_j$, $\beta = \sum_{j=1}^n \beta_j \, dx_j$, and (g^{jk}) is the

MSC2020: 35R30.
Keywords: inverse boundary problem, nonlinear Schrödinger equation, conformally transversally anisotropic manifold, Gaussian beams.

matrix inverse of (g_{jk}) with $g=\sum_{j,k=1}^n g_{jk}\,dx_j\,dx_k$, we have

$$\langle \alpha,\beta\rangle_g=\sum_{j,k=1}^n g^{jk}\alpha_j\beta_k.$$

We also have

$$d_A^*=d^*-i\langle \bar{A},\cdot\rangle_g.$$

In local coordinates, we see that

$$d^*v=-\sum_{j,k=1}^n |g|^{-1/2}\partial_{x_j}(|g|^{1/2}g^{jk}v_k), \tag{1-1}$$

where $|g|=\det(g_{jk})$ and $v=\sum_{j=1}^n v_j\,dx_j$.

In this paper we shall consider 1-forms and scalar functions depending holomorphically on a parameter $z\in\mathbb{C}$. Specifically, let $A:M\times\mathbb{C}\to T^*M$ and $V:M\times\mathbb{C}\mapsto\mathbb{C}$ satisfy the following conditions:

(A_i) The map $\mathbb{C}\ni z\mapsto A(\,\cdot\,,z)$ is holomorphic with values in $C^{1,1}(M,T^*M)$, the space of 1-forms with complex-valued $C^{1,1}(M)$ coefficients.

(V_i) The map $\mathbb{C}\ni z\mapsto V(\,\cdot\,,z)$ is holomorphic with values in $C^{1,1}(M)$.

(V_{ii}) $V(x,0)=0$, for all $x\in M$.

Here $C^{1,1}(M)$ is the space of C^1 functions on M with a Lipschitz gradient.

It follows from (A_i), (V_i), and (V_{ii}) that A and V can be expanded into the power series

$$A(x,z)=\sum_{k=0}^\infty A_k(x)\frac{z^k}{k!},\quad A_k(x):=\partial_z^k A(x,0)\in C^{1,1}(M,T^*M), \tag{1-2}$$

converging in the $C^{1,1}(M,T^*M)$ topology, and

$$V(x,z)=\sum_{k=1}^\infty V_k(x)\frac{z^k}{k!},\quad V_k(x):=\partial_z^k V(x,0)\in C^{1,1}(M), \tag{1-3}$$

converging in the $C^{1,1}(M)$ topology.

Let us introduce the nonlinear magnetic Schrödinger operator defined by

$$\begin{aligned}L_{A,V}u&=d^*_{\overline{A(\cdot,u)}}\,d_{A(\cdot,u)}u+V(\,\cdot\,,u)\\&=-\Delta_g u+d^*(iA(\,\cdot\,,u)u)-i\langle A(\,\cdot\,,u),du\rangle_g+\langle A(\,\cdot\,,u),A(\,\cdot\,,u)\rangle_g u+V(\,\cdot\,,u),\end{aligned} \tag{1-4}$$

for $u\in C^\infty(M)$. Notice that the first-order linearization of $L_{A,V}$ is the standard linear magnetic Schrödinger operator $d^*_{\overline{A_0}}d_{A_0}+V_1$. Furthermore, we also assume that $A_0\in C^\infty(M,T^*M)$, $V_1\in C^\infty(M)$, and that

(i) 0 is not a Dirichlet eigenvalue of the operator $d^*_{\overline{A_0}}d_{A_0}+V_1$.

Consider the Dirichlet problem for the nonlinear magnetic Schrödinger operator

$$\begin{cases}L_{A,V}u=0 & \text{in } M^{\text{int}},\\ u|_{\partial M}=f.\end{cases} \tag{1-5}$$

It is shown in Theorem B.1 that under the above assumptions, there exist $\delta > 0$ and $C > 0$ such that when $f \in B_\delta(\partial M) := \{f \in C^{2,\alpha}(\partial M) : \|f\|_{C^{2,\alpha}(\partial M)} < \delta\}$, $0 < \alpha < 1$, the problem (1-5) has a unique solution $u = u_f \in C^{2,\alpha}(M)$ satisfying $\|u\|_{C^{2,\alpha}(M)} < C\delta$. Here $C^{2,\alpha}(M)$ stands for the standard Hölder space of functions on M. Associated to the problem (1-5), we define the Dirichlet-to-Neumann map

$$\Lambda_{A,V} f = \partial_\nu u_f|_{\partial M}, \tag{1-6}$$

where $f \in B_\delta(\partial M)$ and ν is the unit outer normal to the boundary.

The inverse problem that we are interested in is whether the knowledge of the Dirichlet-to-Neumann map $\Lambda_{A,V}$ determines the nonlinear magnetic and electric potentials, A and V, respectively.

When $A = 0$ and $V(x, z) = V_1(x)z$, the inverse problem for the linear Schrödinger operator $-\Delta_g + V_1$ is related to the Calderón problem, which has been the object of intense study but remains open in the case of a general smooth Riemannian manifold (M, g) of dimension $n \geq 3$ with smooth boundary. Let us mention that the unique determination of the potential V_1 from the knowledge of the Dirichlet-to-Neumann map Λ_{0,V_1} was established in [Sylvester and Uhlmann 1987] in the Euclidean setting, in [Isozaki 2004] for hyperbolic manifolds, and in [Kohn and Vogelius 1984; Lassas and Uhlmann 2001; Lee and Uhlmann 1989] in the analytic case. The uniqueness in the inverse boundary problem for the linear magnetic Schrödinger operator $d^*_{\bar{A}_0} d_{A_0} + V_1$ up to a suitable gauge transformation was obtained in [Nakamura et al. 1995] in the Euclidean setting; see also [Krupchyk and Uhlmann 2014]. Going beyond these settings, the most general uniqueness results were obtained in the case when the manifold (M, g) is conformally transversally anisotropic and the transversal manifold satisfies some additional assumptions. Following [Dos Santos Ferreira et al. 2009; 2016], let us recall the definition of a conformally transversally anisotropic manifold.

Definition 1.1. A compact smooth oriented Riemannian manifold (M, g) of dimension $n \geq 3$ with smooth boundary is said to be conformally transversally anisotropic if there exists an $(n-1)$-dimensional smooth compact Riemannian manifold (M_0, g_0) with smooth boundary such that $M \Subset \mathbb{R} \times M_0$ and $g = c(e \oplus g_0)$, where e is the Euclidean metric on $\mathbb{R}$ and c is a positive smooth function on M.

In the case when (M, g) is conformally transversally anisotropic, assuming that the transversal manifold (M_0, g_0) is simple in the sense that the boundary ∂M_0 is strictly convex and, for any point $p \in M_0$, the exponential map $\exp_p$ with its maximal domain of definition in $T_p M_0$ is a diffeomorphism onto M_0, the global uniqueness for the inverse boundary problem for the linear magnetic Schrödinger equation up to a gauge was proven in [Dos Santos Ferreira et al. 2009]; see also [Krupchyk and Uhlmann 2018]. Note that the geodesic ray transform on functions and 1-forms is invertible on simple manifolds; see [Anikonov 1978; Muhometov 1977].

These uniqueness results were strengthened in [Dos Santos Ferreira et al. 2016], where the global uniqueness in the inverse boundary problem for the linear Schrödinger equation was established under the assumption that the geodesic ray transform on the transversal manifold is injective. Similar results for the inverse boundary problem for the linear magnetic Schrödinger equation were obtained in [Cekić 2017; Krupchyk and Uhlmann 2018]. The injectivity of the geodesic ray transform is open in general, and has only been established under certain geometric assumptions. In particular, the injectivity of the geodesic ray transform is proven in [Stefanov et al. 2018; Uhlmann and Vasy 2016] when M_0 has strictly convex boundary and is foliated by strictly convex hypersurfaces, and in [Guillarmou 2017; Guillarmou et al. 2021] when M_0 has a hyperbolic trapped set and no conjugate points. As an example of the latter, one can consider a negatively curved manifold M_0. We refer to [Dos Santos Ferreira et al. 2020] where the linearized anisotropic Calderón problem was studied on a transversally anisotropic manifold under certain mild conditions on the transversal manifold related to the geometry of pairs of intersecting geodesics.

Turning the attention to inverse problems for nonlinear PDEs, it was discovered in [Kurylev et al. 2018] that nonlinearity can be helpful in solving inverse problems for hyperbolic equations; see also [Feizmohammadi et al. 2021; Lassas et al. 2018]. Similar phenomena for inverse problems for semilinear elliptic PDEs have been revealed in [Feizmohammadi and Oksanen 2020; Lassas et al. 2021a]; see also [Krupchyk and Uhlmann 2020a; 2020b; Lai and Zhou 2020; Lassas et al. 2021b]. A common feature of all of the aforementioned works is that the presence of a nonlinearity allows one to solve inverse problems for nonlinear equations in cases where the corresponding inverse problem in the linear setting is open.

In particular, the inverse boundary problem for the nonlinear Schrödinger equation

$$L_{0,V}u = -\Delta_g u + V(\cdot\,, u) = 0$$

on a conformally transversally anisotropic manifold (M, g) of dimension $n \geq 3$ was studied in [Feizmohammadi and Oksanen 2020; Lassas et al. 2021a], and the following result was obtained: if V satisfies the assumptions (V_i), (V_{ii}), and

(V_{iii}) $\partial_z V(x, 0) = \partial_z^2 V(x, 0) = 0$, for all $x \in M$,

then the knowledge of the Dirichlet-to-Neumann map $\Lambda_{0,V}$ determines V in $M \times \mathbb{C}$ uniquely. Notice that remarkably there are no assumptions on the transversal manifold in this result while the inverse problem for the linear Schrödinger equation is still open in this generality. The proof of this result relies on higher-order linearizations of the Dirichlet-to-Neumann map, which allow one to reduce the inverse problem to the following density result; see [Lassas et al. 2021a].

Proposition 1.2. *Let (M, g) be a conformally transversally anisotropic manifold of dimension $n \geq 3$, and let $q \in C^{1,1}(M)$. If*

$$\int_M q u_1 u_2 u_3 u_4 \, dV_g = 0, \tag{1-7}$$

for all harmonic functions $u_j \in C^\infty(M)$, $j = 1, 2, 3, 4$, then $q \equiv 0$.

The purpose of this paper is to extend the aforementioned result of [Feizmohammadi and Oksanen 2020; Lassas et al. 2021a] to the nonlinear magnetic Schrödinger equation $L_{A,V}u = 0$ given by (1-4). To state our result, similarly to the assumption (V_{iii}) on the potential V, we shall also suppose that the nonlinear magnetic potential A satisfies

(A_{ii}) $A(x, 0) = \partial_z A(x, 0) = 0$, for all $x \in M$.

Our main result is as follows.

Theorem 1.3. *Let (M, g) be a conformally transversally anisotropic manifold of dimension $n \ge 3$. Let $A^{(1)}, A^{(2)} : M \times \mathbb{C} \to T^*M$ and $V^{(1)}, V^{(2)} : M \times \mathbb{C} \mapsto \mathbb{C}$ satisfy the assumptions (A_i), (A_{ii}), and (V_i), (V_{ii}), (V_{iii}), respectively. If $\Lambda_{A^{(1)},V^{(1)}} = \Lambda_{A^{(2)},V^{(2)}}$ then $A^{(1)} = A^{(2)}$ and $V^{(1)} = V^{(2)}$ in $M \times \mathbb{C}$.*

Remark 1.4. Let us point out that there are no assumptions on the transversal manifold in Theorem 1.3, whereas the corresponding inverse boundary problem for the linear magnetic Schrödinger operator is still open in this generality.

Remark 1.5. Notice that as opposed to the inverse boundary problem for the linear magnetic Schrödinger equation, where one can determine the magnetic potential up to a gauge transformation only, in our nonlinear setting the unique determination of both potentials is possible, due to the assumptions (A_i), (A_{ii}), and (V_i), (V_{ii}), (V_{iii}), which imply that the first-order linearization of the nonlinear equation is given by $-\Delta_g u = 0$, rather than by the linear magnetic Schrödinger equation.

Similarly to [Feizmohammadi and Oksanen 2020; Lassas et al. 2021a], the proof of Theorem 1.3 relies on higher-order linearizations of the Dirichlet-to-Neumann map $\Lambda_{A,V}$, as well as a suitable consequence of the following density result, which may be of some independent interest.

Proposition 1.6. *Let (M, g) be a conformally transversally anisotropic manifold of dimension $n \ge 3$, and let $A \in C^{1,1}(M, T^*M)$ be a 1-form. If*

$$\int_M \langle A, d(u_1 u_2 u_3)\rangle_g u_4 \, dV_g = 0, \tag{1-8}$$

for all harmonic functions $u_j \in C^\infty(M)$, $j = 1, 2, 3, 4$, then $A \equiv 0$.

The starting point in the proof of Proposition 1.6 consists of showing that the boundary traces of the 1-form A, as well as of its normal derivative, vanish, as a consequence of the integral identity (1-8). This allows us to extend A by zero to $\mathbb{R} \times M_0 \setminus M$, while preserving its regularity. The proof of Proposition 1.6 then follows the strategy of the proof of Proposition 1.2 established in [Lassas et al. 2021a]. Specifically, we construct harmonic functions to be used in (1-8), based on suitable Gaussian beams quasimodes associated to two nontangential intersecting geodesics on the transversal manifold M_0. Using the freedom of working with four harmonic functions, we construct a pair of harmonic functions based on a Gaussian beam quasimode v and its complex conjugate $\bar{v}$, concentrated near one geodesic, and another pair of harmonic functions based on a Gaussian beam quasimode w and its complex conjugate $\overline{w}$, concentrated near the other geodesic. The product $d(v\bar{v}w)\overline{w}$ is supported near the finitely many points of intersections of these geodesics, and the product does not have high oscillations. This makes it possible to conclude that $A = 0$, using both nonstationary as well as stationary phase arguments (the Laplace method).

Remark 1.7. Our regularity assumption on A in Proposition 1.6 is motivated by the fact that the continuity of the zero extension of A to $\mathbb{R} \times M_0 \setminus M$ is needed for a rough stationary phase argument and the Lipschitz continuity of the gradient of the zero extension of A is needed for a nonstationary phase argument in the proof of Proposition 1.6.

Returning to the proof of Theorem 1.3, let us mention that due to the assumptions (A_{ii}) and (V_{ii}), (V_{iii}), only the linearizations of the Dirichlet-to-Neumann map of order ≥ 3 become useful when recovering the nonlinear potentials $A(x,z)$ and $V(x,z)$. Considering the m-th order linearization, $m\ge 3$, leads to the integral identity

$$\int_M ((m+1)i\langle A, d(u_1\cdots u_m)\rangle_g u_{m+1} - (mid^*(A)+V)u_1\cdots u_{m+1})\,dV_g = 0, \tag{1-9}$$

where $A = A^{(1)}_{m-1} - A^{(2)}_{m-1}$ and $V = V^{(1)}_m - V^{(2)}_m$, which is valid for any harmonic function $u_l \in C^{2,\alpha}(M)$ with $l = 1,\dots,m+1$. Setting $u_1 = \cdots = u_{m-3} = 1$ in (1-9) gives the identity

$$(m+1)i\int_M \langle A,\ d(u_{m-2}u_{m-1}u_m)\rangle_g u_{m+1}\,dV_g = \int_M (mid^*(A)+V)u_{m-2}u_{m-1}u_m u_{m+1})\,dV_g. \tag{1-10}$$

To proceed, we first show that (1-10) implies that $A|_{\partial M} = 0$ and $\partial_\nu A|_{\partial M} = 0$, and then use a consequence of Proposition 1.6 to obtain that $A\equiv 0$; see Corollary 4.1 below. To recover V, we substitute $A = 0$ in (1-10) and rely on Proposition 1.2.

Remark 1.8. The assumptions (A_i), (A_{ii}), (V_i), (V_{ii}), and (V_{iii}) in Theorem 1.3 are made precisely so that the higher-order linearizations of the Dirichlet-to-Neumann map $\Lambda_{A,V}$ lead to the integral identities (1-9) involving at least four harmonic functions, and the freedom of working with four harmonic functions allows one to solve the inverse boundary problem without any assumption on the transversal manifold; see also [Lassas et al. 2021a].

Let us point out that inverse boundary problems for the nonlinear magnetic Schrödinger equation in the Euclidean space, both in the case of full and partial data, have been studied in [Lai and Zhou 2020]. The density of certain products of gradients of harmonic functions in the Euclidean space has been recently established in [Cârstea and Feizmohammadi 2021], when solving an inverse boundary problem for certain anisotropic quasilinear elliptic equations.

Finally, let us remark that inverse boundary problems for nonlinear elliptic PDEs have been studied extensively in the literature. We refer to [Cârstea and Feizmohammadi 2021; Cârstea et al. 2019; Feizmohammadi and Oksanen 2020; Hervas and Sun 2002; Isakov and Nachman 1995; Isakov and Sylvester 1994; Kang and Nakamura 2002; Krupchyk and Uhlmann 2020a; 2020b; Lai and Zhou 2020; Lassas et al. 2021a; 2021b; Sun 1996; 2004; 2010, Sun and Uhlmann 1997].

The paper is organized as follows. In Section 2 we recall the construction of harmonic functions on a conformally transversally anisotropic manifold based on Gaussian beams quasimodes constructed on $\mathbb{R}\times M_0$ and localized near nontangential geodesics on the transversal manifold M_0. For the convenience of the reader, in Section 3 we provide a proof of Proposition 1.6 in a simplified setting. Section 4 is devoted to the proof of Proposition 1.6 in the general case. The proof of Theorem 1.3 occupies Section 5. Appendix A discusses a standard rough version of stationary phase needed in the proof of Proposition 1.6. In Appendix B, we show the well-posedness of the Dirichlet problem for the nonlinear magnetic Schrödinger equation, in the case of small boundary data. The determination of the first-order boundary traces of a scalar function and a 1-form, via suitable orthogonality relations involving harmonic functions on the manifold M, is presented in Appendix C. Finally, Appendix D discusses some basic properties of geodesics which are used in the body of the paper.

2. Gaussian beams quasimodes and construction of harmonic functions

Let (M, g) be a conformally transversally anisotropic manifold so that $(M, g) \Subset (\mathbb{R} \times M_0, c(e \oplus g_0))$. Let us write $x = (x_1, x')$ for local coordinates in $\mathbb{R} \times M_0$. Note that $\phi(x) = \pm\alpha x_1$, $\alpha > 0$, is a limiting Carleman weight for $-h^2\Delta_g$; see [Dos Santos Ferreira et al. 2009].

Letting $\tilde{g} = e \oplus g_0$, we have

$$c^{(n+2)/4} \circ (-\Delta_g) \circ c^{-(n-2)/4} = -\Delta_{\tilde{g}} + q, \tag{2-1}$$

where

$$q = -c^{(n+2)/4}\Delta_g(c^{-(n-2)/4});$$

see [Dos Santos Ferreira et al. 2016]. Here $q \in C^\infty(\mathbb{R} \times M_0; \mathbb{R})$. It follows from (2-1) that in order to construct harmonic functions on (M, g) based on Gaussian beams quasimodes, we shall need to have Gaussian beams quasimodes for the Schrödinger operator $-\Delta_{\tilde{g}} + q$, conjugated by an exponential weight corresponding to the limiting Carleman weight ϕ. Our quasimodes will be constructed on the manifold $\mathbb{R} \times M_0$ and will be localized to nontangential geodesics on the transversal manifold M_0. A unit speed geodesic $\gamma : [-S_1, S_2] \to M_0$, $0 < S_1, S_2 < \infty$, is called nontangential if $\gamma(-S_1), \gamma(S_2) \in \partial M_0$, $\dot{\gamma}(-S_1), \dot{\gamma}(S_2)$ are nontangential vectors to ∂M_0, and $\gamma(t) \in M_0^{\text{int}}$ for all $-S_1 < t < S_2$; see [Dos Santos Ferreira et al. 2016]. As in [Lassas et al. 2021a], it will be convenient to normalize our quasimodes in $L^4(M_0)$, as later we shall have to deal with products of four such quasimodes. We shall need the following essentially well-known result, see [Feizmohammadi and Oksanen 2020, Section 4.1]; see also [Dos Santos Ferreira et al. 2016; Lassas et al. 2021a].

Proposition 2.1. *Let $\alpha > 0$, and let $\tau = s + i\lambda$, $s \geq 1$, with $\lambda \in \mathbb{R}$ fixed. Then for any $k \in \mathbb{N}$ and $R \geq 1$, there exist $N \in \mathbb{N}$ and families of Gaussian beam quasimodes $v_1(\,\cdot\,; s), v_2(\,\cdot\,; s) \in C^\infty(\mathbb{R} \times M_0)$ such that*

$$\begin{aligned} \|e^{-\alpha\tau x_1}(-\Delta_{\tilde{g}} + q)e^{\alpha\tau x_1} v_1(\,\cdot\,; s)\|_{H^k((I\times M_0)^{\text{int}})} &= \mathcal{O}(s^{-R}), \\ \|e^{\alpha\tau x_1}(-\Delta_{\tilde{g}} + q)e^{-\alpha\tau x_1} v_2(\,\cdot\,; s)\|_{H^k((I\times M_0)^{\text{int}})} &= \mathcal{O}(s^{-R}), \end{aligned} \tag{2-2}$$

and

$$\|v_j(\,\cdot\,; s)\|_{L^4(I\times M_0)} = \mathcal{O}(1), \quad \|v_j(\,\cdot\,; s)\|_{L^\infty(I\times M_0)} = \mathcal{O}(1)s^{(n-2)/8}, \quad j = 1, 2, \tag{2-3}$$

as $s \to \infty$. Here $I \subset \mathbb{R}$ is an arbitrary bounded interval. The local structure of the quasimodes is as follows: Let $p \in \gamma([-S_1, S_2])$ and let $t_1 < \cdots < t_P$ be the times in $[-S_1, S_2]$ when $\gamma(t_l) = p$. In a sufficiently small neighborhood U of p, the quasimode v_j is a finite sum

$$v_j|_U = v_j^{(1)} + \cdots + v_j^{(P)}.$$

Each $v_j^{(l)}$ has the form

$$v_1^{(l)} = s^{(n-2)/8} e^{i\alpha\tau\varphi^{(l)}} a^{(l)}, \quad v_2^{(l)} = s^{(n-2)/8} e^{i\alpha\tau\varphi^{(l)}} b^{(l)}, \quad l = 1, \ldots, P,$$

where $\varphi = \varphi^{(l)} \in C^\infty(\overline{U}; \mathbb{C})$ satisfies, for t close to t_l,

$$\varphi(\gamma(t)) = t, \quad \nabla\varphi(\gamma(t)) = \dot{\gamma}(t), \quad \operatorname{Im}(\nabla^2\varphi(\gamma(t))) \geq 0, \quad \operatorname{Im}(\nabla^2\varphi)|_{\dot{\gamma}(t)^\perp} > 0,$$

and $a^{(l)}, b^{(l)} \in C^\infty(\mathbb{R}\times \overline{U})$ are of the form

$$a^{(l)}(x_1,t,y;s)=\Big(\sum_{j=0}^{N}\tau^{-j}a_j^{(l)}\Big)\chi\Big(\frac{y}{\delta'}\Big),\quad b^{(l)}(x_1,t,y;s)=\Big(\sum_{j=0}^{N}\tau^{-j}b_j^{(l)}\Big)\chi\Big(\frac{y}{\delta'}\Big),$$

where $a_0^{(l)}=b_0^{(l)}$ is independent of x_1 and the potential q,

$$a_0^{(l)}(t,y)=a_{00}^{(l)}(t)+\mathcal{O}(|y|),\qquad a_{00}^{(l)}(t)\neq 0,\quad \textit{for all } t,$$
$$a_1^{(l)}(x_1,t,y)=a_{10}^{(l)}(x_1,t)+\mathcal{O}(|y|),\quad b_1^{(l)}(x_1,t,y)=b_{10}^{(l)}(x_1,t)+\mathcal{O}(|y|).$$

Here $a_{10}^{(l)}(x_1,t)=e^{f^{(l)}(t)}\tilde a_{10}^{(l)}(x_1,t)$ and $b_{10}^{(l)}(x_1,t)=e^{f^{(l)}(t)}\tilde b_{10}^{(l)}(x_1,t)$, where $f^{(l)}$ is independent of the potential q, and further we have that $\tilde a_{10}^{(l)}$ and $\tilde b_{10}^{(l)}$ satisfy the equations

$$(\partial_{x_1}+i\partial_t)\tilde a_{10}^{(l)}=\frac{1}{\alpha}\Big(-\frac{1}{2}e^{-f^{(l)}}(\Delta_{\tilde g}a_0^{(l)})|_{y=0}+C_0^{(l)}q(x_1,t,0)\Big),$$
$$(\partial_{x_1}-i\partial_t)\tilde b_{10}^{(l)}=\frac{1}{\alpha}\Big(\frac{1}{2}e^{-f^{(l)}}(\Delta_{\tilde g}a_0^{(l)})|_{y=0}-C_0^{(l)}q(x_1,t,0)\Big),$$

where $C_0^{(l)}\neq 0$ is a constant, independent of the potential q. Here (t,y) are the Fermi coordinates for γ for t close to t_l, $\chi\in C_0^\infty(\mathbb{R}^{n-2})$ is such that $0\le\chi\le 1$, $\chi=1$ for $|y|\le\frac{1}{4}$ and $\chi=0$ for $|y|\ge\frac{1}{2}$, and $\delta'>0$ is a fixed number that can be taken arbitrarily small.

Remark 2.2. In the special case when the conformal factor c is equal to 1, we have $q=0$, $g=\tilde g$, and

$$e^{\mp\alpha\tau x_1}\circ(-\Delta_g)\circ e^{\pm\alpha\tau x_1}=-\Delta_g\mp 2\alpha\partial_{x_1}-(\alpha\tau)^2.$$

Thus, we can take the Gaussian beams quasimodes in (2-2) to be equal, $v_1=v_2$, and independent of x_1.

Next we shall construct harmonic functions on (M,g) based on the Gaussian beams quasimodes of Proposition 2.1. To that end, we shall use the approach of [Dos Santos Ferreira et al. 2009], based on Carleman estimates with limiting Carleman weights. The construction is standard, see [Dos Santos Ferreira et al. 2016; Lassas et al. 2021a], and is presented here for the convenience of the reader only.

Assume, as we may, that (M,g) is embedded in a compact smooth manifold (N,g) without boundary of the same dimension. Our starting point is the following Carleman estimates for the Schrödinger operator, which is established in [Dos Santos Ferreira et al. 2009, Lemma 4.3].

Proposition 2.3. *Let $q\in C^\infty(M)$. Then given any $t\in\mathbb{R}$, we have for all $h>0$ small enough and all $u\in C_0^\infty(M^{\text{int}})$ that*

$$h\|u\|_{H^t_{\text{scl}}(N)}\le C\|e^{\phi/h}(-h^2\Delta+h^2q)e^{-\phi/h}u\|_{H^t_{\text{scl}}(N)},\quad C>0. \tag{2-4}$$

Here $H^t(N)$, $t\in\mathbb{R}$, is the standard Sobolev space, equipped with the natural semiclassical norm

$$\|u\|_{H^t_{\text{scl}}(N)}=\|(1-h^2\Delta_g)^{t/2}u\|_{L^2(N)}.$$

Using a standard argument, see [Dos Santos Ferreira et al. 2009], we convert the Carleman estimate (2-4) into the following solvability result.

Proposition 2.4. *Let $t \in \mathbb{R}$. If $h > 0$ is small enough, then for any $v \in H^t(M^{\text{int}})$, there is a solution $u \in H^t(M^{\text{int}})$ of the equation*

$$e^{\phi/h}(-h^2\Delta + h^2 q)e^{-\phi/h}u = v \quad \text{in } M^{\text{int}}$$

which satisfies

$$\|u\|_{H^t_{\text{scl}}(M^{\text{int}})} \le \frac{C}{h}\|v\|_{H^t_{\text{scl}}(M^{\text{int}})}.$$

Here

$$H^t(M^{\text{int}}) = \{V|_{M^{\text{int}}} : V \in H^t(N)\}, \quad t \in \mathbb{R},$$

with the norm

$$\|v\|_{H^t_{\text{scl}}(M^0)} = \inf_{V \in H^t_{\text{scl}}(N), v = V|_{M^{\text{int}}}} \|V\|_{H^t_{\text{scl}}(N)}.$$

Let $\alpha > 0$, and let

$$\tau = s + i\lambda \quad \text{with } 1 \le s = \frac{1}{h}, \quad \lambda \in \mathbb{R}, \quad \lambda \text{ fixed.}$$

In view of (2-1), to construct suitable harmonic functions on (M, g), we shall find complex geometric optics solution to the equation

$$(-\Delta_{\tilde{g}} + q)\tilde{u} = 0 \quad \text{in } M^{\text{int}} \tag{2-5}$$

having the form

$$\tilde{u}_1 = e^{\alpha\tau x_1}(v_1 + r_1) \quad \text{and} \quad \tilde{u}_2 = e^{-\alpha\tau x_1}(v_2 + r_2),$$

where v_1 and v_2 are the Gaussian beam quasimodes given in Proposition 2.1, and r_1 and r_2 are the remainder terms. Thus, $\tilde{u}_1$ is a solution of (2-5) provided that

$$e^{-\alpha x_1/h}(-h^2\Delta_{\tilde{g}} + h^2 q)e^{\alpha x_1/h}(e^{i\alpha\lambda x_1}r_1) = -e^{i\alpha\lambda x_1}e^{-\alpha\tau x_1}(-h^2\Delta_{\tilde{g}} + h^2 q)e^{\alpha\tau x_1}v_1. \tag{2-6}$$

For any $k \in \mathbb{N}$ and $R \ge 1$ arbitrarily large, Propositions 2.4 and 2.1 imply that there is $r_1 \in H^k(M^{\text{int}})$ such that

$$\|r_1\|_{H^k_{\text{scl}}(M^{\text{int}})} \le \mathcal{O}(h^{-1})\|e^{-\alpha\tau x_1}(-h^2\Delta_{\tilde{g}} + h^2 q)e^{\alpha\tau x_1}v_1\|_{H^k_{\text{scl}}(M^{\text{int}})} = \mathcal{O}(h^{R-1}),$$

and therefore, for any K, there is R large enough so that

$$\|r_1\|_{H^k(M^{\text{int}})} \le h^{-k}\|r_1\|_{H^k_{\text{scl}}(M^{\text{int}})} = \mathcal{O}(h^K).$$

Similarly, one can construct r_2. This together with (2-1) gives the following result concerning the construction of harmonic functions on (M, g) based on Gaussian beams quasimodes.

Proposition 2.5. *Let $\alpha > 0$, and let $\tau = s + i\lambda$, $s = 1/h$, with $\lambda \in \mathbb{R}$ being fixed. For all k, K, and $h > 0$ small enough, there are $u_1, u_2 \in H^k(M^{\text{int}})$ solutions of $-\Delta_g u_j = 0$ in M^{int} having the form*

$$u_1 = e^{\alpha\tau x_1}c^{-(n-2)/4}(v_1 + r_1) \quad \text{and} \quad u_2 = e^{-\alpha\tau x_1}c^{-(n-2)/4}(v_2 + r_2),$$

where $v_1 = v_1(\,\cdot\,; s)$, $v_2 = v_2(\,\cdot\,; s) \in C^\infty(\mathbb{R} \times M_0)$ are the Gaussian beam quasimodes from Proposition 2.1, and $r_1, r_2 \in H^k(M^{\text{int}})$ are such that $\|r_j\|_{H^k(M^{\text{int}})} = \mathcal{O}(h^K)$ as $h \to 0$.

Remark 2.6. Taking $k > \frac{1}{2}n + 3$ and using the Sobolev embedding $H^k(M^{\text{int}}) \subset C^3(M)$, we see that $u_j \in C^3(M)$ with

$$\|r_j\|_{C^3(M)} = \mathcal{O}(h^K),$$

as $h \to 0$, $j = 1, 2$.

3. Proof of Proposition 1.6 in a simplified setting

The proof of Proposition 1.6 will follow along the lines of the proof of [Lassas et al. 2021a, Proposition 5.1]. Before we prove Proposition 1.6 in the general case, let us explain the main ideas in a simplified setting.

Let us assume that each point $p \in M_0^{\text{int}}$ is the unique intersection point of two distinct nontangential non-self-intersecting geodesics γ and η. Assume furthermore that the conformal factor c equals 1. As we shall see below, in this simplified setting the continuity of A suffices, and therefore to extend A by 0 to the continuous form on $\mathbb{R} \times M_0 \setminus M$, we only need to show $A|_{\partial M} = 0$. This follows by taking $u_2 = u_3 = 1$ in (1-8) and applying Proposition C.3.

In view of Proposition C.5, we see that (1-8) also holds for all harmonic functions $u_j \in C^{2,\alpha}(M)$, $0 < \alpha < 1$, $j = 1, \dots, 4$.

Let $s = 1/h$, and let $\lambda \in \mathbb{R}$ be fixed. Our choice of the harmonic functions below will be similar to [Lassas et al. 2021a]. Specifically, using Proposition 2.5 and Remark 2.6, we see that there exist harmonic functions $u_j \in C^3(M)$, $j = 1, \dots, 4$, on (M, g) of the form

$$\begin{aligned} u_1 &= e^{-(s+i\lambda)x_1}(v + r_1), & u_2 &= \overline{e^{(s+i\lambda)x_1}(v + r_2)}, \\ u_3 &= e^{-sx_1}(w + r_3), & u_4 &= \overline{e^{sx_1}(w + r_4)}, \end{aligned} \tag{3-1}$$

where

$$\|r_j\|_{C^1(M)} = \mathcal{O}(s^{-K}), \tag{3-2}$$

as $s \to \infty$, $K \gg 1$, and $v = v(\,\cdot\,; s)$, $w = w(\,\cdot\,; s) \in C^\infty(M_0)$ are Gaussian beams quasimodes concentrating near the geodesics η and γ, respectively, constructed in Proposition 2.1; see also Remark 2.2. We have

$$v(x'; s) = s^{(n-2)/8} e^{i(s+i\lambda)\varphi(x')} a(x'; s) \quad \text{and} \quad w(x'; s) = s^{(n-2)/8} e^{is\psi(x')} b(x'; s), \tag{3-3}$$

where

$$\begin{aligned} &\varphi(\eta(t)) = t, & &\nabla\varphi(\eta(t)) = \dot{\eta}(t), & &\operatorname{Im}(\nabla^2\varphi(\eta(t))) \geq 0, & &\operatorname{Im}(\nabla^2\varphi)|_{\dot{\eta}(t)^\perp} > 0, \\ &\psi(\gamma(\tau)) = \tau, & &\nabla\psi(\gamma(\tau)) = \dot{\gamma}(\tau), & &\operatorname{Im}(\nabla^2\psi(\gamma(\tau))) \geq 0, & &\operatorname{Im}(\nabla^2\psi)|_{\dot{\gamma}(\tau)^\perp} > 0, \end{aligned} \tag{3-4}$$

and

$$a(t, y; s) = \left(\sum_{j=0}^{N} \tau^{-j} a_j\right) \chi\left(\frac{y}{\delta'}\right), \quad b(\tau, z; s) = \left(\sum_{j=0}^{N} \tau^{-j} b_j\right) \chi\left(\frac{z}{\delta'}\right), \tag{3-5}$$

where

$$\begin{aligned} a_0(t, y) &= a_{00}^{(l)}(t) + \mathcal{O}(|y|), & a_{00}(t) &\neq 0, & &\text{for all } t, \\ b_0(\tau, z) &= a_{00}(\tau) + \mathcal{O}(|z|), & b_{00}(\tau) &\neq 0, & &\text{for all } \tau. \end{aligned} \tag{3-6}$$

Here (t, y) and (τ, z) are the Fermi coordinates for the geodesics η and γ, $\chi \in C_0^\infty(\mathbb{R}^{n-2})$ is such that $0 \le \chi \le 1$, $\chi = 1$ for $|y| \le \frac{1}{4}$ and $\chi = 0$ for $|y| \ge \frac{1}{2}$, and $\delta' > 0$ is a fixed number that can be taken arbitrarily small. We also have

$$\|v\|_{L^4(M_0)} = \|w\|_{L^4(M_0)} = \mathcal{O}(1), \quad \|v\|_{L^\infty(M_0)} = \|w\|_{L^\infty(M_0)} = \mathcal{O}(s^{(n-2)/8}), \tag{3-7}$$

as $s \to \infty$. Similarly, we find that

$$\begin{aligned} &\|s^{(n-2)/8} e^{i(s+i\lambda)\varphi} \nabla a\|_{L^4(M_0)} = \|s^{(n-2)/8} e^{is\psi} \nabla b\|_{L^4(M_0)} = \mathcal{O}(1), \\ &\|\nabla v\|_{L^4(M_0)} = \mathcal{O}(s), \qquad \|\nabla w\|_{L^4(M_0)} = \mathcal{O}(s), \\ &\|\nabla v\|_{L^\infty(M_0)} = \mathcal{O}(s^{(n+6)/8}), \quad \|\nabla w\|_{L^\infty(M_0)} = \mathcal{O}(s^{(n+6)/8}), \end{aligned} \tag{3-8}$$

as $s \to \infty$.

Now it follows from (3-1) that

$$(u_1 u_2 u_3)(x) = e^{-2i\lambda x_1 - s x_1}(|v(x')|^2 w(x') + R(x)),$$

where

$$R = |v|^2 r_3 + (w + r_3)(v\bar{r}_2 + \bar{v} r_1 + r_1 \bar{r}_2).$$

Using (3-2), (3-7), and (3-8), we see that

$$\|R\|_{C^1(M)} = \mathcal{O}(s^{-L}), \tag{3-9}$$

where L is large depending on K. Hence, we have

$$\partial_{x_1}(u_1 u_2 u_3) = e^{-2i\lambda x_1 - s x_1}[(-2i\lambda - s)(|v|^2 w + R) + \partial_{x_1} R],$$

and therefore, using (3-9), (3-2), and (3-7), we get

$$\partial_{x_1}(u_1 u_2 u_3) u_4 = -s e^{-2i\lambda x_1} |v|^2 |w|^2 + \mathcal{O}_{L^1(M)}(1), \tag{3-10}$$

as $s \to \infty$. We also get

$$\partial_{x_k}(u_1 u_2 u_3) = e^{-2i\lambda x_1 - s x_1}(\partial_{x_k}(|v|^2 w) + \partial_{x_k}(R))$$

for $k = 2, \dots, n$, and therefore, (3-9), (3-2), (3-7), and (3-8) yield

$$\partial_{x_k}(u_1 u_2 u_3) u_4 = e^{-2i\lambda x_1} \partial_{x_k}(|v|^2 w) \overline{w} + \mathcal{O}_{L^1(M)}(1), \tag{3-11}$$

as $s \to \infty$. Writing $A = (A_1, A')$ and using (3-10) and (3-11), we conclude that

$$\langle A, d(u_1 u_2 u_3) \rangle_g u_4 = e^{-2i\lambda x_1}(-s A_1 |v|^2 |w|^2 + \langle A', d_{x'}(|v|^2 w) \overline{w} \rangle_{g_0}) + \mathcal{O}_{L^1(M)}(1), \tag{3-12}$$

as $s \to \infty$. It follows from (1-8) with the help of (3-12) that

$$\int_M e^{-2i\lambda x_1}(-s A_1 |v|^2 |w|^2 + \langle A', d_{x'}(|v|^2 w) \overline{w} \rangle_{g_0}) \, dV_g = \mathcal{O}(1), \tag{3-13}$$

as $s \to \infty$.

Extending A by zero to $\mathbb{R} \times M_0 \setminus M$, and denoting the extension again by A, we now see that $A \in C(\mathbb{R} \times M_0, T^*(\mathbb{R} \times M_0))$ as $A|_{\partial M} = 0$. Denoting the partial Fourier transform of A in the x_1 variable by $\hat{A}(\lambda, x')$, we get from (3-13) that

$$\int_{M_0} (-s\hat{A}_1(2\lambda, \cdot)|v|^2|w|^2 + \langle \hat{A}'(2\lambda, \cdot), d_{x'}(|v|^2 w)\overline{w}\rangle_{g_0})\, dV_{g_0} = \mathcal{O}(1), \tag{3-14}$$

as $s \to \infty$. Since v and w can be chosen to be supported in arbitrarily small but fixed neighborhoods of η and γ, respectively, and since η and γ only intersect at p, the products $|v|^2|w|^2$ and $d_{x'}(|v|^2 w)\overline{w}$ concentrate in a small neighborhood U of p. Using (3-3) and (3-5), we see that in U,

$$\begin{aligned} |v|^2|w|^2 &= s^{(n-2)/2} e^{-2s(\operatorname{Im}\varphi + \operatorname{Im}\psi)} e^{-2\lambda \operatorname{Re}\varphi} (|a_0|^2|b_0|^2 + \mathcal{O}_{L^\infty(M_0)}(1/s)) \\ &= s^{(n-2)/2} e^{-2s(\operatorname{Im}\varphi + \operatorname{Im}\psi)} e^{-2\lambda \operatorname{Re}\varphi} |a_0|^2|b_0|^2 + \mathcal{O}_{L^1(M_0)}(1/s), \end{aligned} \tag{3-15}$$

and

$$\begin{aligned} d_{x'}(|v|^2 w)\overline{w} = s^{(n-2)/2} e^{-2s(\operatorname{Im}\varphi + \operatorname{Im}\psi)} e^{-2\lambda \operatorname{Re}\varphi} \big[&(is(2i\, d\operatorname{Im}\varphi + d\psi)(|a_0|^2|b_0|^2 + \mathcal{O}_{L^\infty(M_0)}(1/s)) \\ &- 2\lambda(d\operatorname{Re}\varphi)|a|^2|b|^2 + d_{x'}(|a|^2 b)\bar{b}\big] \\ = s^{(n-2)/2} e^{-2s(\operatorname{Im}\varphi + \operatorname{Im}\psi)} e^{-2\lambda \operatorname{Re}\varphi} &is(2i\, d\operatorname{Im}\varphi + d\psi)|a_0|^2|b_0|^2 + \mathcal{O}_{L^1(M_0)}(1), \end{aligned} \tag{3-16}$$

as $s \to \infty$. Substituting (3-15) and (3-16) into (3-14) and dividing by $s^{1/2}$, we obtain

$$s^{(n-1)/2} \int_U (-\hat{A}_1(2\lambda, \cdot) + i\langle \hat{A}'(2\lambda, \cdot), 2i\, d\operatorname{Im}\varphi + d\psi\rangle_{g_0}) e^{-2\lambda \operatorname{Re}\varphi} |a_0|^2|b_0|^2 e^{-s\Psi}\, dV_{g_0} = \mathcal{O}(s^{-1/2}), \tag{3-17}$$

as $s \to \infty$, where

$$\Psi = 2(\operatorname{Im}\varphi + \operatorname{Im}\psi).$$

It follows from (3-4) that

$$\Psi(p) = 0, \quad d\Psi(p) = 0, \quad \nabla^2\Psi(p) > 0,$$

where the later inequality is a consequence of the fact that the Hessians of $\operatorname{Im}\varphi$ and $\operatorname{Im}\psi$ at p are positive definite in the directions orthogonal to η and γ, respectively.

Let us now denote by $z = (z_1, \dots, z_{n-1})$ the geodesic normal coordinates in (M_0, g_0) with the origin at p. Then

$$g_0(z) = 1 + \mathcal{O}(|z|^2), \tag{3-18}$$

see [Petersen 2006, Chapter 2, Section 8, p. 56], and $dV_{g_0} = |g_0(z)|^{1/2}\, dz$. Passing to the limit as $s \to \infty$ in (3-17) and using the rough version of the stationary phase Lemma A.1, as well as (3-18), we obtain

$$(-\hat{A}_1(2\lambda, p) + i\hat{A}'(2\lambda, p)(\dot{\gamma}(t_0))) e^{-2\lambda \operatorname{Re}\varphi(p)} |a_{00}(p)|^2 |b_{00}(p)|^2 = 0,$$

where $p = \gamma(t_0)$, for all $\lambda \in \mathbb{R}$. As $a_{00}(p) \neq 0$, $b_{00}(p) \neq 0$, and λ is arbitrary, we see that

$$-A_1(x_1, p) + iA'(x_1, p)(\dot{\gamma}(t_0)) = 0,$$

which is equivalent to

$$(iA_1, A')(x_1, p)(1, \dot{\gamma}(t_0)) = 0. \tag{3-19}$$

Here we may replace $\dot\gamma(t_0)$ by $-\dot\gamma(t_0)$. Thus, (3-19) gives that $A_1(x_1, p) = 0$, and since the point (x_1, p) is an arbitrary point in $\mathbb{R} \times M_0$, we get $A_1 \equiv 0$. Hence, we only need to show that the 1-form $A'(x_1, \cdot\,)$ vanishes on M_0, knowing that

$$A'(x_1, p)(\dot\gamma(t_0)) = 0. \tag{3-20}$$

To that end, we assume without loss of generality that $v_1 = \dot\gamma(t_0) = (1, 0, \dots, 0) \in \mathbb{R}^{n-1}$, and consider the small perturbations of v_1 given by

$$v_2 = \frac{1}{\sqrt{1+\varepsilon^2}}(1, \varepsilon, 0, \dots, 0), \quad \dots, \quad v_{n-1} = \frac{1}{\sqrt{1+\varepsilon^2}}(1, 0, \dots, 0, \varepsilon), \tag{3-21}$$

for $\varepsilon > 0$ small. The unit vectors $v_1, \dots, v_{n-1}$ are linearly independent, and thus, they span the tangent space T_pM_0. By Proposition D.2, for $\varepsilon > 0$ sufficiently small, the unit speed geodesic γ_{p,v_j} through (p, v_j), $j = 2, \dots, n-1$, is nontangential between boundary points, does not have self-intersections, and intersects η at the point p only. Applying the discussion above with $\gamma = \gamma_{p,v_j}$, we obtain that $A'(x_1, p)(v_j) = 0$, $j = 2, \dots, n-1$. This together with (3-20) gives that $A'(x_1, p) = 0$. The proof of Proposition 1.6 in the simplified case is complete.

4. Proof of Proposition 1.6 in the general setting

In the case of a general transversal manifold M_0, the nontangential geodesics γ and η might have self-intersections and may intersect more than in one point, which complicates the proof. To proceed we shall follow [Lassas et al. 2021a] and introduce additional parameters in the construction of harmonic functions. Furthermore, we shall implement the presence of the conformal factor c which is assumed to be equal to 1 in [Lassas et al. 2021a].

Let us proceed to discuss the choice of two nontangential geodesics to be used when constructing Gaussian beams quasimodes. When doing so let us first observe that arguing as in the proof of Theorem 1.2 of [Salo 2017], we may assume that (M_0, g_0) has a strictly convex boundary. An application of [Salo 2017, Lemma 3.1] gives therefore that there exists a null set E in (M_0, g_0) such that all points in $M_0 \setminus E$ lie on some nontangential geodesic joining boundary points. Fix a point $y_0 \in M_0^{\text{int}} \setminus E$ and let $\gamma : [-S_1, S_2] \to M_0$, $0 < S_1, S_2 < \infty$, be a unit speed nontangential geodesic such that $\gamma(0) = y_0$. Then by Proposition D.1, moving the point y_0 along γ a little and reparametrizing the geodesic, if necessary, there exists a small neighborhood $W \subset S_{y_0}M_0$ of $w_0 = \dot\gamma(0)$ such that for every $w \in W$, $w \neq w_0$, the unit speed geodesic $\eta : [-T_1, T_2] \to M_0$, $0 < T_1, T_2 < \infty$, such that $\eta(0) = y_0$ and $\dot\eta(0) = w$ is also nontangential, and γ and η do not intersect each other at the boundary of M_0. Notice that γ and η are distinct and are not reverses of each other. As we shall see below, the fact that γ and η do not intersect each other at the boundary of M_0 allows us to avoid the use of stationary and nonstationary phase on the boundary of M_0.

By [Lassas et al. 2021a], we know that γ and η can intersect only finitely many times. Let us denote by $p_1, \dots, p_N \in M_0^{\text{int}}$ the distinct intersection points of γ and η. For each r, $r = 1, \dots, N$, let $t_1^{(r)} < \cdots < t_{P_r}^{(r)}$ be the times in $[-T_1, T_2]$ when $\eta(t_j^r) = p_r$, and let $\tau_1^{(r)} < \cdots < \tau_{Q_r}^{(r)}$ be the times in $[-S_1, S_2]$ when $\gamma(\tau_j^{(r)}) = p_r$. Let U_r be a small neighborhood of p_r, $r = 1, \dots, N$.

Choosing harmonic functions. First it follows from Proposition C.5 that (1-8) continues to hold for all harmonic functions $u_j \in C^{2,\alpha}(M),\ 0<\alpha<1,\ j=1,\dots,4$.

Let $s\ge 1$ and let $L>0,\ \lambda,\mu\in\mathbb{R}$ be fixed. By Proposition 2.5 and Remark 2.6, there are harmonic functions $u_j\in C^3(M)$ of the form

$$\begin{aligned} u_1&=e^{(s+i\mu)x_1}c^{-(n-2)/4}(v_1+r_1), & u_2&=\overline{e^{-(s+i\mu)x_1}c^{-(n-2)/4}(v_2+r_2)},\\ u_3&=e^{-L(s+i\lambda)x_1}c^{-(n-2)/4}(w_1+r_3), & u_4&=\overline{e^{L(s+i\lambda)x_1}c^{-(n-2)/4}(w_2+r_4)}, \end{aligned}\tag{4-1}$$

where

$$\|r_j\|_{C^1(M)}=\mathcal{O}(s^{-K}),\tag{4-2}$$

as $s\to\infty,\ K\gg 1$, and $v_j\in C^\infty(\mathbb{R}\times M_0),\ j=1,2$, and $w_j\in C^\infty(\mathbb{R}\times M_0),\ j=1,2$, are the Gaussian beam quasimodes constructed in Proposition 2.1 and associated to the nontangential geodesics η and γ, respectively, such that

$$\operatorname{supp}(v_j(\,\cdot\,;s))\subset\mathbb{R}\times\text{small neigh}(\eta)\quad\text{and}\quad\operatorname{supp}(w_j(\,\cdot\,;s))\subset\mathbb{R}\times\text{small neigh}(\gamma).\tag{4-3}$$

Notice that here we follow [Lassas et al. 2021a], and the minor differences are as follows: in order to incorporate the presence of the conformal factor our Gaussian beams quasimodes are constructed on all of $\mathbb{R}\times M_0$ rather than on M_0 as in that work, and the parameters μ and λ are real.

Let us now recall a local description of the quasimodes v_j and w_j near the intersection points p_r of γ and η. In doing so, let us fix p to be one of the intersection points p_r and let us set $U=U_r$. In the open set U, the quasimodes v_j are of the form

$$v_j|_U=\sum_{k=1}^{P}v_j^{(k)},\quad j=1,2,\tag{4-4}$$

where $t_1<\dots<t_P$ are the times in $[-T_1,T_2]$ when $\eta(t_k)=p$. Each $v_1^{(k)}$ and $v_2^{(k)}$ in (4-4) has the form

$$v_1^{(k)}=s^{(n-2)/8}e^{i(s+i\mu)\varphi^{(k)}}a^{(k)},\quad v_2^{(k)}=s^{(n-2)/8}e^{i(s+i\mu)\varphi^{(k)}}b^{(k)},\quad k=1,\dots,P,\tag{4-5}$$

where $\varphi=\varphi^{(k)}\in C^\infty(\overline{U};\mathbb{C})$ satisfies, for t close to t_k,

$$\varphi(\eta(t))=t,\quad \nabla\varphi(\eta(t))=\dot\eta(t),\quad \operatorname{Im}(\nabla^2\varphi(\eta(t)))\ge 0,\quad \operatorname{Im}(\nabla^2\varphi)|_{\dot\eta(t)^\perp}>0,\tag{4-6}$$

and each $a^{(k)},b^{(k)}\in C^\infty(\mathbb{R}\times\overline{U})$ is of the form

$$a^{(k)}(x_1,t,y;s)=\biggl(\sum_{j=0}^{N}\tau^{-j}a_j^{(k)}\biggr)\chi\Bigl(\frac{y}{\delta'}\Bigr),\quad b^{(k)}(x_1,t,y;s)=\biggl(\sum_{j=0}^{N}\tau^{-j}b_j^{(k)}\biggr)\chi\Bigl(\frac{y}{\delta'}\Bigr),\tag{4-7}$$

where $a_0^{(k)}=b_0^{(k)}$ is independent of x_1 and

$$a_0^{(k)}(t,y)=a_{00}^{(k)}(t)+\mathcal{O}(|y|),\quad a_{00}^{(k)}(t)\ne 0,\quad\text{for all }t.\tag{4-8}$$

Here (t,y) are the Fermi coordinates for η for t close to t_k, $\chi\in C_0^\infty(\mathbb{R}^{n-2})$ is such that $0\le\chi\le 1,\ \chi=1$ for $|y|\le\frac14$ and $\chi=0$ for $|y|\ge\frac12$, and $\delta'>0$ is a fixed number that can be taken arbitrarily small.

Furthermore, in U, the quasimodes w_1 and w_2 are finite sums

$$w_j|_U = \sum_{k=1}^{Q} w_j^{(k)}, \quad j = 1, 2, \tag{4-9}$$

where $\tau_1 < \cdots < \tau_Q$ are the times in $[-S_1, S_2]$ when $\gamma(\tau_k) = p$. Each $w_1^{(k)}$ and $w_2^{(k)}$ in (4-9) has the form

$$w_1^{(k)} = s^{(n-2)/8} e^{Li(s+i\lambda)\psi^{(k)}} c^{(k)}, \quad w_2^{(k)} = s^{(n-2)/8} e^{Li(s+i\lambda)\psi^{(k)}} d^{(k)}, \quad k = 1, \dots, Q, \tag{4-10}$$

where each $\psi = \psi^{(k)} \in C^\infty(\overline{U}; \mathbb{C})$ satisfies, for τ close to τ_k,

$$\psi(\gamma(\tau)) = \tau, \quad \nabla\psi(\gamma(\tau)) = \dot{\gamma}(\tau), \quad \operatorname{Im}(\nabla^2\psi(\gamma(\tau))) \geq 0, \quad \operatorname{Im}(\nabla^2\psi)|_{\dot{\gamma}(\tau)^\perp} > 0, \tag{4-11}$$

and each $c^{(k)}, d^{(k)} \in C^\infty(\mathbb{R} \times \overline{U})$ is of the form

$$c^{(k)}(x_1, \tau, z; s) = \biggl(\sum_{j=0}^{N} \tau^{-j} c_j^{(k)}\biggr) \chi\Bigl(\frac{z}{\delta'}\Bigr), \quad d^{(k)}(x_1, \tau, z; s) = \biggl(\sum_{j=0}^{N} \tau^{-j} d_j^{(k)}\biggr) \chi\Bigl(\frac{z}{\delta'}\Bigr), \tag{4-12}$$

where $c_0^{(k)} = d_0^{(k)}$ is independent of x_1 and

$$c_0^{(k)}(\tau, z) = c_{00}^{(k)}(\tau) + \mathcal{O}(|z|), \quad c_{00}^{(k)}(\tau) \neq 0, \quad \text{for all } \tau. \tag{4-13}$$

Here (τ, z) are the Fermi coordinates for γ for t close to t_k.

We also have

$$\begin{aligned}
\|v_j\|_{L^4(M)} &= \mathcal{O}(1), & \|\nabla v_j\|_{L^4(M)} &= \mathcal{O}(s),\\
\|w_j\|_{L^4(M)} &= \mathcal{O}(1), & \|\nabla w_j\|_{L^4(M)} &= \mathcal{O}(s),\\
\|v_j\|_{L^\infty(M)} &= \mathcal{O}(s^{(n-2)/8}), & \|\nabla v_j\|_{L^\infty(M)} &= \mathcal{O}(s^{(n+6)/8}),\\
\|w_j\|_{L^\infty(M)} &= \mathcal{O}(s^{(n-2)/8}), & \|\nabla w_j\|_{L^\infty(M)} &= \mathcal{O}(s^{(n+6)/8}),
\end{aligned} \tag{4-14}$$

as $s \to \infty$, $j = 1, 2$.

Now it follows from (4-1) that

$$u_1 u_2 u_3 = e^{(-Ls+2i\mu-Li\lambda)x_1} c^{-3(n-2)/4} (v_1 \bar{v}_2 w_1 + \tilde{R}), \tag{4-15}$$

where

$$\tilde{R} = r_3 v_1 \bar{v}_2 + (w_1 + r_3)(v_1 \bar{r}_2 + \bar{v}_2 r_1 + r_1 \bar{r}_2).$$

Using (4-2) and (4-14), we get

$$\|\tilde{R}\|_{C^1(M)} = \mathcal{O}(s^{-L}), \tag{4-16}$$

where L is large. Hence, we have

$$\begin{aligned}
\partial_{x_1}(u_1 u_2 u_3) = e^{(-Ls+2i\mu-Li\lambda)x_1} \bigl[&(-Ls + 2i\mu - Li\lambda) c^{-3(n-2)/4} (v_1 \bar{v}_2 w_1 + \tilde{R})\\
&+ \partial_{x_1}(c^{-3(n-2)/4})(v_1 \bar{v}_2 w_1 + \tilde{R}) + c^{-3(n-2)/4} (\partial_{x_1}(v_1 \bar{v}_2 w_1) + \partial_{x_1} \tilde{R})\bigr],
\end{aligned}$$

and therefore, in view of (4-16), (4-2), and (4-14), we get

$$\partial_{x_1}(u_1 u_2 u_3) u_4 = e^{2i(\mu-L\lambda)x_1} c^{-(n-2)} [-Ls v_1 \bar{v}_2 w_1 \bar{w}_2 + \partial_{x_1}(v_1 \bar{v}_2 w_1) \bar{w}_2] + \mathcal{O}_{L^1(M)}(1), \tag{4-17}$$

as $s \to \infty$. We also have from (4-15) that

$$\partial_{x_k}(u_1u_2u_3) = e^{(-Ls+2i\mu-Li\lambda)x_1}[c^{-3(n-2)/4}(\partial_{x_k}(v_1\bar{v}_2w_1) + \partial_{x_k}\tilde{R}) + \partial_{x_k}(c^{-3(n-2)/4})(v_1\bar{v}_2w_1 + \tilde{R})],$$

for $k = 2, \dots, n$, and therefore, in view of (4-16), (4-2), and (4-14), we get

$$\partial_{x_k}(u_1u_2u_3)u_4 = e^{2i(\mu-L\lambda)x_1}c^{-(n-2)}\partial_{x_k}(v_1\bar{v}_2w_1)\overline{w}_2 + \mathcal{O}_{L^1(M)}(1), \tag{4-18}$$

as $s \to \infty$.

For future reference, we also note that

$$u_1u_2u_3u_4 = e^{2i(\mu-L\lambda)x_1}c^{-(n-2)}(v_1\bar{v}_2w_1\overline{w}_2 + \tilde{R}w_2 + (v_1\bar{v}_2w_1 + \tilde{R})\bar{r}_2) = \mathcal{O}_{L^1(M)}(1), \tag{4-19}$$

as $s \to \infty$.

Using (4-17) and (4-18), we obtain

$$\begin{aligned}\langle A, d(u_1u_2u_3)\rangle_g u_4 = e^{2i(\mu-L\lambda)x_1}c^{1-n}\big(&A_1(-Lsv_1\bar{v}_2w_1\overline{w}_2 + \partial_{x_1}(v_1\bar{v}_2w_1)\overline{w}_2) \\ &+ \langle A', d_{x'}(v_1\bar{v}_2w_1)\rangle_{g_0}\overline{w}_2\big) + \mathcal{O}_{L^1(M)}(1),\end{aligned} \tag{4-20}$$

as $s \to \infty$.

It follows from (1-8) in view of (4-20) that

$$\int_M (A_1(-Lsv_1\bar{v}_2w_1\overline{w}_2+\partial_{x_1}(v_1\bar{v}_2w_1)\overline{w}_2)+\langle A', d_{x'}(v_1\bar{v}_2w_1)\rangle_{g_0}\overline{w}_2)e^{2i(\mu-L\lambda)x_1}c^{1-n}\,dV_g = \mathcal{O}(1), \tag{4-21}$$

as $s \to 0$.

Now taking $u_2 = u_3 = 1$ in (1-8) and applying Proposition C.3, we obtain that $A|_{\partial M} = 0$ and $\partial_\nu A|_{\partial M} = 0$. Let us extend A by zero to $(\mathbb{R} \times M_0) \setminus M$ and denote this extension by A again. Since $A \in C^{1,1}(M, T^*M)$ and $A|_{\partial M} = 0$, $\partial_\nu A|_{\partial M} = 0$, we see that $A \in C^{1,1}(\mathbb{R} \times M_0, T^*(\mathbb{R} \times M_0))$. Now (4-21) implies that

$$\begin{aligned}\int_{\mathbb{R}\times M_0} &(A_1(-Lsv_1\bar{v}_2w_1\overline{w}_2 + \partial_{x_1}(v_1\bar{v}_2w_1)\overline{w}_2) + \langle A', d_{x'}(v_1\bar{v}_2w_1)\rangle_{g_0}\overline{w}_2) \\ &\times e^{2i(\mu-L\lambda)x_1}c^{1-n}\,dV_g = \mathcal{O}(1),\end{aligned} \tag{4-22}$$

as $s \to 0$. In view of (4-3), (4-22) gives

$$\begin{aligned}\sum_{r=1}^{N}\int_{\mathbb{R}\times U_r} &(A_1(-Lsv_1\bar{v}_2w_1\overline{w}_2 + \partial_{x_1}(v_1\bar{v}_2w_1)\overline{w}_2) + \langle A', d_{x'}(v_1\bar{v}_2w_1)\rangle_{g_0}\overline{w}_2) \\ &\times e^{2i(\mu-L\lambda)x_1}c^{1-n}\,dV_g = \mathcal{O}(1),\end{aligned} \tag{4-23}$$

as $s \to 0$, where the U_r are sufficiently small neighborhoods of the points p_r of the intersections of γ and η. Using (4-4), (4-5), (4-7), (4-10), and (4-12), we obtain that in U_r,

$$v_1\bar{v}_2w_1\overline{w}_2 = s^{(n-2)/2}\sum_{1\le k,l\le P_r}\;\sum_{1\le m,j\le Q_r} e^{is\Psi^r_{klmj}}e^{\Phi^r_{klmj}}a_0^{(k),r}\,\overline{a_0^{(l),r}}\,c_0^{(m),r}\,\overline{c_0^{(j),r}} + \mathcal{O}_{L^1(I\times M_0)}(1/s), \tag{4-24}$$

where

$$\Psi^r_{klmj} = \varphi^{(k),r} - \overline{\varphi^{(l),r}} + L\psi^{(m),r} - L\overline{\psi^{(j),r}}, \tag{4-25}$$

$$\Phi^r_{klmj} = -\mu\varphi^{(k),r} - \mu\overline{\varphi^{(l),r}} - L\lambda\psi^{(m),r} - L\lambda\overline{\psi^{(j),r}}, \tag{4-26}$$

and $I \subset \mathbb{R}$ is a bounded interval. Recall that all $a_0^{(k),r}$ and $c^{(m),r}$ are independent of x_1. This fact also implies that

$$\partial_{x_1}(v_1 \bar{v}_2 w_1)\overline{w}_2 = \mathcal{O}_{L^1(I\times M_0)}(1/s). \tag{4-27}$$

Using (4-4), (4-5), (4-7), (4-10), and (4-12), we also get that in U_r,

$$\begin{aligned} d_{x'}(v_1 \bar{v}_2 w_1)\overline{w}_2 = s^{(n-2)/2} \sum_{1\le k,l\le P_r} \sum_{1\le m,j\le Q_r} & is(d\varphi^{(k),r} - d\overline{\varphi^{(l),r}} + L\, d\psi^{(m),r}) \\ & \times e^{is\Psi^r_{klmj}} e^{\Phi^r_{klmj}} a_0^{(k),r} \overline{a_0^{(l),r}} c_0^{(m),r} \overline{c_0^{(j),r}} + \mathcal{O}_{L^1(I\times M_0)}(1). \end{aligned} \tag{4-28}$$

Substituting (4-24), (4-27), and (4-28) into (4-23), using that $dV_g = c^{n/2}\, dx_1\, dV_{g_0}$, and dividing (4-23) by $s^{1/2}$, we obtain

$$s^{(n-1)/2} \sum_{r=1}^{N} \sum_{1\le k,l\le P_r} \sum_{1\le m,j\le Q_r} \int_{U_r} B^r_{klmj} e^{is\Psi^r_{klmj}}\, dV_{g_0} = \mathcal{O}(s^{-1/2}), \tag{4-29}$$

where

$$\begin{aligned} B^r_{klmj} = [-L\widehat{A_1 c^{1-n/2}}(2(\mu - L\lambda), \cdot) + i\langle \widehat{A' c^{1-n/2}}(2(\mu - L\lambda), \cdot)\, d\varphi^{(k),r} - d\overline{\varphi^{(l),r}} + L\, d\psi^{(m),r}\rangle_{g_0}] & \\ \times e^{\Phi^r_{klmj}} a_0^{(k),r} \overline{a_0^{(l),r}} c_0^{(m),r} \overline{c_0^{(j),r}}. & \end{aligned} \tag{4-30}$$

Notice that the occurrence of the factor $s^{(n-1)/2}$ is natural here, in view of a subsequent application of the stationary phase method, in its rough version, to the integral in the left-hand side of (4-29).

Choosing L. The argument below follows [Lassas et al. 2021a] closely and is presented here for completeness and the convenience of the reader only. We claim that $L > 0$ can be chosen sufficiently large but fixed so that $d\Psi^r_{klmj}(p_r) = 0$ for all points p_r, $1 \le r \le N$, if and only if $k = l$ and $m = j$. Indeed, it follows from (4-25) that

$$\begin{aligned} \nabla\Psi^r_{klmij}(p_r) &= (\nabla\varphi^{(k),r} - \nabla\overline{\varphi^{(l),r}} + L\nabla\psi^{(m),r} - L\nabla\overline{\psi^{(j),r}})(p_r) \\ &= \dot{\eta}(t^r_k) - \dot{\eta}(t^r_l) + L\dot{\gamma}(\tau^r_m) - L\dot{\gamma}(\tau^r_j). \end{aligned} \tag{4-31}$$

If $k = l$ and $m = j$, (4-31) implies that $\nabla\Psi^r_{klmij}(p_r) = 0$ for all $1 \le r \le N$. Now since the geodesic γ is nontangential, and therefore not closed, we have $\dot{\gamma}(\tau^r_m) - \dot{\gamma}(\tau^r_j) \ne 0$, for all $m \ne j$, for all r, $1 \le r \le N$. Let

$$\alpha = \min\{|\dot{\gamma}(\tau^r_m) - \dot{\gamma}(\tau^r_j)| : m \ne j,\ 1 \le m, j \le Q_r,\ 1 \le r \le N\} > 0.$$

Then in view of the fact that η is a unit speed geodesic, it follows from (4-31) that for all r, $1 \le r \le N$, and for all $m \ne j$,

$$|\nabla\Psi^r_{klmj}(p_r)| \ge L\alpha - 2 \ge 1, \tag{4-32}$$

provided that $L \ge 3/\alpha$. Hence, if $d\Psi^r_{klmj}(p_r) = 0$ then $m = j$, and therefore, (4-31) implies that

$$\nabla\Psi^r_{klmk}(p_r) = \dot{\eta}(t^r_k) - \dot{\eta}(t^r_l). \tag{4-33}$$

This completes the proof of the claim since $\dot{\eta}(t^r_k) - \dot{\eta}(t^r_l) \ne 0$ for all $k \ne l$ and all r, $1 \le r \le N$.

In what follows we choose $L \geq 3/\alpha$. Furthermore, it follows from (4-32) and (4-33) that for such L, there exists $\beta > 0$ such that

$$|\nabla \Psi^r_{klmj}(p_r)| \geq \beta > 0, \tag{4-34}$$

for $(k, l, m, j) \in \{(k, l, m, j) : 1 \leq k, l \leq P_r,\ 1 \leq m, j \leq Q_r\} \setminus \{(k, l, m, j) : k = l,\ m = j\},\ 1 \leq r \leq N$.

Returning to (4-29), we write the integral there as

$$I = s^{(n-1)/2} \sum_{r=1}^{N} \sum_{1\leq k,l\leq P_r} \sum_{1\leq m,j\leq Q_r} \int_{U_r} B^r_{klmj} e^{is\Psi^r_{klmj}}\, dV_{g_0} = \sum_{r=1}^{N} (I^r_1 + I^r_2), \tag{4-35}$$

where

$$\begin{aligned} I^r_1 &= s^{(n-1)/2} \sum_{1\leq k\leq P_r} \sum_{1\leq m\leq Q_r} \int_{U_r} B^r_{kkmm} e^{is\Psi^r_{kkmm}}\, dV_{g_0}, \\ I^r_2 &= s^{(n-1)/2} \sum_{1\leq k\neq l\leq P_r} \sum_{1\leq m\neq j\leq Q_r} \int_{U_r} B^r_{klmj} e^{is\Psi^r_{klmj}}\, dV_{g_0}. \end{aligned} \tag{4-36}$$

Rough stationary phase calculation. Here the analysis is concerned with the integrals I^r_1. It follows from (4-25) that

$$\Psi^r_{kkmm} = 2i(\operatorname{Im} \varphi^{(k),r} + L \operatorname{Im} \psi^{(m),r}),$$

and therefore, $d\Psi^r_{kkmm}(p_r) = 0$, $\Psi^r_{kkmm}(p_r) = 0$, and $\operatorname{Im} \nabla^2 \Psi^r_{kkmm}(p_r) > 0$, where $p_r \in M_0^{\text{int}}$ is the point of intersection of γ and η. Note that $U_r \subset M_0^{\text{int}}$, and hence, there will be no contributions from the boundary.

Let us denote by $z = (z_1, \dots, z_{n-1})$ the geodesic normal coordinates in (M_0, g_0) with origin at p_r. Writing $dV_{g_0} = |g_0(z)|^{1/2}\, dz$, applying Lemma A.1, and using (4-30) and (4-26), we obtain that

$$\begin{aligned} &\lim_{s\to\infty} s^{(n-1)/2} \int_{U_r} B^r_{kkmm} e^{is\Psi^r_{kkmm}}\, dV_{g_0} \\ &\quad = \lim_{s\to\infty} s^{(n-1)/2} \int_{\text{neigh}(0,\mathbb{R}^{n-1})} B^r_{kkmm}(z) |g_0(z)|^{1/2} e^{is\Psi^r_{kkmm}(z)}\, dz = C^r_{kkmm} B^r_{kkmm}(p_r) \\ &\quad = C^r_{kkmm}[-L\widehat{A_1 c^{1-n/2}}(2(\mu - L\lambda), p_r) + iL\widehat{A' c^{1-n/2}}(2(\mu - L\lambda), p_r)(\dot{\gamma}(\tau^r_m))] \\ &\qquad\qquad \times e^{-2\mu t^r_k - 2L\lambda\tau^r_m} |a^{(k),r}_{00}(p_r)|^2 |c^{(m),r}_{00}(p_r)|^2, \end{aligned} \tag{4-37}$$

where

$$C^r_{kkmm} = \frac{(2\pi)^{(n-1)/2}}{(\det \operatorname{Im} \nabla^2 \Psi^r_{kkmm}(p_r))^{1/2}} > 0.$$

Here we also used that

$$\varphi^{(k),r}(p_r) = t^r_k \quad \text{and} \quad \psi^{(m),r}(p_r) = \tau^r_m.$$

Thus, we see from (4-36) and (4-37) that

$$\begin{aligned} \lim_{s\to\infty} I^r_1 = \sum_{1\leq k\leq P_r} \sum_{1\leq m\leq Q_r} C^r_{kkmm}[-L\widehat{A_1 c^{1-n/2}}(2(\mu - L\lambda), p_r) + iL\widehat{A' c^{1-n/2}}(2(\mu - L\lambda), p_r)(\dot{\gamma}(\tau^r_m))] \\ \times e^{-2\mu t^r_k - 2L\lambda\tau^r_m} |a^{(k),r}_{00}(p_r)|^2 |c^{(m),r}_{00}(p_r)|^2. \end{aligned} \tag{4-38}$$

Nonstationary phase calculation. Here the analysis is concerned with I_2^r in (4-36). It follows from (4-25) that

$$\Psi^r_{klmj} = \tilde{\Psi}^r_{klmj} + i\operatorname{Im}\varphi^{(k),r} + i\operatorname{Im}\varphi^{(l),r} + Li\operatorname{Im}\psi^{(m),r} + Li\operatorname{Im}\psi^{(j),r}, \tag{4-39}$$

where

$$\tilde{\Psi}^r_{klmj} = \operatorname{Re}\varphi^{(k),r} - \operatorname{Re}\varphi^{(l),r} + L\operatorname{Re}\psi^{(m),r} - L\operatorname{Re}\psi^{(j),r} \in C^\infty \tag{4-40}$$

is real such that $|\nabla\tilde{\Psi}^r_{klmj}(p_r)| = |\nabla\Psi^r_{klmj}(p_r)| \geq \beta > 0$ provided $L > 3/\alpha$ in view of (4-34).

Let us denote by $z = (z_1, \dots, z_{n-1})$ the geodesic normal coordinates in (M_0, g_0) with origin at p. Motivated by (4-30) and (4-39), we set

$$f(z) = [-L\widehat{A_1 c^{1-n/2}}(2(\mu - L\lambda), z) + i\langle\widehat{A' c^{1-n/2}}(2(\mu - L\lambda), z)d\varphi^{(k),r} - d\overline{\varphi^{(l),r}} + Ld\psi^{(m),r}\rangle_{g_0}] \\ \times e^{\Phi^r_{klmj}}|g_0(z)|^{1/2} \in C^{1,1}_0(M_0),$$

and

$$\hat{a}_0^{(k),r} = s^{(n-2)/8}e^{-s\operatorname{Im}\varphi^{(k),r}}a_0^{(k),r}, \quad \hat{c}_0^{(m),r} = s^{(n-2)/8}e^{-s\operatorname{Im}\psi^{(m),r}}c_0^{(m),r}. \tag{4-41}$$

Thus,

$$I^r_{2,klmj} := s^{(n-1)/2}\int_{U_r} B^r_{klmj}e^{is\Psi^r_{klmj}}\,dV_g = s^{1/2}\int_{\operatorname{neigh}(0,\mathbb{R}^{n-1})} f(z)\hat{a}_0^{(k),r}\,\overline{\hat{a}_0^{(l),r}}\,\hat{c}_0^{(m),r}\,\overline{\hat{c}_0^{(j),r}}\,e^{is\tilde{\Psi}^r_{klmj}(z)}\,dz. \tag{4-42}$$

Note that f is independent of s, and

$$\|\hat{a}_0^{(k),r}\|_{L^4(M_0)} = \mathcal{O}(1), \quad \|\hat{c}_0^{(m),r}\|_{L^4(M_0)} = \mathcal{O}(1), \tag{4-43}$$

as $s \to \infty$. We next claim that

$$\|\nabla\hat{a}_0^{(k),r}\|_{L^4(M_0)} = \mathcal{O}(s^{1/2}), \quad \|\nabla\hat{c}_0^{(m),r}\|_{L^4(M_0)} = \mathcal{O}(s^{1/2}), \tag{4-44}$$

as $s \to \infty$; see [Lassas et al. 2021a]. Let us recall the argument briefly. It is enough to show the first bound in (4-44). To that end, we have from (4-41) that

$$\nabla\hat{a}_0^{(k),r} = s^{(n-2)/8}e^{-s\operatorname{Im}\varphi^{(k),r}}(-s(\nabla\operatorname{Im}\varphi^{(k),r})a_0^{(k),r} + \nabla a_0^{(k),r}). \tag{4-45}$$

It suffices to control the first term in the right-hand side of (4-45), and to this end we note that in the Fermi coordinates (t, y), associated with the geodesic η, we have

$$|\nabla\operatorname{Im}\varphi^{(k),r}(t, y)| = \mathcal{O}(|y|) \tag{4-46}$$

and

$$\operatorname{Im}\varphi^{(k),r}(t, y) \geq c|y|^2, \tag{4-47}$$

for some $c > 0$; see (4-6). Thus, using (4-46) and (4-47), we get

$$\|s^{(n-2)/8}e^{-s\operatorname{Im}\varphi^{(k),r}}s(\nabla\operatorname{Im}\varphi^{(k),r})a_0^{(k),r}\|_{L^4(M_0)} = \mathcal{O}(s^{(n-2)/8}s)\left(\int_{|y|\leq 1/2} e^{-4s\operatorname{Im}\varphi^{(k),r}}|y|^4\,dy\right)^{1/4} = \mathcal{O}(s^{1/2}).$$

This bound together with (4-45) shows the first bound in (4-44). Similarly to (4-44), we also have

$$\|\partial^\alpha\hat{a}_0^{(k),r}\|_{L^4(M_0)} = \mathcal{O}(s^{|\alpha|/2}), \quad \|\partial^\alpha\hat{c}_0^{(m),r}\|_{L^4(M_0)} = \mathcal{O}(s^{|\alpha|/2}), \quad \text{for all } \alpha, \tag{4-48}$$

as $s \to \infty$. Furthermore, as $\delta' > 0$ can be chosen as small as we wish, we see that $\hat{a}_0^{(k),r}$ and $\hat{c}_0^{(k),r}$ have compact support in U_r.

Letting

$$L = \frac{\nabla \tilde{\Psi}^r_{klmij} \cdot \nabla}{i|\nabla \tilde{\Psi}^r_{klmij}|^2},$$

we have $L(e^{is\tilde{\Psi}^r_{klmij}}) = se^{is\tilde{\Psi}^r_{klmij}}$. Integrating by parts in (4-42), we get

$$I^r_{2,klmj} = s^{-1/2} \int_{\text{neigh}(0,\mathbb{R}^{n-1})} e^{is\tilde{\Psi}^r_{klmj}(z)} L^t \big(f(z)\hat{a}_0^{(k),r}\, \overline{\hat{a}_0^{(l),r}}\, \hat{c}_0^{(m),r}\, \overline{\hat{c}_0^{(j),r}}\big)\, dz,$$

where $L^t = -L - \operatorname{div} L$. Now in view of (4-40) and (4-43), we see that

$$s^{-1/2} \left| \int_{\text{neigh}(0,\mathbb{R}^{n-1})} e^{is\tilde{\Psi}^r_{klmj}(z)} (\operatorname{div} L)\big(f(z)\hat{a}_0^{(k),r}\, \overline{\hat{a}_0^{(l),r}}\, \hat{c}_0^{(m),r}\, \overline{\hat{c}_0^{(j),r}}\big)\, dz \right| = \mathcal{O}(s^{-1/2}),$$

and in view of (4-44),

$$s^{-1/2} \left| \int_{\text{neigh}(0,\mathbb{R}^{n-1})} e^{is\tilde{\Psi}^r_{klmj}(z)} f(z) \nabla \big(\hat{a}_0^{(k),r}\, \overline{\hat{a}_0^{(l),r}}\, \hat{c}_0^{(m),r}\, \overline{\hat{c}_0^{(j),r}}\big)\, dz \right| = \mathcal{O}(1),$$

as $s \to \infty$. As f is independent of s, we see, after one integration by parts in (4-42), that $I^r_{2,klmj} = \mathcal{O}(1)$. Since ∇f is Lipschitz, we can integrate by parts a second time, and using (4-48), we get

$$I^r_{2,klmj} = \mathcal{O}(s^{-1/2}), \tag{4-49}$$

as $s \to \infty$. Notice that it is precisely here that we need the assumption that our 1-form A is an element of $C_0^{1,1}(\mathbb{R} \times M_0, T^*(\mathbb{R} \times M_0))$.

We get, in view of (4-36) and (4-49),

$$I^r_2 = \mathcal{O}(s^{-1/2}), \tag{4-50}$$

as $s \to \infty$.

Completion of the proof. Passing to the limit $s \to \infty$ in (4-29) and using (4-35), (4-36), (4-38), and (4-50), we obtain

$$\sum_{r=1}^{N} \sum_{k=1}^{P_r} \sum_{m=1}^{Q_r} C^r_{kkmm} \big[-L\widehat{A_1 c^{1-n/2}}(2(\mu - L\lambda), p_r) + iL\widehat{A' c^{1-n/2}}(2(\mu - L\lambda), p_r)(\dot{\gamma}(\tau^r_m))\big] \\ \times e^{-2\mu t^r_k - 2L\lambda\tau^r_m} |a_{00}^{(k),r}(p_r)|^2 |c_{00}^{(m),r}(p_r)|^2 = 0. \tag{4-51}$$

Next we would like to determine each term in the sum in (4-51) separately. To do this, we shall follow [Lassas et al. 2021a]. First choosing $\mu = (1 - L)\lambda$, we get

$$\sum_{r=1}^{N} \sum_{k=1}^{P_r} \sum_{m=1}^{Q_r} \big[-L\widehat{A_1 c^{1-n/2}}(2\lambda(1 - 2L), p_r) + iL\widehat{A' c^{1-n/2}}(2\lambda(1 - 2L), p_r)(\dot{\gamma}(\tau^r_m))\big] \\ \times C^r_{kkmm} e^{-2\lambda[L(\tau^r_m - t^r_k) + t^r_k]} |a_{00}^{(k),r}(p_r)|^2 |c_{00}^{(m),r}(p_r)|^2 = 0. \tag{4-52}$$

It is shown in [Lassas et al. 2021a] that for all $L \geq 1$ sufficiently large,

$$L(\tau_{m_1}^{r_1} - t_{k_1}^{r_1}) + t_{k_1}^{r_1} \neq L(\tau_{m_2}^{r_2} - t_{k_2}^{r_2}) + t_{k_2}^{r_2} \tag{4-53}$$

when $(r_1, k_1, m_1) \neq (r_2, k_2, m_2)$, and fixing $L \geq 3/\alpha$ large enough, we may assume in what follows that (4-53) holds. We shall next need Lemma 5.2 from [Lassas et al. 2021a] which can be stated as follows: let $f_1, \dots, f_N \in \mathcal{E}'(\mathbb{R})$ be such that for some distinct real numbers $a_1, \dots, a_N$, one has

$$\sum_{j=1}^{N} \hat{f}_j(\lambda) e^{a_j \lambda} = 0, \quad \lambda \in \mathbb{R},$$

then $f_1 = \cdots = f_N = 0$. Applying this result, we get for all r, k, m, λ,

$$(-\widehat{A_1 c^{1-n/2}}(2\lambda(1-2L), p_r) + i\widehat{A' c^{1-n/2}}(2\lambda(1-2L), p_r)(\dot{\gamma}(\tau_m^r)))C_{kkmm}^r |a_{00}^{(k),r}(p_r)|^2 |c_{00}^{(m),r}(p_r)|^2 = 0,$$

and as $C_{kkmm}^r \neq 0$, $a_{00}^{(k),r}(p_r) \neq 0$, and $c_{00}^{(m),r}(p_r) \neq 0$, we get, taking the inverse Fourier transform in x_1,

$$-A_1(x_1, p_r) + iA'(x_1, p_r)(\dot{\gamma}(\tau_m^r)) = 0,$$

for all $x_1 \in \mathbb{R}$, p_r, and τ_m^r. Since y_0 was one of the points p_r, and $\gamma(\tau_m^r) = y_0$, we know

$$(iA_1, A')(x_1, y_0)(1, \dot{\gamma}(\tau_m^r)) = 0. \tag{4-54}$$

Here we may replace $\dot{\gamma}(\tau_m^r)$ by $-\dot{\gamma}(\tau_m^r)$, and thus, (4-54) implies that $A_1(x_1, y_0) = 0$, for all $x_1 \in \mathbb{R}$ and almost all $y_0 \in M_0$, and therefore, by continuity, $A_1 \equiv 0$. Hence, we are left with proving that the 1-form $A'(x_1, \cdot)$ vanishes on M_0 from the fact that

$$A'(x_1, y_0)(\dot{\gamma}(\tau_m^r)) = 0. \tag{4-55}$$

To proceed we assume without loss of generality that $v_1 := \dot{\gamma}(\tau_m^r) = (1, 0, \dots, 0) \in \mathbb{R}^{n-1}$ and consider its small perturbations $v_2, \dots, v_{n-1}$ given by (3-21). The unit vectors $v_1, \dots, v_{n-1}$ are linearly independent, and therefore, they span the tangent space $T_{y_0}M_0$. By Proposition D.1, for $\varepsilon > 0$ sufficiently small, the unit speed geodesic γ_{y_0,v_j}, $j = 2, \dots, n-1$, through (y_0, v_j) is nontangential between boundary points, and γ and γ_{y_0,v_j} do not intersect each other at the boundary of M_0. Applying the discussion above with $\eta = \gamma$ and $\gamma = \gamma_{y_0,v_j}$, we get

$$A'(x_1, y_0)(v_j) = 0, \quad j = 2, \dots, n-1. \tag{4-56}$$

It follows from (4-55) and (4-56) that the 1-form $A'(x_1, y_0)$ equals 0, and therefore, $A' \equiv 0$. This completes the proof of Proposition 1.6 in the general setting.

In the course of the proof of Proposition 1.6, we also proved the following result.

Corollary 4.1. *Let (M, g) be a conformally transversally anisotropic manifold of dimension $n \geq 3$. Let $A \in C^{1,1}(M, T^*M)$ be a 1-form such that $A|_{\partial M} = 0$ and $\partial_\nu A|_{\partial M} = 0$. If*

$$\int_M \langle A, d(u_1 u_2 u_3) \rangle_g u_4 \, dV_g = \mathcal{O}(1),$$

as $s \to \infty$, for all harmonic functions $u_l \in C^3(M)$, $l = 1, \dots, 4$, of the form (4-1), *then $A \equiv 0$.*

5. Proof of Theorem 1.3

Let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_m) \in \mathbb{C}^m$, $m \geq 3$, and consider the Dirichlet problem (1-5) with

$$f = \sum_{k=1}^{m} \varepsilon_k f_k, \quad f_k \in C^{2,\alpha}(\partial M), \quad k = 1, \dots, m.$$

Then for all $|\varepsilon|$ sufficiently small, the problem (1-5) has a unique small solution $u(\,\cdot\,, \varepsilon) \in C^{2,\alpha}(M)$, which depends holomorphically on $\varepsilon \in \operatorname{neigh}(0, \mathbb{C}^m)$; see Theorem B.1.

We shall use an induction argument on $m \geq 3$ to show that all the coefficients A_m and V_m in (1-2) and (1-3), see also (1-5), can be determined from the Dirichlet-to-Neumann map $\Lambda_{A,V}$ given in (1-6).

First, let $m = 3$, and let us proceed to carry out a third-order linearization of the Dirichlet-to-Neumann map. Let $u_j = u_j(x, \varepsilon)$ be the unique small solution of the Dirichlet problem

$$\begin{cases} -\Delta_g u_j + i d^*\big(\sum_{k=2}^{\infty} A_k^{(j)}(x)(u_j^k/k!)u_j\big) - i\big\langle \sum_{k=2}^{\infty} A_k^{(j)}(x)(u_j^k/k!), du_j\big\rangle_g & \\ \quad + \big\langle \sum_{k=2}^{\infty} A_k^{(j)}(x)(u_j^k/k!), \sum_{k=2}^{\infty} A_k^{(j)}(x)(u_j^k/k!)\big\rangle_g u_j + \sum_{k=3}^{\infty} V_k^{(j)}(x)(u_j^k/k!) = 0 & \text{in } M, \\ u_j = \varepsilon_1 f_1 + \varepsilon_2 f_2 + \varepsilon_3 f_3 & \text{on } \partial M, \end{cases} \tag{5-1}$$

for $j = 1, 2$. Differentiating (5-1) with respect to ε_l, $l = 1, 2, 3$, and using that $u_j(x, 0) = 0$, we get

$$\begin{cases} -\Delta_g v_j^{(l)} = 0 & \text{in } M, \\ v_j^{(l)} = f_l & \text{on } \partial M, \end{cases} \tag{5-2}$$

where $v_j^{(l)} = \partial_{\varepsilon_l} u_j|_{\varepsilon=0}$. By the uniqueness and the elliptic regularity for the Dirichlet problem (5-2), we have that $v^{(l)} := v_1^{(l)} = v_2^{(l)} \in C^{2,\alpha}(M)$, $l = 1, 2, 3$; see [Gilbarg and Trudinger 1983, Theorem 6.15].

Applying $\partial_{\varepsilon_k}\partial_{\varepsilon_l}|_{\varepsilon=0}$, $k, l = 1, 2, 3$, to (5-1), we next get

$$\begin{cases} -\Delta_g w_j^{(k,l)} = 0 & \text{in } M, \\ w_j^{(k,l)} = 0 & \text{on } \partial M, \end{cases} \tag{5-3}$$

where $w_j^{(k,l)} = \partial_{\varepsilon_k}\partial_{\varepsilon_l} u_j|_{\varepsilon=0}$, and therefore, $w_j^{(k,l)} = 0$ for all $j = 1, 2$ and $k, l = 1, 2, 3$. Finally, applying $\partial_{\varepsilon_1}\partial_{\varepsilon_2}\partial_{\varepsilon_3}|_{\varepsilon=0}$ to (5-1), we obtain the third-order linearization

$$\begin{cases} -\Delta_g w_j + 3i d^*(A_2^{(j)} v^{(1)} v^2 v^{(3)}) - i\langle A_2^{(j)}, d(v^{(1)} v^{(2)} v^{(3)})\rangle_g + V_3^{(j)} v^{(1)} v^{(2)} v^{(3)} = 0 & \text{in } M, \\ w_j = 0 & \text{on } \partial M, \end{cases} \tag{5-4}$$

where $w_j = \partial_{\varepsilon_1}\partial_{\varepsilon_2}\partial_{\varepsilon_3} u_j|_{\varepsilon=0}$. Using that

$$d^*(Av) = (d^*A)v - \langle A, dv\rangle_g, \tag{5-5}$$

for any 1-form A and a function v, we can rewrite (5-4) as

$$\begin{cases} -\Delta_g w_j - 4i\langle A_2^{(j)}, d(v^{(1)} v^{(2)} v^{(3)})\rangle_g + (3i d^*(A_2^{(j)}) + V_3^{(j)}) v^{(1)} v^{(2)} v^{(3)} = 0 & \text{in } M, \\ w_j = 0 & \text{on } \partial M. \end{cases} \tag{5-6}$$

The fact that

$$\Lambda_{A^{(1)},V^{(1)}}(\varepsilon_1 f_1+\varepsilon_2 f_2+\varepsilon_3 f_3)=\Lambda_{A^{(2)},V^{(2)}}(\varepsilon_1 f_1+\varepsilon_2 f_2+\varepsilon_3 f_3)$$

for all small ε and all $f_j\in C^{2,\alpha}(\partial M)$ implies that $\partial_\nu u_1|_{\partial M}=\partial_\nu u_2|_{\partial M}$. Therefore, an application of $\partial_{\varepsilon_1}\partial_{\varepsilon_2}\partial_{\varepsilon_3}|_{\varepsilon=0}$ yields $\partial_\nu w_1|_{\partial M}=\partial_\nu w_2|_{\partial M}$. Multiplying (5-6) by $v^{(4)}\in C^{2,\alpha}(M)$ harmonic in (M,g) and applying Green's formula, we get

$$\int_M \bigl(4i\langle A, d(v^{(1)}v^{(2)}v^{(3)})\rangle_g v^{(4)}-(3id^*(A)+V)v^{(1)}v^{(2)}v^{(3)}v^{(4)}\bigr)\,dV_g=0, \tag{5-7}$$

for all $v^{(l)}\in C^{2,\alpha}(M)$ harmonic in (M,g), $l=1,\dots,4$. Here $A=A_2^{(1)}-A_2^{(2)}$ and $V=V_3^{(1)}-V_3^{(2)}$. An application of Proposition C.4 implies that $A|_{\partial M}=0$ and $\partial_\nu A|_{\partial M}=0$.

Choosing $v^{(l)}=u_l\in C^3(M)$, $l=1,\dots,4$, to be harmonic functions of the form (4-1), and using (4-19), we first observe that (5-7) implies that

$$\int_M \langle A, d(u_1u_2u_3)\rangle_g u_4\,dV_g=\mathcal{O}(1),$$

as $s\to\infty$. By Corollary 4.1, we get $A\equiv 0$, and therefore, $A_2^{(1)}=A_2^{(2)}$. Substituting $A=0$ into (5-7), we get

$$\int_M V v^{(1)}v^{(2)}v^{(3)}v^{(4)}\,dV_g=0,$$

for all harmonic functions $v^{(l)}\in C^{2,\alpha}(M)$, $l=1,\dots,4$. Using Proposition 1.2, we obtain that $V=0$, and thus, $V_3^{(1)}=V_3^{(2)}$.

Let $m\geq 4$ and assume that

$$A_k=A_k^{(1)}=A_k^{(2)},\quad \text{for } k=2,\dots,m-2,\qquad V_k=V_k^{(1)}=V_k^{(2)},\quad \text{for } k=3,\dots,m-1. \tag{5-8}$$

To show that $A_{m-1}^{(1)}=A_{m-1}^{(2)}$ and $V_m^{(1)}=V_m^{(2)}$, we shall perform the m-th order linearization of the Dirichlet-to-Neumann map. To that end, let $u_j=u_j(x,\varepsilon)$ be the unique small solution of the Dirichlet problem

$$\begin{cases} -\Delta_g u_j+id^*\bigl(\sum_{k=2}^\infty A_k^{(j)}(x)(u_j^k/k!)u_j\bigr)-i\bigl\langle\sum_{k=2}^\infty A_k^{(j)}(x)(u_j^k/k!), du_j\bigr\rangle_g & \\ \quad +\bigl\langle\sum_{k=2}^\infty A_k^{(j)}(x)(u_j^k/k!), \sum_{k=2}^\infty A_k^{(j)}(x)(u_j^k/k!)\bigr\rangle_g u_j+\sum_{k=3}^\infty V_k^{(j)}(x)(u_j^k/k!)=0 & \text{in } M,\\ u_j=\varepsilon_1 f_1+\dots+\varepsilon_m f_m & \text{on } \partial M,\end{cases} \tag{5-9}$$

for $j=1,2$. We would like to apply $\partial_{\varepsilon_1}\cdots\partial_{\varepsilon_m}|_{\varepsilon=0}$ to (5-9). First we observe that

$$\partial_{\varepsilon_1}\cdots\partial_{\varepsilon_m}\biggl(id^*\biggl(\sum_{k=m}^\infty A_k^{(j)}(x)\frac{u_j^k}{k!}u_j\biggr)-i\biggl\langle\sum_{k=m}^\infty A_k^{(j)}(x)\frac{u_j^k}{k!}, du_j\biggr\rangle_g+\sum_{k=m+1}^\infty V_k^{(j)}(x)\frac{u_j^k}{k!}\biggr)$$

is a sum of terms, each of them containing positive powers of u_j which vanish when $\varepsilon=0$. The only term in the expression for $\partial_{\varepsilon_1}\cdots\partial_{\varepsilon_m}(V_m^{(j)}(x)u_j^m/m!)$ which does not contain a positive power of u_j

is $V_m^{(j)}(x)\partial_{\varepsilon_1}u_j\cdots\partial_{\varepsilon_m}u_j$. Furthermore, the only term in the expression for

$$\partial_{\varepsilon_1}\cdots\partial_{\varepsilon_m}\biggl(id^*\biggl(A_{m-1}^{(j)}\frac{u_j^m}{(m-1)!}\biggr)\biggr)$$

which does not contain a positive power of u_j is $mid^*(A_{m-1}^{(j)}\partial_{\varepsilon_1}u_j\cdots\partial_{\varepsilon_m}u_j)$. The only terms in

$$\partial_{\varepsilon_1}\cdots\partial_{\varepsilon_m}\biggl\langle A_{m-1}^{(j)}\frac{u_j^{m-1}}{(m-1)!},du_j\biggr\rangle_g$$

which do not contain a positive power of u_j can be written as $\langle A_{m-1}^{(j)},d(\partial_{\varepsilon_1}u_j\cdots\partial_{\varepsilon_m}u_j)\rangle_g$. The expression

$$\partial_{\varepsilon_1}\cdots\partial_{\varepsilon_m}\biggl(id^*\biggl(\sum_{k=2}^{m-2}A_k^{(j)}(x)\frac{u_j^k}{k!}u_j\biggr)-i\biggl\langle\sum_{k=2}^{m-2}A_k^{(j)}(x)\frac{u_j^k}{k!},du_j\biggr\rangle_g+\sum_{k=3}^{m-1}V_k^{(j)}(x)\frac{u_j^k}{k!}\biggr)$$

is independent of $j=1,2$, in view of (5-8) and the fact that it contains only derivatives of u_j of the form $\partial^k_{\varepsilon_{l_1},\dots,\varepsilon_{l_k}}u_j|_{\varepsilon=0}$ with $k=1,\dots,m-2$ and $\varepsilon_{l_1},\dots,\varepsilon_{l_k}\in\{\varepsilon_1,\dots,\varepsilon_m\}$. Here we use the fact that

$$\partial^k_{\varepsilon_{l_1},\dots,\varepsilon_{l_k}}u_1|_{\varepsilon=0}=\partial^k_{\varepsilon_{l_1},\dots,\varepsilon_{l_k}}u_2|_{\varepsilon=0}$$

for $k=1,\dots,m-1$ and $\varepsilon_{l_1},\dots,\varepsilon_{l_k}\in\{\varepsilon_1,\dots,\varepsilon_m\}$. This follows by applying the operators $\partial^k_{\varepsilon_{l_1},\dots,\varepsilon_{l_k}}|_{\varepsilon=0}$ to (5-9), using (5-8) and the unique solvability of the Dirichlet problem for the Laplacian.

The terms in the expression for

$$\partial_{\varepsilon_1}\cdots\partial_{\varepsilon_m}\biggl(\biggl\langle\sum_{k=2}^{\infty}A_k^{(j)}(x)\frac{u_j^k}{k!},\sum_{k=2}^{\infty}A_k^{(j)}(x)\frac{u_j^k}{k!}\biggr\rangle_g u_j\biggr)$$

which do not contain a positive power of u_j, only contain $A_2^{(j)},\dots,A_{m-3}^{(j)}$, and only derivatives of u_j of the form $\partial^k_{\varepsilon_{l_1},\dots,\varepsilon_{l_k}}u_j|_{\varepsilon=0}$ with $k=1,\dots,m-4$ and $\varepsilon_{l_1},\dots,\varepsilon_{l_k}\in\{\varepsilon_1,\dots,\varepsilon_m\}$, which are independent of $j=1,2$.

Hence, the m-th order linearization has the form

$$\begin{cases}-\Delta_g w_j+mid^*(A_{m-1}^{(j)}v^{(1)}\cdots v^{(m)})-i\langle A_{m-1}^{(j)},d(v^{(1)}\cdots v^{(m)})\rangle_g+V_m^{(j)}v^{(1)}\cdots v^{(m)}=H_m & \text{in } M,\\ w_j=0 & \text{on } \partial M,\end{cases}$$

where $w_j=\partial_{\varepsilon_1}\cdots\partial_{\varepsilon_m}u_j|_{\varepsilon=0}$ and H_m is known and independent of $j=1,2$. Using (5-5), the previous system can be written as

$$\begin{cases}-\Delta_g w_j-(m+1)i\langle A_{m-1}^{(j)},d(v^{(1)}\cdots v^{(m)})\rangle_g+(mid^*(A_{m-1}^{(j)})+V_m^{(j)})v^{(1)}\cdots v^{(m)}=H_m & \text{in } M,\\ w_j=0 & \text{on } \partial M.\end{cases}$$

Proceeding as in the case $m=3$, we see that

$$\int_M((m+1)i\langle A,d(v^{(1)}\cdots v^{(m)})\rangle_g v^{(m+1)}-(mid^*(A)+V)v^{(1)}\cdots v^{(m+1)})\,dV_g=0,$$

for any $v^{(l)}\in C^{2,\alpha}(M)$ harmonic, $l=1,\dots,m+1$. Here $A=A_{m-1}^{(1)}-A_{m-1}^{(2)}$ and $V=V_m^{(1)}-V_m^{(2)}$. Setting $v^{(1)}=\cdots=v^{(m-3)}=1$ and arguing as in the case $m=3$, we complete the proof of Theorem 1.3.

Appendix A: A rough stationary phase argument

We need the following rough version of the stationary phase; see [Lassas et al. 2021a].

Lemma A.1. *Let $\Psi \in C^\infty(\mathbb{R}^n; \mathbb{R})$ be such that*

$$\Psi(0) = 0, \quad \Psi'(0) = 0, \quad \text{and} \quad \Psi''(0) > 0. \tag{A-1}$$

Let $V \subset \mathbb{R}^n$ be a sufficiently small neighborhood of zero, and let $a \in C(\overline{V})$. Then

$$\lim_{s\to\infty} s^{n/2} \int_V e^{-s\Psi(z)} a(z)\, dz = \frac{(2\pi)^{n/2}}{(\det \Psi''(0))^{1/2}} a(0). \tag{A-2}$$

Proof. Taylor expanding the phase function Ψ and using (A-1), we get

$$\Psi(z) = \tfrac{1}{2}\Psi''(0)z\cdot z + \mathcal{O}(|z|^3),$$

and therefore,

$$\Psi(z) \geq c|z|^2, \tag{A-3}$$

with some $c > 0$, for all $z \in V$, a sufficiently small neighborhood of zero. Making the change of variables $z \mapsto s^{1/2}z$ in the integral in (A-2) and using the dominated convergence theorem, we obtain that

$$\begin{aligned} \lim_{s\to\infty} s^{n/2} \int_V e^{-s\Psi(z)} a(z)\, dz &= \lim_{s\to\infty} \int_{s^{1/2}V} e^{-s\Psi(z/s^{1/2})} a(z/s^{1/2})\, dz \\ &= \biggl(\int_{\mathbb{R}^n} e^{-\Psi''(0)z\cdot z/2}\, dz\biggr) a(0) = \frac{(2\pi)^{n/2}}{(\det \Psi''(0))^{1/2}} a(0). \end{aligned}$$

Here we use the following consequence of (A-3),

$$|\chi_{s^{1/2}V} e^{-s\Psi(z/s^{1/2})} a(z/s^{1/2})| \leq \mathcal{O}(1) e^{-c|z|^2} \in L^1(\mathbb{R}^n),$$

where $\chi_{s^{1/2}V}$ is the characteristic function of the set $s^{1/2}V$. Thus, (A-2) follows. □

Appendix B: Well-posedness of the Dirichlet problem for a nonlinear magnetic Schrödinger equation

The purpose of this appendix is to show the well-posedness of the Dirichlet problem for a nonlinear magnetic Schrödinger equation with small boundary data. The argument is standard, see [Krupchyk and Uhlmann 2020a; Lassas et al. 2021a], and is given here for completeness and the convenience of the reader.

Let (M, g) be a smooth compact Riemannian manifold of dimension $n \geq 2$ with smooth boundary. Let $C^{k,\alpha}(M)$ stand for the Hölder space on M, where $k \in \mathbb{N} \cup \{0\}$ and $0 < \alpha < 1$; see [Hörmander 1976, Appendix A]. Let us note that $C^{k,\alpha}(M)$ is an algebra under pointwise multiplication, and

$$\|uv\|_{C^{k,\alpha}(M)} \leq C(\|u\|_{C^{k,\alpha}(M)} \|v\|_{L^\infty(M)} + \|u\|_{L^\infty(M)} \|v\|_{C^{k,\alpha}M)}), \quad u, v \in C^{k,\alpha}(M); \tag{B-1}$$

see [Hörmander 1976, Theorem A.7].

Consider the Dirichlet problem for the nonlinear magnetic Schrödinger operator

$$\begin{cases} L_{A,V}u=0 & \text{in } M,\\ u=f & \text{on } \partial M,\end{cases} \tag{B-2}$$

where $L_{A,V}$ is given in (1-4). Here the 1-form A mapping $M\times\mathbb{C}$ to T^*M and the function V mapping $M\times\mathbb{C}$ to $\mathbb{C}$ satisfy the following conditions:

(A) The map $\mathbb{C}\ni z\mapsto A(\cdot\,,z)$ is holomorphic with values in $C^{1,\alpha}(M,T^*M)$, the space of 1-forms with complex-valued $C^{1,\alpha}(M)$ coefficients.

(V_i) The map $\mathbb{C}\ni z\mapsto V(\cdot\,,z)$ is holomorphic with values in $C^{\alpha}(M)$.

(V_{ii}) $V(x,0)=0$, for all $x\in M$.

The condition (V_{ii}) guarantees that $u=0$ is a solution to (B-2) when $f=0$. It follows from (A), (V_i), and (V_{ii}) that A and V can be expanded into the power series

$$A(x,z)=\sum_{k=0}^{\infty}A_k(x)\frac{z^k}{k!},\qquad A_k(x):=\partial_z^kA(x,0)\in C^{1,\alpha}(M,T^*M), \tag{B-3}$$

converging in the $C^{1,\alpha}(M,T^*M)$ topology, and

$$V(x,z)=\sum_{k=1}^{\infty}V_k(x)\frac{z^k}{k!},\qquad V_k(x):=\partial_z^kV(x,0)\in C^{\alpha}(M), \tag{B-4}$$

converging in the $C^{\alpha}(M)$ topology. We also assume that $A_0\in C^{\infty}(M,T^*M)$ and $V_1\in C^{\infty}(M)$. Let us assume furthermore that

(i) 0 is not a Dirichlet eigenvalue of the operator $d_{\overline{A_0}}^*d_{A_0}+V_1$.

Under all of the assumptions above, we have the following result.

Theorem B.1. *There exist $\delta>0$ and $C>0$ such that for any*

$$f\in B_\delta(\partial M):=\{f\in C^{2,\alpha}(\partial M):\|f\|_{C^{2,\alpha}(\partial M)}<\delta\},$$

the problem (B-2) *has a solution $u=u_f\in C^{2,\alpha}(M)$ which satisfies*

$$\|u\|_{C^{2,\alpha}(M)}\le C\|f\|_{C^{2,\alpha}(\partial M)}.$$

The solution u is unique within the class $\{u\in C^{2,\alpha}(M):\|u\|_{C^{2,\alpha}(M)}<C\delta\}$ and it depends holomorphically on $f\in B_\delta(\partial M)$. Furthermore, the map

$$B_\delta(\partial M)\to C^{1,\alpha}(M),\qquad f\mapsto\partial_\nu u_f|_{\partial M}$$

is holomorphic.

Proof. We shall follow [Lassas et al. 2021a], see also [Krupchyk and Uhlmann 2020a], and in order to prove this result we shall rely on the implicit function theorem for holomorphic maps between complex Banach spaces; see [Pöschel and Trubowitz 1987, p. 144]. To that end, we let

$$B_1=C^{2,\alpha}(\partial M),\qquad B_2=C^{2,\alpha}(M),\qquad\text{and}\qquad B_3=C^{\alpha}(M)\times C^{2,\alpha}(\partial M),$$

and introduce the map

$$F : B_1 \times B_2 \to B_3, \quad F(f, u) = (L_{A,V} u, u|_{\partial M} - f). \tag{B-5}$$

Let us verify that the map F indeed enjoys the mapping properties given in (B-5). To that end, let $u \in C^{2,\alpha}(M)$ and note first that $-\Delta_g u \in C^\alpha(M)$. Let us check that $A(\,\cdot\,, u(\,\cdot\,)) \in C^{1,\alpha}(M, T^*M)$. By Cauchy's estimates, the coefficients A_k in (B-3) satisfy

$$\|A_k\|_{C^{1,\alpha}(M,T^*M)} \le \frac{k!}{R^k} \sup_{|z|=R} \|A(\,\cdot\,, z)\|_{C^{1,\alpha}(M,T^*M)}, \quad R > 0, \tag{B-6}$$

for all $k = 0, 1, \dots$. Using (B-1) and (B-6), we obtain

$$\left\| \frac{A_k}{k!} u^k \right\|_{C^{1,\alpha}(M,T^*M)} \le \frac{C^k}{R^k} \|u\|^k_{C^{1,\alpha}(M)} \sup_{|z|=R} \|A(\,\cdot\,, z)\|_{C^{1,\alpha}(M,T^*M)}, \tag{B-7}$$

for all $k = 0, 1, \dots$. Choosing $R = 2C\|u\|_{C^{1,\alpha}(M)}$, it follows from (B-7) that the series $\sum_{k=0}^\infty A_k(x)u^k/k!$ converges in $C^{1,\alpha}(M, T^*M)$, and thus, $A(\,\cdot\,, u(\,\cdot\,)) \in C^{1,\alpha}(M, T^*M)$. Similarly, $V(\,\cdot\,, u(\,\cdot\,)) \in C^\alpha(M)$; see also [Krupchyk and Uhlmann 2020a]. Hence, using (1-4), we see that $L_{A,V} u \in C^\alpha(M)$.

We next claim that the map F in (B-5) is holomorphic. To this end, we first note that F is locally bounded as F is continuous in (f, u). Thus, it suffices to show that F is weakly holomorphic; see [Pöschel and Trubowitz 1987, p. 133]. In doing so, let $(f_0, u_0), (f_1, u_1) \in B_1 \times B_2$, and let us prove that the map

$$\lambda \mapsto F((f_0, u_0) + \lambda(f_1, u_1))$$

is holomorphic in $\mathbb{C}$ with values in B_3. It suffices to check that the map $\lambda \mapsto A(x, u_0(x) + \lambda u_1(x))$ is holomorphic in $\mathbb{C}$ with values in $C^{1,\alpha}(M, T^*M)$, as the fact that the map $\lambda \mapsto V(x, u_0(x) + \lambda u_1(x))$ is holomorphic in $\mathbb{C}$ with values in $C^\alpha(M)$ can be proved similarly; see [Krupchyk and Uhlmann 2020a]. The holomorphy of $\lambda \mapsto A(x, u_0(x) + \lambda u_1(x))$ follows from the fact that in view of (B-7), the series

$$\sum_{k=0}^\infty \frac{A_k}{k!}(u_0 + \lambda u_1)^k$$

converges in $C^{1,\alpha}(M, T^*M)$, locally uniformly in $\lambda \in \mathbb{C}$.

We have $F(0, 0) = 0$, and the partial differential $\partial_u F(0, 0) : B_2 \to B_3$ is given by

$$\partial_u F(0, 0)v = (d^*_{A_0} d_{A_0} v + V_1 v, v|_{\partial M}).$$

By the assumption (i), we have that the map $\partial_u F(0, 0) : B_2 \to B_3$ is a linear isomorphism; see [Gilbarg and Trudinger 1983, Theorem 6.15].

An application of the implicit function theorem, see [Pöschel and Trubowitz 1987, p. 144], allows us to conclude that there exists $\delta > 0$ and a unique holomorphic map $S : B_\delta(\partial M) \to C^{2,\alpha}(M)$ such that $S(0) = 0$ and $F(f, S(f)) = 0$ for all $f \in B_\delta(\partial M)$. Setting $u = S(f)$ and noting that S is Lipschitz continuous with $S(0) = 0$, we see that

$$\|u\|_{C^{2,\alpha}(M)} \le C\|f\|_{C^{2,\alpha}(\partial M)}. \qquad \square$$

Appendix C: First-order boundary determination of potentials

When proving Theorem 1.3 and Proposition 1.6, an important step consists in determining the boundary values, as well as the normal derivatives, of a scalar function and a 1-form, via suitable orthogonality relations involving harmonic functions on the manifold. The purpose of this section is to carry out this step. In doing so, we shall rely on the methods developed in [Brown 2001; Brown and Salo 2006], with suitable modifications in [Guillarmou and Tzou 2011, Appendix], where the boundary values of a scalar potential and a vector field are recovered. The main contribution of this section is that we push the methods a little further, in order to recover the first-order normal derivatives of the potential and the 1-form under limited regularity assumptions; see also [Alessandrini et al. 2018]. We would like to mention the works [Brown and Salo 2006; García and Zhang 2016, Appendix], where the gradient of a C^1-conductivity at the boundary of a Euclidean domain is recovered; see also [Alessandrini 1990; Caro and Garcia 2017; Caro and Meroño 2020]. We refer to [Kohn and Vogelius 1984; Lee and Uhlmann 1989; Nakamura et al. 1995; Sylvester and Uhlmann 1988], where the entire Taylor series at the boundary of C^∞-coefficients are recovered.

To proceed, we shall need the following density result for the space of L^2-harmonic functions; see also [Choe et al. 2004, Corollary 2.14] for a different approach in the Euclidean setting.

Proposition C.1. *Let (M, g) be a smooth compact Riemannian manifold of dimension $n \ge 2$ with smooth boundary. The set of harmonic functions on M^{int} that are smooth up to the boundary is dense in the space of L^2-harmonic functions in the L^2 topology.*

Proof. Let $u \in L^2(M)$ be harmonic, i.e., $-\Delta_g u = 0$ in M^{int}. Then by the partial hypoellipticity of the Laplacian, see [Eskin 2011, Theorem 26.1], we have $f = u|_{\partial M} \in H^{-1/2}(\partial M)$. There exists therefore a sequence $f_j \in C^\infty(\partial M)$, $j = 1, 2, \dots$, such that $\|f_j - f\|_{H^{-1/2}(\partial M)} \to 0$, as $j \to \infty$. The Dirichlet problem

$$\begin{cases} -\Delta_g u_j = 0 & \text{in } M^{\text{int}}, \\ u_j|_{\partial M} = f_j, \end{cases}$$

has a unique solution $u_j \in H^1(M)$, and by the boundary elliptic regularity, $u_j \in C^\infty(M)$. By [Eskin 2011, Theorem 26.3], we get

$$\|u_j - u\|_{L^2(M)} \le C\|f_j - f\|_{H^{-1/2}(\partial M)} \to 0,$$

as $j \to \infty$, establishing the proposition. □

Our first boundary determination result follows. While this result is not used in this work, the construction of a family of harmonic functions given in the proof is needed for the proof of Proposition C.3 below. Furthermore, we state this result and provide the proof for completeness and the convenience of the reader.

Proposition C.2. *Let (M, g) be a conformally transversally anisotropic manifold of dimension $n \ge 3$, and let $V \in C^{1,1}(M)$. If*

$$\int_M V u_1 u_2 \, dV_g = 0, \tag{C-1}$$

for all harmonic functions $u_1, u_2 \in C^\infty(M)$, then $V|_{\partial M} = 0$ and $\partial_\nu V|_{\partial M} = 0$.

Proof. By Proposition C.1, we see that (C-1) continues to hold for all harmonic functions $u_1, u_2 \in L^2(M)$. To proceed, we shall follow [Brown 2001; Brown and Salo 2006], constructing a family of functions, whose boundary values have a highly oscillatory behavior while becoming increasingly concentrated near a given point on the boundary of M. To convert such functions to harmonic functions, we follow the idea of [Guillarmou and Tzou 2011, Appendix] and rely on a Carleman estimate for the conjugated Laplacian with a gain of two derivatives, established in [Salo and Tzou 2009, Lemma 2.1] in the Euclidean case and in [Krupchyk and Uhlmann 2018, Proposition 2.2] in the conformally transversally anisotropic case.

Let $x_0 \in \partial M$ and let $(x_1, \dots, x_n)$ be the boundary normal coordinates centered at x_0 so that in these coordinates, $x_0 = 0$, the boundary ∂M is given by $\{x_n = 0\}$, and M^{int} is given by $\{x_n > 0\}$. We have, see [Lee and Uhlmann 1989],

$$g(x', x_n) = \sum_{\alpha,\beta=1}^{n-1} g_{\alpha\beta}(x)\, dx_\alpha\, dx_\beta + (dx_n)^2, \tag{C-2}$$

and we may also assume that the coordinates $x' = (x_1, \dots, x_{n-1})$ are chosen so that

$$g^{\alpha\beta}(x', 0) = \delta^{\alpha\beta} + \mathcal{O}(|x'|^2), \quad 1 \le \alpha, \ \ \beta \le n-1; \tag{C-3}$$

see [Petersen 2006, Chapter 2, Section 8, p. 56]. Therefore,

$$g^{\alpha\beta}(x', x_n) = g^{\alpha\beta}(x', 0) + \mathcal{O}(x_n) = \delta^{\alpha\beta} + \mathcal{O}(|x'|^2) + \mathcal{O}(x_n). \tag{C-4}$$

In view of (C-3), we have

$$-\Delta_g = D_{x_n}^2 + \sum_{\alpha,\beta=1}^{n-1} g^{\alpha\beta}(x) D_{x_\alpha} D_{x_\beta} + f(x) D_{x_n} + R(x, D_{x'}), \tag{C-5}$$

where f is a smooth function and R is a differential operator of order 1 in x' with smooth coefficients; see [Lee and Uhlmann 1989]. Notice that in the local coordinates, $T_{x_0}\partial M = \mathbb{R}^{n-1}$, equipped with the Euclidean metric. The unit tangent vector τ is then given by $\tau = (\tau', 0)$, where $\tau' \in \mathbb{R}^{n-1}$, $|\tau'| = 1$. Associated to the tangent vector τ' is the covector $\xi'_\alpha = \sum_{\beta=1}^{n-1} g_{\alpha\beta}(0)\tau'_\beta = \tau'_\alpha \in T^*_{x_0}\partial M$.

Let $\eta \in C_0^\infty(\mathbb{R}^n; \mathbb{R})$ be such that $\operatorname{supp}(\eta)$ is in a small neighborhood of 0, and

$$\int_{\mathbb{R}^{n-1}} \eta(x', 0)^2\, dx' = 1. \tag{C-6}$$

Let $\frac{1}{3} \le \alpha \le \frac{1}{2}$. Following [Brown and Salo 2006], in the boundary normal coordinates, we set

$$v_0(x) = \eta\left(\frac{x}{\lambda^\alpha}\right) e^{i(\tau'\cdot x' + i x_n)/\lambda}, \quad 0 < \lambda \ll 1, \tag{C-7}$$

so that $v_0 \in C^\infty(M)$, with $\operatorname{supp}(v_0)$ in an $\mathcal{O}(\lambda^\alpha)$ neighborhood of $x_0 = 0$. Here τ' is viewed as a covector.

A direct computation

$$\|v_0\|^2_{L^2(M)} = \mathcal{O}(1) \int_{|x| \le c\lambda^\alpha,\, x_n \ge 0} e^{-2x_n/\lambda}\, dx'\, dx_n = \mathcal{O}(\lambda^{\alpha(n-1)}) \int_0^\infty e^{-2t} \lambda\, dt = \mathcal{O}(\lambda^{\alpha(n-1)+1}), \tag{C-8}$$

as λ → 0, shows that

$$\|v_0\|_{L^2(M)} = \mathcal{O}(\lambda^{\alpha(n-1)/2+1/2}). \tag{C-9}$$

Following [Guillarmou and Tzou 2011, Appendix], we shall construct a harmonic function $u \in L^2(M)$ of the form

$$u = v_0 + r,$$

and therefore, we need to find $r \in L^2(M)$ satisfying

$$\Delta_g r = -\Delta_g v_0 \quad \text{in } M^{\text{int}}. \tag{C-10}$$

To that end, we shall rely on the following Carleman estimate for the conjugated Laplacian with a gain of two derivatives established in [Salo and Tzou 2009, Lemma 2.1; Krupchyk and Uhlmann 2018, Proposition 2.2]: for all $0 < h \ll 1$ and all $v \in C_0^\infty(M^{\text{int}})$, we have

$$\|v\|_{H^2_{\text{scl}}(M^{\text{int}})} \le \frac{C}{h} \|e^{\varphi/h} \circ (-h^2\Delta_g) \circ e^{-\varphi/h} v\|_{L^2(M)}. \tag{C-11}$$

Here the limiting Carleman weight $\varphi(x)$ equals x_1. Using a standard argument, one can convert the Carleman estimate (C-11) into a solvability result. Applying this solvability result with $h > 0$ small but fixed, we see that there exists a solution $r \in L^2(M)$ of (C-10) such that

$$\|r\|_{L^2(M)} \le C \|\Delta_g v_0\|_{H^{-2}(M^{\text{int}})}. \tag{C-12}$$

Next we claim that

$$\|\Delta_g v_0\|_{H^{-2}(M^{\text{int}})} = \mathcal{O}(\lambda^{\alpha(n-3)/2+3/2}), \quad \tfrac{1}{3} \le \alpha \le \tfrac{1}{2}, \tag{C-13}$$

as λ → 0. In order to prove (C-13), we first compute the Euclidean Laplacian acting on v_0:

$$\begin{aligned}\Delta v_0 &= e^{i(\tau'\cdot x' + i x_n)/\lambda}\left[\lambda^{-2\alpha}(\Delta\eta)\left(\frac{x}{\lambda^\alpha}\right) + 2i\lambda^{-\alpha-1}(\nabla\eta)\left(\frac{x}{\lambda^\alpha}\right)\cdot(\tau', i) - \lambda^{-2}(\tau', i)\cdot(\tau', i)\eta\left(\frac{x}{\lambda^\alpha}\right)\right] \\ &= e^{i(\tau'\cdot x' + i x_n)/\lambda}\left[\lambda^{-2\alpha}(\Delta\eta)\left(\frac{x}{\lambda^\alpha}\right) + 2i\lambda^{-\alpha-1}(\nabla\eta)\left(\frac{x}{\lambda^\alpha}\right)\cdot(\tau', i)\right],\end{aligned} \tag{C-14}$$

where we have used that $(\tau', i)\cdot(\tau', i) = 0$. The second term in the right-hand side of (C-14) has the worst growth as $\alpha \to 0$ and we will analyze it. The first term in the right-hand side of (C-14) can be treated in a similar fashion. To that end, we note that the second term in the right-hand side of (C-14) has the form

$$\lambda^{-\alpha-1}\chi\left(\frac{x}{\lambda^\alpha}\right)e^{i(\tau'\cdot x' + i x_n)/\lambda},$$

where $\chi \in C^\infty(\mathbb{R}^n)$ is supported in a small neighborhood of 0, and we can proceed similarly to [Guillarmou and Tzou 2011, Appendix]. Setting

$$L = \frac{\nabla\bar{\phi}\cdot\nabla}{i|\nabla\phi|^2} = \frac{1}{2i}\nabla\bar{\phi}\cdot\nabla, \quad \phi = \tau'\cdot x' + i x_n,$$

we get $Le^{i(\tau'\cdot x'+ix_n)/\lambda}=\lambda^{-1}e^{i(\tau'\cdot x'+ix_n)/\lambda}$. Letting $\psi\in C_0^\infty(M^{\text{int}})$ and integrating by parts twice using the operator L, we obtain

$$\lambda^{-\alpha-1}\int_M \chi\left(\frac{x}{\lambda^\alpha}\right)\psi(x)e^{i(\tau'\cdot x'+ix_n)/\lambda}\,dV_g \\ =\lambda^{-\alpha-1}\lambda^2\int_M (L)^2\left(\chi\left(\frac{x}{\lambda^\alpha}\right)\psi(x)|g(x)|^{1/2}\right)e^{i(\tau'\cdot x'+ix_n)/\lambda}\,dx, \tag{C-15}$$

since the transpose L^t equals $-L$. The term in the right-hand side of (C-15), where the bound cannot be improved integrating by parts further, will occur when the operator $(L)^2$ falls on ψ, and in this case, using the Cauchy–Schwarz inequality and a computation similar to (C-8), we get

$$\left|\lambda^{-\alpha+1}\int_M \chi\left(\frac{x}{\lambda^\alpha}\right)e^{i(\tau'\cdot x'+ix_n)/\lambda}(L)^2(\psi(x))\,dV_g\right| \\ \le \lambda^{-\alpha+1}\left\|\chi\left(\frac{x}{\lambda^\alpha}\right)e^{i(\tau'\cdot x'+ix_n)/\lambda}\right\|_{L^2(M)}\|\psi\|_{H^2(M^{\text{int}})}\le \mathcal{O}(\lambda^{\alpha(n-3)/2+3/2})\|\psi\|_{H^2(M^{\text{int}})}. \tag{C-16}$$

Proceeding similarly, integrating by parts using the operator L, if needed, we can bound all the other terms in (C-15) with the same bound as in (C-16). Therefore, it follows from (C-14) and (C-16) that for $0<\alpha\le\frac{1}{2}$, we have

$$\|\Delta v_0\|_{H^{-2}(M^{\text{int}})}=\mathcal{O}(\lambda^{\alpha(n-3)/2+3/2}), \tag{C-17}$$

as $\lambda\to 0$. To get the bound (C-13) for the Laplace–Beltrami operator, we notice that in view of (C-3), (C-5), and (C-17), we have to bound

$$\sum_{\alpha,\beta=1}^{n-1}(g^{\alpha\beta}(x)-\delta^{\alpha\beta})D_{x_\alpha}D_{x_\beta}v_0+f(x)D_{x_n}v_0+R(x,D_{x'})v_0 \tag{C-18}$$

in $H^{-2}(M^{\text{int}})$. Let us proceed to bound the first term. To that end, we compute

$$D_{x_\alpha}D_{x_\beta}v_0=e^{i(\tau'\cdot x'+ix_n)/\lambda}\left[\lambda^{-2\alpha}(D_{x_\alpha}D_{x_\beta}\eta)\left(\frac{x}{\lambda^\alpha}\right)+\lambda^{-1-\alpha}(D_{x_\alpha}\eta)\left(\frac{x}{\lambda^\alpha}\right)\tau_\beta\right. \\ \left.+\lambda^{-1-\alpha}(D_{x_\beta}\eta)\left(\frac{x}{\lambda^\alpha}\right)\tau_\alpha+\lambda^{-2}\tau_\alpha\tau_\beta\eta\left(\frac{x}{\lambda^\alpha}\right)\right]. \tag{C-19}$$

The worst growth as $\lambda\to 0$ is in the fourth term in (C-19), and therefore, in view of (C-18), we proceed to bound

$$\lambda^{-2}(g^{\alpha\beta}-\delta^{\alpha\beta})\chi\left(\frac{x}{\lambda^\alpha}\right)e^{i(\tau'\cdot x'+ix_n)/\lambda},\qquad \chi(x)=\tau_\alpha\tau_\beta\eta(x),$$

in $H^{-2}(M^{\text{int}})$. The other terms in the first term in (C-18) can be bounded similarly. As before, integrating by parts twice using the operator L, we get

$$\lambda^{-2}\int_M(g^{\alpha\beta}-\delta^{\alpha\beta})\chi\left(\frac{x}{\lambda^\alpha}\right)e^{i(\tau'\cdot x'+ix_n)/\lambda}\psi\,dV_g \\ =\int_M(L)^2\left((g^{\alpha\beta}-\delta^{\alpha\beta})\chi\left(\frac{x}{\lambda^\alpha}\right)\psi|g(x)|^{1/2}\right)e^{i(\tau'\cdot x'+ix_n)/\lambda}\,dx. \tag{C-20}$$

The term in the right-hand side of (C-20) where the bound cannot be improved occurs when the operator $(L)^2$ falls on ψ, and in this case, using the Cauchy–Schwarz inequality, (C-4), and a computation similar to (C-8), we get

$$
\begin{aligned}
&\left|\int_M (g^{\alpha\beta}-\delta^{\alpha\beta})\chi\left(\frac{x}{\lambda^\alpha}\right)e^{i(\tau'\cdot x'+ix_n)/\lambda}(L)^2\psi\,dV_g\right| \\
&\qquad\le \left(\int_M (\mathcal{O}(|x'|^4)+\mathcal{O}(x_n^2))\chi^2\left(\frac{x}{\lambda^\alpha}\right)e^{-2x_n/\lambda}\,dV_g\right)^{1/2}\|\psi\|_{H^2(M^{\mathrm{int}})} \\
&\qquad\le \left(\mathcal{O}(\lambda^{2\alpha}\lambda^{\alpha(n-1)/2+1/2})+\mathcal{O}(\lambda^{\alpha(n-1)/2})\left(\int_0^\infty x_n^2e^{-2x_n/\lambda}\,dx_n\right)^{1/2}\right)\|\psi\|_{H^2(M^{\mathrm{int}})} \\
&\qquad= (\mathcal{O}(\lambda^{\alpha(n+3)/2+1/2})+\mathcal{O}(\lambda^{\alpha(n-1)/2+3/2}))\|\psi\|_{H^2(M^{\mathrm{int}})}. \qquad\text{(C-21)}
\end{aligned}
$$

The growth in λ in (C-21) is smaller than or equal to that in the desired bound (C-13) provided that $\alpha\ge\frac{1}{3}$. Proceeding similarly integrating by parts, using the operator L if needed, we can bound all the other terms in (C-20) by the bound which is the same or better than

$$\mathcal{O}(\lambda^{\alpha(n-3)/2+3/2})\|\psi\|_{H^2(M^{\mathrm{int}})}.$$

Thus, using this and in view of (C-18)–(C-21), we conclude that

$$\left\|\sum_{\alpha,\beta=1}^{n-1}(g^{\alpha\beta}(x)-\delta^{\alpha\beta})D_{x_\alpha}D_{x_\beta}v_0\right\|_{H^{-2}(M^{\mathrm{int}})}=\mathcal{O}(\lambda^{\alpha(n-3)/2+3/2}), \qquad\text{(C-22)}$$

provided that $\frac{1}{3}\le\alpha\le\frac{1}{2}$. Finally, as $R(x,D_{x'})$ is a differential operator of order 1 in x', similarly, we get

$$\|f(x)D_{x_n}v_0+R(x,D_{x'})v_0\|_{H^{-2}(M^{\mathrm{int}})}=\mathcal{O}(\lambda^{\alpha(n-1)/2+3/2}), \qquad\text{(C-23)}$$

which is better than the desired bound (C-13). Hence, combining (C-17), (C-22), and (C-23), we get (C-13).

Now it follows from (C-12) and (C-13) that

$$\|r\|_{L^2(M)}=\mathcal{O}(\lambda^{\alpha(n-3)/2+3/2}),\quad \tfrac{1}{3}\le\alpha\le\tfrac{1}{2}, \qquad\text{(C-24)}$$

as $\lambda\to 0$. Notice that the bound for r in L^2 is better than the bound for v_0 in L^2; see (C-9).

Letting

$$u_1=v_0+r,\quad u_2=\overline{v_0+r}, \qquad\text{(C-25)}$$

in (C-1) and multiplying (C-1) by $\lambda^{-\alpha(n-1)-1}$, we get

$$0=\lambda^{-\alpha(n-1)-1}\int_M V(v_0+r)(\bar v_0+\bar r)\,dV_g=\lambda^{-\alpha(n-1)-1}(I_1+I_2+I_3). \qquad\text{(C-26)}$$

Here

$$I_1=\int_M V|v_0|^2\,dV_g,\quad I_2=\int_M V(v_0\bar r+\bar v_0 r)\,dV_g,\quad\text{and}\quad I_3=\int_M V|r|^2\,dV_g.$$

Using (C-9) and (C-24), we obtain

$$\lambda^{-\alpha(n-1)-1}|I_2|\le\mathcal{O}(\lambda^{-\alpha(n-1)-1})\|v_0\|_{L^2(M)}\|r\|_{L^2(M)}=\mathcal{O}(\lambda^{1-\alpha}), \qquad\text{(C-27)}$$

and

$$\lambda^{-\alpha(n-1)-1}|I_3| \le \mathcal{O}(\lambda^{-\alpha(n-1)-1})\|r\|^2_{L^2(M)} = \mathcal{O}(\lambda^{2-2\alpha}), \tag{C-28}$$

as $\lambda \to 0$. Using (C-7), (C-6), the fact that V is continuous, and making the change of variables $y' = x'/\lambda^\alpha$ and $y_n = x_n/\lambda$, we get

$$\begin{aligned}\lim_{\lambda\to 0} \lambda^{-\alpha(n-1)-1} I_1 &= \lim_{\lambda\to 0} \int_{\mathbb{R}^{n-1}} \int_0^\infty V(\lambda^\alpha y', \lambda y_n)\eta^2(y', \lambda^{1-\alpha} y_n) e^{-2y_n} |g(\lambda^\alpha y', \lambda y_n)|^{1/2}\, dy'\, dy_n \\ &= V(0)|g(0)|^{1/2} \int_0^{+\infty} e^{-2y_n}\, dy_n = \frac{1}{2} V(0). \end{aligned} \tag{C-29}$$

Passing to the limit $\lambda \to 0$ in (C-26) and using (C-27)–(C-29), we obtain $V(0) = 0$, showing that $V|_{\partial M} = 0$. Notice that here we can consider any α, $\frac{1}{3} \le \alpha \le \frac{1}{2}$.

Next we would like to prove that $\partial_\nu V|_{\partial M} = 0$. To that end, as before, we let $x_0 \in \partial M$ and consider boundary normal coordinates centered at x_0. As $V \in C^{1,1}$ and $V(x', 0) = 0$, using the fundamental theorem of calculus and integrating by parts, we have for x near $x_0 = 0$,

$$\begin{aligned} V(x', x_n) &= \int_0^1 \frac{d}{dt} V(x', t x_n)\, d(t-1) = V'_{x_n}(x', 0) x_n + \int_0^1 (1-t) \frac{d^2}{dt^2} V(x', t x_n) \\ &= V'_{x_n}(x', 0) x_n + \int_0^1 (1-t) V''_{x_n x_n}(x', t x_n) x_n^2\, dt = V'_{x_n}(x', 0) x_n + \mathcal{O}(x_n^2). \end{aligned} \tag{C-30}$$

Now substituting u_1 and u_2 as given by (C-25) into (C-1), multiplying (C-1) by $\lambda^{-\alpha(n-1)-2}$, and then using (C-30), we get

$$0 = \lambda^{-\alpha(n-1)-2} \int_M V(v_0 + r)(\bar{v}_0 + \bar{r})\, dV_g = \lambda^{-\alpha(n-1)-2}(I_{1,1} + I_{1,2} + I_2 + I_3). \tag{C-31}$$

Here

$$\begin{aligned} I_{1,1} &= \int_M V'_{x_n}(x', 0) x_n |v_0|^2\, dV_g, \quad & I_{1,2} &= \int_M \mathcal{O}(x_n^2)|v_0|^2\, dV_g, \\ I_2 &= \int_M V(v_0 \bar{r} + \bar{v}_0 r)\, dV_g, & I_3 &= \int_M V|r|^2\, dV_g. \end{aligned}$$

Using (C-7) and (C-6), making the change of variables $y' = x'/\lambda^\alpha$ and $y_n = x_n/\lambda$, and using that V'_{x_n} is continuous, we obtain

$$\begin{aligned}\lim_{\lambda\to 0} \lambda^{-\alpha(n-1)-2} I_{1,1} &= \lim_{\lambda\to 0} \int_{\mathbb{R}^{n-1}} \int_0^\infty V'_{x_n}(\lambda^\alpha y', 0)\eta^2(y', \lambda^{1-\alpha} y_n) y_n e^{-2y_n} |g(\lambda^\alpha y', \lambda y_n)|^{1/2}\, dy'\, dy_n \\ &= V'_{x_n}(0)|g(0)|^{1/2} \int_0^{+\infty} y_n e^{-2y_n}\, dy_n = \frac{1}{4} V'_{x_n}(0). \end{aligned} \tag{C-32}$$

Using (C-7), we get

$$\lambda^{-\alpha(n-1)-2}|I_{1,2}| \le \mathcal{O}(\lambda^{-\alpha(n-1)-2}) \int_{|x| \le c\lambda^\alpha,\, x_n \ge 0} x_n^2 e^{-2x_n/\lambda}\, dx'\, dx_n = \mathcal{O}(\lambda). \tag{C-33}$$

Using (C-24), we see that

$$\lambda^{-\alpha(n-1)-2}|I_3| \le \mathcal{O}(\lambda^{-\alpha(n-1)-2})\|r\|^2_{L^2(M)} = \mathcal{O}(\lambda^{1-2\alpha}) = o(1), \tag{C-34}$$

as $\lambda \to 0$, provided that $\alpha < \frac{1}{2}$.

In view of (C-7) and (C-30), we have

$$\|Vv_0\|_{L^2(M)} = \left(\int_{|x|\le c\lambda^\alpha, x_n\ge 0} \mathcal{O}(x_n^2)e^{-2x_n/\lambda}\,dx'\,dx_n\right)^{1/2} = \mathcal{O}(\lambda^{\alpha(n-1)/2+3/2}),$$

and therefore, using (C-24), we obtain

$$\lambda^{-\alpha(n-1)-2}|I_2| \le \mathcal{O}(\lambda^{-\alpha(n-1)-2})\|r\|_{L^2(M)}\|Vv_0\|_{L^2(M)} = \mathcal{O}(\lambda^{1-\alpha}). \tag{C-35}$$

Passing to the limit $\lambda \to 0$ in (C-31), and using (C-32), (C-33), (C-23), and (C-35), we get $V'_{x_n}(0) = 0$ provided that α is a fixed number satisfying $\frac{1}{3} \le \alpha < \frac{1}{2}$. This shows that $\partial_\nu V|_{\partial M} = 0$. □

In order to prove Proposition 1.6, we shall need the following boundary determination result.

Proposition C.3. *Let (M, g) be a conformally transversally anisotropic manifold of dimension $n \ge 3$. Let $A \in C^{1,1}(M, T^*M)$ be a 1-form. If*

$$\int_M \langle A, du_1\rangle_g u_2\,dV_g = 0, \tag{C-36}$$

for all harmonic functions $u_1, u_2 \in C^\infty(M)$, then $A|_{\partial M} = 0$ and $\partial_\nu A|_{\partial M} = 0$.

Proof. First by Proposition C.1, we see that (C-36) holds for all harmonic functions $u_2 \in L^2(M)$. To prove this result, we shall test the integral identity (C-36) with harmonic functions $u_2 \in L^2(M)$, constructed in Proposition C.2, of the form

$$u_2 = \overline{v_0 + r}. \tag{C-37}$$

Since for u_1 we need estimates in $H^1(M^{\text{int}})$, we shall construct u_1 following [Brown 2001; Brown and Salo 2006]; see also [Krupchyk and Uhlmann 2018, Appendix A]. We let

$$u_1 = v_0 + r_1, \tag{C-38}$$

where $r_1 \in H_0^1(M^{\text{int}})$ is a solution to the Dirichlet problem

$$\begin{cases} -\Delta_g r_1 = \Delta_g v_0 & \text{in } M, \\ r_1|_{\partial M} = 0. \end{cases} \tag{C-39}$$

Note that by boundary elliptic regularity, $r_1 \in C^\infty(M)$, and therefore, $u_1 \in C^\infty(M)$.

Applying the Lax–Milgram lemma to (C-39), we get

$$\|r_1\|_{H_0^1(M^{\text{int}})} \le C\|\Delta_g v_0\|_{H^{-1}(M^{\text{int}})}. \tag{C-40}$$

Similarly to the bound (C-13), one can show that

$$\|\Delta_g v_0\|_{H^{-1}(M^{\text{int}})} = \mathcal{O}(\lambda^{\alpha(n-3)/2+1/2}), \quad \tfrac{1}{3} \le \alpha \le \tfrac{1}{2};$$

see also [Krupchyk and Uhlmann 2018, Appendix A]. This bound together with (C-40) implies that

$$\|r_1\|_{H^1(M^{\text{int}})} = \mathcal{O}(\lambda^{\alpha(n-3)/2+1/2}), \quad \tfrac{1}{3} \le \alpha \le \tfrac{1}{2}, \tag{C-41}$$

as λ → 0.

We shall also need the bound

$$\|dv_0\|_{L^2(M)} = \mathcal{O}(\lambda^{\alpha(n-1)/2-1/2}), \tag{C-42}$$

as λ → 0, which is in view of (C-7) implied by the estimate

$$\|dv_0\|_{L^2(M)} \le \mathcal{O}(1)\biggl(\int_{|x|\le c\lambda^\alpha,\, x_n\ge 0} \lambda^{-2} e^{-2x_n/\lambda}\, dx'\, dx_n\biggr)^{1/2} = \mathcal{O}(\lambda^{\alpha(n-1)/2-1/2}).$$

Now substituting u_1 and u_2 given by (C-38) and (C-37), respectively, into (C-36) and multiplying (C-36) by $\lambda^{-\alpha(n-1)}$, we get

$$0 = \lambda^{-\alpha(n-1)} \int_M \langle A, dv_0 + dr_1\rangle_g (\bar v_0 + \bar r)\, dV_g = \lambda^{-\alpha(n-1)}(I_1 + I_2 + I_3), \tag{C-43}$$

where

$$I_1 = \int_M \langle A, dv_0\rangle_g \bar v_0\, dV_g, \quad I_2 = \int_M \langle A, dr_1\rangle_g (\bar v_0 + \bar r)\, dV_g, \quad \text{and} \quad I_3 = \int_M \langle A, dv_0\rangle_g \bar r\, dV_g.$$

First using (C-7), we write

$$I_1 = I_{1,1} + I_{1,2},$$

where

$$I_{1,1} = i\lambda^{-1} \int_M \langle A, \tau'\cdot dx' + i\,dx_n\rangle_g \eta^2\Bigl(\frac{x}{\lambda^\alpha}\Bigr) e^{-2x_n/\lambda}\, dV_g,$$
$$I_{1,2} = \lambda^{-\alpha} \int_M \Bigl\langle A, (d\eta)\Bigl(\frac{x}{\lambda^\alpha}\Bigr)\Bigr\rangle_g \eta\Bigl(\frac{x}{\lambda^\alpha}\Bigr) e^{-2x_n/\lambda}\, dV_g.$$

Using (C-2), and making the change of variables $y' = x'/\lambda^\alpha$ and $y_n = x_n/\lambda$, we get

$$\begin{aligned}\lim_{\lambda\to 0} \lambda^{-\alpha(n-1)} I_{1,1} &= i \lim_{\lambda\to 0} \int_{\mathbb{R}^{n-1}} \int_0^{+\infty} |g(\lambda^\alpha y', \lambda y_n)|^{1/2} \eta^2(y', \lambda^{1-\alpha} y_n) e^{-2y_n} \\ &\qquad \times \Bigl(\sum_{\alpha,\beta=1}^{n-1} g^{\alpha\beta}(\lambda^\alpha y', \lambda y_n) A_\alpha(\lambda^\alpha y', \lambda y_n)\tau'_\beta + A_n(\lambda^\alpha y', \lambda y_n) i\Bigr)\, dy'\, dy_n \\ &= i\Bigl(\sum_{\alpha,\beta=1}^{n-1} g^{\alpha\beta}(0) A_\alpha(0)\tau'_\beta + A_n(0) i\Bigr) |g(0)|^{1/2} \int_0^{+\infty} e^{-2y_n}\, dy_n \\ &= \frac{i}{2}\langle A(0), (\tau', i)\rangle. \end{aligned} \tag{C-44}$$

Estimating similarly as in (C-8), we get

$$\lambda^{-\alpha(n-1)}|I_{1,2}| \le \mathcal{O}(\lambda^{-\alpha n}) \Bigl\|(d\eta)\Bigl(\frac{x}{\lambda^\alpha}\Bigr)\Bigr\|_{L^2(M)} \Bigl\|\eta\Bigl(\frac{x}{\lambda^\alpha}\Bigr) e^{-2x_n/\lambda}\Bigr\|_{L^2(M)} = \mathcal{O}(\lambda^{(1-\alpha)/2}). \tag{C-45}$$

Using (C-9), (C-24), and (C-41), we see that

$$\lambda^{-\alpha(n-1)}|I_2| \le \mathcal{O}(\lambda^{-\alpha(n-1)})\|dr_1\|_{L^2(M)}\|v_0+r\|_{L^2(M)} = \mathcal{O}(\lambda^{1-\alpha}). \tag{C-46}$$

Finally, using (C-42) and (C-24), we obtain

$$\lambda^{-\alpha(n-1)}|I_3| \le \mathcal{O}(\lambda^{-\alpha(n-1)})\|dv_0\|_{L^2(M)}\|r\|_{L^2(M)} = \mathcal{O}(\lambda^{1-\alpha}). \tag{C-47}$$

Passing to the limit $\lambda \to 0$ in (C-43) and using (C-44)–(C-47), we conclude that $\langle A(0), (\tau', i)\rangle = 0$. Now changing τ' to $-\tau'$, we see that $A_n(0) = 0$, and therefore, $\langle A'(0), \tau'\rangle = 0$, where $A' = (A_1, \dots, A_{n-1})$. As $\tau' \in \mathbb{R}^{n-1}$ is an arbitrary tangent vector to ∂M at $x_0 = 0$, we get $A'(0) = 0$. This shows that $A|_{\partial M} = 0$.

Next we shall show that $\partial_\nu A|_{\partial M} = 0$. To that end, as before, we let $x_0 \in \partial M$ and consider the boundary normal coordinates centered at x_0. Applying computations similar to (C-30) to each component of A, we get

$$A(x', x_n) = (A'_{1x_n}, \dots, A'_{nx_n})(x', 0)x_n + \mathcal{O}(x_n^2) = \partial_{x_n} A(x', 0)x_n + \mathcal{O}(x_n^2). \tag{C-48}$$

Substituting u_1 and u_2 given by (C-38) and (C-37) into (C-36), and multiplying (C-36) by $\lambda^{-\alpha(n-1)-1}$, we have in view of (C-48),

$$0 = \lambda^{-\alpha(n-1)-1}\int_M \langle A, dv_0 + dr_1\rangle_g(\bar{v}_0 + \bar{r})\, dV_g = \lambda^{-\alpha(n-1)-1}(I_{1,1} + I_{1,2} + I_2 + I_3 + I_4), \tag{C-49}$$

where

$$I_{1,1} = \int_M \langle \partial_{x_n} A(x', 0)x_n, dv_0\rangle_g \bar{v}_0\, dV_g, \quad I_{1,2} = \int_M \langle \mathcal{O}(x_n^2), dv_0\rangle_g \bar{v}_0\, dV_g,$$

$$I_2 = \int_M \langle A, dr_1\rangle_g \bar{v}_0\, dV_g, \quad I_3 = \int_M \langle A, dr_1\rangle_g \bar{r}\, dV_g, \quad I_4 = \int_M \langle A, dv_0\rangle_g \bar{r}\, dV_g.$$

In view of (C-7) we write

$$I_{1,11} = i\lambda^{-1}\int_M \langle \partial_{x_n} A(x', 0)x_n, \tau' \cdot dx' + i\,dx_n\rangle_g \eta^2\Big(\frac{x}{\lambda^\alpha}\Big)e^{-2x_n/\lambda}\, dV_g,$$

$$I_{1,12} = \lambda^{-\alpha}\int_M \Big\langle \partial_{x_n} A(x', 0)x_n, (d\eta)\Big(\frac{x}{\lambda^\alpha}\Big)\Big\rangle_g \eta\Big(\frac{x}{\lambda^\alpha}\Big)e^{-2x_n/\lambda}\, dV_g.$$

Using (C-2), and making the change of variables $y' = x'/\lambda^\alpha$ and $y_n = x_n/\lambda$, we get

$$\begin{aligned}
\lim_{\lambda\to 0}\lambda^{-\alpha(n-1)-1}I_{1,11} &= i\lim_{\lambda\to 0}\int_{\mathbb{R}^{n-1}}\int_0^{+\infty} |g(\lambda^\alpha y', \lambda y_n)|^{1/2} y_n \eta^2(y', \lambda^{1-\alpha}y_n)e^{-2y_n} \\
&\qquad \times\Big(\sum_{\alpha,\beta=1}^{n-1} g^{\alpha\beta}(\lambda^\alpha y', \lambda y_n)\partial_{x_n} A_\alpha(\lambda^\alpha y', 0)\tau'_\beta + \partial_{x_n} A_n(\lambda^\alpha y', 0)i\Big)\, dy'\, dy_n \\
&= i\Big(\sum_{\alpha,\beta=1}^{n-1} g^{\alpha\beta}(0)\partial_{x_n} A_\alpha(0)\tau'_\beta + \partial_{x_n} A_n(0)i\Big)|g(0)|^{1/2}\int_0^{+\infty} y_n e^{-2y_n}\, dy_n \\
&= \frac{i}{4}\langle \partial_{x_n} A(0), (\tau', i)\rangle.
\end{aligned} \tag{C-50}$$

Estimating similarly as in (C-8), we get

$$\lambda^{-\alpha(n-1)-1}|I_{1,12}| \le \mathcal{O}(\lambda^{-\alpha n-1}) \left\|(d\eta)\left(\frac{x}{\lambda^\alpha}\right)\right\|_{L^2(M)} \left\|x_n\eta\left(\frac{x}{\lambda^\alpha}\right)e^{-2x_n/\lambda}\right\|_{L^2(M)} = \mathcal{O}(\lambda^{(1-\alpha)/2}). \tag{C-51}$$

Using (C-42) and estimating similarly as in (C-8), we obtain

$$\lambda^{-\alpha(n-1)-1}|I_{1,2}| \le \mathcal{O}(\lambda^{-\alpha(n-1)-1})\|dv_0\|_{L^2(M)}\|x_n^2 v_0\|_{L^2(M)} = \mathcal{O}(\lambda). \tag{C-52}$$

Using (C-41) and (C-48), we get

$$\lambda^{-\alpha(n-1)-1}|I_2| \le \mathcal{O}(\lambda^{-\alpha(n-1)-1})\|dr_1\|_{L^2(M)}\|x_n v_0\|_{L^2(M)} = \mathcal{O}(\lambda^{1-\alpha}). \tag{C-53}$$

Using (C-41) and (C-24), we have

$$\lambda^{-\alpha(n-1)-1}|I_3| \le \mathcal{O}(\lambda^{-\alpha(n-1)-1})\|dr_1\|_{L^2(M)}\|r\|_{L^2(M)} = \mathcal{O}(\lambda^{1-2\alpha}) = o(1), \tag{C-54}$$

as $\lambda \to 0$, provided that $\alpha < \frac{1}{2}$.

Using (C-48), (C-24), and the fact that

$$\|x_n dv_0\|_{L^2(M)} = \mathcal{O}(\lambda^{\alpha(n-1)/2+1/2}),$$

we obtain

$$\lambda^{-\alpha(n-1)-1}|I_4| \le \mathcal{O}(\lambda^{-\alpha(n-1)-1})\|x_n dv_0\|_{L^2(M)}\|r\|_{L^2(M)} = \mathcal{O}(\lambda^{1-\alpha}). \tag{C-55}$$

Let us fix $\frac{1}{3} \le \alpha < \frac{1}{2}$. Passing to the limit $\lambda \to 0$ in (C-49) and using (C-50)–(C-55), we conclude that $\langle \partial_{x_n} A(0), (\tau', i)\rangle = 0$, and therefore, $\partial_{x_n} A(0) = 0$. This shows that $\partial_\nu A|_{\partial M} = 0$. □

Finally, in order to prove Theorem 1.3 we shall need the following boundary determination result.

Proposition C.4. *Let (M, g) be a conformally transversally anisotropic manifold of dimension $n \ge 3$. Let $A \in C^{1,1}(M, T^*M)$ be a 1-form and $V \in C^{1,1}(M)$. If*

$$\int_M (4i\langle A, d(u_1u_2u_3)\rangle_g u_4 - (3id^*(A) + V)u_1u_2u_3u_4)\, dV_g = 0 \tag{C-56}$$

for all harmonic functions $u_j \in C^{2,\alpha}(M)$, $j = 1, \dots 4$, then $A|_{\partial M} = 0$ and $\partial_\nu A|_{\partial M} = 0$.

Proof. We also have

$$\int_M (4i\langle A, d(u_2u_3u_4)\rangle_g u_1 - (3id^*(A) + V)u_1u_2u_3u_4)\, dV_g = 0. \tag{C-57}$$

Subtracting (C-57) from (C-56), we get

$$\int_M \langle A, d(u_1u_2u_3)\rangle_g u_4\, dV_g - \int_M \langle A, d(u_2u_3u_4)\rangle_g u_1\, dV_g = 0. \tag{C-58}$$

Letting $u_3 = u_4 = 1$, (C-58) gives

$$\int_M \langle A, du_1\rangle_g u_2\, dV_g = 0 \tag{C-59}$$

for all harmonic functions $u_1, u_2 \in C^{2,\alpha}(M)$, and therefore for all harmonic functions $u_1, u_2 \in C^\infty(M)$. The result follows by an application of Proposition C.3. □

When proving Proposition 1.6, we shall also need the following standard density result.

Proposition C.5. *Let (M, g) be a smooth compact Riemannian manifold of dimension $n \geq 2$ with smooth boundary. The set of harmonic functions in M^{int} that are smooth up to the boundary is dense in the space of $C^{2,\alpha}(M)$-harmonic functions, $0 < \alpha < 1$, in the $C^{2,\beta}(M)$ topology, for $0 < \beta < \alpha$.*

Proof. The proof follows along the lines of the proof of Proposition C.1. Indeed, let $u \in C^{2,\alpha}(M)$ be harmonic in M^{int} and let $f = u|_{\partial M} \in C^{2,\alpha}(\partial M)$. Let $0 < \beta < \alpha$, and by density, there exists $f_j \in C^\infty(\partial M)$ such that $\|f_j - f\|_{C^{2,\beta}(\partial M)} \to 0$, as $j \to \infty$; see [Hörmander 1976, Theorem A.10]. The Dirichlet problem

$$\begin{cases} -\Delta_g u_j = 0 & \text{in } M^{\text{int}}, \\ u_j|_{\partial M} = f_j, \end{cases}$$

has a unique solution $u_j \in C^{2,\alpha}(M)$, and by elliptic regularity, we have $u_j \in C^\infty(M)$. Using the fact that $C^{2,\alpha}(M) \subset C^{2,\beta}(M)$ and the following bound for the solution to the Dirichlet problem for the Laplacian, see [Gilbarg and Trudinger 1983, Section 6.3, p. 109],

$$\|u_j - u\|_{C^{2,\beta}(M)} \leq C\|f_j - f\|_{C^{2,\beta}(\partial M)} \to 0,$$

we get the claim. □

Appendix D: Some facts about nontangential geodesics

When proving Proposition 1.6, in order to avoid the use of stationary and nonstationary phase arguments on the boundary of the manifold, we shall need the following result concerning nontangential geodesics which was kindly proven for us by Gabriel Paternain.

Proposition D.1. *Let (M_0, g_0) be a smooth compact Riemannian manifold of dimension $n \geq 2$ with smooth boundary, and let γ be a unit speed nontangential geodesic on M_0 between boundary points. Then for each point $y_0 = \gamma(t_0) \in M_0^{\text{int}}$, except for finitely many, there exists a small neighborhood*

$$W \subset S_{y_0}M_0 = \{w \in T_{y_0}M_0 : |w|_g = 1\}$$

of $w_0 = \dot\gamma(t_0)$ such that for every $w \in W$, $w \neq w_0$, the unit speed geodesic η on M_0 passing through (y_0, w) is also nontangential between boundary points, and γ and η do not intersect each other at the boundary of M_0.

Proof. Let us first notice that the property of a geodesic being nontangential is stable under small perturbations of the initial conditions, in view of the C^∞-dependence of the geodesic flow on the initial conditions. Let $y_0 = \gamma(t_0) \in M_0^{\text{int}}$. Reparametrizing the geodesic γ if necessary, we may assume that $\gamma : [-S_1, S_2] \to M_0$, $0 < S_1, S_2 < \infty$, is such that $\gamma(0) = y_0$ and $\dot\gamma(0) = w_0$. Let us consider the map

$$F_{y_0} : \text{neigh}(w_0, S_{y_0}M_0) \to \text{neigh}(\gamma(S_2), \partial M_0), \quad F_{y_0}(w) = \pi(\varphi_{\tau(y_0,w)}(y_0, w)), \tag{D-1}$$

where $\tau(y_0, w)$ is the exit time of the geodesic $\gamma_{y_0,w}$ through (y_0, w), $\varphi_t : SM_0 \to SM_0$, $t \in \mathbb{R}$, is the geodesic flow, given by

$$\varphi_t(y, w) = (\gamma_{y,w}(t), \dot\gamma_{y,w}(t)), \tag{D-2}$$

and $\pi : SM_0 \to M_0$, $\pi(y, w) = y$ is the canonical projection.

The exit time $\tau(y_0, w)$ depends smoothly on w, in view of the implicit function theorem and the fact that the geodesic γ is nontangential. The map F_{y_0} is therefore smooth, and we have $F_{y_0}(w_0) = \gamma(S_2)$.

Let us now compute the differential of F_{y_0} at w_0 acting on a vector $\eta \in T_{w_0} S_{y_0} M_0$. To that end, consider a curve $w : (-a, a) \to S_{y_0} M_0$ such that $w(0) = w_0$ and $\dot{w}(0) = \eta$, and by the chain rule, we get

$$\begin{aligned} F'_{y_0}(w_0)\eta &= \frac{d}{ds}\bigg|_{s=0} F_{y_0}(w(s)) = \frac{d}{ds}\bigg|_{s=0} \pi(\varphi_{\tau(y_0, w(s))}(y_0, w(s))) \\ &= d\pi(\varphi_{\tau(y_0,w_0)}(y_0, w_0))\left(\frac{d}{dt}\bigg|_{t=\tau(y_0,w_0)} \varphi_t(y_0, w_0)\frac{\partial\tau}{\partial w}(y_0, w_0)\cdot\eta + \frac{\partial\varphi_{\tau(y_0,w_0)}}{\partial w}(y_0, w_0)\eta\right). \end{aligned} \tag{D-3}$$

To proceed, we recall some facts about the geometry of the tangent bundle following [Paternain 1999]. First, letting

$$V(y, w) = \ker(d\pi(y, w)) \subset T_{(y,w)} SM_0$$

be the vertical fiber of TSM_0 at (y, w), see [Paternain 1999, Section 1.3.1], we have the splitting

$$T_{(y,w)} SM_0 = H(y, w) \oplus V(y, w),$$

where $H(y, w)$ is the horizontal fiber of TSM_0 at (y, w); see [Paternain 1999, Section 1.3, p. 13]. Both $V(y, w)$ and $H(y, w)$ can be identified with $S_y M_0$, and for $\xi \in T_{(y,w)} SM_0$, we write $\xi = (\xi^h, \xi^v)$, where $\xi^h, \xi^v \in S_y M_0$ are the corresponding horizontal and vertical parts of ξ. Let $X : SM_0 \to TSM_0$ be the geodesic vector field given by

$$X(\varphi_t(y, w)) = \frac{d}{dt}\varphi_t(y, w). \tag{D-4}$$

It follows from [Paternain 1999, Section 1.3, p. 13] that we have

$$X(y, w) = (w, 0). \tag{D-5}$$

Now in view of the above splitting, we have $(0, \eta) \in V(y_0, w_0)$, and therefore, we get

$$\frac{\partial\varphi_{\tau(y_0,w_0)}}{\partial w}(y_0, w_0)\eta = d\varphi_{\tau(y_0,w_0)}(y_0, w_0)(0, \eta). \tag{D-6}$$

Using the fact that $\tau(y_0, w_0) = S_2$, (D-2), and (D-4)–(D-6), we obtain from (D-3) that

$$\begin{aligned} F'_{y_0}(w_0)\eta &= d\pi(\gamma(S_2), \dot\gamma(S_2))\left(X(\gamma(S_2), \dot\gamma(S_2))\frac{\partial\tau}{\partial w}(y_0, w_0)\cdot\eta + d\varphi_{\tau(y_0,w_0)}(y_0, w_0)(0, \eta)\right) \\ &= \dot\gamma(S_2)\frac{\partial\tau}{\partial w}(y_0, w_0)\cdot\eta + d\pi(\gamma(S_2), \dot\gamma(S_2))(d\varphi_{S_2}(y_0, w_0)(0, \eta)). \end{aligned} \tag{D-7}$$

Now by [Paternain 1999, Lemma 1.40], see also [Ilmavirta 2020, Theorem 11.2], for the differential of the geodesic flow we get that

$$d\varphi_{S_2}(y_0, w_0)(0, \eta) = (J_{(0,\eta)}(S_2), \dot J_{(0,\eta)}(S_2)), \tag{D-8}$$

where $J_{(0,\eta)}$ is the Jacobi field along the geodesic $t \mapsto \pi(\varphi_t(y_0, w_0)) = \gamma(t)$ with the initial conditions

$$J_{(0,\eta)}(0) = 0, \quad \dot J_{(0,\eta)} = \eta. \tag{D-9}$$

Using [Ilmavirta 2020, Exercise 5.9], (D-9), and the fact that $\eta \in T_{w_0}S_{y_0}M_0$, we have

$$\langle \dot\gamma(S_2), J_{(0,\eta)}(S_2)\rangle = \langle \dot\gamma(0), J_{(0,\eta)}(0)\rangle + S_2\langle \dot\gamma(0), \dot J_{(0,\eta)}(0)\rangle = S_2\langle w_0, \eta\rangle = 0, \tag{D-10}$$

showing that the Jacobi field $J_{(0,\eta)}$ is normal to γ. It follows from (D-7) and (D-8) that

$$F'_{y_0}(w_0)\eta = \dot\gamma(S_2)\frac{\partial\tau}{\partial w}(y_0, w_0)\cdot\eta + J_{(0,\eta)}(S_2). \tag{D-11}$$

Using (D-11) and the orthogonally (D-10), we see that if $F'_{y_0}(w_0)$ has a nontrivial kernel, then there exists $\eta \neq 0$ such $J_{(0,\eta)}(S_2) = 0$, and therefore, the points y_0 and $\gamma(S_2)$ are conjugate points along γ; see [Ilmavirta 2020, Definition 7.3]. Thus, $F'_{y_0}(w_0)$ is bijective as long as y_0 is not a conjugate point to $\gamma(S_2)$ along γ.

By the inverse function theorem, F_{y_0} is a local diffeomorphism if y_0 is not a conjugate point to $\gamma(S_2)$ along γ.

Hence, if y_0 is not a conjugate point to $\gamma(S_2)$ and $\gamma(-S_1)$ along γ, there exists a small neighborhood $W \subset S_{y_0}M_0$ of w_0 such that for every $w \in W$, $w \neq w_0$, the unit speed geodesic $\eta : [-T_1, T_2] \to M_0$, $0 < T_1, T_2 < \infty$, such that $\eta(0) = y_0$ and $\dot\eta(0) = w$ is also nontangential between boundary points, and γ and η do not intersect each other at the boundary of M_0. Using the fact that γ can only self-intersect at y_0 finitely many times, see [Kenig and Salo 2013, Lemma 7.2], by choosing W sufficiently small so that the corresponding finitely many tangent vectors of γ and their negatives do not belong to W, we achieve that the geodesics η and γ are distinct and are not reverses of each other.

To conclude the proof, we recall from [do Carmo 1992, p. 248] that

$$\{p \in \gamma([-S_1, S_2]) : p \text{ is conjugate to } \gamma(-S_1) \text{ or } \gamma(S_2)\}$$

is discrete, and since M_0 is compact, it is finite. This completes the proof of the claim. □

When proving Proposition 1.6 in the simplified setting, we shall need some basic facts about nontangential geodesics. These facts are known, see [Dos Santos Ferreira et al. 2020, Section 3], and are presented here for completeness and the convenience of the reader.

Proposition D.2. *Let (M_0, g_0) be a smooth compact Riemannian manifold of dimension $n \geq 2$ with smooth boundary.*

(i) *Let γ be a unit speed non-self-intersecting nontangential geodesic on M_0, and let $y_0 = \gamma(t_0) \in M_0^{\text{int}}$. Then there exists a small neighborhood W of $w_0 = \dot\gamma(t_0)$ in $S_{y_0}M_0$ such that for every $w \in W$, the unit speed geodesic $\gamma_{y_0,w}$ passing through (y_0, w) is nontangential between boundary points and does not have self-intersections.*

(ii) *Let γ and η be unit speed non-self-intersecting nontangential geodesics on M_0 with the only point of intersection $y_0 = \gamma(t_0) = \eta(s_0) \in M_0^{\text{int}}$. Then there exists a small neighborhood W of $w_0 = \dot\gamma(t_0)$ in $S_{y_0}M_0$ such that for every $w \in W$, the unit speed geodesic $\gamma_{y_0,w}$ passing through (y_0, w) is nontangential between boundary points, does not have self-intersections, and intersects η at the point y_0 only.*

Proof. Here we follow [Dos Santos Ferreira et al. 2020, Section 3]. Let us prove (i). Reparametrizing the geodesic γ if necessary, we may assume that $\gamma : [-S_1, S_2] \to M_0$, $0 < S_1, S_2 < \infty$, is such that $\gamma(0) = y_0$ and $\dot\gamma(0) = w_0$. First the property of a geodesic being nontangential is stable under small perturbations of the initial conditions, in view of C^∞-dependence of the geodesic flow on the initial conditions. Assume the contrary: there is a sequence $w_k \to w_0$ in $S_{y_0}M_0$ as $k \to \infty$ such that there are times $t_k < s_k$ when the corresponding geodesic $\gamma_{y_0,w_k} : [-S_1(k), S_2(k)] \to M_0$ with $\gamma_{y_0,w_k}(0) = y_0$, $\dot\gamma_{y_0,w_k}(0) = w_k$ self-intersects:

$$a_k := \gamma_{y_0,w_k}(t_k) = \gamma_{y_0,w_k}(s_k). \tag{D-12}$$

Note that the sequences $-S_1(k)$ and $S_2(k)$ approach $-S_1$ and S_2, respectively, as $k \to \infty$. Therefore, the sequences t_k and s_k are bounded, and passing to subsequences, we may assume that $t_k \to t_0$ and $s_k \to s_0$. Letting $k \to \infty$ in (D-12), we get $\gamma(t_0) = \gamma(s_0)$. Since γ does not have self-intersections we obtain $t_0 = s_0$.

As all geodesics γ_{y_0,w_k} are nontangential, it follows from (D-12) that $a_k \in M_0^{\text{int}}$. As M_0 is compact, it has a positive injectivity radius $\operatorname{Inj}(M_0) > 0$. Here we have extended M_0 to a closed manifold to speak about the injectivity radius and the boundary will not cause any problems as $a_k \in M_0^{\text{int}}$. Now (D-12) implies that

$$s_k \ge t_k + 2\operatorname{Inj}(M_0),$$

and therefore, $s_0 - t_0 \ge 2\operatorname{Inj}(M_0) > 0$, which is a contradiction. Hence, (i) follows.

To prove (ii), first reparametrizing the geodesics γ and η if necessary, we may assume that the map $\gamma : [-S_1, S_2] \to M_0$, $0 < S_1, S_2 < \infty$, is such that $\gamma(0) = y_0$ and $\dot\gamma(0) = w_0$, and $\eta : [-T_1, T_2] \to M_0$, $0 < T_1, T_2 < \infty$, is such that $\eta(0) = y_0$. By (i), there exists a small neighborhood W of w_0 in $S_{y_0}M_0$ such that for every $w \in W$, the unit speed geodesic $\gamma_{y_0,w}$ such that $\gamma_{y_0,w}(0) = y_0$ and $\dot\gamma_{y_0,w}(0) = w$ is nontangential between boundary points and does not have self-intersections. We shall show that the neighborhood W can be made smaller so that every $\gamma_{y_0,w}$ intersects η at the point y_0 only. Let us assume the opposite: there is a sequence $w_k \to w_0$ in $S_{y_0}M_0$ as $k \to \infty$ such that there are times $t_k \ne 0$, $s_k \ne 0$ when the corresponding geodesic γ_{y_0,w_k} intersects η:

$$\gamma_{y_0,w_k}(t_k) = \eta(s_k). \tag{D-13}$$

Note that here we used that γ_{y_0,w_k} and η do not have self-intersections. We also have

$$\gamma_{y_0,w_k}(0) = \eta(0) = y_0. \tag{D-14}$$

Passing to subsequences, we have that $t_k \to t_0$ and $s_k \to s_0$. Thus, it follows from (D-13) that $\gamma(t_0) = \eta(s_0)$, and therefore, as γ and η do not self-intersect and y_0 is the only point of their intersection, we get $t_0 = s_0 = 0$. In view of (D-13) we have

$$\gamma_{y_0,w_k}(t_k) = \eta(s_k) \to \eta(0) = y_0 \in M_0^{\text{int}},$$

and thus, for k sufficiently large, $\gamma_{y_0,w_k}(t_k) = \eta(s_k) \in M_0^{\text{int}}$. This together with (D-14) gives

$$|t_k| > \operatorname{Inj}(M_0) > 0 \quad \text{and} \quad |s_k| > \operatorname{Inj}(M_0) > 0$$

for k sufficiently large, otherwise the geodesics γ_{y_0,w_k} and η would intersect at a geodesic ball centered at y_0, which is a contradiction. Thus, (ii) follows. □

Acknowledgements

We are very grateful to the referee for helpful comments and suggestions. Krupchyk is deeply grateful to Gabriel Paternain for the generous help with the proof of Proposition D.1. Krupchyk would also like to thank Ali Feizmohammadi and Lauri Oksanen for helpful discussions, and Giovanni Alessandrini for pointing out the reference [Alessandrini et al. 2018]. The research of Krupchyk is partially supported by the NSF (DMS 1815922, DMS 2109199). The research of Uhlmann is partially supported by the NSF, a Walker Professorship at the University of Washington, and a Si Yuan Professorship at IAS, HKUST. Part of the work was supported by NSF grant DMS-1440140 while Krupchyk and Uhlmann were in residence at MSRI in Berkeley, California, during the Fall 2019 semester.

References

[Alessandrini 1990] G. Alessandrini, "Singular solutions of elliptic equations and the determination of conductivity by boundary measurements", *J. Differential Equations* **84**:2 (1990), 252–272. MR Zbl

[Alessandrini et al. 2018] G. Alessandrini, M. V. de Hoop, R. Gaburro, and E. Sincich, "Lipschitz stability for a piecewise linear Schrödinger potential from local Cauchy data", *Asymptot. Anal.* **108**:3 (2018), 115–149. MR Zbl

[Anikonov 1978] Y. E. Anikonov, Некоторые методы исследования многомерных обратных задач для дифференциалных уравнений, Nauka, Novosibirsk, Russia, 1978. MR Zbl

[Brown 2001] R. M. Brown, "Recovering the conductivity at the boundary from the Dirichlet to Neumann map: a pointwise result", *J. Inverse Ill-Posed Probl.* **9**:6 (2001), 567–574. MR Zbl

[Brown and Salo 2006] R. M. Brown and M. Salo, "Identifiability at the boundary for first-order terms", *Appl. Anal.* **85**:6-7 (2006), 735–749. MR Zbl

[do Carmo 1992] M. P. do Carmo, *Riemannian geometry*, Birkhäuser, Boston, 1992. MR Zbl

[Caro and Garcia 2017] P. Caro and A. Garcia, "The Calderón problem with corrupted data", *Inverse Problems* **33**:8 (2017), art. id. 085001. MR Zbl

[Caro and Meroño 2020] P. Caro and C. J. Meroño, "The observational limit of wave packets with noisy measurements", *SIAM J. Math. Anal.* **52**:5 (2020), 5196–5212. MR Zbl

[Cârstea and Feizmohammadi 2021] C. I. Cârstea and A. Feizmohammadi, "An inverse boundary value problem for certain anisotropic quasilinear elliptic equations", *J. Differential Equations* **284** (2021), 318–349. MR Zbl

[Cârstea et al. 2019] C. I. Cârstea, G. Nakamura, and M. Vashisth, "Reconstruction for the coefficients of a quasilinear elliptic partial differential equation", *Appl. Math. Lett.* **98** (2019), 121–127. MR Zbl

[Cekić 2017] M. Cekić, "Calderón problem for connections", *Comm. Partial Differential Equations* **42**:11 (2017), 1781–1836. MR Zbl

[Choe et al. 2004] B. R. Choe, H. Koo, and H. Yi, "Projections for harmonic Bergman spaces and applications", *J. Funct. Anal.* **216**:2 (2004), 388–421. MR Zbl

[Dos Santos Ferreira et al. 2009] D. Dos Santos Ferreira, C. E. Kenig, M. Salo, and G. Uhlmann, "Limiting Carleman weights and anisotropic inverse problems", *Invent. Math.* **178**:1 (2009), 119–171. MR Zbl

[Dos Santos Ferreira et al. 2016] D. Dos Santos Ferreira, Y. Kurylev, M. Lassas, and M. Salo, "The Calderón problem in transversally anisotropic geometries", *J. Eur. Math. Soc.* **18**:11 (2016), 2579–2626. MR Zbl

[Dos Santos Ferreira et al. 2020] D. Dos Santos Ferreira, Y. Kurylev, M. Lassas, T. Liimatainen, and M. Salo, "The linearized Calderón problem in transversally anisotropic geometries", *Int. Math. Res. Not.* **2020**:22 (2020), 8729–8765. MR Zbl

[Eskin 2011] G. Eskin, *Lectures on linear partial differential equations*, Grad. Stud. Math. **123**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl

[Feizmohammadi and Oksanen 2020] A. Feizmohammadi and L. Oksanen, "An inverse problem for a semi-linear elliptic equation in Riemannian geometries", *J. Differential Equations* **269**:6 (2020), 4683–4719. MR Zbl

[Feizmohammadi et al. 2021] A. Feizmohammadi, M. Lassas, and L. Oksanen, "Inverse problems for nonlinear hyperbolic equations with disjoint sources and receivers", *Forum Math. Pi* **9** (2021), art. id. e10. MR Zbl

[García and Zhang 2016] A. García and G. Zhang, "Reconstruction from boundary measurements for less regular conductivities", *Inverse Problems* **32**:11 (2016), art. id. 115015. MR Zbl

[Gilbarg and Trudinger 1983] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Grundlehren der Math. Wissenschaften **224**, Springer, 1983. MR Zbl

[Guillarmou 2017] C. Guillarmou, "Lens rigidity for manifolds with hyperbolic trapped sets", *J. Amer. Math. Soc.* **30**:2 (2017), 561–599. MR Zbl

[Guillarmou and Tzou 2011] C. Guillarmou and L. Tzou, "Calderón inverse problem with partial data on Riemann surfaces", *Duke Math. J.* **158**:1 (2011), 83–120. MR Zbl

[Guillarmou et al. 2021] C. Guillarmou, M. Mazzucchelli, and L. Tzou, "Boundary and lens rigidity for non-convex manifolds", *Amer. J. Math.* **143**:2 (2021), 533–575. MR Zbl

[Hervas and Sun 2002] D. Hervas and Z. Sun, "An inverse boundary value problem for quasilinear elliptic equations", *Comm. Partial Differential Equations* **27**:11-12 (2002), 2449–2490. MR Zbl

[Hörmander 1976] L. Hörmander, "The boundary problems of physical geodesy", *Arch. Ration. Mech. Anal.* **62**:1 (1976), 1–52. MR Zbl

[Ilmavirta 2020] J. Ilmavirta, "Geometry of geodesics", preprint, 2020. arXiv 2008.00073

[Isakov and Nachman 1995] V. Isakov and A. I. Nachman, "Global uniqueness for a two-dimensional semilinear elliptic inverse problem", *Trans. Amer. Math. Soc.* **347**:9 (1995), 3375–3390. MR Zbl

[Isakov and Sylvester 1994] V. Isakov and J. Sylvester, "Global uniqueness for a semilinear elliptic inverse problem", *Comm. Pure Appl. Math.* **47**:10 (1994), 1403–1410. MR Zbl

[Isozaki 2004] H. Isozaki, "Inverse spectral problems on hyperbolic manifolds and their applications to inverse boundary value problems in Euclidean space", *Amer. J. Math.* **126**:6 (2004), 1261–1313. MR Zbl

[Kang and Nakamura 2002] K. Kang and G. Nakamura, "Identification of nonlinearity in a conductivity equation via the Dirichlet-to-Neumann map", *Inverse Problems* **18**:4 (2002), 1079–1088. MR Zbl

[Kenig and Salo 2013] C. Kenig and M. Salo, "The Calderón problem with partial data on manifolds and applications", *Anal. PDE* **6**:8 (2013), 2003–2048. MR Zbl

[Kohn and Vogelius 1984] R. Kohn and M. Vogelius, "Determining conductivity by boundary measurements", *Comm. Pure Appl. Math.* **37**:3 (1984), 289–298. MR Zbl

[Krupchyk and Uhlmann 2014] K. Krupchyk and G. Uhlmann, "Uniqueness in an inverse boundary problem for a magnetic Schrödinger operator with a bounded magnetic potential", *Comm. Math. Phys.* **327**:3 (2014), 993–1009. MR Zbl

[Krupchyk and Uhlmann 2018] K. Krupchyk and G. Uhlmann, "Inverse problems for magnetic Schrödinger operators in transversally anisotropic geometries", *Comm. Math. Phys.* **361**:2 (2018), 525–582. MR Zbl

[Krupchyk and Uhlmann 2020a] K. Krupchyk and G. Uhlmann, "Partial data inverse problems for semilinear elliptic equations with gradient nonlinearities", *Math. Res. Lett.* **27**:6 (2020), 1801–1824. MR Zbl

[Krupchyk and Uhlmann 2020b] K. Krupchyk and G. Uhlmann, "A remark on partial data inverse problems for semilinear elliptic equations", *Proc. Amer. Math. Soc.* **148**:2 (2020), 681–685. MR Zbl

[Kurylev et al. 2018] Y. Kurylev, M. Lassas, and G. Uhlmann, "Inverse problems for Lorentzian manifolds and non-linear hyperbolic equations", *Invent. Math.* **212**:3 (2018), 781–857. MR Zbl

[Lai and Zhou 2020] R.-Y. Lai and T. Zhou, "Partial data inverse problems for nonlinear magnetic Schrödinger equations", 2020. To appear in *Math. Res. Lett.* arXiv 2007.02475

[Lassas and Uhlmann 2001] M. Lassas and G. Uhlmann, "On determining a Riemannian manifold from the Dirichlet-to-Neumann map", *Ann. Sci. École Norm. Sup.* (4) **34**:5 (2001), 771–787. MR Zbl

[Lassas et al. 2018] M. Lassas, G. Uhlmann, and Y. Wang, "Inverse problems for semilinear wave equations on Lorentzian manifolds", *Comm. Math. Phys.* **360**:2 (2018), 555–609. MR Zbl

[Lassas et al. 2021a] M. Lassas, T. Liimatainen, Y.-H. Lin, and M. Salo, "Inverse problems for elliptic equations with power type nonlinearities", *J. Math. Pures Appl.* (9) **145** (2021), 44–82. MR Zbl

[Lassas et al. 2021b] M. Lassas, T. Liimatainen, Y.-H. Lin, and M. Salo, "Partial data inverse problems and simultaneous recovery of boundary and coefficients for semilinear elliptic equations", *Rev. Mat. Iberoam.* **37**:4 (2021), 1553–1580. MR Zbl

[Lee and Uhlmann 1989] J. M. Lee and G. Uhlmann, "Determining anisotropic real-analytic conductivities by boundary measurements", *Comm. Pure Appl. Math.* **42**:8 (1989), 1097–1112. MR Zbl

[Muhometov 1977] R. G. Muhometov, "The problem of recovery of a two-dimensional Riemannian metric and integral geometry", *Dokl. Akad. Nauk SSSR* **232**:1 (1977), 32–35. In Russian; translated in *Soviet Math. Dokl.* **18**:1 (1977), 27–31. MR Zbl

[Nakamura et al. 1995] G. Nakamura, Z. Q. Sun, and G. Uhlmann, "Global identifiability for an inverse problem for the Schrödinger equation in a magnetic field", *Math. Ann.* **303**:3 (1995), 377–388. MR Zbl

[Paternain 1999] G. P. Paternain, *Geodesic flows*, Progr. Math. **180**, Birkhäuser, Boston, 1999. MR Zbl

[Petersen 2006] P. Petersen, *Riemannian geometry*, 2nd ed., Grad. Texts in Math. **171**, Springer, 2006. MR Zbl

[Pöschel and Trubowitz 1987] J. Pöschel and E. Trubowitz, *Inverse spectral theory*, Pure Appl. Math. **130**, Academic Press, Boston, 1987. MR Zbl

[Salo 2017] M. Salo, "The Calderón problem and normal forms", preprint, 2017. arXiv 1702.02136

[Salo and Tzou 2009] M. Salo and L. Tzou, "Carleman estimates and inverse problems for Dirac operators", *Math. Ann.* **344**:1 (2009), 161–184. MR Zbl

[Stefanov et al. 2018] P. Stefanov, G. Uhlmann, and A. Vasy, "Inverting the local geodesic X-ray transform on tensors", *J. Anal. Math.* **136**:1 (2018), 151–208. MR Zbl

[Sun 1996] Z. Sun, "On a quasilinear inverse boundary value problem", *Math. Z.* **221**:2 (1996), 293–305. MR Zbl

[Sun 2004] Z. Sun, "Inverse boundary value problems for a class of semilinear elliptic equations", *Adv. Appl. Math.* **32**:4 (2004), 791–800. MR Zbl

[Sun 2010] Z. Sun, "An inverse boundary-value problem for semilinear elliptic equations", *Electron. J. Differential Equations* **2010** (2010), art. id. 37. MR Zbl

[Sun and Uhlmann 1997] Z. Sun and G. Uhlmann, "Inverse problems in quasilinear anisotropic media", *Amer. J. Math.* **119**:4 (1997), 771–797. MR Zbl

[Sylvester and Uhlmann 1987] J. Sylvester and G. Uhlmann, "A global uniqueness theorem for an inverse boundary value problem", *Ann. of Math.* (2) **125**:1 (1987), 153–169. MR Zbl

[Sylvester and Uhlmann 1988] J. Sylvester and G. Uhlmann, "Inverse boundary value problems at the boundary: continuous dependence", *Comm. Pure Appl. Math.* **41**:2 (1988), 197–219. MR Zbl

[Uhlmann and Vasy 2016] G. Uhlmann and A. Vasy, "The inverse problem for the local geodesic ray transform", *Invent. Math.* **205**:1 (2016), 83–120. MR Zbl

Received 2 Feb 2021. Revised 26 Oct 2021. Accepted 4 Feb 2022.

KATYA KRUPCHYK: katya.krupchyk@uci.edu
Department of Mathematics, University of California, Irvine, CA, United States

GUNTHER UHLMANN: gunther@math.washington.edu
Department of Mathematics, University of Washington, Seattle, WA, United States

ANALYSIS AND PDE
Vol. 16 (2023), No. 8, pp. 1869–1884
DOI: 10.2140/apde.2023.16.1869

DISCRETE VELOCITY BOLTZMANN EQUATIONS IN THE PLANE: STATIONARY SOLUTIONS

LEIF ARKERYD AND ANNE NOURI

We prove the existence of stationary mild solutions for normal discrete velocity Boltzmann equations in the plane with no pair of colinear interacting velocities and given ingoing boundary values. We remove an important restriction from a previous paper that all velocities point into the same half-space. A key property is L^1 compactness of integrated collision frequency for a sequence of approximations. This is proven using the Kolmogorov–Riesz theorem, which here replaces the L^1 compactness of velocity averages in the continuous velocity case, not available when the velocities are discrete.

1. Introduction

The Boltzmann equation is the fundamental mathematical model in the kinetic theory of gases. Replacing its continuum of velocities with a discrete set of velocities is a simplification, preserving the essential features of free flow and quadratic collision term. Besides this fundamental aspect, the discrete equations can approximate the Boltzmann equation with any given accuracy [Palczewski et al. 1997; Fainsilber et al. 2006; Mischler 1997], and are thereby useful for approximations and numerics. In the quantum realm they can also be more directly connected to microscopic quasi/particle models. A discrete velocity model of a kinetic gas is a system of partial differential equations having the form,

$$\frac{\partial f_i}{\partial t}(t,z)+v_i\cdot\nabla_z f_i(t,z)=Q_i(f,f)(t,z),\qquad t>0,\ z\in\Omega,\ 1\le i\le p,$$

where $f_i(t,z)$, $1\le i\le p$, are phase space densities at time t, position z and velocities v_i. The spatial domain is Ω. The given discrete velocities are v_i, $1\le i\le p$. For $f=(f_i)_{1\le i\le p}$, the collision operator $Q=(Q_i)_{1\le i\le p}$ with gain part Q^+, loss part Q^-, and collision frequency ν, is given by

$$Q_i(f,f)=\sum_{j,l,m=1}^{p}\Gamma_{ij}^{lm}(f_lf_m-f_if_j)=Q_i^+(f,f)-Q_i^-(f,f),$$

$$Q_i^+(f,f)=\sum_{j,l,m=1}^{p}\Gamma_{ij}^{lm}f_lf_m,\quad Q_i^-(f,f)=f_i\nu_i(f),\quad \nu_i(f)=\sum_{j,l,m=1}^{p}\Gamma_{ij}^{lm}f_j,\quad i=1,\dots,p.$$

The collision coefficients satisfy

$$\Gamma_{ij}^{lm}=\Gamma_{ji}^{lm}=\Gamma_{lm}^{ij}\ge 0. \tag{1-1}$$

MSC2020: 60K35, 82C40, 82C99.
Keywords: stationary Boltzmann equation, discrete coplanar velocities, normal model.

If a collision coefficient Γ_{ij}^{lm} is nonzero, then the conservation laws for momentum and energy,

$$v_i + v_j = v_l + v_m, \quad |v_i|^2 + |v_j|^2 = |v_l|^2 + |v_m|^2, \tag{1-2}$$

are satisfied. We call a pair of velocities (v_i, v_j) interacting if for some $(l, m) \in \{1, \dots, p\}^2$ we have $\Gamma_{ij}^{lm} > 0$. The discrete velocity model (DVM) is called normal (see [Cercignani 1985]) if any solution of the equations

$$\Psi(v_i) + \Psi(v_j) = \Psi(v_l) + \Psi(v_m),$$

where the indices $(i, j; l, m)$ take all possible values satisfying $\Gamma_{ij}^{lm} > 0$, is given by

$$\Psi(v) = a + b \cdot v + c|v|^2$$

for some constants $a, c \in \mathbb{R}$ and $b \in \mathbb{R}^d$. We consider

$$\text{the generic case of normal coplanar velocity sets} \\ \text{with no pair of colinear interacting velocities } (v_i, v_j). \tag{1-3}$$

The case is generic. Indeed, consider a normal velocity set such that, for some interacting velocities (v_i, v_j), v_i and v_j are colinear. Then there exists an arbitrary small vector v_0 such that the velocity set $(v_i + v_0)_{1 \le i \le p}$ is normal and with no colinear interacting velocities. The paper considers stationary solutions to normal coplanar discrete velocity models satisfying (1-3), in a strictly convex bounded open subset $\Omega \subset \mathbb{R}^2$, with C^2 boundary $\partial\Omega$ and given boundary inflow. Denote by $n(Z)$ the inward normal to $Z \in \partial\Omega$. Denote the v_i-ingoing (resp. v_i-outgoing) part of the boundary by

$$\partial\Omega_i^+ = \{Z \in \partial\Omega : v_i \cdot n(Z) > 0\} \quad (\text{resp. } \partial\Omega_i^- = \{Z \in \partial\Omega : v_i \cdot n(Z) < 0\}).$$

Let

$$s_i^+(z) = \inf\{s > 0 : z - s v_i \in \partial\Omega_i^+\}, \quad s_i^-(z) = \inf\{s > 0 : z + s v_i \in \partial\Omega_i^-\}, \quad z \in \Omega.$$

Write

$$z_i^+(z) = z - s_i^+(z) v_i \quad (\text{resp. } z_i^-(z) = z + s_i^-(z) v_i) \tag{1-4}$$

for the ingoing (resp. outgoing) point on $\partial\Omega$ of the characteristics through z in direction v_i.

The stationary boundary value problem

$$v_i \cdot \nabla f_i(z) = Q_i(f, f)(z), \quad z \in \Omega, \tag{1-5}$$

$$f_i(z) = f_{bi}(z), \quad z \in \partial\Omega_i^+, \ 1 \le i \le p, \tag{1-6}$$

is considered in L^1 in one of the following equivalent forms [DiPerna and Lions 1989]: the exponential multiplier form,

$$\begin{aligned} f_i(z) = {} & f_{bi}(z_i^+(z)) e^{-\int_0^{s_i^+(z)} \nu_i(f)(z_i^+(z) + s v_i)\, ds} \\ & + \int_0^{s_i^+(z)} Q_i^+(f, f)(z_i^+(z) + s v_i) e^{-\int_s^{s_i^+(z)} \nu_i(f)(z_i^+(z) + r v_i)\, dr}\, ds, \quad \text{a.a. } z \in \Omega, \ 1 \le i \le p, \end{aligned} \tag{1-7}$$

the mild form,

$$f_i(z) = f_{bi}(z_i^+(z)) + \int_0^{s_i^+(z)} Q_i(f,f)(z_i^+(z)+sv_i)\,ds, \quad \text{a.a. } z\in\Omega,\ 1\le i\le p, \tag{1-8}$$

the renormalized form,

$$v_i\cdot\nabla\ln(1+f_i)(z) = \frac{Q_i(f,f)}{1+f_i}(z), \quad z\in\Omega, \qquad f_i(z)=f_{bi}(z), \quad z\in\partial\Omega_i^+,\ 1\le i\le p, \tag{1-9}$$

in the sense of distributions. Denote by $L^1_+(\Omega)$ the set of nonnegative integrable functions on Ω. For a distribution function $f=(f_i)_{1\le i\le p}$, define its entropy (resp. entropy dissipation) by

$$\sum_{i=1}^p \int_\Omega f_i \ln f_i(z)\,dz, \quad \biggl(\text{resp. } \sum_{i,j,l,m=1}^p \Gamma_{ij}^{lm}\int_\Omega (f_l f_m - f_i f_j)\ln\frac{f_l f_m}{f_i f_j}(z)\,dz\biggr).$$

The main result of the paper is:

Theorem 1.1. *Consider a coplanar normal discrete velocity model and a nonnegative ingoing boundary value f_b with mass and entropy inflows bounded,*

$$\int_{\partial\Omega_i^+} v_i\cdot n(z)\, f_{bi}(1+\ln f_{bi})(z)\,d\sigma(z) < +\infty, \quad 1\le i\le p.$$

For the boundary value problem (1-5)–(1-6) *satisfying* (1-3), *there exists a stationary mild solution in $(L^1_+(\Omega))^p$ with finite mass and entropy-dissipation.*

Given $i\in\{1,\dots,p\}$, if $\Gamma_{ij}^{lm}=0$ for all j, l and m, then f_i equals its ingoing boundary value, and the rest of the system can be solved separately. Such i's are not present in the following discussion. Most mathematical results for stationary discrete velocity models of the Boltzmann equation have been obtained in one space dimension. An overview is given in [Płatkowski and Illner 1988]. Half-space problems [Bernhoff 2012] and weak shock waves [Bernhoff and Bobylev 2007] for discrete velocity models have also been studied. A discussion of normal discrete velocity models, i.e., conserving nothing but mass, momentum and energy, can be found in [Bobylev et al. 2010]. In two dimensions, special classes of solutions to the Broadwell model are given in [Bobylev and Toscani 1996; Bobylev 1996; Ilyin 2014]. The Broadwell model, not included in the present results, is a four-velocity model, with $v_1+v_2=v_3+v_4=0$ and v_1, v_3 orthogonal. A detailed study of the stationary Broadwell equation in a rectangle with comparison to a Carleman-like system is given in [Bobylev 1996], as well as a discussion of (in-)compressibility aspects. A main result in [Cercignani et al. 1988] is the existence of continuous solutions to the two-dimensional stationary Broadwell model with continuous boundary data for a rectangle. The paper [Arkeryd and Nouri 2020b] solves that problem in an L^1-setting. The proof uses in an essential way the constancy of the sums f_1+f_2 and f_3+f_4 along characteristics, which no longer holds in the present paper. For every normal model, there is a priori control of entropy dissipation, mass and entropy flows through the boundary. From there, the main difficulties are to prove that for a sequence of approximations, weak L^1 compactness holds and the limit of the collision operator equals the collision operator of the limit. In [Arkeryd and Nouri 2020a], weak L^1 compactness of a sequence of approximations was obtained with assumption (1-3) together with the assumption that all velocities v_i point out into the same half-plane. In this paper we keep assumption (1-3), remove the second assumption and provide a new proof of weak L^1 compactness of approximations using (1-3). Assumption (1-3) is also crucial for proving L^1 compactness of the integrated collision frequencies, which is important for the convergence procedure. Our paper also differs from [Arkeryd and Nouri 2020a] in the limit procedure. The frame of the limit procedure in that paper is the splitting into "good" and "bad" characteristics following the approach in our earlier stationary continuous velocity papers [Arkeryd and Nouri 1995; 1999]. Here we have instead utilized sub- and supersolutions used in the classical evolutionary frame for renormalized solutions to the Boltzmann equation [DiPerna and Lions 1989]. For the continuous velocity evolutionary Boltzmann equation, the compactness properties of the collision frequency use in an essential way the averaging lemma, which is not available for the discrete velocity Boltzmann model. In the present paper, the compactness properties are proven by the Kolmogorov–Riesz theorem. Also the argument used in the stationary paper [Arkeryd and Nouri 1995] in the continuous velocity case for obtaining control of entropy, hence weak L^1 compactness of a sequence of approximations from the control of entropy dissipation, does not work in a discrete velocity case because the number of velocities is finite. The proof starts in Section 2 from bounded approximations. In Section 3, L^1 compactness properties of the approximations are proven. Section 4 is devoted to the proof of Theorem 1.1.

2. Approximations

Denote by $\mathbb{N}^* = \mathbb{N}\setminus\{0\}$ and by $a \wedge b$ the minimum of two real numbers a and b. Let μ_α be a smooth mollifier in $\mathbb{R}^2$ with support in the ball centered at the origin of radius α. Outside the boundary the function to be convolved with μ_α, is continued in the normal direction by its boundary value. Let $\tilde{\mu}_k$ be a smooth mollifier on $\partial\Omega$ in a ball of radius $1/k$. Define

$$f_{bi}^k = \Big(f_{bi}(\,\cdot\,)\wedge\frac{k}{2}\Big)*\tilde{\mu}_k, \quad 1\le i\le p,\ k\in\mathbb{N}^*.$$

The lemma introduces a primary approximated boundary value problem with damping and convolutions.

Lemma 2.1. *For any $\alpha > 0$ and $k\in\mathbb{N}^*$, there is a solution $F^{\alpha,k}\in(L^1_+(\Omega))^p$ to*

$$\alpha F_i^{\alpha,k}+v_i\cdot\nabla F_i^{\alpha,k}=\sum_{j,l,m=1}^{p}\Gamma_{ij}^{lm}\Big(\frac{F_l^{\alpha,k}}{1+F_l^{\alpha,k}/k}\,\frac{F_m^{\alpha,k}*\mu_\alpha}{1+F_m^{\alpha,k}*\mu_\alpha/k}-\frac{F_i^{\alpha,k}}{1+F_i^{\alpha,k}/k}\,\frac{F_j^{\alpha,k}*\mu_\alpha}{1+F_j^{\alpha,k}*\mu_\alpha/k}\Big), \tag{2-1}$$

$$F_i^{\alpha,k}(z)=f_{bi}^k(z), \quad z\in\partial\Omega_i^+,\ 1\le i\le p. \tag{2-2}$$

Proof of Lemma 2.1. For a proof of Lemma 2.1 we refer to the second section in [Arkeryd and Nouri 2020a]. Let $k\in\mathbb{N}^*$ be given. Each component of $F^{\alpha,k}$ is bounded by a multiple of k^2. Therefore $(F^{\alpha,k})_{\alpha\in]0,1[}$ is weakly compact in $(L^1(\Omega))^p$. For a subsequence, the convergence is strong in $(L^1(\Omega))^p$ as stated in the following lemma. □

Lemma 2.2. *There is a sequence $(\beta(q))_{q\in\mathbb{N}}$ tending to zero when $q\to+\infty$ and a function $F^k\in L^1$ such that $(F^{\beta(q),k})_{q\in\mathbb{N}}$ strongly converges in $(L^1(\Omega))^p$ to F^k when $q\to+\infty$.*

Proof of Lemma 2.2. For a proof of Lemma 2.2 we refer to Lemma 3.1 in [Arkeryd and Nouri 2020a]. Define

$$Q_i^{+k}=\sum_{j,l,m=1}^{p}\Gamma_{ij}^{lm}\frac{F_l^k}{1+F_l^k/k}\,\frac{F_m^k}{1+F_m^k/k},\quad \nu_i^k=\sum_{j,l,m=1}^{p}\Gamma_{ij}^{lm}\frac{F_j^k}{(1+F_i^k/k)(1+F_j^k/k)},\tag{2-3}$$

$$Q_i^k=Q_i^{+k}-F_i^k\nu_i^k,\quad 1\le i\le p,\tag{2-4}$$

and denote by $\widetilde{D}_k$ the entropy production term of the approximations,

$$\widetilde{D}_k=\sum_{i,j,l,m=1}^{p}\Gamma_{ij}^{lm}\biggl(\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}-\frac{F_i^k}{1+F_i^k/k}\frac{F_j^k}{1+F_j^k/k}\biggr)\ln\frac{F_l^kF_m^k(1+F_i^k/k)(1+F_j^k/k)}{(1+F_l^k/k)(1+F_m^k/k)F_i^kF_j^k}.\tag{2-5}$$

All throughout the paper, c_b denotes constants that may vary from line to line but is independent of parameters tending to $+\infty$ or to zero. □

Lemma 2.3. *F^k is a nonnegative solution to*

$$v_i\cdot\nabla F_i^k=Q_i^{+k}-F_i^k\nu_i^k,\tag{2-6}$$

$$F_i^k(z)=f_{bi}^k(z),\quad z\in\partial\Omega_i^+,\ 1\le i\le p.\tag{2-7}$$

Solutions $(F^k)_{k\in\mathbb{N}^*}$ to (2-6)–(2-7) *have mass and entropy dissipation bounded from above uniformly with respect to k. Moreover their outgoing flows at the boundary are controlled as follows*:

$$\sum_{i=1}^{p}\int_{\partial\Omega_i^-,F_i^k\le k}|v_i\cdot n(Z)|F_i^k\ln F_i^k(Z)\,d\sigma(Z)+\ln\frac{k}{2}\int_{\partial\Omega_i^-,F_i^k>k}|v_i\cdot n(Z)|F_i^k\,d\sigma(Z)\le c_b.\tag{2-8}$$

Proof of Lemma 2.3. Passing to the limit when $q\to+\infty$ in (2-1)–(2-2) written for $F^{\beta(q),k}$, implies that F^k is a solution in $(L_+^1(\Omega))^p$ to (2-6)–(2-7). For a proof of the rest of Lemma 2.3, we refer to Lemma 3.2 in [Arkeryd and Nouri 2020a]. □

3. On compactness of sequences of approximations

This section is devoted to proving L^1 compactness properties of the approximations. In Proposition 3.1, weak L^1 compactness of $(F^k)_{k\in\mathbb{N}^*}$ is proven. Lemma 3.2 splits Ω into a set of i-characteristics with arbitrary small measure and its complement, where both the approximations and their integrated collision frequencies are bounded. In Lemma 3.3, the strong L^1 compactness of integrated collision frequency is proven.

Proposition 3.1. *The sequence $(F^k)_{k\in\mathbb{N}^*}$ solution to* (2-6)–(2-7) *is weakly compact in L^1.*

Proof of Proposition 3.1. By Lemma 2.3, $(F^k)_{k\in\mathbb{N}^*}$ is uniformly bounded in $(L^1(\Omega))^p$. Given (2-8) and the bound

$$F_i^k(z)\le F_i^k(z+s_i^-(z)v_i)\exp\biggl(\Gamma\sum_{j\in J_i}\int_{-s_i^+(z)}^{s_i^-(z)}F_j(z+rv_i)\,dr\biggr),\quad z\in\Omega,\ i\in\{1,\dots,p\},\tag{3-1}$$

on F^k, the weak L^1 compactness of $(F^k)_{k\in\mathbb{N}^*}$ will follow from the uniform boundedness in $L^\infty(\partial\Omega_i^+)$ of

$$\left(\int_0^{s_i^-(Z)} F_j(Z+rv_i)\,dr\right)_{j\in J_i,k\in\mathbb{N}}, \tag{3-2}$$

where J_i denotes the set $\{j\in\{1,\dots,p\} : (v_i, v_j)$ are interacting velocities$\}$. By (1-3), there exists $\eta>0$ such that, for all interacting velocities (v_i, v_j),

$$|\sin(\widehat{v_i, v_j})|>\eta. \tag{3-3}$$

Let $i\in\{1,\dots,p\}$ and $Z\in\partial\Omega_i^+$. Multiply the equation satisfied by F_j^k by $(v_i^\perp\cdot v_j)/|v_i|$ and integrate it on one of the half domains defined by the segment $[Z, Z+s_i^-(Z)v_i]$. Summing over $j\in\{1,\dots,p\}$ implies that

$$\sum_{j=1}^p \sin^2(\widehat{v_i, v_j})\int_0^{s_i^-(Z)} F_j^k(Z+sv_i)\,ds\le c_b,\quad Z\in\partial\Omega_i^+. \tag{3-4}$$

Together with (3-3), this leads to the control of (3-2). □

Recall the exponential multiplier form for the approximations $(F^k)_{k\in\mathbb{N}^*}$,

$$F_i^k(z)=f_{bi}^k(z_i^+(z))e^{-\int_{-s_i^+(z)}^0 \nu_i^k(z+sv_i)\,ds} + \int_{-s_i^+(z)}^0 Q_i^{+k}(z+sv_i)e^{-\int_s^0 \nu_i^k(F^k)(z+rv_i)\,dr}\,ds,\quad \text{a.a. } z\in\Omega,\ 1\le i\le p, \tag{3-5}$$

with ν_i^k and Q_i^{+k} defined in (2-3). An i-characteristics is a segment of points $[Z-s_i^+(Z)v_i, Z]$, where $Z\in\partial\Omega_i^-$. Define $\Gamma=\max_{i,j,l,m}\Gamma_{ij}^{lm}$.

Lemma 3.2. *For $i\in\{1,\dots,p\}$, $k\in\mathbb{N}^*$ and $\epsilon>0$, there is a subset $\Omega_i^{k,\epsilon}$ of i-characteristics of Ω with measure smaller than $c_b\epsilon$ such that for any $z\in\Omega\setminus\Omega_i^{k,\epsilon}$*

$$F_i^k(z)\le\frac{1}{\epsilon^2}\exp\left(\frac{p\Gamma}{\epsilon^2}\right),\quad \int_{-s_i^+(z)}^{s_i^-(z)}\nu_i^k(z+sv_i)\,ds\le\frac{p\Gamma}{\epsilon^2}. \tag{3-6}$$

Proof of Lemma 3.2. By the strict convexity of Ω, there are for every $i\in\{1,\dots,p\}$ two points of $\partial\Omega$, denoted by $\widetilde{Z}_i$ and $\bar{Z}_i$, such that

$$v_i\cdot n(\widetilde{Z}_i)=v_i\cdot n(\bar{Z}_i)=0.$$

Let $\tilde{l}_i$ (resp. $\bar{l}_i$) be the largest boundary arc included in $\partial\Omega_i^-$ with one endpoint $\widetilde{Z}_i$ (resp. $\bar{Z}_i$) such that

$$-\epsilon\le v_i\cdot n(Z)\le 0,\quad Z\in\tilde{l}_i\cup\bar{l}_i. \tag{3-7}$$

Let J_i be the subset of $\{1,\dots,p\}$ such that,

$$\text{for some } (l,m)\in\{1,\dots,p\}^2,\quad \Gamma_{ij}^{lm}>0,\quad j\in J_i. \tag{3-8}$$

It follows from the exponential form of F_i^k that

$$F_i^k(z)\le F_i^k(z+s_i^-(z)v_i)\exp\left(\Gamma\sum_{j\in J_i}\int_{-s_i^+(z)}^{s_i^-(z)} F_j(z+rv_i)\,dr\right),\quad z\in\Omega. \tag{3-9}$$

The boundedness of the mass flow of $(F_i^k)_{k\in\mathbb{N}^*}$ across $\partial\Omega_i^-$ is

$$\int_{\partial\Omega_i^-} |v_i\cdot n(Z)|F_i^k(Z)\,d\sigma(Z)\le c_b,\quad k\in\mathbb{N}^*. \tag{3-10}$$

It follows from (3-7)–(3-10) that the measure of the set

$$\Big\{Z\in\partial\Omega_i^-\cap\tilde{l}_i^c\cap\bar{l}_i^c : F_i^k(Z)>\frac{1}{\epsilon^2}\Big\}$$

is smaller than $c_b\epsilon$. The boundedness of the mass of $(F_j^k)_{k\in\mathbb{N}^*}$ can be written

$$\int_\Omega F_j^k(z)\,dz=\int_{\partial\Omega_i^-}|v_i\cdot n(Z)|\bigg(\int_{-s_i^+(Z)}^0 F_j^k(Z+rv_i)\,dr\bigg)\,d\sigma(Z)\le c_b,\quad j\in J_i.$$

Hence the measure of the set

$$\bigg\{Z\in\partial\Omega_i^-\cap\tilde{l}_i^c\cap\bar{l}_i^c : \int_{-s_i^+(Z)}^0 F_j^k(Z+rv_i)\,dr>\frac{1}{\epsilon^2}\bigg\},\quad j\in J_i,$$

is smaller than $c_b\epsilon$. Consequently, the measure of the set of $Z\in\partial\Omega_i^-\cap\tilde{l}_i^c\cap\bar{l}_i^c$ outside of which

$$F_i^k(Z)\le\frac{1}{\epsilon^2}\quad\text{and}\quad\int_{-s_i^+(Z)}^0 F_j^k(Z+rv_i)\,dr\le\frac{1}{\epsilon^2},\quad j\in J_i,$$

is bounded by $c_b\epsilon$. Together with (3-9), this implies that the measure of the complement of the set of $Z\in\partial\Omega_i^-$ such that

$$F_i^k(z)\le\frac{1}{\epsilon^2}\exp\bigg(\frac{p\Gamma}{\epsilon^2}\bigg)\quad\text{and}\quad\int_{-s_i^+(z)}^{s_i^-(z)}\nu_i^k(z+rv_i)\,dr\le\frac{p\Gamma}{\epsilon^2}$$

for $z=Z+sv_i,\ s\in[-s_i^+(Z),0]$, is bounded by $c_b\epsilon$. With it $c_b\epsilon$ is a bound for the measure of the complement, denoted by $\Omega_i^{k,\epsilon}$, of the set of i-characteristics in Ω such that for all points z on the i-characteristics, (3-6) holds. □

Given $i\in\{1,\dots,p\}$ and $\epsilon>0$, let $\chi_i^{k,\epsilon}$ denote the characteristic function of the complement of $\Omega_i^{k,\epsilon}$. The following lemma proves the compactness in $L^1(\Omega)$ of the k-sequence of integrated collision frequencies.

Lemma 3.3. *The sequences*

$$\bigg(\int_{-s_i^+(z)}^0 \nu_i^k(z+sv_i)\,ds\bigg)_{k\in\mathbb{N}^*},\quad 1\le i\le p,$$

are strongly compact in $L^1(\Omega)$.

Proof of Lemma 3.3. Take $\Gamma_{ij}^{lm}>0$. By (1-3), v_i and v_j span $\mathbb{R}^2$. Denote by (a,b) the corresponding coordinate system, (a^-,a^+) defined by

$$a^-=\min\{a\in\mathbb{R}:(a,b)\in\Omega\text{ for some }b\},\quad a^+=\max\{a\in\mathbb{R}:(a,b)\in\Omega\text{ for some }b\},$$

and by D the Jacobian of the change of variables $z \to (a, b)$. The uniform bound for the mass of $(F^k)_{k\in\mathbb{N}^*}$ proven in Lemma 2.3, implies

$$\left(\int_\Omega \int_{-s_i^+(z)}^0 \nu_i^k(z+sv_i)\,ds\,dz\right)_{k\in\mathbb{N}^*}$$

is bounded in L^1 uniformly with respect to k. Indeed, for some $(b^-(a), b^+(a))$, $a \in [a^-, a^+]$,

$$\begin{aligned}\int_\Omega \int_{-s_i^+(z)}^0 F_j^k(z+sv_i)\,ds\,dz &= D\int_{a^-}^{a^+}\int_{b^-(a)}^{b^+(a)}\int_{-s_i^+(bv_j)}^{a} F_j^k(bv_j+sv_i)\,ds\,db\,da\\ &\le D\int_{a^-}^{a^+}\int_{b^-(a)}^{b^+(a)}\int_{-s_i^+(bv_j)}^{s_i^-(bv_j)} F_j^k(bv_j+sv_i)\,ds\,db\,da\\ &\le c\int_\Omega F_j^k(z)dz, \qquad j\in J_i.\end{aligned}$$

By the Kolmogorov–Riesz theorem [Kolmogorov 1931; Riesz 1933], the compactness of

$$\left(\int_{-s_i^+(z)}^0 \nu_i^k(z+sv_i)\,ds\right)_{k\in\mathbb{N}^*}$$

will follow from its translational equicontinuity in $L^1(\Omega)$. Equicontinuity in the direction v_i, and in the direction v_j with the mild form (1-8) for F_j^k, come naturally. Here the assumption (1-3) becomes crucial. The sequence

$$\left(\int_{-s_i^+(z)}^0 F_j^k(z+sv_i)\,ds\right)_{k\in\mathbb{N}^*}, \qquad j\in J_i, \tag{3-11}$$

is translationally equicontinuous in the v_i-direction. Indeed, $s_i^+(z+hv_i) = s_i^+(z)+h$ so that, denoting by $I(0,h)$ the interval with endpoints 0 and h and using the uniform bound on the mass of $(F_j^k)_{k\in\mathbb{N}^*}$,

$$\begin{aligned}\int_\Omega\left|\int_{-s_i^+(z+hv_i)}^0 F_j^k(z+hv_i+sv_i)\,ds - \int_{-s_i^+(z)}^0 F_j^k(z+sv_i)\,ds\right|dz &= \int_\Omega\int_{s\in I(0,h)} F_j^k(z+sv_i)\,ds\,dz\\ &\le c|h|.\end{aligned}$$

Let us prove the translational equicontinuity of (3-11) in the v_j-direction. By the weak L^1 compactness of $(F_j^k)_{k\in\mathbb{N}^*}$, it is sufficient to prove the translational equicontinuity in the v_j-direction of

$$\left(\int_{s_i^+(z)}^0 \chi_j^{k,\epsilon}F_j^k(z+sv_i)\,ds\right)_{k\in\mathbb{N}^*}.$$

Expressing $F_j^k(z+hv_j+sv_i)$ (resp. $F_j^k(z+sv_i)$) as integral along its v_j-characteristics, it holds that

$$\left|\int_{-s_i^+(z+hv_j)}^0 \chi_j^{k,\epsilon}F_j^k(z+hv_j+sv_i)\,ds - \int_{-s_i^+(z)}^0 \chi_j^{k,\epsilon}F_j^k(z+sv_i)\,ds\right| \le |A_{ij}^k(z,h)|+|B_{ij}^k(z,h)|,$$

where

$$A_{ij}^k(z,h) = \int_{-s_i^+(z+hv_j)}^0 \chi_j^{k,\epsilon} f_{bj}^k(z_j^+(z+hv_j+sv_i))\,ds - \int_{-s_i^+(z)}^0 \chi_j^{k,\epsilon} f_{bj}^k(z_j^+(z+sv_i))\,ds,$$

and

$$\begin{aligned}B^k_{ij}(z,h)=\int_{-s_i^+(z+hv_j)}^{0}\int_{-s_j^+(z+hv_j+sv_i)}^{0}&\chi_j^{k,\epsilon}\,Q_j^k(z+hv_j+sv_i+rv_j)\,dr\,ds\\&-\int_{-s_i^+(z)}^{0}\int_{-s_j^+(z+sv_i)}^{0}\chi_j^{k,\epsilon}\,Q_j^k(z+sv_i+rv_j)\,dr\,ds,\end{aligned}$$

with Q_i^k defined in (2-3). Denote by $(z_j^+(z_i^+(z)), z_j^+(z_i^+(z+hv_j)))$ the boundary arc with endpoints $z_j^+(z_i^+(z))$ and $z_j^+(z_i^+(z+hv_j))$ and of length tending to zero with h. Performing the change of variables $s\to Z=z_j^+(z+hv_j+sv_i)$ (resp. $s\to Z=z_j^+(z+sv_i)$) in the first (resp. second) term of $A^k_{ij}(z,h)$, and using that the sequence $(f^k_{bi})_{k\in\mathbb{N}^*}$ is bounded by f_{bi}, it holds that

$$\lim_{h\to 0}\int_\Omega |A^k_{ij}(z,h)|\,dz=0, \tag{3-12}$$

uniformly with respect to k. Moreover, for some $\omega_h(z)\subset\Omega$ of measure or order $|h|$ uniformly with respect to $z\in\Omega$,

$$B^k_{ij}(z,h)=\int_{\omega_h(z)}\chi_j^{k,\epsilon}\,Q_j^k(Z)\,dZ. \tag{3-13}$$

The sequence $(\chi_j^{k,\epsilon}Q_j^k)_{k\in\mathbb{N}^*}$ is weakly compact in L^1. Indeed,

$$\begin{aligned}\chi_j^{k,\epsilon}Q_j^k&\le\frac{1}{\ln\Lambda}\widetilde{D}_k+\Gamma\Lambda\Big(\sum_{i\in J_j}F_i^k\Big)(\chi_j^{k,\epsilon}F_j^k)\\&\le\frac{1}{\ln\Lambda}\widetilde{D}_k+\frac{\Gamma\Lambda}{\epsilon^2}\exp\Big(\frac{p\Gamma}{\epsilon^2}\Big)\Big(\sum_{i\in J_j}F_i^k\Big),\qquad\Lambda>1,\end{aligned} \tag{3-14}$$

with $(\widetilde{D}_k)_{k\in\mathbb{N}^*}$ uniformly bounded in L^1 and $(F_i^k)_{k\in\mathbb{N}^*}$ weakly compact in L^1. Hence,

$$\lim_{h\to 0}\int_\Omega |B^k_{ij}(z,h)|\,dz=0,\qquad\text{uniformly with respect to } k.$$ □

4. The passage to the limit in the approximations

Let f be the weak L^1 limit of a subsequence of the solutions $(F^k)_{k\in\mathbb{N}^*}$ to (2-6)–(2-7), still denoted by $(F^k)_{k\in\mathbb{N}^*}$. For proving that f is a mild solution of (1-5)–(1-6), it is sufficient to prove that, for any $\eta>0$ and $i\in\{1,\dots,p\}$, there is a set X_i^η of i-characteristics with complementary set of measure smaller than $c\eta$, such that

$$\begin{aligned}\int_\Omega\varphi\chi_i^\eta f_i(z)\,dz=\int_\Omega&\varphi\chi_i^\eta f_{bi}(z_i^+(z))\,dz\\&+\int_\Omega\int_{-s_i^+(z)}^0(\varphi\chi_i^\eta Q_i(f,f)+\chi_i^\eta f_i v_i\cdot\nabla\varphi)(z+sv_i)\,ds\,dz,\qquad\varphi\in C^1(\bar\Omega),\end{aligned} \tag{4-1}$$

where χ_i^η denotes the characteristic function of X_i^η. Define the set X_i^η as follows. For every $\epsilon>0$, pass to the limit when $k\to+\infty$ in

$$\chi_i^{k,\epsilon}F_i^k(z)\le\chi_i^{k,\epsilon}F_i^k(z_i^-(z))\,\exp\Big(\int_{-s_i^+(z)}^{s_i^-(z)}\nu_i^k(z+sv_i)\,ds\Big),\qquad\text{a.a. } z\in\Omega,\ k\in\mathbb{N}^*, \tag{4-2}$$

and use the weak L^1 compactness of $(\chi_i^{k,\epsilon} F_i^k)_{k\in\mathbb{N}^*}$, the weak L^1 compactness and the uniform boundedness in L^∞ of $(\chi_i^{k,\epsilon} F_i^k(z_i^-(z)))_{k\in\mathbb{N}^*}$, and the strong L^1 compactness of

$$\left(\int_{-s_i^+(z)}^{s_i^-(z)} \nu_i^k(z+s v_i)\,ds\right)_{k\in\mathbb{N}^*}.$$

It implies

$$F_i^\epsilon(z) \le F_i^\epsilon(z_i^-(z)) \exp\left(\int_{-s_i^+(z)}^{s_i^-(z)} \nu_i(f)(z+s v_i)\,ds\right), \quad \text{a.a. } z\in\Omega,\ \epsilon\in\,]0,1[,$$

where F_i^ϵ is the limit of a subsequence of $(\chi_i^{k,\epsilon} F_i^k)_{k\in\mathbb{N}^*}$ and $\nu_i(f)=\sum_{j,l,m=1}^p \Gamma_{ij}^{lm} f_j$. By the monotonicity in ϵ of $(F^\epsilon)_{\epsilon\in]0,1[}$ (resp. $(F^\epsilon(z_i^-(z)))_{\epsilon\in]0,1[}$) and the uniform boundedness of their masses, it holds that

$$f_i(z) \le f_i(z_i^-(z)) \exp\left(\int_{-s_i^+(z)}^{s_i^-(z)} \nu_i(f)(z+s v_i)\,ds\right), \quad \text{a.a. } z\in\Omega.$$

From here the proof follows the lines of the proof of Lemma 3.2, so that given $\eta>0$, there is a set X_i^η of i-characteristics, with complementary set of measure smaller than $c\eta$, such that

$$f_i(z) \le \frac{1}{\eta} e^{p\Gamma/\eta} \quad \text{and} \quad \int_{-s_i^+(z)}^{s_i^-(z)} \nu_i(f)(z+s v_i)\,ds \le \frac{p\Gamma}{\eta}, \quad \text{a.a. } z\in X_\eta. \tag{4-3}$$

Denote by $C_+^1(\bar\Omega)$ the subspace of nonnegative functions of $C^1(\bar\Omega)$.

Lemma 4.1. *The function f is a subsolution of* (1-5)–(1-6), *i.e.*,

$$\int_\Omega \varphi \chi_i^\eta f_i(z)\,dz \le \int_\Omega \varphi f_{bi}(z_i^+(z))\,dz + \int_\Omega \int_{-s_i^+(z)}^0 \chi_i^\eta f_i\, v_i\cdot\nabla\varphi(z+s v_i)\,ds\,dz \\ + \int_\Omega \int_{-s_i^+(z)}^0 \varphi Q_i(f,f)(z+s v_i)\,ds\,dz, \quad 1\le i\le p,\ \varphi\in C_+^1(\bar\Omega). \tag{4-4}$$

Proof of Lemma 4.1. Let $i\in\{1,\dots,p\}$ and $\varphi\in C_+^1(\bar\Omega)$ be given. Write the mild form of $\varphi\chi_i^\eta\chi_i^{k,\epsilon}F_i^k$ and integrate it on Ω. This yields

$$\int_\Omega \varphi\chi_i^\eta\chi_i^{k,\epsilon} F_i^k(z)\,dz = \int_\Omega \varphi\chi_i^\eta\chi_i^{k,\epsilon} f_{bi}^k(z_i^+(z))\,dz + \int_\Omega\int_{-s_i^+(z)}^0 \chi_i^\eta\chi_i^{k,\epsilon} F_i^k v_i\cdot\nabla\varphi(z+s v_i)\,ds\,dz \\ + \int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta\chi_i^{k,\epsilon}(Q_i^{+k}-F_i^k\nu_i^k)(z+s v_i)\,ds\,dz. \tag{4-5}$$

By the weak L^1 compactness of $(F_i^k)_{k\in\mathbb{N}^*}$ and the linearity with respect to $\chi_i^{k,\epsilon}F_i^k$ of the first line of (4-5), its passage to the limit when $k\to+\infty$ is straightforward. Let us pass to the limit when $k\to+\infty$ in any term of the loss term of (4-5), denoted by $\Gamma_{ij}^{lm}L^k$, where

$$L^k := \int_\Omega \chi_i^\eta\chi_i^{k,\epsilon}(z) \int_{-s_i^+(z)}^0 \varphi \frac{F_i^k}{1+F_i^k/k}\,\frac{F_j^k}{1+F_j^k/k}(z+s v_i)\,ds\,dz, \quad j\in J_i, \tag{4-6}$$

and J_i is defined in (3-8). By integration by parts, L_k equals

$$\begin{aligned}\int_\Omega \int_{-s_i^+(z)}^0 \chi_i^\eta \chi_i^{k,\epsilon}(\varphi(Q_i^{+k} - F_i^k \nu_i^k) + (v_i\cdot\nabla\varphi)F_i^k)(z+sv_i)\\ \times\left(\int_s^0 \chi_i^{k,\epsilon}\frac{F_j^k}{(1+F_i^k/k)(1+F_j^k/k)}(z+rv_i)\,dr\right)ds\,dz\\ +\int_\Omega \chi_i^\eta\chi_i^{k,\epsilon}\varphi\frac{f_{bi}^k}{1+f_{bi}^k/k}(z_i^+(z))\int_{-s_i^+(z)}^0\frac{F_j^k}{1+F_j^k/k}(z+sv_i)\,ds\,dz.\end{aligned}\tag{4-7}$$

Denote by (a,b) the coordinate system in the (v_i,v_j) basis, $(a^-,a^+)\in\mathbb{R}^2$ and $(b^-(a),b^+(a))\in\mathbb{R}^2$ for every $a\in\,]a^-,a^+[$, such that

$$\Omega=\{av_i+bv_j : a\in\,]a^-,a^+[,\ b\in\,]b^-(a),b^+(a)[\}.\tag{4-8}$$

The first term in L^k can be written as $\int_{a^-}^{a^+} l^k(a)\,da$ with l^k defined as

$$\begin{aligned}l^k(a)=\int_{b^-(a)}^{b^+(a)}\int_{-s_i(bv_j)}^{a}\chi_i^\eta\chi_i^{k,\epsilon}\big(\varphi(Q_i^{+k}-F_i^k\nu_i^k)+(v_i\cdot\nabla\varphi)F_i^k\big)(sv_i+bv_j)\\ \times\left(\int_s^a\chi_i^{k,\epsilon}\frac{F_j^k}{(1+F_i^k/k)(1+F_j^k/k)}(rv_i+bv_j)\,dr\right)ds\,db.\end{aligned}\tag{4-9}$$

For each rational number a, the sequence of functions

$$(b,s)\in[b^-(a),b^+(a)]\times[-s_i^+(bv_j),a]\to\chi_i^\eta\chi_i^{k,\epsilon}\big(\varphi(Q_i^{+k}-F_i^k\nu_i^k)+(v_i\cdot\nabla\varphi)F_i^k\big)(sv_i+bv_j)$$

is weakly compact in L^1, whereas

$$(b,s)\to\int_s^a\chi_i^{k,\epsilon}\frac{F_j^k}{(1+F_i^k/k)(1+F_j^k/k)}(rv_i+bv_j)\,dr$$

is by Lemma 3.3 strongly compact in L^1, and by Lemma 3.2 uniformly bounded in L^∞. The convergence follows for any rational number a. With a diagonal process, there is a subsequence of (l^k), still denoted by (l^k), converging for any rational a. Moreover,

$$\lim_{h\to0}(l^k(a+h)-l^k(a))=0,\tag{4-10}$$

uniformly with respect to k and a, by the weak L^1 compactness of

$$\big(\chi_i^\eta\chi_i^{k,\epsilon}(\varphi(Q_i^{+k}-F_i^k\nu_i^k)+(v_i\cdot\nabla\varphi)F_i^k)\big)_{k\in\mathbb{N}^*}\quad\text{and}\quad(F_j^k)_{k\in\mathbb{N}^*}.$$

Thus (l^k) is a uniform converging sequence on $[a^-,a^+]$. The second term in L^k can be treated analogously, $(\chi_i^{k,\epsilon}f_{bi}^k)_{k\in\mathbb{N}^*}$ being uniformly bounded in L^∞. The convergence follows. In order to determine the limit of L^k when $k\to+\infty$, note that

$$\chi_i^\eta\chi_i^{k,\epsilon}(\varphi(Q_i^{+k}-F_i^k\nu_i^k)+(v_i\cdot\nabla\varphi)F_i^k)=v_i\cdot\nabla(\chi_i^\eta\chi_i^{k,\epsilon}\varphi F_i^k),$$

which weakly converges in L^1 to $v_i \cdot \nabla(\chi_i^\eta \varphi F_i^\epsilon)$ when $k \to +\infty$. Hence

$$\lim_{k\to+\infty} L^k = \int_\Omega \int_{-s_i^+(z)}^0 v_i \cdot \nabla(\chi_i^\eta \varphi F_i^\epsilon)(z+sv_i)\left(\int_s^0 f_j(z+rv_i)\,dr\right) ds\,dz \\ + \int_\Omega \chi_i^\eta \varphi f_{bi}(z_i^+(z))\left(\int_{-s_i^+(z)}^0 f_j(z+sv_i)\,ds\right) dz.$$

By a backwards integration by parts,

$$\lim_{k\to+\infty} L^k = \int_\Omega \int_{-s_i^+(z)}^0 \varphi \chi_i^\eta F_i^\epsilon f_j(z+sv_i)\,ds\,dz. \tag{4-11}$$

In order to prove (4-4), let us prove that each

$$\Gamma_{ij}^{lm} \int_\Omega \int_{-s_i^+(z)}^0 \varphi \chi_i^\eta \chi_i^{k,\epsilon} \frac{F_l^k}{1+F_l^k/k} \frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz, \qquad j \in J_i, \tag{4-12}$$

term from Q_i^{+k} in (4-5) converges when $k \to +\infty$ to a limit smaller than

$$\Gamma_{ij}^{lm} \int_\Omega \int_{-s_i^+(z)}^0 \varphi \chi_i^\eta F_l^{\epsilon'} f_m(z+sv_i)\,ds\,dz + \alpha(\epsilon'), \quad \epsilon' \in\,]0,1[, \text{ with } \lim_{\epsilon'\to 0} \alpha(\epsilon') = 0. \tag{4-13}$$

Take $\Gamma_{ij}^{lm} = 1$, $j \in J_i$, for simplicity. Let $(\mu_{1/n})_{n\in\mathbb{N}^*}$ be the sequence of mollifiers defined at the beginning of Section 2 for $\alpha = 1/n$, and split (4-12) into

$$\int_\Omega \int_{-s_i^+(z)}^0 \varphi(\chi_i^\eta * \mu_{1/n}) \chi_l^{k,\epsilon'} \chi_i^{k,\epsilon} \frac{F_l^k}{1+F_l^k/k} \frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz \\ + \int_\Omega \int_{-s_i^+(z)}^0 \varphi(\chi_i^\eta * \mu_{1/n})(1-\chi_l^{k,\epsilon'}) \chi_i^{k,\epsilon} \frac{F_l^k}{1+F_l^k/k} \frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz \\ + \int_\Omega \int_{-s_i^+(z)}^0 \varphi(\chi_i^\eta - (\chi_i^\eta * \mu_{1/n})) \chi_i^{k,\epsilon} \frac{F_l^k}{1+F_l^k/k} \frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz \\ \le \int_\Omega \int_{-s_i^+(z)}^0 \varphi(\chi_i^\eta * \mu_{1/n}) \chi_l^{k,\epsilon'} \frac{F_l^k}{1+F_l^k/k} \frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz \\ + \frac{c}{\ln\Lambda} + \frac{c\Lambda}{\epsilon^2} e^{p\Gamma/\epsilon^2} \sum_{j\in J_i} \left(\int_{\Omega_l^{k,\epsilon'}} F_j^k(z)\,dz + \int_\Omega \varphi |\chi_i^\eta - (\chi_i^\eta * \mu_{1/n})| F_j^k(z)\,dz \right) \\ \le \int_\Omega \int_{-s_i^+(z)}^0 \varphi(\chi_i^\eta * \mu_{1/n}) \chi_l^{k,\epsilon'} \frac{F_l^k}{1+F_l^k/k} \frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz \\ + \frac{c}{\ln\Lambda} + \frac{c\Lambda}{\epsilon^2} e^{p\Gamma/\epsilon^2} \left(\Lambda'\epsilon' + \frac{1}{\ln\Lambda'} + \frac{1}{\ln(k/2)} + \widetilde{\Lambda} \| \chi_i^\eta - (\chi_i^\eta * \mu_{1/n}) \|_{L^1} + \frac{1}{\ln\widetilde{\Lambda}} \right) \quad \text{by (2-8)–(3-1),} \\ \Lambda > 1, \quad \Lambda' > 1, \quad \widetilde{\Lambda} > 1, \quad \epsilon' > 0. \tag{4-14}$$

Denote by D the Jacobian of the change of variables $z \to (a, b)$. For some smooth function A, and any integrable function g,

$$\begin{aligned}\int_\Omega \int_{-s_i^+(z)}^0 g(z+sv_i)ds dz &= D \int_{b^-}^{b^+} \int_{a^-(b)}^{a^+(b)} \int_{-s_i^+(bv_j)}^{a} g(sv_i + bv_j) ds\, da\, db \\ &= D \int_{b^-}^{b^+} \int_{-s_i^+(bv_j)}^{a^+(b)} (a^+(b) - \max\{a^-(b), s\}) g(sv_i + bv_j)\, ds\, db \\ &= \int_\Omega A(\alpha, \gamma) g(\alpha v_l + \gamma v_m)\, d\alpha\, d\gamma.\end{aligned}$$

Hence,

$$\begin{aligned}\lim_{k\to+\infty} \iint_{-s_i^+(z)}^0 \varphi(\chi_i^\eta * \mu_{1/n}) \chi_l^{k,\epsilon'} \frac{F_l^k}{1+F_l^k/k} \frac{F_m^k}{1+F_m^k/k} (z+sv_i)\, ds\, dz \\ = \int_\Omega \int_{-s_i^+(z)}^0 \varphi(\chi_i^\eta * \mu_{1/n}) F_l^{\epsilon'} f_m(z+sv_i)\, ds\, dz, \quad \epsilon' \in]0,1[. \quad (4\text{-}15)\end{aligned}$$

For $\widetilde{\Lambda}$ large enough, pass to the limit when $k \to +\infty$ and $n \to +\infty$ in (4-14). Up to subsequences, the weak L^1 limits F_i^ϵ and $F_i^{\epsilon'}$ of $(\chi_i^{k,\epsilon} F_i^k)_{k\in\mathbb{N}^*}$ and $(\chi_i^{k,\epsilon'} F_i^k)_{k\in\mathbb{N}^*}$ when $k \to +\infty$ satisfy

$$\begin{aligned}\int_\Omega \varphi\chi_i^\eta F_i^\epsilon(z)\, dz \le \int_\Omega \varphi\chi_i^\eta f_{bi}^k(z_i^+(z))\, dz + \int_\Omega \int_{-s_i^+(z)}^0 \chi_i^\eta F_i^\epsilon v_i \cdot \nabla\varphi(z+sv_i)\, ds\, dz \\ + \int_\Omega \int_{-s_i^+(z)}^0 \varphi\chi_i^\eta (Q_i^+(F^{\epsilon'}, f) - F_i^\epsilon \nu_i(f))(z+sv_i)\, ds\, dz \\ + \frac{c}{\ln\Lambda} + \frac{c\Lambda}{\epsilon^2} e^{p\Gamma/\epsilon^2}\left(\Lambda'\epsilon' + \frac{1}{\ln\Lambda'}\right), \quad (\epsilon,\epsilon') \in]0,1[^2,\ \Lambda > 1,\ \Lambda' > 1. \quad (4\text{-}16)\end{aligned}$$

Choose Λ large enough, ϵ small enough, Λ' large enough, ϵ' small enough, in this order. The passage to the limit when $\epsilon \to 0$ and $\epsilon' \to 0$ in (4-16) results from the monotone convergence theorem, the family $(F^\epsilon)_{\epsilon\in]0,1[}$ being nondecreasing, with mass uniformly bounded, together with the mass of $(\chi_i^\eta Q_i^+(F^{\epsilon'}, f))_{\epsilon'\in]0,1[}$ and $(\chi_i^\eta F_i^{\epsilon'} \nu_i(f))_{\epsilon'\in]0,1[}$. Consequently, (4-4) holds. □

Lemma 4.2. *The function f is a solution to* (1-5)–(1-6).

Proof of Lemma 4.2. For proving Lemma 4.2, it remains to prove that

$$\begin{aligned}\int_\Omega \varphi\chi_i^\eta f_i(z)\, dz \ge \int_\Omega \varphi\chi_i^\eta f_{bi}(z_i^+(z))\, dz + \int_\Omega \int_{-s_i^+(z)}^0 \chi_i^\eta f_i v_i \cdot \nabla\varphi(z+sv_i)\, ds\, dz \\ + \int_\Omega \int_{-s_i^+(z)}^0 \varphi\chi_i^\eta Q_i(f,f)(z+sv_i)\, ds\, dz, \quad 1 \le i \le p,\ \varphi \in C_+^1(\bar{\Omega}). \quad (4\text{-}17)\end{aligned}$$

For $\beta > 0$, start from the equation for $\varphi\chi_i^\eta F_i^k$ written in renormalized form,

$$\begin{aligned}\beta^{-1}\varphi\chi_i^\eta \ln(1+\beta F_i^k)(z) - \beta^{-1}\varphi\chi_i^\eta \ln(1+\beta f_{bi}^k)(z_i^+(z)) \\ + \int_{-s_i^+(z)}^0 \beta^{-1}\chi_i^\eta \ln(1+\beta F_i^k)\, v_i \cdot \nabla\varphi(z+sv_i)\, ds = \int_{-s_i^+(z)}^0 \frac{\varphi\chi_i^\eta (Q_i^{+k} - F_i^k \nu_i^k)}{1+\beta F_i^k}(z+sv_i)\, ds. \quad (4\text{-}18)\end{aligned}$$

It holds

$$\beta^{-1}\ln(1+\beta x) < x, \quad \beta \in]0,1[\qquad \text{and} \qquad \lim_{\beta\to 0}\beta^{-1}\ln(1+\beta x) = x, \quad x > 0.$$

Hence in weak L^1 the sequence $(\beta^{-1}\ln(1+\beta F_i^k))_{k\in\mathbb{N}^*}$ converges modulo a subsequence to a function $F^\beta \le f$ when $k \to +\infty$. The mass of the limit increases to the mass of f, when $\beta \to 0$. This gives in the final limit $\beta \to 0$ for the left-hand side of (4-18)

$$\varphi\chi_i^\eta f_i(z) - \varphi\chi_i^\eta f_{bi}(z_i^+(z)) - \int_{-s_i^+(z)}^0 \chi_i^\eta f_i\, v_i \cdot \nabla\varphi(z+sv_i)\,ds. \tag{4-19}$$

Using analogous arguments as for the limit of the loss term in Lemma 4.1, it holds that

$$\lim_{k\to+\infty}\Gamma_{ij}^{lm}\int_\Omega\int_{-s_i^+(z)}^0 \frac{\varphi\chi_i^\eta F_i^k F_j^k}{1+\beta F_i^k}(z+sv_i)\,ds\,dz \\ = \Gamma_{ij}^{lm}\int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta\Big(\text{weak } L^1 \lim_{k\to+\infty}\frac{F_i^k}{1+\beta F_i^k}\Big) f_j(z+sv_i)\,ds\,dz, \quad j \in J_i.$$

But

$$\text{weak } L^1 \lim_{k\to+\infty}\frac{F_i^k}{1+\beta F_i^k} \le \text{weak } L^1 \lim_{k\to+\infty} F_i^k,$$

and

$$\int_\Omega \text{weak } L^1 \lim_{k\to+\infty}\frac{F_i^k}{1+\beta F_i^k}(z)\,dz \quad \text{increases to} \quad \int_\Omega \text{weak } L^1 \lim_{k\to+\infty} F_i^k(z)\,dz$$

when $\beta \to 0$. Hence

$$\lim_{\beta\to 0}\lim_{k\to+\infty}\Gamma_{ij}^{lm}\int_\Omega\int_{-s_i^+(z)}^0 \frac{\varphi\chi_i^\eta F_i^k F_j^k}{1+\beta F_i^k}(z+sv_i)\,ds\,dz = \Gamma_{ij}^{lm}\int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta f_i f_j(z+sv_i)\,ds\,dz. \tag{4-20}$$

For the gain term and any $(l,m) \in \{1,\dots,p\}^2$ such that $\Gamma_{ij}^{lm} > 0$ for some $j \in \{1,\dots,p\}$,

$$\int_\Omega\int_{-s_i^+(z)}^0 \frac{\varphi\chi_i^\eta}{1+\beta F_i^k}\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz \\ \ge \int_\Omega\int_{-s_i^+(z)}^0 \frac{\varphi\chi_i^\eta\chi_l^{k,\epsilon}}{1+\beta F_i^k}\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz \\ = \int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta\chi_l^{k,\epsilon}\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz \\ - \int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta\chi_l^{k,\epsilon}\frac{\beta F_i^k}{1+\beta F_i^k}\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz \\ \ge \int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta\chi_l^{k,\epsilon}\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz \\ - c\Lambda\sum_{j\in J_i}\int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta\chi_l^{k,\epsilon}\frac{\beta(F_i^k)^2F_j^k}{1+\beta F_i^k}(z+sv_i)\,ds\,dz - \frac{c}{\ln\Lambda}, \quad \Lambda > 1,\ \epsilon \in]0,1[. \tag{4-21}$$

It holds

$$\lim_{k\to+\infty}\int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta\chi_l^{k,\epsilon}\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz=\int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta F_l^\epsilon f_m^k(z+sv_i)\,ds\,dz. \tag{4-22}$$

Choose Λ large enough and split the domain of integration of every $j\in J_i$ term in (4-21) into

$$\{F_i^k\le\Lambda'\}\cup\left\{F_i^k>\Lambda' \text{ and } F_i^kF_j^k>\widetilde{\Lambda}\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}\right\}$$
$$\cup\left\{F_i^k>\Lambda' \text{ and } F_i^kF_j^k\le\widetilde{\Lambda}\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}\right\},\quad \Lambda'>1,\ \widetilde{\Lambda}>1.$$

It holds that

$$\int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta\chi_l^{k,\epsilon}\frac{\beta(F_i^k)^2F_j^k}{1+\beta F_i^k}(z+sv_i)\,ds\,dz$$
$$\le c\left(\beta(\Lambda')^2+\frac{1}{\ln\widetilde{\Lambda}}+\frac{\widetilde{\Lambda}}{\epsilon^2}e^{p\Gamma/\epsilon^2}\int_{F_i^k>\Lambda'}F_m^k(z)\,dz\right),\quad \beta\in\,]0,1[,\ \Lambda'>0,\ \widetilde{\Lambda}>1. \tag{4-23}$$

The last term in (4-23) tends to zero when $\widetilde{\Lambda}\to+\infty$, $\Lambda'\to+\infty$, $\beta\to 0$ in this order, uniformly with respect to k. Consequently,

$$\lim_{\beta\to0}\lim_{k\to+\infty}\int_\Omega\int_{-s_i^+(z)}^0 \frac{\varphi\chi_i^\eta}{1+\beta F_i^k}\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz\ge\int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta F_l^\epsilon f_m(z+sv_i)\,ds\,dz.$$

This holds for every $\epsilon>0$. Hence

$$\lim_{\beta\to0}\lim_{k\to+\infty}\int_\Omega\int_{-s_i^+(z)}^0 \frac{\varphi\chi_i^\eta}{1+\beta F_i^k}\frac{F_l^k}{1+F_l^k/k}\frac{F_m^k}{1+F_m^k/k}(z+sv_i)\,ds\,dz\ge\int_\Omega\int_{-s_i^+(z)}^0 \varphi\chi_i^\eta f_l f_m(z+sv_i)\,ds\,dz.$$

And so, (4-17) holds. Together with (4-4), this proves (4-1). □

References

[Arkeryd and Nouri 1995] L. Arkeryd and A. Nouri, "A compactness result related to the stationary Boltzmann equation in a slab, with applications to the existence theory", *Indiana Univ. Math. J.* **44**:3 (1995), 815–839. MR Zbl

[Arkeryd and Nouri 1999] L. Arkeryd and A. Nouri, "On the stationary Povzner equation in $\mathbb{R}^n$", *J. Math. Kyoto Univ.* **39**:1 (1999), 115–153. MR Zbl

[Arkeryd and Nouri 2020a] L. Arkeryd and A. Nouri, "On stationary solutions to normal, coplanar discrete Boltzmann equation models", *Commun. Math. Sci.* **18**:8 (2020), 2215–2234. MR Zbl

[Arkeryd and Nouri 2020b] L. Arkeryd and A. Nouri, "Stationary solutions to the two-dimensional Broadwell model", *Doc. Math.* **25** (2020), 2023–2048. MR Zbl

[Bernhoff 2012] N. Bernhoff, "Half-space problem for the discrete Boltzmann equation: condensing vapor flow in the presence of a non-condensable gas", *J. Stat. Phys.* **147**:6 (2012), 1156–1181. MR Zbl

[Bernhoff and Bobylev 2007] N. Bernhoff and A. Bobylev, "Weak shock waves for the general discrete velocity model of the Boltzmann equation", *Commun. Math. Sci.* **5**:4 (2007), 815–832. MR Zbl

[Bobylev 1996] A. V. Bobylev, "Exact solutions of discrete kinetic models and stationary problems for the plane Broadwell model", *Math. Methods Appl. Sci.* **19**:10 (1996), 825–845. MR Zbl

[Bobylev and Toscani 1996] A. V. Bobylev and G. Toscani, "Two-dimensional half-space problems for the Broadwell discrete velocity model", *Contin. Mech. Thermodyn.* **8**:5 (1996), 257–274. MR Zbl

[Bobylev et al. 2010] A. Bobylev, M. Vinerean, and Å. Windfäll, "Discrete velocity models of the Boltzmann equation and conservation laws", *Kinet. Relat. Models* **3**:1 (2010), 35–58. MR Zbl

[Cercignani 1985] C. Cercignani, "Sur des critères d'existence globale en théorie cinétique discrète", *C. R. Acad. Sci. Paris Sér. I Math.* **301**:3 (1985), 89–92. MR Zbl

[Cercignani et al. 1988] C. Cercignani, R. Illner, and M. Shinbrot, "A boundary value problem for the two-dimensional Broadwell model", *Comm. Math. Phys.* **114**:4 (1988), 687–698. MR Zbl

[DiPerna and Lions 1989] R. J. DiPerna and P.-L. Lions, "On the Cauchy problem for Boltzmann equations: global existence and weak stability", *Ann. of Math.* (2) **130**:2 (1989), 321–366. MR Zbl

[Fainsilber et al. 2006] L. Fainsilber, P. Kurlberg, and B. Wennberg, "Lattice points on circles and discrete velocity models for the Boltzmann equation", *SIAM J. Math. Anal.* **37**:6 (2006), 1903–1922. MR Zbl

[Ilyin 2014] O. V. Ilyin, "Symmetries, the current function, and exact solutions for Broadwell's two-dimensional stationary kinetic model", *Teoret. Mat. Fiz* **179**:3 (2014), 350–359. In Russian; translated in *Theoret. Math. Phys.* **179**:3 (2014), 679–688. Zbl

[Kolmogorov 1931] A. Kolmogoroff, "Über Kompaktheit der Funktionenmengen bei der Konvergenz im Mittel", *Nachr. Ges. Wiss. Göttingen Fachgruppe* **9** (1931), 60–63. Zbl

[Mischler 1997] S. Mischler, "Convergence of discrete-velocity schemes for the Boltzmann equation", *Arch. Ration. Mech. Anal.* **140**:1 (1997), 53–77. MR Zbl

[Palczewski et al. 1997] A. Palczewski, J. Schneider, and A. V. Bobylev, "A consistency result for a discrete-velocity model of the Boltzmann equation", *SIAM J. Numer. Anal.* **34**:5 (1997), 1865–1883. MR Zbl

[Płatkowski and Illner 1988] T. Płatkowski and R. Illner, "Discrete velocity models of the Boltzmann equation: a survey on the mathematical aspects of the theory", *SIAM Rev.* **30**:2 (1988), 213–255. MR Zbl

[Riesz 1933] M. Riesz, "Sur les ensembles compacts de fonctions sommables", *Acta Sci. Math.* (*Szeged*) **6** (1933), 136–142. Zbl

Received 28 May 2021. Revised 29 Nov 2021. Accepted 14 Feb 2022.

LEIF ARKERYD: arkeryd@chalmers.se
Mathematical Sciences, Göteborg, Sweden

ANNE NOURI: anne.nouri@univ-amu.fr
Aix-Marseille University, CNRS, I2M UMR 7373, Marseille, France

ANALYSIS AND PDE
Vol. 16 (2023), No. 8, pp. 1885–1954
DOI: 10.2140/apde.2023.16.1885

BOSONS IN A DOUBLE WELL: TWO-MODE APPROXIMATION AND FLUCTUATIONS

ALESSANDRO OLGIATI, NICOLAS ROUGERIE AND DOMINIQUE SPEHNER

We study the ground state for many interacting bosons in a double-well potential, in a joint limit where the particle number and the distance between the potential wells both go to infinity. Two single-particle orbitals (one for each well) are macroscopically occupied, and we are concerned with deriving the corresponding effective Bose–Hubbard Hamiltonian. We prove an energy expansion, including the two-mode Bose–Hubbard energy and two independent Bogoliubov corrections (one for each potential well), and a variance bound for the number of particles falling inside each potential well. The latter is a signature of a correlated ground state in that it violates the central limit theorem.

1. Introduction

The mathematical study of macroscopic limits of many-body quantum mechanics has made sizable progress in recent years [Ammari 2013; Benedikter et al. 2016; Golse 2016; Lieb et al. 2005; Rougerie 2014; 2015; 2020; Schlein 2013; Spohn 1991]. The situation that is most understood is the mean-field limit of many weak interparticle interactions. Following Boltzmann's original picture of molecular chaos [Golse 2016; Spohn 1980; 1991; Gallagher et al. 2013; Mischler 2011; Pulvirenti and Simonella 2016; Jabin 2014], an independent particles picture emerges, wherein statistical properties of the system are computed from a nonlinear PDE. This is based on interparticle correlations being negligible at leading order, which, for bosonic systems, comes about through the macroscopic occupancy of a single one-body state (orbital, mode).

MSC2020: primary 35Q40, 81V73, 82B10; secondary 47A75, 49S05.
Keywords: many-body quantum mechanics, Bose–Einstein condensation, Bose–Hubbard model, Bogoliubov theory.

In this paper we consider a particular example where, by contrast, correlations play a leading role, through the occupation of two one-body states. Namely, we consider the mean-field limit of a large bosonic system in a symmetric double-well potential. In the joint limit $N \to \infty$, $L \to \infty$ (large particle number, large interwell separation) there is one macroscopically occupied one-body state (orbital) for each well. In a previous work [Rougerie and Spehner 2018], two of us have shown that, when the tunneling energy across the potential barrier is $o(N^{-1})$, the ground state of the N-body Hamiltonian H_N exhibits strong interparticle correlations, in the sense that the variance of the particle number in each well is much smaller than $\sqrt{N}$ (the central limit theorem does not hold).

Here we extend this result to cases where the tunneling energy goes like $N^{-\delta}$ with any $\delta > 0$. This in particular includes the much more intricate case where $\delta < 1$ and the tunneling energy thus cannot be neglected as in [Rougerie and Spehner 2018]. We also prove that the ground state energy of H_N is close to the ground state energy of a simpler effective Bose–Hubbard Hamiltonian. Our energy estimates include the contributions of order $O(1)$ described by a generalized Bogoliubov Hamiltonian, which we show to be given by the sum of the Bogoliubov energies associated to each well, up to errors $o(1)$.

The main feature of the symmetric double well situation is the fact that the N-body state of particles that macroscopically occupy the two main orbitals is in general nontrivial. This is to be compared with the case of complete Bose–Einstein condensation in a single orbital, in which the energy of the condensate is a purely one-body quantity, obtained from the ground state of a suitable nonlinear Schrödinger (NLS) equation. We note that our system, although two modes are occupied to the leading order, is physically very different from a two-component Bose–Einstein condensate [Michelangeli and Olgiati 2017; Anapolitanos et al. 2017; Michelangeli et al. 2019], in which two distinct bosonic species macroscopically occupy one mode each. Rather, it is closer to the case of a single-species fragmented condensate [Dimonte et al. 2021].

The effective theory for our double-well system is obtained by projecting the full Hamiltonian on the subspace spanned by the two appropriate modes (one for each well, identified via NLS theory). Such a projection is known in the physics literature as the two-mode approximation. After some further simplifications this leads to the two-mode Bose–Hubbard Hamiltonian

$$H_{\mathrm{BH}} = \frac{T}{2}(a_1^\dagger a_2 + a_2^\dagger a_1) + g(a_1^\dagger a_1^\dagger a_1 a_1 + a_2^\dagger a_2^\dagger a_2 a_2), \tag{1-1}$$

with $a_j^\dagger$, a_j the standard bosonic creation/annihilation operators associated with the two modes. The first term describes hopping of particles through the double-well's energy barrier, with $T < 0$ the tunneling energy. The second term (with $g > 0$ an effective coupling constant) is the pair interaction energy of particles in each well.

We aim at deriving the above from the full many-body Schrödinger Hamiltonian for N bosons in mean-field scaling ($N \to \infty$, λ fixed)

$$H_N := \sum_{j=1}^{N} (-\Delta_j + V_{\mathrm{DW}}(x_j)) + \frac{\lambda}{N-1} \sum_{1 \leqslant i < j \leqslant N} w(x_i - x_j) \tag{1-2}$$

acting on the Hilbert space ($d = 1, 2, 3$ is the spatial dimension)

$$\mathfrak{H}^N := \bigotimes_{\mathrm{sym}}^{N} L^2(\mathbb{R}^d) \simeq L^2_{\mathrm{sym}}(\mathbb{R}^{dN}). \tag{1-3}$$

Here $V_{\rm DW}$ and w are, respectively, the double-well external potential and the repulsive pair-interaction potential (precise assumptions will be stated below). We study the ground-state problem: the lowest eigenvalue and associated eigenfunction of H_N.

The main new feature that we tackle is that $V_{\rm DW}$ is chosen to depend on a large parameter L in the manner

$$V_{\rm DW}(x) := \min(|x-x_L|^s, |x+x_L|^s), \quad s \geqslant 2,\ |x_L| = \frac{L}{2}. \tag{1-4}$$

This is a simple model for a symmetric trap with two global minima at $x = \pm x_L$. In the limit $L \to \infty$ both the distance between the minima and the height of the in-between energy barrier diverge. As a consequence, the mean-field Hartree energy functional obtained in the standard way by testing with an iid ansatz (pure Bose–Einstein condensate)

$$\mathcal{E}^{\rm H}[u] := \frac{1}{N}\langle u^{\otimes N}|H_N|u^{\otimes N}\rangle \tag{1-5}$$

has two orthogonal low-lying energy states, denoted by u_+, u_- (u_+ being the ground state). Their energies are separated by a tunneling term

$$T = T(L) \xrightarrow[L\to\infty]{} 0.$$

All other energy modes are separated from u_+, u_- by an energy gap independent of L. This picture is mathematically vindicated by semiclassical methods [Dimassi and Sjöstrand 1999; Helffer 1988]. For the model at hand we refer to [Olgiati and Rougerie 2021], whose estimates we use as an input in the sequel. One can show that

$$u_1 := \frac{u_+ + u_-}{\sqrt{2}}, \quad u_2 := \frac{u_+ - u_-}{\sqrt{2}} \tag{1-6}$$

are well-localized in one potential well each. These are the modes entering the Bose–Hubbard Hamiltonian (1-1). If we denote by P the orthogonal projection onto the subspace spanned by u_+, u_- (or equivalently u_1, u_2), the Bose–Hubbard description basically amounts to restricting all available one-body states to $PL^2(\mathbb{R}^d)$

$$H_{\rm BH} \simeq (P)^{\otimes N} H_N (P)^{\otimes N} - E_0 \tag{1-7}$$

acting on $\bigotimes_{\rm sym}^N (PL^2(\mathbb{R}^d))$. Here E_0 is a mean-field energy reference, and the appropriate choice of g in (1-1) is

$$g = \frac{\lambda}{2(N-1)} \iint_{\mathbb{R}^d\times\mathbb{R}^d} |u_1(x)|^2 w(x-y)|u_1(y)|^2\,dx\,dy.$$

The tunneling energy T is essentially the gap between the Hartree energies of u_+ and u_-, which goes to 0 superexponentially fast when $L \to \infty$ (see below).

A salient feature of the Bose–Hubbard ground state is that it satisfies[1]

$$\left\langle \left(a_j^\dagger a_j - \frac{N}{2}\right)^2 \right\rangle_{\rm BH} \ll N, \quad j = 1, 2, \tag{1-8}$$

[1] $\langle\,\cdot\,\rangle_{\rm BH}$ denotes expectation in the Bose–Hubbard ground state.

in the limit $N \to \infty$, $L \to \infty$, where $a_j^\dagger a_j$ is the operator counting the number of particles occupying the mode $j = 1, 2$. This is *number squeezing*, a signature of strong correlations. Actually, the problem being invariant under the exchange of the modes[2] we certainly have

$$\langle a_j^\dagger a_j \rangle_{\mathrm{BH}} = \frac{N}{2}, \quad j = 1, 2.$$

Thus what (1-8) says is that the standard deviation from this mean does not satisfy the central limit theorem. Hence the events "particle n lives in the j-th well", $n = 1, \dots, N$, are measurably *not* independent. Such an estimate is governed by energy estimates precise to order $o(1)$ in the limit $N \to \infty$, $L \to \infty$. In the usual mean-field limit with a single well (L fixed), an energy correction of order $O(1)$ arises, due to quantum fluctuations [Seiringer 2011; Grech and Seiringer 2013; Dereziński and Napiórkowski 2014; Lewin et al. 2015; Nam and Seiringer 2015; Boccato et al. 2019; 2020]. This also occurs in our setting, due to the (small) occupancy of modes orthogonal to u_1, u_2. This is conveniently described by a Bogoliubov Hamiltonian, which is quadratic in creation/annihilation operators. The latter has a ground-state energy E^{Bog}, which is of order $O(1)$ in the joint limit (we will give more precise definitions below). Denoting by

$$E(N) := \inf \sigma(H_N), \quad E_{\mathrm{BH}} := \inf \sigma(H_{\mathrm{BH}}) \tag{1-9}$$

the lowest eigenvalues of the full Hamiltonian and its two-mode approximation respectively, our main energy estimate takes the form

$$|E(N) - E_0 - E_{\mathrm{BH}} - E^{\mathrm{Bog}}| \to 0 \tag{1-10}$$

in the limit $N \to \infty$, $T \to 0$, provided $0 < \lambda$ is small enough (independently of N and T). This implies number squeezing

$$\left\langle \left(a_j^\dagger a_j - \frac{N}{2} \right)^2 \right\rangle_{\Psi_{\mathrm{gs}}} \ll N, \quad j = 1, 2, \tag{1-11}$$

in the true ground state Ψ_{gs} of (1-2) ($\langle \cdot \rangle_{\Psi_{\mathrm{gs}}}$ denotes expectation in this state). To avoid some technicalities we assume that λ is fixed and $T = N^{-\delta}$ with some arbitrary $\delta > 0$. In essence the above results, however, only require $N \to \infty$, $T \ll \lambda$. They are thus optimal in the sense that the opposite regime $N \to \infty$, $T \gtrsim \lambda$ (for fixed λ this implies $L \lesssim 1$, see (2-13)) corresponds to the usual mean-field situation for a fixed potential, where a central limit theorem holds [Rademacher and Schlein 2019]. This is called "Rabi regime" in the physics literature; see [Rougerie and Spehner 2018, Section 1.3] for more details. The ground state of the system is expected to be approximated by a Bose–Einstein condensate

$$\Psi_{\mathrm{gs}} \approx u_+^{\otimes N} \approx \left(\frac{u_1 + u_2}{\sqrt{2}} \right)^{\otimes N}, \tag{1-12}$$

with a variance of order N for the number of particles in the modes u_1 and u_2. The aforementioned techniques dealing with the single-well problem allow to prove the (appropriately rigorous version of the) first approximation in (1-12), with u_+ the Hartree ground state. When T, L are fixed however, there does not seem to be a sharp mathematical way to define the privileged modes u_1, u_2 and actually prove the second approximation in (1-12) in a well-defined scaling regime.

[2] Equivalent to a reflection around the double-well's peak.

In [Rougerie and Spehner 2018], estimates (1-10)–(1-11) have been proved (essentially) in the restricted regime $T \ll N^{-1}$. When $T \gtrsim N^{-1}$, the tunneling contribution to the energy becomes relevant for the order of precision we aim at, and we cannot just separate the contributions of each well as in [Rougerie and Spehner 2018]. Instead we prove that the two wells are coupled only via the dynamics in the two-mode subspace, which we isolate from quantum fluctuations. We need to monitor both the number of excited particles and the variance of the occupation numbers of the low-lying modes. Roughly speaking the former is controlled by the Bogoliubov Hamiltonian and the latter by the Bose–Hubbard one. The main difficulty is however that these quantities are a priori coupled at the relevant order of the energy expansion because of the nontrivial dynamics in the two-mode subspace. More specifically we have to control processes where an exchange of particles between the modes u_+ and u_- mediates the excitation of particles out of the two-mode subspace.

In the next section we state our main results precisely and provide a more extended sketch of the proof, before proceeding to the details in the rest of the paper. As a final comment before that, we hope that future investigations will allow us to prove something about the low-lying excitation spectrum of the system at hand. We expect two types of excited eigenvalues, yielding essentially independent contributions: those coming from the excited states of the Bose–Hubbard Hamiltonian (1-1) and those coming from the generalized Bogoliubov Hamiltonian defined in Section 3B. The latter actually commutes with a shift operator, so that one might expect H_N to have some "almost continuous" spectrum in the sense of very close eigenvalues in the limit $N \to \infty$ (with spacing $o_N(1)$).

2. Main statements

2A. ***The double-well Hamiltonian.*** We consider the action of the Hamiltonian

$$H_N = \sum_{j=1}^{N}(-\Delta_j + V_{\mathrm{DW}}(x_j)) + \frac{\lambda}{N-1}\sum_{1\leqslant i<j\leqslant N} w(x_i - x_j),$$

already introduced in (1-2), on the space $\mathfrak{H}^N = L^2_{\mathrm{sym}}(\mathbb{R}^{dN})$, $d = 1, 2, 3$. The coupling constant proportional to $(N-1)^{-1}$ in (1-2) formally makes the contributions from the two sums in H_N of the same order in N. We introduced a further fixed coupling constant $\lambda > 0$. For simplicity we make liberal assumptions on the data of the problem, which we do not claim to be optimal for the results we will prove.

Assumption 2.1 (the interaction potential). We assume that w is a radial bounded function with compact support. We also suppose that it is positive and of positive type, that is, with $\hat{w}$ the Fourier transform,

$$w(x) \geqslant 0 \quad \text{a.e.} \qquad \text{and} \qquad \hat{w}(k) \geqslant 0 \quad \text{a.e.} \tag{2-1}$$

Assumption 2.2 (the double-well potential). Let $L > 0$ and

$$x_L := \Big(\frac{L}{2}, 0, \dots, 0\Big) \in \mathbb{R}^d, \quad -x_L = \Big(-\frac{L}{2}, 0, \dots, 0\Big) \in \mathbb{R}^d$$

represent the centers of the wells. We define

$$V_{\mathrm{DW}}(x) = \min\{V(x - x_L), V(x + x_L)\}, \tag{2-2}$$

with

$$V(x) = |x|^s, \quad s \geqslant 2\,. \tag{2-3}$$

Note that, since w is radial, the choice of two wells with centers on the x_1-axis is without loss of generality. To model two deep and well-separated wells, we shall let the interwell distance diverge

$$L = 2|x_L| \xrightarrow[N\to\infty]{} \infty.$$

Low-lying energy modes (see [Olgiati and Rougerie 2021] for more details)**.** Given a one-body function $u \in L^2(\mathbb{R}^d)$, its Hartree energy (1-5) reads

$$\mathcal{E}^{\mathrm{H}}[u] := \int_{\mathbb{R}^d} |\nabla u(x)|^2\,dx + \int_{\mathbb{R}^d} V_{\mathrm{DW}}(x)|u(x)|^2\,dx + \frac{\lambda}{2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} w(x-y)|u(x)|^2|u(y)|^2\,dx\,dy. \tag{2-4}$$

We define u_+ to be the minimizer of $\mathcal{E}^{\mathrm{H}}$ at unit mass, i.e.,

$$\mathcal{E}^{\mathrm{H}}[u_+] = \inf\left\{\mathcal{E}^{\mathrm{H}}[u]\;\middle|\; u \in H^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d, V_{\mathrm{DW}}(x)\,dx),\ \int_{\mathbb{R}^d}|u|^2 = 1\right\}. \tag{2-5}$$

Its existence follows from standard arguments. By a convexity argument such a minimizer must be unique up to a constant phase, which can be fixed so as to ensure $u_+ > 0$, which we henceforth do; see, e.g., [Lieb and Loss 1997, Theorem 11.8].

The mean-field Hamiltonian

$$h_{\mathrm{MF}} := -\Delta + V_{\mathrm{DW}} + \lambda w * |u_+|^2 \tag{2-6}$$

is the functional derivative of $\mathcal{E}^{\mathrm{H}}$ at u_+, seen as a self-adjoint operator on $L^2(\mathbb{R}^d)$. Since V_{DW} grows at infinity, h_{MF} has compact resolvent, and therefore a complete basis of eigenvectors. The Euler–Lagrange equation for the energy minimization problem reads

$$h_{\mathrm{MF}}u_+ = \mu_+ u_+, \tag{2-7}$$

with the chemical potential/Lagrange multiplier

$$\mu_+ = \mathcal{E}^{\mathrm{H}}[u_+] + \frac{\lambda}{2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} w(x-y)|u_+(x)|^2|u_+(y)|^2\,dx\,dy. \tag{2-8}$$

By standard arguments, μ_+ is the lowest eigenvalue of h_{MF}, corresponding to the nondegenerate eigenfunction u_+.

We next define u_- to be the first excited normalized eigenvector of h_{MF}, i.e.,

$$h_{\mathrm{MF}}u_- = \mu_- u_-, \tag{2-9}$$

where $\mu_- > \mu_+$ satisfies

$$\mu_- = \inf\left\{\langle u, h_{\mathrm{DW}}u\rangle \;\middle|\; u \in \mathcal{D}(h_{\mathrm{MF}}),\ \int_{\mathbb{R}^d}\bar{u}u_+ = 0,\ \int_{\mathbb{R}^d}|u|^2 = 1\right\}. \tag{2-10}$$

It follows from the arguments of [Olgiati and Rougerie 2021] that u_- is nondegenerate.

Since h_{DW} is a double-well Hamiltonian, all its eigenvectors are mainly localized [Helffer 1988; Dimassi and Sjöstrand 1999] around the two centers $\pm x_L$. As a consequence, the two linear combinations

$$u_1 = \frac{u_+ + u_-}{\sqrt{2}}, \quad u_2 = \frac{u_+ - u_-}{\sqrt{2}} \tag{2-11}$$

are mainly localized, respectively, in the left and right wells. These are the low-energy modes whose role was anticipated above.

Tunneling parameter. The gap $\mu_- - \mu_+$ of h_{MF} is closely related to the magnitude of the tunneling effect between wells. Indeed,

$$\mu_- - \mu_+ = \langle (u_- - u_+), h_{\mathrm{MF}}(u_- + u_+)\rangle = 2\langle u_2, h_{\mathrm{MF}} u_1\rangle,$$

and, as said, u_1 and u_2 are mainly localized, respectively, in the right and left wells. To quantify this we define the semiclassical Agmon distance [Agmon 1982; Dimassi and Sjöstrand 1999; Helffer 1988] associated to the one-well potential V

$$A(x) = \int_0^{|x|} \sqrt{V(r')}\, dr' = \frac{|x|^{1+s/2}}{1+s/2}. \tag{2-12}$$

We then set

$$T := e^{-2A(L/2)}. \tag{2-13}$$

As we will recall in Theorem A.1 below, we essentially have

$$\mu_- - \mu_+ \simeq T. \tag{2-14}$$

We will work in the regime

$$N \to \infty, \quad \lambda \text{ fixed}, T \ll 1 \text{ or, equivalently, } L \gg 1. \tag{2-15}$$

2B. ***Second quantization and effective Hamiltonians.*** The many-body Hilbert space $\mathfrak{H}^N$ is the N-th sector of the bosonic Fock space

$$\mathfrak{F} := \bigoplus_{n=0}^{\infty} L^2(\mathbb{R}^d)^{\otimes_{\mathrm{sym}} n} \tag{2-16}$$

on which we define the usual algebra of bosonic creation and annihilation operators (see Section 3 for the precise definition) whose commutation relations are

$$[a_u, a_v^\dagger] = \langle u, v\rangle_{L^2}, \quad [a_u, a_v] = [a_u^\dagger, a_v^\dagger] = 0, \quad u, v \in L^2(\mathbb{R}^d). \tag{2-17}$$

Given a generic one-body orbital $u \in L^2(\mathbb{R}^d)$ we introduce the particle number operator

$$\mathcal{N}_u := a_u^\dagger a_u$$

whose action on $\mathfrak{H}^N$ is

$$\mathcal{N}_u = \sum_{j=1}^N |u\rangle\langle u|_j. \tag{2-18}$$

Here $|u\rangle\langle u|_j$ acts as the orthogonal projection $|u\rangle\langle u|$ on the j-th variable and as the identity on all other variables.

One can extend the Hamiltonian H_N to $\mathfrak{F}$ as

$$H_N = \sum_{m,n\geqslant 1} h_{mn}\, a_m^\dagger a_n + \frac{\lambda}{2(N-1)} \sum_{m,n,p,q\geqslant 1} w_{mnpq}\, a_m^\dagger a_n^\dagger a_p a_q, \tag{2-19}$$

whose restriction on the N-th sector coincides with (1-2). The notation above is

$$\begin{aligned} h_{mn} &:= \langle u_m, (-\Delta + V_{\mathrm{DW}}) u_n\rangle, \\ w_{mnpq} &:= \langle u_m \otimes u_n, w\, u_p \otimes u_q\rangle \end{aligned} \tag{2-20}$$

for an orthonormal basis $(u_n)_{n\in\mathbb{N}}$ of $L^2(\mathbb{R}^d)$, with $a_n^\dagger$, a_n the associated creation and annihilation operators.

Two-mode energy in the low-energy subspace. Let P be the orthogonal projector onto the linear span of (u_+, u_-) (or, equivalently, (u_1, u_2)). We define the two-mode Hamiltonian

$$H_{2\text{-mode}} := P^{\otimes N} H_N P^{\otimes N} \tag{2-21}$$

and the associated ground state energy

$$E_{2\text{-mode}} := \inf\left\{ \langle \Psi_N | H_{2\text{-mode}} | \Psi_N\rangle,\ \Psi_N \in \bigotimes_{\mathrm{sym}}^N (PL^2(\mathbb{R}^d)),\ \int_{\mathbb{R}^{dN}} |\Psi_N|^2 = 1 \right\}. \tag{2-22}$$

Later we will discuss the relationship between the above and

$$E_{\mathrm{BH}} := \inf \sigma(H_{\mathrm{BH}}), \tag{2-23}$$

the bottom of the spectrum of the Bose–Hubbard Hamiltonian

$$H_{\mathrm{BH}} := \frac{\mu_+ - \mu_-}{2}(a_1^\dagger a_2 + a_2^\dagger a_1) + \frac{\lambda}{2(N-1)} w_{1111}(a_1^\dagger a_1^\dagger a_1 a_1 + a_2^\dagger a_2^\dagger a_2 a_2) \tag{2-24}$$

on the space $\bigotimes_{\mathrm{sym}}^N (PL^2(\mathbb{R}^d))$. As discussed in Section 4, H_{BH} is obtained from H_N by retaining only terms corresponding to the subspace spanned by u_+, u_- (equivalently u_1, u_2) in (2-19) and making a few further simplifications.

Bogoliubov energy of excitations. We will adopt the following notation for a spectral decomposition of h_{MF}:

$$h_{\mathrm{MF}} = \mu_+ |u_+\rangle\langle u_+| + \mu_- |u_-\rangle\langle u_-| + \sum_{m\geqslant 3} \mu_m |u_m\rangle\langle u_m|. \tag{2-25}$$

As stated in Theorem A.1(vi) (proved in [Olgiati and Rougerie 2021]) an appropriate choice of the u_m, with $m \geqslant 3$, ensures that the modes (compare with (2-11))

$$u_{r,\alpha} := \frac{u_{2\alpha+1} + u_{2\alpha+2}}{\sqrt{2}} \quad \text{and} \quad u_{\ell,\alpha} := \frac{u_{2\alpha+1} - u_{2\alpha+2}}{\sqrt{2}}, \tag{2-26}$$

with $\alpha \geqslant 1$, are (mostly) localized, respectively, in the right and left half-space. They pairwise generate the spectral subspaces of h_{MF} corresponding to $\mu_{2\alpha+1}$ and $\mu_{2\alpha+2}$. We will always use either the basis of $L^2(\mathbb{R}^d)$ from (2-25) or that from (2-26) (with the addition of u_+, u_- or u_r, u_ℓ). Since all these functions solve, or are linear combinations of functions that solve, an elliptic equation with real coefficients, we can (and will) always assume that they are real-valued functions. We also define

$$P_r := \sum_{\alpha\geqslant 1} |u_{r,\alpha}\rangle\langle u_{r,\alpha}| \quad P_\ell := \sum_{\alpha\geqslant 1} |u_{\ell,\alpha}\rangle\langle u_{\ell,\alpha}|, \tag{2-27}$$

and

$$\mathrm{Tr}_\perp(A) := \sum_{m\geqslant 3}\langle u_m, Au_m\rangle, \quad \mathrm{Tr}_{\perp,r}(A) := \sum_{\alpha\geqslant 1}\langle u_{r,\alpha}, Au_{r,\alpha}\rangle, \quad \mathrm{Tr}_{\perp,\ell}(A) := \sum_{\alpha\geqslant 1}\langle u_{\ell,\alpha}, Au_{\ell,\alpha}\rangle. \tag{2-28}$$

Then the Bogoliubov energy is given as

$$\begin{aligned} E^{\mathrm{Bog}} := -\tfrac{1}{2}\mathrm{Tr}_{\perp,r}\big[D_r + \lambda P_r K_{11} P_r - \sqrt{D_r^2 + 2\lambda D_r^{1/2} P_r K_{11} P_r D_r^{1/2}}\,\big] \\ - \tfrac{1}{2}\mathrm{Tr}_{\perp,\ell}\big[D_\ell + \lambda P_\ell K_{22} P_\ell - \sqrt{D_\ell^2 + 2\lambda D_\ell^{1/2} P_\ell K_{22} P_\ell D_\ell^{1/2}}\,\big], \end{aligned} \tag{2-29}$$

where

$$D_r := P_r(h_{\mathrm{MF}} - \mu_+)P_r, \quad D_\ell := P_\ell(h_{\mathrm{MF}} - \mu_+)P_\ell \tag{2-30}$$

and K_{11} and K_{22} are the two operators on $L^2(\mathbb{R}^d)$ defined by

$$\langle v, K_{11}u\rangle = \tfrac{1}{2}\langle v\otimes u_1, w\, u_1\otimes u\rangle, \quad \langle v, K_{22}u\rangle = \tfrac{1}{2}\langle v\otimes u_2, w\, u_2\otimes v\rangle.$$

The quantity E^{Bog} is essentially the sum of the lowest eigenvalues of two independent bosonic quadratic Hamiltonians acting on the left and right modes respectively (compare with the explicit formulas in [Grech and Seiringer 2013] and see [Bach and Bru 2016; Bruneau and Dereziński 2007; Dereziński 2017] and references therein for further literature). It will turn out to (asymptotically) coincide with the bottom of the spectrum of the full Bogoliubov Hamiltonian (3-18), i.e., the part of H_N that contains exactly two creators/annihilators for excited modes u_m with $m \geqslant 3$. That the traces in (2-29) are finite is not a priori obvious, and will be part of the proof. The two summands in the right-hand side of (2-29) coincide thanks to the symmetry of the system under reflections around the $x_1 = 0$ axis. Each summand also coincides, as $T \to 0$, with the bottom of the spectrum of the Bogoliubov Hamiltonian for particles occupying one-well excited modes above a one-well Hartree minimizer, centered either in x_L or $-x_L$, used in [Rougerie and Spehner 2018].

2C. *Main theorems.*

Theorem 2.3 (variance and energy of the ground state). *Assume that, as $N \to \infty$, $T \sim N^{-\delta}$ for some fixed $\delta > 0$. Let Ψ_{gs} be the unique (up to a phase) ground state of H_N. There exists $\lambda_0 > 0$ such that, for all* $0 < \lambda \leqslant \lambda_0$,

$$\lim_{N\to\infty} \frac{1}{N}\langle(\mathcal{N}_1 - \mathcal{N}_2)^2\rangle_{\Psi_{\mathrm{gs}}} = 0 \tag{2-31}$$

and

$$\lim_{N\to\infty} |E(N) - E_{2\text{-mode}} - E^{\mathrm{Bog}}| = 0. \tag{2-32}$$

A few comments:

(1) We believe the result holds without the smallness condition on λ. The precise condition we need is that the left-hand side of (8-26) be bounded below by a constant, which we could so far prove only for small λ.

(2) As part of the proof we find

$$\langle \mathcal{N}_1 + \mathcal{N}_2 \rangle_{\Psi_{\mathrm{gs}}} = \langle \mathcal{N}_{u_+} + \mathcal{N}_{u_-} \rangle_{\Psi_{\mathrm{gs}}} = N + O(1).$$

Since u_1 and u_2 are obtained one from the other by reflecting across $\{x_1 = 0\}$ and the full problem is invariant under such a reflection, this implies

$$\langle \mathcal{N}_1 \rangle_{\Psi_{\mathrm{gs}}} = \langle \mathcal{N}_2 \rangle_{\Psi_{\mathrm{gs}}} \simeq \frac{N}{2} + O(1), \tag{2-33}$$

so that we can reformulate (2-31) as

$$\langle (\mathcal{N}_1 - \langle \mathcal{N}_1 \rangle)^2 \rangle_{\Psi_{\mathrm{gs}}} \ll N.$$

(3) Central limit theorems are known to hold for mean-field bosonic systems in one-well-like situations [Buchholz et al. 2014; Rademacher and Schlein 2019]. For $T \gtrsim 1$ we recover such a situation: a single Bose–Einstein condensate in the state u_+ with Bogoliubov corrections on top, captured by a quasifree (gaussian) state. This would essentially lead to

$$\left\langle \left(\mathcal{N}_1 - \frac{N}{2} \right)^2 \right\rangle_{u_+^{\otimes N}} \simeq \langle \mathcal{N}_{u_1}^2 \rangle_{u_+^{\otimes N}} - (\langle \mathcal{N}_{u_1} \rangle_{u_+^{\otimes N}})^2 \simeq \frac{N}{4}.$$

The estimate (2-31) is a significant departure from this situation: correlations within the two-mode subspace are strong enough to reduce the variance significantly.

We also have estimates clarifying the nature of the main terms captured by our energy asymptotics in Theorem 2.3:

Proposition 2.4 (main terms in the two-mode energy). *Assume that, as $N \to \infty$, $T \sim N^{-\delta}$ for some fixed $\delta > 0$. Then we have that, for any fixed $\varepsilon > 0$,*

$$\left| E_{2\text{-mode}} - N h_{11} + \frac{\lambda N^2}{4(N-1)} (2 w_{1122} - w_{1212}) - E_{\mathrm{BH}} \right| \leqslant C_\varepsilon \max(T^{1/2-\varepsilon}, N^{-1+\varepsilon\delta}), \tag{2-34}$$

where $E_{2\text{-mode}}$ and E_{BH} are defined respectively in (2-22) *and* (2-23). *Moreover*

$$\left| E_{\mathrm{BH}} - \left(\frac{\lambda N^2}{4(N-1)} w_{1111} - \frac{\lambda N}{2(N-1)} w_{1111} + (\mu_+ - \mu_-) \frac{N}{2} \right) \right| \leqslant C_\varepsilon \max(T^{1/2-\varepsilon}, N^{-1+\varepsilon\delta}). \tag{2-35}$$

A few comments:

(1) We expect the remainders in the right-hand sides of (2-34) and (2-35) to be essentially sharp and to be part of the expansion of the full many-body energy $E(N)$. They lead to a variance bounded as (essentially)

$$\frac{1}{N} \langle (\mathcal{N}_1 - \mathcal{N}_2)^2 \rangle_{\mathrm{BH}} \leqslant C \max(T^{1/2}, N^{-1})$$

in the Bose–Hubbard ground state. Deriving such estimates at the level of the full many-body ground state would require that we improve our method of proof.

(2) The reference energy $N h_{11}$, N times the minimal one-well energy with no interactions, is usually subtracted from the Bose–Hubbard Hamiltonian as a basic energy reference and we follow this convention. The other terms appearing in the left-hand side of (2-34), which produce an energy shift between $E_{2\text{-mode}}$

and E_{BH}, are interaction energies due to particles tunneling through the double well's peak (not included in the Bose–Hubbard model). Depending on the parameter regime and possible improvements of some of our bounds, they may or may not be smaller than the other relevant terms. Since we can isolate them exactly in our energy expansions, we keep track of them as exact expressions, but they are not very relevant to the main thrust of the argument.

(3) The three main terms we isolate in the Bose–Hubbard energy are more interesting. The first one, $\lambda N^2/(4(N-1))w_{1111}$ is a one-well mean-field interaction energy. This is the leading order for any reasonable two-mode state, independently of its details. The second term $-\lambda N/(2(N-1))w_{1111}$, however, is a reduction of the interaction energy due to the suppressed variance of the true ground state. We had captured it before [Rougerie and Spehner 2018] in a reduced parameter regime. It is in any case larger than our biggest error term, which we show is $o(1)$. The last term $(\mu_+ - \mu_-)(N/2)$ is the tunneling contribution, not captured in [Rougerie and Spehner 2018]. When $\delta < 1$, i.e., $T \gg N^{-1}$, it is larger than our main error term.

2D. ***Sketch of proof.*** The general strategy is to group the various contributions to H_N in the second quantized formulation (2-19), much as in the derivation of Bogoliubov's theory in [Seiringer 2011; Grech and Seiringer 2013; Lewin et al. 2015; Dereziński and Napiórkowski 2014]. We use a basis of $L^2(\mathbb{R}^d)$ as discussed around (2-25) and distinguish between:

- Terms that contain only creators/annihilators corresponding to the two-mode subspace $\operatorname{span}(u_+, u_-)$. After some simplifications they yield the two-mode energy $E_{2\text{-mode}}$, which we prove controls the variance (2-31); see Section 4.
- Linear terms that contain exactly one creator/annihilator corresponding to the excited subspace $\operatorname{span}(u_+, u_-)^\perp$. These should be negligible in the final estimate.
- Quadratic terms that contain exactly two creators/annihilators corresponding to the excited subspace. In those we replace the creators/annihilators of the two-mode subspace by numbers, which leads to a Bogoliubov-like Hamiltonian acting on $\ell^2(\mathfrak{F}^\perp)$, where $\mathfrak{F}^\perp$ is the bosonic Fock space generated by the excited modes.
- Cubic and quartic terms that contain at least three creators/annihilators corresponding to the excited subspace. These can be neglected due to the low occupancy of said subspace.

To bring these heuristics to fruition we need a priori bounds (see Section 6) on:

- The number of excited particles and their kinetic energy.
- A joint moment of the number and kinetic energy of the excited particles.
- The variance of particle numbers in the low-lying subspace.

The first bounds follow from Onsager's lemma (see [Rougerie 2020, Section 2.1] and references therein) supplemented by our estimates on the Hartree problem in [Olgiati and Rougerie 2021]. We also obtain

$$\langle \mathcal{N}_{u_-} \rangle \leqslant C \min(N, T^{-1}) \tag{2-36}$$

at this stage, which we use later in the proof. For the second estimate, we start with the strategy of [Seiringer 2011; Grech and Seiringer 2013], but in our case the variance in the low-lying subspace enters the bound. Combining with a first rough energy estimate proves that the left side of (2-31) is bounded independently of N and T, which can then be used to close the second estimate.

With these estimates at hand we can deal efficiently with the quadratic, cubic and quartic terms mentioned above. The Bogoliubov Hamiltonian acting only on the excited space is introduced via a partial isometry $\mathcal{U}_N : \mathfrak{H}^N \mapsto \ell^2(\mathfrak{F}^\perp)$ that we conjugate the difference $H_N - H_{2\text{-mode}}$ with; see Section 3. This generalizes the excitation map introduced in [Lewin et al. 2015]. That the Bogoliubov Hamiltonian acts on $\ell^2(\mathfrak{F}^\perp)$ and not just $\mathfrak{F}^\perp$ keeps track of the population imbalance in the two-mode subspace. Relying on estimates from [Olgiati and Rougerie 2021] we can then split all the excited modes into a left and right part as in (2-26) and neglect couplings between left and right modes. After some further manipulations, this reduces the full Bogoliubov Hamiltonian to two independent ones acting on $\mathfrak{F}(P_\ell L^2(\mathbb{R}^d))$ and $\mathfrak{F}(P_r L^2(\mathbb{R}^d))$, the bosonic Fock spaces generated by the left and right modes respectively; see (2-27). Their ground energies yield the E_{Bog} energy entering the statement.

The part of the proof we find the most difficult is the treatment of linear terms. In the one-well case they are negligible [Seiringer 2011; Grech and Seiringer 2013; Lewin et al. 2015; Dereziński and Napiórkowski 2014] as a consequence of the optimality of the low-energy subspace.[3] Cancellations of this form also occur in our setting (see (5-23) below), using that $h_{\text{MF}} u_\pm = \mu_\pm u_\pm \perp u_m$ if $m \geqslant 3$ and that $\| |u_+| - |u_-| \| \lessapprox T^{1/2}$ as shown in [Olgiati and Rougerie 2021]. More complicated linear terms appear, however, an example being proportional to (with a_m an annihilator on the excited subspace, $m \geqslant 3$)

$$\frac{\lambda}{2(N-1)} a_+^\dagger (a_+^\dagger a_- + a_-^\dagger a_+) a_m.$$

Using our a priori bounds (think of a_m as being $O(1)$), the above would be $o(1)$ if the result (2-31) was known a priori for

$$a_+^\dagger a_- + a_-^\dagger a_+ = \mathcal{N}_1 - \mathcal{N}_2.$$

That terms of this type finally turn out to be negligible is a signature not of the optimal choice of the low-lying two-mode subspace, which we used already, but of the particular Bose–Hubbard ground state within it, witnessed by its small expectation of $N^{-1}(\mathcal{N}_1 - \mathcal{N}_2)^2$.

To eliminate these extra linear terms, we will "complete a square" by defining (see Section 7) shifted creation and annihilation operators for the excited modes. In terms of those the combination of quadratic and linear terms is a new quadratic Hamiltonian corrected by a remainder term $\propto \lambda^2 N^{-1}(\mathcal{N}_1 - \mathcal{N}_2)^2$, depending on the variance operator. The latter we can absorb in $H_{2\text{-mode}}$ for small enough coupling constant λ. Another remainder comes from the fact that the shifted operators satisfy the canonical commutation relations (CCR) only approximately, so that the diagonalization of the new quadratic Hamiltonian is more involved. After we have decoupled the contributions of the two wells by estimating cross-terms in the resulting expressions, we can rely on ideas from [Grech and Seiringer 2013] to handle that aspect, for we have a precise control on the commutators of the shifted operators.

[3] They are the second quantization of the functional derivative of the Hartree energy at the minimizer.

3. Mapping to the space of excitations

We will use the second quantization formalism, calling $\mathfrak{F}$ the Fock space associated to $L^2(\mathbb{R}^d)$, and $a^\dagger(f)$, $a(f)$ the creation and annihilation operators associated to $f \in L^2(\mathbb{R}^d)$. We refer the reader to, e.g., [Gustafson and Sigal 2011, Section 18] for precise definitions. We will adopt the notation

$$a_+^\sharp := a^\sharp(u_+), \quad a_-^\sharp := a^\sharp(u_-), \quad a_m^\sharp := a^\sharp(u_m),$$
$$a_{r,\alpha}^\sharp := a^\sharp(u_{r,\alpha}) = \frac{a_{2\alpha+1}^\sharp + a_{2\alpha+2}^\sharp}{\sqrt{2}}, \quad a_{\ell,\alpha}^\sharp := a^\sharp(u_{\ell,\alpha}) = \frac{a_{2\alpha+1}^\sharp - a_{2\alpha+2}^\sharp}{\sqrt{2}}$$

for $\sharp \in \{\,\cdot\,,\dagger\}$, where u_+, u_-, u_m, $u_{r,\alpha}$, and $u_{\ell,\alpha}$ with $m, \alpha \in \mathbb{N}\setminus\{0\}$ are the modes introduced in Section 2. We will denote by $\mathrm{d}\Gamma(A)$ the second quantization of a k-body operator, and by $\mathcal{N}_m = a_m^\dagger a_m$ the number operator for the m-th mode. We furthermore define the number operator for modes beyond u_+ and u_- (or u_1 and u_2)

$$\mathcal{N}_\perp := \sum_{m\geqslant 3} \mathcal{N}_m. \tag{3-1}$$

As anticipated in Section 2, the Hamiltonian (1-2) can be written as, in the notation we introduced,[4]

$$\begin{aligned} H_N &= \mathrm{d}\Gamma(-\Delta + V_{\mathrm{DW}}) + \frac{\lambda}{(N-1)}\mathrm{d}\Gamma(w) \\ &= \sum_{m,n\geqslant 1} h_{mn}\, a_m^\dagger a_n + \frac{\lambda}{2(N-1)} \sum_{m,n,p,q\geqslant 1} w_{mnpq}\, a_m^\dagger a_n^\dagger a_p a_q. \end{aligned} \tag{3-2}$$

Two-mode Hamiltonian. The part of H_N in which summations are restricted to the first two indices will play a major role.

Definition 3.1 (two-mode Hamiltonian). We define

$$H_{2\text{-mode}} := \sum_{m,n\in\{1,2\}} h_{mn}\, a_m^\dagger a_n + \frac{\lambda}{2(N-1)} \sum_{m,n,p,q\in\{1,2\}} w_{mnpq}\, a_m^\dagger a_n^\dagger a_p a_q \tag{3-3}$$

as an operator on the N-body space $\mathfrak{H}^N$.

There are a few differences between $H_{2\text{-mode}}$ and the Bose–Hubbard Hamiltonian H_{BH} from (2-24):

- H_{BH} is defined on the N-body space generated by the modes u_1 and u_2 only; that is, $\bigotimes_{\mathrm{sym}}^N(PL^2(\mathbb{R}^d))$. This is equivalent to identifying $\mathcal{N}_1 + \mathcal{N}_2 = N$ when working with $H_{2\text{-mode}}$.
- All quartic terms of (3-3) that contain both $a_1^\sharp$ and $a_2^\sharp$ are neglected in H_{BH}.
- $H_{2\text{-mode}}$ contains the one-well noninteracting terms proportional to h_{11} and h_{22}. They will give the energy Nh_{11} appearing in (2-34).
- The coefficient of $a_1^\dagger a_2 + a_2^\dagger a_1$ in (3-3) will turn out to be a perturbation of the $(\mu_+ - \mu_-)/2$ of H_{BH}. The same for the coefficient of the quartic terms.

The difference between $H_{2\text{-mode}}$ and H_{BH} is not a priori small. We will often work with $H_{2\text{-mode}}$, and discuss in Section 4 its relation with H_{BH}.

[4] We are considering w as the two-body observable corresponding to the multiplication by the function $w(x-y)$.

3A. ***Excitation space.*** The energy of the fraction of particles that occupy $\{u_m\}_{m\geqslant 3}$ needs to be separately monitored. To this end, it will be useful to consider the second quantization of operators restricted to the orthogonal complement of u_1 and u_2. We define the projections

$$P := |u_+\rangle\langle u_+| + |u_-\rangle\langle u_-| = |u_1\rangle\langle u_1| + |u_2\rangle\langle u_2|, \quad P^\perp := \mathbb{1} - P = \sum_{m\geqslant 3} |u_m\rangle\langle u_m|. \tag{3-4}$$

For self-adjoint operators A on $\mathfrak{H}$ and B on $\mathfrak{H}\otimes\mathfrak{H}$ we define

$$\mathrm{d}\Gamma_\perp(A) := \mathrm{d}\Gamma(P^\perp A P^\perp) = \sum_{m,n\geqslant 3} \langle u_m, A u_n\rangle a_m^\dagger a_n, \tag{3-5}$$

$$\mathrm{d}\Gamma_\perp(B) := \mathrm{d}\Gamma(P^\perp\otimes P^\perp B P^\perp\otimes P^\perp) = \sum_{m,n,p,q\geqslant 3} \langle u_m\otimes u_n, B u_p\otimes u_q\rangle a_m^\dagger a_n^\dagger a_p a_q. \tag{3-6}$$

In this notation,

$$\mathcal{N}_\perp = \mathrm{d}\Gamma_\perp(\mathbb{1}).$$

Let us introduce the Hilbert space decomposition induced by P and $P^\perp$

$$\mathfrak{H}^N = \left(\mathrm{span}\{u_+\}\oplus\mathrm{span}\{u_-\}\oplus\bigoplus_{m\geqslant 3}^{\infty}\mathrm{span}\{u_m\}\right)^{\otimes_{\mathrm{sym}}N} = \left(\mathrm{span}\{u_1\}\oplus\mathrm{span}\{u_2\}\oplus\bigoplus_{m\geqslant 3}^{\infty}\mathrm{span}\{u_m\}\right)^{\otimes_{\mathrm{sym}}N}. \tag{3-7}$$

Accordingly, any $\psi_N\in\mathfrak{H}^N$ can be uniquely expanded in the form

$$\psi_N = \sum_{s=0}^{N} \sum_{d=-N+s,\,-N+s+2,\,\dots}^{\dots,\,N-s-2,\,N-s} u_1^{\otimes(N-s+d)/2}\otimes_{\mathrm{sym}} u_2^{\otimes(N-s-d)/2}\otimes_{\mathrm{sym}}\Phi_{s,d} \tag{3-8}$$

for suitable

$$\Phi_{s,d}\in(\{u_1,u_2\}^\perp)^{\otimes_{\mathrm{sym}}s}.$$

The index s represents the number of excited particles, i.e., those living in the orthogonal of $\mathrm{span}(u_1,u_2)$. The index d is the difference[5] between the number of particles in u_1 and the number of particles in u_2. Notice that (3-8) defines $\Phi_{s,d}$ only for those pairs of integers (s,d) such that $(N-s+d)/2$ is an integer.

For each fixed d, the collection of functions $\{\Phi_{s,d}\}_{0\leqslant s\leqslant N}$ identifies a vector in the truncated Fock space

$$\mathfrak{F}_\perp^{\leqslant N} := \bigoplus_{s=0}^{N}(\{u_1,u_2\}^\perp)^{\otimes_{\mathrm{sym}}s}\subset\mathfrak{F}_\perp\subset\mathfrak{F}. \tag{3-9}$$

Replicating the construction for all d we naturally arrive at the following definition.

Definition 3.2 (excitation space). We define the full space of excitations as

$$\ell^2(\mathfrak{F}_\perp) := \bigoplus_{s\in\mathbb{N},d\in\mathbb{Z}}(\{u_1,u_2\}^\perp)^{\otimes_{\mathrm{sym}}s} = \bigoplus_{d\in\mathbb{Z}}\mathfrak{F}_\perp. \tag{3-10}$$

A generic $\Phi\in\ell^2(\mathfrak{F}_\perp)$ is of the form

$$\Phi = \bigoplus_{s\in\mathbb{N},d\in\mathbb{Z}}\Phi_{s,d} \quad\text{such that}\quad \Phi_{s,d}\in(\{u_1,u_2\}^\perp)^{\otimes_{\mathrm{sym}}s} \quad\text{and}\quad \sum_{s,d}\|\Phi_{s,d}\|_{L^2}^2 < +\infty.$$

[5] It will be clear from the context when d stands for this difference or the physical space dimension.

We will adopt capital letters (as in Φ) to indicate excitation vectors in $\ell^2(\mathfrak{F}_\perp)$, while reserving small letters (as in ψ_N) for N-body wave-functions in $\mathfrak{H}^N$.

There is a natural operator mapping an N-body wave-function to its excitation content as in (3-8). We define it by generalizing ideas from [Lewin et al. 2015] (see [Rougerie 2020, Definition 5.10] and subsequent discussion for review):

Definition 3.3 (excitation map). Given any $\psi_N \in \mathfrak{H}^N$, consider its expansion (3-8). We define the excitation map as the operator

$$\mathcal{U}_N : \mathfrak{H}^N \to \ell^2(\mathfrak{F}_\perp), \quad \text{acting as } \mathcal{U}_N \psi_N = \bigoplus_{\substack{0 \leqslant s \leqslant N,\, |d| \leqslant N-s, \\ (N-s+d)/2 \in \mathbb{N}}} \Phi_{s,d}. \tag{3-11}$$

It is easy to check that $\mathcal{U}_N$ is a partial isometry from $\mathfrak{H}^N$ into $\ell^2(\mathfrak{F}_\perp)$; i.e., it acts unitarily if $\mathcal{U}_N^*$ is restricted to $\operatorname{Ran} \mathcal{U}_N$. In order to isolate the contributions to the energy that come from excited particles, we will conjugate the Hamiltonian H_N (or rather $H_N - H_{2\text{-mode}}$) with the unitary $\mathcal{U}_N$. This boils down to having formulas describing the action of $\mathcal{U}_N$ on creation and annihilation operators. We keep the same notation for the operators $a_m^\sharp$ with $m \geqslant 3$ after conjugation with $\mathcal{U}_N$, that is,

$$\mathcal{U}_N a_m^\dagger a_n \mathcal{U}_N^* = a_m^\dagger a_n, \quad m, n \geqslant 3.$$

We do the same for the operator representing the number of excitations, which, on $\ell^2(\mathfrak{F}_\perp)$, acts according to

$$\mathcal{N}_\perp \Phi = \bigoplus_{s \in \mathbb{N}, d \in \mathbb{Z}} s \Phi_{s,d}. \tag{3-12}$$

The difference $\mathcal{N}_1 - \mathcal{N}_2$ on the other hand corresponds to the operator that has the indices d as eigenvalues:

Definition 3.4 (difference operator). The difference operator on $\ell^2(\mathcal{F}_\perp)$ is defined as

$$\mathfrak{D} := \mathcal{U}_N (\mathcal{N}_1 - \mathcal{N}_2) \mathcal{U}_N^\dagger, \quad \text{with action } \mathfrak{D}\Phi = \bigoplus_{s \in \mathbb{N}, d \in \mathbb{Z}} d \Phi_{s,d}. \tag{3-13}$$

We will refer to $\mathfrak{D}^2$ (or $(\mathcal{N}_1 - \mathcal{N}_2)^2$ on $\mathfrak{H}^N$) as the *variance* operator.

We also need the unitary operator that shifts the index d by one unit.

Definition 3.5 (shift operator). We define the unitary operator

$$\Theta : \ell^2(\mathfrak{F}_\perp) \to \ell^2(\mathfrak{F}_\perp), \quad \text{with action } (\Theta\Phi)_{s,d} = \Phi_{s,d-1}. \tag{3-14}$$

As an immediate consequence of the above definitions we have, for any $m \geqslant 3$,

$$\begin{aligned} &[\mathfrak{D}, \Theta] = \Theta, \\ &[a_m, \Theta] = [a_m^\dagger, \Theta] = 0, \\ &[\mathfrak{D}, a_m] = [\mathfrak{D}, a_m^\dagger] = 0, \end{aligned} \tag{3-15}$$

which will be useful in the sequel. It follows from the first commutation relation and the unitarity of Θ that

$$\Theta^* f(\mathfrak{D}) \Theta = f(\Theta^* \mathfrak{D} \Theta) = f(\mathfrak{D} + 1) \tag{3-16}$$

for any smooth real function f (by functional calculus). We record the action of $\mathcal{U}_N$ on operators of the type $a^\dagger a$, needed to conjugate the full Hamiltonian, in the following:

Lemma 3.6 (operators on the excited Fock space). *For any* $m, n \geqslant 3$ *we have*

$$\begin{aligned}
&\mathcal{U}_N a_1^\dagger a_1 \mathcal{U}_N^* = \frac{N-\mathcal{N}_\perp+\mathfrak{D}}{2}, && \mathcal{U}_N a_1^\dagger a_2 \mathcal{U}_N^* = \Theta\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}+1}{2}}\sqrt{\frac{N-\mathcal{N}_\perp-\mathfrak{D}+1}{2}}\,\Theta,\\
&\mathcal{U}_N a_2^\dagger a_2 \mathcal{U}_N^* = \frac{N-\mathcal{N}_\perp-\mathfrak{D}}{2}, && \mathcal{U}_N a_1^\dagger a_m \mathcal{U}_N^* = \Theta\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}+1}{2}}\,a_m,\\
&\mathcal{U}_N a_2^\dagger a_m \mathcal{U}_N^* = \Theta^{-1}\sqrt{\frac{N-\mathcal{N}_\perp-\mathfrak{D}+1}{2}}\,a_m, && \mathcal{U}_N a_m^{\sharp_1} a_n^{\sharp_2} \mathcal{U}_N = a_m^{\sharp_1} a_n^{\sharp_2}
\end{aligned}\tag{3-17}$$

as identities on $\operatorname{Ran}\mathcal{U}_N$, *with* $\sharp_1, \sharp_2 \in \{\cdot\,, \dagger\}$.

Proof. The derivation of the first three identities is similar. We focus on the second one. We have, for $\Phi \in \operatorname{Ran}\mathcal{U}_N$,

$$\begin{aligned}
&a_1^\dagger a_2 \mathcal{U}_N^* \Phi\\
&\quad= \sum_{s=0}^{N} \sum_{d=-N+s,\,-N+s+2,\,\dots}^{\dots,N-s-2,N-s} \sqrt{\frac{N-s+d+2}{2}}\sqrt{\frac{N-s-d}{2}}\, u_1^{\otimes(N-s+d+2)/2} \otimes_{\mathrm{sym}} u_2^{\otimes(N-s-d-2)/2} \otimes_{\mathrm{sym}} \Phi_{s,d}\\
&\quad= \sum_{s=0}^{N} \sum_{d'=-N+s+2,\,-N+s+4,\,\dots}^{\dots,N-s,N-s+2} \sqrt{\frac{N-s+d'}{2}}\sqrt{\frac{N-s-d'+2}{2}}\, u_1^{\otimes(N-s+d')/2} \otimes_{\mathrm{sym}} u_2^{\otimes(N-s-d')/2} \otimes_{\mathrm{sym}} \Phi_{s,d'-2}.
\end{aligned}$$

Thus, acting with $\mathcal{U}_N$ we find

$$\begin{aligned}
(\mathcal{U}_N a_1^\dagger a_2 \mathcal{U}_N^* \Phi)_{s,d'} &= \sqrt{\frac{N-s+d'}{2}}\sqrt{\frac{N-s-d'+2}{2}}\,\Phi_{s,d'-2}\\
&= \left(\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}}{2}}\sqrt{\frac{N-\mathcal{N}_\perp-\mathfrak{D}+2}{2}}\,\Theta^2\Phi\right)_{s,d'}.
\end{aligned}$$

Using the unitarity of Θ, the commutation of Θ with $\mathcal{N}_\perp$ and the identity (3-16), one finds

$$\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}}{2}}\sqrt{\frac{N-\mathcal{N}_\perp-\mathfrak{D}+2}{2}}\,\Theta = \Theta\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}+1}{2}}\sqrt{\frac{N-\mathcal{N}_\perp-\mathfrak{D}+1}{2}}$$

and the second identity in (3-17) follows.

The proofs of the last three identities are basically identical. We focus on the first one. We have

$$\begin{aligned}
a_1^\dagger a_m \mathcal{U}_N^* \Phi &= \sum_{s=1}^{N} \sum_{d=-N+s,\,-N+s+2,\,\dots}^{\dots,N-s-2,N-s} \sqrt{\frac{N-s+d+2}{2}}\, u_1^{\otimes(N-s+d+2)/2} \otimes_{\mathrm{sym}} u_2^{\otimes(N-s-d)/2} \otimes_{\mathrm{sym}} (a_m\Phi)_{s-1,d}\\
&= \sum_{s'=0}^{N-1} \sum_{d=-N+s+1,\,-N+s+3,\,\dots}^{\dots,N-s-1,N-s+1} \sqrt{\frac{N-s'+d'}{2}}\, u_1^{\otimes(N-s'+d')/2} \otimes_{\mathrm{sym}} u_2^{\otimes(N-s'-d')/2} \otimes_{\mathrm{sym}} (a_m\Phi)_{s',d'-1}.
\end{aligned}$$

Acting with $\mathcal{U}_N$ we find

$$(\mathcal{U}_N a_1^\dagger a_m \mathcal{U}_N^* \Phi)_{s',d'} = \sqrt{\frac{N-s'+d'}{2}}(a_m \Phi)_{s',d'-1} = \left(\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}}{2}}\,\Theta a_m \Phi\right)_{s',d'}$$

and the result is again obtained by commuting Θ all the way to the left using (3-15). □

With the above we will be able to conjugate with $\mathcal{U}_N$ each summand in the Hamiltonian (3-2). For example

$$\mathcal{U}_N a_1^\dagger a_1^\dagger a_1 a_m \mathcal{U}_N^* = \mathcal{U}_N a_1^\dagger a_m \mathcal{U}_N^* \,\mathcal{U}_N a_1^\dagger a_1 \,\mathcal{U}_N^* = \Theta\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}}{2}}\,\frac{N-\mathcal{N}_\perp+\mathfrak{D}-1}{2}a_m$$

for any $m \geqslant 3$.

3B. ***Bogoliubov Hamiltonian.*** The Bogoliubov Hamiltonian is a quadratic operator on $\ell^2(\mathcal{F}_\perp)$ that represents the main contribution to the energy inside $\mathcal{U}_N(H_N - H_{2\text{-mode}})\mathcal{U}_N^*$, i.e., after the contribution from the modes u_1 and u_2 has been subtracted. We first define operators K_{11}, K_{22}, $K_{12} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ through their matrix elements

$$\begin{aligned}
\langle v,\, K_{11}u\rangle &= \tfrac{1}{2}\langle v\otimes u_1\,,\ w\, u_1 \otimes u\rangle,\\
\langle v,\, K_{22}u\rangle &= \tfrac{1}{2}\langle v\otimes u_2\,,\ w\, u_2 \otimes u\rangle,\\
\langle v,\, K_{12}u\rangle &= \langle v\otimes u_1\,,\ w\, u_2 \otimes u\rangle.
\end{aligned}$$

Since u_1 and u_2 are real, we have $K_{11} = K_{11}^*$ and $K_{22} = K_{22}^*$. Since w is bounded and $u_1, u_2 \in L^2(\mathbb{R}^d)$, Young's inequality immediately shows that these are bounded operators.

Definition 3.7 (Bogoliubov Hamiltonian). We define the Bogoliubov Hamiltonian as the operator on $\ell^2(\mathfrak{F}_\perp)$ given by

$$\begin{aligned}
\mathbb{H} = &\sum_{m,n\geqslant 3}\Big(-\Delta + V_{\mathrm{DW}} + \frac{\lambda}{2}w * |u_1|^2 + \frac{\lambda}{2}w * |u_2|^2 + \lambda K_{11} + \lambda K_{22} - \mu_+\Big)_{mn} a_m^\dagger a_n\\
&+ \frac{\lambda}{2}\sum_{m,n\geqslant 3}(K_{11})_{mn}(\Theta^{-2}a_m^\dagger a_n^\dagger + \Theta^2 a_m a_n) + \frac{\lambda}{2}\sum_{m,n\geqslant 3}(K_{22})_{mn}(\Theta^{2}a_m^\dagger a_n^\dagger + \Theta^{-2} a_m a_n)\\
&\quad + \frac{\lambda}{2}\sum_{m,n\geqslant 3}(K_{12})_{mn} a_m^\dagger a_n^\dagger + \frac{\lambda}{2}\sum_{m,n\geqslant 3}(K_{12}^*)_{mn} a_m a_n\\
&\qquad + \frac{\lambda}{2}\sum_{m,n\geqslant 3}(K_{12} + w * (u_1u_2))_{mn}\Theta^2 a_m^\dagger a_n + \frac{\lambda}{2}\sum_{m,n\geqslant 3}(K_{12}^* + w * (u_1u_2))_{mn}\Theta^{-2} a_m^\dagger a_n. \qquad (3\text{-}18)
\end{aligned}$$

The above is formally obtained from H_N by:

(1) Considering the parts of H_N in (3-2) that contain *exactly two* $a_m^\sharp$ with $m \geqslant 3$.

(2) Acting with (3-17) to pass to the space $\ell^2(\mathfrak{F}_\perp)$.

(3) Replacing all fractions coming from the right-hand sides of (3-17) by $(N-1)/2$.

This procedure will be made rigorous in Proposition 5.1 below.

A crucial feature of $\mathbb{H}$ is that, if we could ignore the terms coupling modes (mostly) supported in different wells (for example the last two lines of (3-18)), then $\mathbb{H}$ would coincide with the sum of two commuting quadratic Hamiltonians, each depending on one-well modes, as we now explain. We start with the following definition (recall the definition of left and right modes in (2-26)):

Definition 3.8 (Θ-translated right and left creators/annihilators). For any $m, \alpha \geqslant 1$ we define

$$\begin{aligned} b_m &:= \Theta\, a_m, & b_{r,\alpha} &:= \Theta\, a_{r,\alpha}, & b_{\ell,\alpha} &:= \Theta\, a_{\ell,\alpha}, \\ c_m &:= \Theta^{-1}\, a_m, & c_{r,\alpha} &:= \Theta^{-1}\, a_{r,\alpha}, & c_{\ell,\alpha} &:= \Theta^{-1}\, a_{\ell,\alpha} \end{aligned} \tag{3-19}$$

together with their adjoints $b^\dagger_m, b^\dagger_{r,\alpha}, b^\dagger_{\ell,\alpha}, c^\dagger_m, c^\dagger_{r,\alpha}, c^\dagger_{\ell,\alpha}$ (recall that $\Theta^* = \Theta^{-1}$).

It is straightforward to check the commutation relations

$$\begin{aligned} &[b_m, b^\dagger_n] = [c_m, c^\dagger_n] = \delta_{mn}, \quad [b_{r,\alpha}, b^\dagger_{r,\beta}] = [b_{\ell,\alpha}, b^\dagger_{\ell,\beta}] = [c_{r,\alpha}, c^\dagger_{r,\beta}] = [c_{\ell,\alpha}, c^\dagger_{\ell,\beta}] = \delta_{\alpha\beta}, \\ &[b_m, b_n] = [c_m, c_n] = 0, \qquad [b_{r,\alpha}, b_{r,\beta}] = [b_{\ell,\alpha}, b_{\ell,\beta}] = [c_{r,\alpha}, c_{r,\beta}] = [c_{\ell,\alpha}, c_{\ell,\beta}] = 0. \end{aligned} \tag{3-20}$$

The $b^\sharp_{r,\alpha}$ operators will be used to construct the excitation energy of the right well, while the $c^\sharp_{\ell,\alpha}$ will be associated with the left well. No other combination contributes to the energy at the order of precision we aim at. This leads to:

Definition 3.9 (right and left Bogoliubov Hamiltonians). The quadratic Hamiltonians for right and left modes are

$$\mathbb{H}_{\text{right}} := \sum_{\alpha,\beta\geqslant 1} \langle u_{r,\alpha}, (h_{\text{MF}} - \mu_+ + \lambda K_{11}) u_{r,\beta}\rangle b^\dagger_{r,\alpha} b_{r,\beta} + \frac{\lambda}{2} \sum_{\alpha,\beta\geqslant 1} \langle u_{r,\alpha}, K_{11} u_{r,\beta}\rangle (b^\dagger_{r,\alpha} b^\dagger_{r,\beta} + b_{r,\alpha} b_{r,\beta}), \tag{3-21}$$

$$\mathbb{H}_{\text{left}} := \sum_{\alpha,\beta\geqslant 1} \langle u_{\ell,\alpha}, (h_{\text{MF}} - \mu_+ + \lambda K_{22}) u_{\ell,\beta}\rangle c^\dagger_{\ell,\alpha} c_{\ell,\beta} + \frac{\lambda}{2} \sum_{\alpha,\beta\geqslant 1} \langle u_{\ell,\alpha}, K_{22} u_{\ell,\beta}\rangle (c^\dagger_{\ell,\alpha} c^\dagger_{\ell,\beta} + c_{\ell,\alpha} c_{\ell,\beta}). \tag{3-22}$$

Since $\langle u_{r,\alpha}, u_{\ell,\beta}\rangle = 0$ for all α, β, every creator or annihilator of a right mode $b^\sharp_{r,\alpha}$ commutes with every creator or annihilator of a left mode $c^\sharp_{\ell,\alpha}$. The two Hamiltonians above hence correspond (after conjugation with Bogoliubov transformations) to independent harmonic oscillators. One should view $\mathbb{H}_{\text{right}}$ (resp. $\mathbb{H}_{\text{left}}$) as obtained from $\mathbb{H}$ by retaining only those summands in which the $L^2(\mathbb{R}^d)$ scalar products are between $u_{r,\alpha}$ modes (resp. $u_{\ell,\alpha}$ modes). A further difference is the appearance of h_{MF} in (3-21) and (3-22) instead of the operator $-\Delta + V_{\text{DW}} + \lambda w * |u_1|^2/2 + \lambda w * |u_2|^2/2$ that appears in (3-18). This is due to the fact that their difference, proportional to $\mathrm{d}\Gamma_\perp(w * (u_1 u_2))$, will turn out to be negligible. The $b^\dagger b$-part of $\mathbb{H}_{\text{right}}$ is the second quantization of the self-adjoint operator $P_r h_{\text{MF}} P_r$ (and a similar property for the $c^\dagger c$ of $\mathbb{H}_{\text{left}}$).

It follows from the above definitions and the discussion in [Grech and Seiringer 2013, Sections 4 and 5] that our previous definition (2-29) coincides with

$$E_{\text{Bog}} = \inf \sigma_{\ell^2(\mathfrak{F}^\perp)}(\mathbb{H}_{\text{right}}) + \inf \sigma_{\ell^2(\mathfrak{F}^\perp)}(\mathbb{H}_{\text{left}}), \tag{3-23}$$

which we can obtain by acting on the vacuum with two commuting Bogoliubov transformations and taking the expectation value of $\mathbb{H}_{\text{right}} + \mathbb{H}_{\text{left}}$ in the quasifree state thus obtained. More details will be provided in Section 8A below.

4. Bounds on the two-mode Hamiltonian

The aim of this section is to prove lower and upper bounds for the Hamiltonian $H_{2\text{-mode}}$ defined in (3-3). We will also show a bound on the Bose–Hubbard energy and prove Proposition 2.4. We define the operator

$$\mathcal{T} := \frac{\mu_+ - \mu_-}{2} - \frac{\lambda}{N-1} w_{1112} \mathcal{N}_\perp - \frac{\lambda}{N-1} w_{1122} (\mathcal{N}_\perp - 1) \tag{4-1}$$

and the energy constants

$$E_0 = N h_{11} + \frac{\lambda N^2}{4(N-1)} (2 w_{1122} - w_{1212}) \tag{4-2}$$

and

$$\begin{aligned} E_N^w &:= N\biggl(\frac{\lambda N}{4(N-1)} (w_{1111} - 4 w_{1122} + 2 w_{1212}) - \frac{\lambda}{2(N-1)} (w_{1111} + w_{1122}) \biggr), \\ \mu &:= h_{11} + \frac{\lambda}{2} w_{1111} + \frac{\lambda N}{2(N-1)} (w_{1212} - 2 w_{1122}) - \frac{\lambda}{2(N-1)} w_{1122}, \\ U &:= \frac{1}{4} (w_{1111} - w_{1212}). \end{aligned} \tag{4-3}$$

The next lemma gives precise estimates on the magnitude of these quantities.

Lemma 4.1 (w-coefficients and chemical potential). *There exist strictly positive constants c and C independent on N and, for any $\varepsilon > 0$, an N-independent constant $C_\varepsilon > 0$ such that*

$$c \leqslant w_{1111} \leqslant C, \tag{4-4}$$

$$|w_{1112}| \leqslant C_\varepsilon T^{1-\varepsilon}, \tag{4-5}$$

$$0 \leqslant w_{1122} \leqslant C_\varepsilon T^{2-\varepsilon}, \tag{4-6}$$

$$0 \leqslant w_{1212} \leqslant C_\varepsilon T^{1-\varepsilon}, \tag{4-7}$$

where T is given by (2-13). *As a consequence, we have*

$$|\mu - \mu_+| \leqslant C_\varepsilon T^{1-\varepsilon}, \tag{4-8}$$

where μ was defined in (4-3) *and μ_+ is the ground state energy of h_{MF}.*

We postpone the proof of this lemma to Appendix B. As a consequence of Lemma 4.1, the reader should keep in mind the rule-of-thumb estimates

$$\begin{aligned} &\mathcal{T} \simeq \frac{\mu_+ - \mu_-}{2} \quad \text{on the states that will be of interest,} \\ &\mu \simeq \mu_+, \\ &U \simeq \frac{w_{1111}}{4} \geqslant C > 0. \end{aligned}$$

4A. ***Lower bound for $H_{\text{2-mode}}$.*** We shall prove the following:

Proposition 4.2 (expression and lower bound for $H_{\text{2-mode}}$). *We have the exact expression*

$$H_{\text{2-mode}} = E_0 + E_N^w + \mathcal{T}(a_1^\dagger a_2 + a_2^\dagger a_1) - \mu\mathcal{N}_\perp + \frac{\lambda U}{N-1}(\mathcal{N}_1 - \mathcal{N}_2)^2 \\ + \frac{2\lambda}{N-1} w_{1122}\mathcal{N}_-^2 + \frac{\lambda}{4(N-1)}(w_{1111} - 2w_{1122} + w_{1212})\mathcal{N}_\perp^2 \tag{4-9}$$

and the lower bound

$$H_{\text{2-mode}} \geqslant E_0 + E_N^w - \mu_+\mathcal{N}_\perp + N\frac{\mu_+ - \mu_-}{2} + \frac{\lambda U}{N-1}(\mathcal{N}_1 - \mathcal{N}_2)^2 - C_\varepsilon T^{1-\varepsilon}\mathcal{N}_\perp. \tag{4-10}$$

To prove Proposition 4.2 we will use the trivial identities

$$a_1^\dagger(\mathcal{N}_1 + \mathcal{N}_2)a_2 + a_2^\dagger(\mathcal{N}_1 + \mathcal{N}_2)a_1 = (\mathcal{N}_1 + \mathcal{N}_2 - 1)(a_1^\dagger a_2 + a_2^\dagger a_1), \tag{4-11}$$

$$\mathcal{N}_1^2 + \mathcal{N}_2^2 = \frac{(\mathcal{N}_1 + \mathcal{N}_2)^2}{2} + \frac{(\mathcal{N}_1 - \mathcal{N}_2)^2}{2}, \quad \mathcal{N}_1\mathcal{N}_2 = \frac{(\mathcal{N}_1 + \mathcal{N}_2)^2}{4} - \frac{(\mathcal{N}_1 - \mathcal{N}_2)^2}{4}, \tag{4-12}$$

as well as the following lemma.

Lemma 4.3 (an identity in the two-mode subspace).

$$(a_1^\dagger a_2)^2 + (a_2^\dagger a_1)^2 + 2\mathcal{N}_1\mathcal{N}_2 = 2(\mathcal{N}_1 + \mathcal{N}_2)(a_1^\dagger a_2 + a_2^\dagger a_1) - (\mathcal{N}_1 + \mathcal{N}_2)^2 + 4\mathcal{N}_-^2 - (\mathcal{N}_1 + \mathcal{N}_2). \tag{4-13}$$

The proof, a simple computation based on the CCR, is in Appendix B.

Proof of Proposition 4.2. We start by proving (4-9), which is actually just another way of writing (3-3). First, notice that, due to the fact that

$$u_1(-x_1, x_2, \dots, x_d) = u_2(x_1, x_2, \dots, x_d),$$

and since $h = -\Delta + V_{\mathrm{DW}}$ involves a symmetric potential V_{DW} with respect to reflexion about the x_1-axis and since $w(x, y) = w(|x - y|)$, we have the relations

$$h_{11} = h_{22}, \quad w_{1111} = w_{2222}, \quad w_{1112} = w_{2221}.$$

Moreover, since we work with a basis of real-valued functions and $w(x - y) = w(y - x)$, we have

$$h_{12} = h_{21}, \quad w_{mnpq} = w_{mqpn} = w_{pnmq} = w_{nmqp}.$$

Using these relations in (3-3) and collecting all terms, we first rewrite (3-3) as

$$\begin{aligned} H_{\text{2-mode}} = {} & h_{11}(\mathcal{N}_1 + \mathcal{N}_2) + h_{12}(a_1^\dagger a_2 + a_2^\dagger a_1) \\ & + \frac{\lambda}{2(N-1)} w_{1111}(\mathcal{N}_1^2 + \mathcal{N}_2^2 - \mathcal{N}_1 - \mathcal{N}_2) \\ & + \frac{\lambda}{N-1} w_{1112}(a_1^\dagger\mathcal{N}_1 a_2 + a_2^\dagger\mathcal{N}_1 a_1 + a_2^\dagger\mathcal{N}_2 a_1 + a_1^\dagger\mathcal{N}_2 a_2) \\ & + \frac{\lambda}{2(N-1)} w_{1122}[(a_1^\dagger a_2)^2 + (a_2^\dagger a_1)^2 + 2\mathcal{N}_1\mathcal{N}_2] + \frac{\lambda}{N-1} w_{1212}\mathcal{N}_1\mathcal{N}_2. \end{aligned}$$

Moreover, using the identities (4-11), (4-12), Lemma 4.3, and the definition of U from (4-3), we find

$$\begin{aligned} H_{\text{2-mode}} = {} & \Big(h_{11} - \frac{\lambda}{2(N-1)}(w_{1111}+w_{1122})\Big)(\mathcal{N}_1+\mathcal{N}_2) \\ & + \frac{\lambda}{4(N-1)}(w_{1111}-2w_{1122}+w_{1212})(\mathcal{N}_1+\mathcal{N}_2)^2 \\ & + \Big(h_{12} + \frac{\lambda}{N-1}w_{1112}(\mathcal{N}_1+\mathcal{N}_2-1) + \frac{\lambda}{N-1}w_{1122}(\mathcal{N}_1+\mathcal{N}_2)\Big)(a_1^\dagger a_2 + a_2^\dagger a_1) \\ & + \frac{\lambda U}{(N-1)}(\mathcal{N}_1-\mathcal{N}_2)^2 + \frac{2\lambda}{N-1}w_{1122}\mathcal{N}_-^2. \end{aligned} \tag{4-14}$$

The identity $\mathcal{N}_1+\mathcal{N}_2 = N-\mathcal{N}_\perp$ now yields

$$\begin{aligned} H_{\text{2-mode}} = {} & E_0 + E_N^w - \mu\mathcal{N}_\perp + \frac{\lambda}{4(N-1)}(w_{1111}-2w_{1122}+w_{1212})\mathcal{N}_\perp^2 \\ & + \Big(h_{12} + \lambda w_{1112} + \lambda w_{1122} - \frac{\lambda}{N-1}w_{1112}\mathcal{N}_\perp - \frac{\lambda}{N-1}w_{1122}(\mathcal{N}_\perp-1)\Big)(a_1^\dagger a_2 + a_2^\dagger a_1) \\ & + \frac{\lambda U}{(N-1)}(\mathcal{N}_1-\mathcal{N}_2)^2 + \frac{2\lambda}{N-1}w_{1122}\mathcal{N}_-^2, \end{aligned}$$

where E_0 and E_N^w are defined by (4-2) and (4-3), respectively. The constant term $E_0 + E_N^w$ comes from the substitution $\mathcal{N}_1+\mathcal{N}_2 \rightsquigarrow N$ in the first two lines of (4-14). The third term $-\mu\mathcal{N}_\perp$ is the contribution coming from substituting $\mathcal{N}_1+\mathcal{N}_2 \rightsquigarrow -\mathcal{N}_\perp$ and $(\mathcal{N}_1+\mathcal{N}_2)^2 \rightsquigarrow -2N\mathcal{N}_\perp$ in the same lines. The proof of (4-9) is completed by recognizing that the main part of the coefficient of $a_1^\dagger a_2 + a_2^\dagger a_1$ is

$$h_{12}+\lambda w_{1112}+\lambda w_{1122} = \big\langle u_1, \big(-\Delta + V_{\text{DW}} + \tfrac{1}{2}\lambda w*(u_1^2+u_2^2) + \lambda w*(u_1u_2)\big)u_2\big\rangle = \langle u_1, h_{\text{MF}}u_2\rangle = \frac{\mu_+ - \mu_-}{2},$$

having used (2-11) to reconstruct $w*|u_+|^2$. This shows that the operator multiplying $(a_1^\dagger a_2 + a_2^\dagger a_1)$ is the operator $\mathcal{T}$ defined in (4-1), thus proving (4-9).

Let us now prove the lower bound (4-10). We will do so by considering all terms in (4-9) and estimating them from below. The main observation is that since $\mu_+ - \mu_- < 0$, we can use the operator inequalities

$$-N \leqslant a_1^\dagger a_2 + a_2^\dagger a_1 \leqslant \mathcal{N}_1+\mathcal{N}_2 = N-\mathcal{N}_\perp \leqslant N.$$

Thus the term $\mathcal{T}(a_1^\dagger a_2 + a_2^\dagger a_1)$ satisfies

$$\begin{aligned} \mathcal{T}(a_1^\dagger a_2 + a_2^\dagger a_1) &= \Big(\frac{\mu_+ - \mu_-}{2} - \frac{\lambda w_{1112}}{N-1}\mathcal{N}_\perp - \frac{\lambda w_{1122}}{N-1}(\mathcal{N}_\perp - 1)\Big)(a_1^\dagger a_2 + a_2^\dagger a_1) \\ &\geqslant -N\left|\frac{\mu_+ - \mu_-}{2} + \frac{\lambda w_{1122}}{N-1}\right| - \frac{\lambda N}{N-1}|w_{1112}+w_{1122}|\mathcal{N}_\perp, \end{aligned} \tag{4-15}$$

where we used that if two operators A and B commute, $z \in \mathbb{C}$, and $-N \leqslant A \leqslant N$ then $zAB \geqslant -|z|NB$. The first absolute value in the right-hand side is smaller than $(\mu_- - \mu_+)/2$ because $\mu_- - \mu_+ \geqslant c_\varepsilon T^{1+\varepsilon} > 0$ by Theorem A.1, $0 < w_{1122} \leqslant C_\varepsilon T^{2-\varepsilon}$ by (4-5), and $T \ll 1$. Furthermore, due to (4-6) the second absolute value is bounded by $C_\varepsilon T^{1-\varepsilon}$. Thus

$$\mathcal{T}(a_1^\dagger a_2 + a_2^\dagger a_1) \geqslant N\frac{\mu_+ - \mu_-}{2} - C_\varepsilon T^{1-\varepsilon}\mathcal{N}_\perp. \tag{4-16}$$

In order to bound the other terms in (4-9) from below, we first notice that, since $w_{1122} \geqslant 0$,

$$\frac{2\lambda}{N-1} w_{1122}\mathcal{N}_-^2 \geqslant 0. \tag{4-17}$$

For the term $-\mu\mathcal{N}_\perp$ we use (4-8) to write

$$-\mu\mathcal{N}_\perp \geqslant -\mu_+\mathcal{N}_\perp - C_\varepsilon T^{1-\varepsilon}\mathcal{N}_\perp. \tag{4-18}$$

The only term left is that proportional to $\mathcal{N}_\perp^2$. Thanks to the positivity of w_{1111} and w_{1212}, using (4-6) and $\mathcal{N}_\perp \leqslant N$, we have

$$\frac{\lambda}{4(N-1)}(w_{1111} - 2w_{1122} + w_{1212})\mathcal{N}_\perp^2 \geqslant -\frac{\lambda}{2(N-1)} w_{1122}\mathcal{N}_\perp^2 \geqslant -C_\varepsilon T^{2-\varepsilon}\mathcal{N}_\perp. \tag{4-19}$$

Plugging (4-16), (4-17), (4-18), and (4-19) into (4-9) gives (4-10). □

4B. ***Upper bound for $H_{\text{2-mode}}$.*** Let us define the trial function

$$\psi_{\text{gauss}} := \sum_{\substack{-\sigma_N^2 \leqslant d \leqslant \sigma_N^2 \\ N+d \text{ is even}}} c_d\, u_1^{\otimes (N+d)/2} \otimes_{\text{sym}} u_2^{\otimes (N-d)/2}, \tag{4-20}$$

where the symmetrized tensor products are normalized in the above and c_d are gaussian coefficients,

$$c_d := \frac{1}{Z_N} e^{-d^2/4\sigma_N^2}, \quad |d| \leqslant \sigma_N^2, \tag{4-21}$$

with σ_N a variance parameter to be fixed later, such that $1 \leqslant \sigma_N \ll N^{1/2}$, and Z_N a normalization factor ensuring $\|\psi_{\text{gauss}}\| = 1$. We will prove:

Proposition 4.4 (upper bound for $H_{\text{2-mode}}$). *Assume that $T \sim N^{-\delta}$ for some $\delta > 0$. Then, with the choice*

$$\sigma_N^2 = \begin{cases} \sqrt{\mu_- - \mu_+}\, N & \text{if } \delta < 2, \\ C & \text{otherwise}, \end{cases} \tag{4-22}$$

with $C \geqslant 1$ a fixed constant, the trial state ψ_{gauss} defined in (4-20) *satisfies*

$$\langle H_{\text{2-mode}}\rangle_{\psi_{\text{gauss}}} \leqslant E_0 + E_N^w + N\frac{\mu_+ - \mu_-}{2} + C_\varepsilon \max(T^{1/2-\varepsilon}, N^{-1+\varepsilon\delta}). \tag{4-23}$$

We start by computing expectation values with respect to the distribution $|c_d|^2$.

Lemma 4.5 (expectation values for the gaussian trial state). *Let c_d be defined by* (4-21) *if $N+d$ is even and $c_d := 0$ if $N+d$ is odd, where $1 \leqslant \sigma_N \leqslant CN^{1/2}$ and Z_N is fixed so that $\sum_{|d| \leqslant \sigma_N^2} |c_d|^2 = 1$. Then*:

- ***Moments.*** *For any $n \in \mathbb{N}$ we have*

$$\sum_{-\sigma_N^2 \leqslant d \leqslant \sigma_N^2} d^{2n}|c_d|^2 \leqslant C\sigma_N^{2n}, \qquad \sum_{-\sigma_N^2 \leqslant d \leqslant \sigma_N^2} d^{2n+1}|c_d|^2 = 0. \tag{4-24}$$

- ***Tunneling term.*** *For any $\kappa \in \mathbb{Z}$,*

$$\left| \sum_{-\sigma_N^2 \leqslant d \leqslant \sigma_N^2 - \kappa} c_d c_{d+\kappa} - 1 \right| \leqslant \frac{C}{\sigma_N^2}. \tag{4-25}$$

Proof. The equality in (4-24) is trivial because of the odd symmetry $d \mapsto -d$. To prove the inequality in (4-24), we note that if $f(x)$ is a differentiable function in $L^1([0,\infty[)$ having a single relative extremum at x_{m}, which is a maximum, then

$$\sum_{0 \leqslant d \leqslant \sigma_N^2} f(d) \leqslant \int_0^\infty f(x)\,dx + f(\lfloor x_{\mathrm{m}} \rfloor) + f(\lfloor x_{\mathrm{m}} \rfloor + 1),$$

where $\lfloor x \rfloor$ denotes the integer part of x. Taking $f(d) = d^{2n} e^{-d^2/2\sigma_N^2}$, which is maximum at $x_{\mathrm{m}} = \sqrt{2n}\,\sigma_N$, we deduce that

$$\sum_{0 \leqslant d \leqslant \sigma_N^2} f(d) \leqslant \sigma_N^{2n+1} \int_0^\infty u^{2n} e^{-u^2/2}\,du + C\sigma_N^{2n}. \tag{4-26}$$

The desired result then follows from the even symmetry $d \mapsto -d$ and from the following lower bound on Z_N:

$$Z_N^2 = \sum_{\substack{|d| \leqslant \sigma_N^2 \\ N+d \text{ is even}}} e^{-d^2/2\sigma_N^2} \geqslant \sum_{\substack{|d| \leqslant \sigma_N \\ N+d \text{ is even}}} e^{-d^2/2\sigma_N^2} \geqslant \sigma_N e^{-1/2}. \tag{4-27}$$

Let us prove (4-25). We have

$$c_d c_{d+\kappa} = c_d^2 e^{-(2\kappa d + \kappa^2)/4\sigma_N^2}.$$

Using the inequality $0 \leqslant e^{-x} - 1 + x \leqslant Cx^2$ valid for any $x \in [-\log(2C), \log(2C)]$ and extending for convenience the definition (4-21) of c_d for $d = \sigma_N^2 + 1, \dots, \sigma_N^2 + \kappa$, we get

$$0 \leqslant \sum_{|d| \leqslant \sigma_N^2} \left(c_d c_{d+\kappa} - c_d^2 + \frac{2\kappa d + \kappa^2}{4\sigma_N^2} c_d^2 \right) \leqslant C \sum_{|d| \leqslant \sigma_N^2} \frac{(2\kappa d + \kappa^2)^2}{16\sigma_N^4} c_d^2 \leqslant \frac{C}{\sigma_N^2},$$

where the last step follows from the estimates in (4-24) proven above. Recalling that $\sum_{|d| \leqslant \sigma_N^2} c_d^2 = 1$, this gives

$$\left| \sum_{|d| \leqslant \sigma_N^2} c_d c_{d+\kappa} - 1 \right| \leqslant \frac{C}{\sigma_N^2}$$

from which we obtain

$$\left| \sum_{-\sigma_N^2 \leqslant d \leqslant \sigma_N^2 - \kappa} c_d c_{d+\kappa} - 1 \right| \leqslant \left| \sum_{|d| \leqslant \sigma_N^2} c_d c_{d+\kappa} - 1 \right| + \frac{C}{Z_N^2} e^{-\sigma_N^2/2} \leqslant \frac{C}{\sigma_N^2}.$$

This proves (4-25). □

Proof of Proposition 4.4. We take the trial state ψ_{gauss} from (4-20) with $1 \leqslant \sigma_N \ll N^{1/2}$ to be suitably optimized at the end. We will compute the expectation value of all terms in (4-9) on ψ_{gauss}. First of all, notice that

$$\mathcal{N}_\perp \psi_{\text{gauss}} = 0,$$

which allows us to neglect all $\mathcal{N}_\perp$ and $\mathcal{N}_\perp{}^2$ terms in (4-9). Hence,

$$\langle H_{2\text{-mode}}\rangle_{\psi_{\mathrm{gauss}}} = E_0 + E_N^w + \left(\frac{\mu_+ - \mu_-}{2} + \frac{\lambda}{N-1} w_{1122}\right)\langle a_1^\dagger a_2 + a_2^\dagger a_1\rangle_{\psi_{\mathrm{gauss}}} + \frac{\lambda U}{N-1}\langle(\mathcal{N}_1 - \mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gauss}}} + \frac{2\lambda}{N-1} w_{1122}\langle \mathcal{N}_-^2\rangle_{\psi_{\mathrm{gauss}}}. \tag{4-28}$$

Let us evaluate the three expectation values on the right-hand side. We have

$$\langle a_1^\dagger a_2 + a_2^\dagger a_1\rangle_{\psi_{\mathrm{gauss}}} = 2 \sum_{-\sigma_N^2 \leqslant d \leqslant \sigma_N^2 - 2} c_d c_{d+2} \sqrt{\frac{N+d+2}{2}\frac{N-d}{2}}.$$

Since $|d| \leqslant \sigma_N^2 \ll N$, we can expand the square root around $d = 0$. We get

$$\left|\langle a_1^\dagger a_2 + a_2^\dagger a_1\rangle_{\psi_{\mathrm{gauss}}} - N \sum_{-\sigma_N^2 \leqslant d \leqslant \sigma_N^2 - 2} c_d c_{d+2}\right| \leqslant N \sum_{-\sigma_N^2 \leqslant d \leqslant \sigma_N^2 - 2} c_d c_{d+2} \left|\sqrt{1 + \frac{2}{N} - \frac{d^2}{N^2} - \frac{2d}{N^2}} - 1\right|$$
$$\leqslant N \sum_{-\sigma_N^2 \leqslant d \leqslant \sigma_N^2 - 2} c_d c_{d+2} \left|\frac{2}{N} - \frac{d^2}{N^2} - \frac{2d}{N^2}\right|. \tag{4-29}$$

We distinguish between two cases:

- If $1 \leqslant \sigma_N^2 \leqslant 2\sqrt{N}$ the second line of (4-29) is bounded by a constant. Indeed
$$\left|\frac{2}{N} - \frac{d^2}{N^2} - \frac{2d}{N^2}\right| \leqslant \frac{3}{N} \quad \text{for } |d| \leqslant 2\sqrt{N}$$
and
$$c_d c_{d+2} \leqslant e\, c_d^2 \quad \text{for } |d| \leqslant \sigma_N^2,$$
and we recall that $\sum_{|d| \leqslant \sigma_N^2} c_d^2 = 1$.
- If $\sigma_N^2 > 2\sqrt{N}$, we split the sum in the second line of (4-29) into a sum running from $-2\sqrt{N}$ to $2\sqrt{N}$ and a remaining sum. Taking advantage of the last two bounds, the expression in this second line is less than
$$C \sum_{|d| \leqslant 2\sqrt{N}} c_d^2 + NC \sum_{2\sqrt{N} < |d| \leqslant \sigma_N^2} c_d^2.$$

The first sum in the right-hand side is bounded by 1. The second sum can be bounded as follows. Setting $d_N = \lfloor 2\sqrt{N} \rfloor$, we have

$$\sum_{2\sqrt{N} < |d| \leqslant \sigma_N^2} c_d^2 = \frac{2}{Z_N^2} \sum_{2\sqrt{N} < d \leqslant \sigma_N^2} \exp\left\{-\frac{(d-d_N)^2}{2\sigma_N^2} - \frac{d d_N}{\sigma_N^2} + \frac{d_N^2}{2\sigma_N^2}\right\}$$
$$\leqslant \frac{2}{Z_N^2} \exp\left\{-\frac{d_N^2}{2\sigma_N^2}\right\} \sum_{0 \leqslant d' \leqslant \sigma_N^2} \exp\left\{-\frac{(d')^2}{2\sigma_N^2}\right\} \leqslant 2e^{-N/\sigma_N^2}.$$

Hence, in all cases one has

$$\left|\langle a_1^\dagger a_2 + a_2^\dagger a_1\rangle_{\psi_{\mathrm{gauss}}} - N \sum_{-\sigma_N^2 \leqslant d \leqslant \sigma_N^2 - 2} c_d c_{d+2}\right| \leqslant C + CNe^{-N/\sigma_N^2}. \tag{4-30}$$

Combining this result with (4-25), we get

$$|\langle a_1^\dagger a_2 + a_2^\dagger a_1\rangle_{\psi_{\text{gauss}}} - N| \leqslant C + \frac{CN}{\sigma_N^2} + CNe^{-N/\sigma_N^2}. \tag{4-31}$$

For the variance term in (4-28) we immediately have, using (4-24),

$$\langle (\mathcal{N}_1 - \mathcal{N}_2)^2\rangle_{\psi_{\text{gauss}}} = \sum_{|d|\leqslant\sigma_N^2} d^2|c_d|^2 \leqslant C\sigma_N^2. \tag{4-32}$$

Finally, since $\mathcal{N}_-^2 \leqslant N\mathcal{N}_-$ on $\mathfrak{H}^N$ and $\mathcal{N}_- = (\mathcal{N}_1 + \mathcal{N}_2 - a_1^\dagger a_2 - a_2^\dagger a_1)/2$, we have by (4-31)

$$\langle \mathcal{N}_-^2\rangle_{\psi_{\text{gauss}}} \leqslant \frac{N}{2}(N - \langle a_1^\dagger a_2 + a_2^\dagger a_1\rangle_{\psi_{\text{gauss}}}) \leqslant CN\left(1 + \frac{N}{\sigma_N^2} + Ne^{-N/\sigma_N^2}\right). \tag{4-33}$$

Plugging (4-31), (4-32), and (4-33) into (4-28), and recalling the estimates (4-4), (4-6), and (4-7) for the w_{mnpq} coefficients and our assumption $1 \leqslant \sigma_N \ll N^{1/2}$, we find

$$\langle H_{2\text{-mode}}\rangle_{\psi_{\text{gauss}}} \leqslant E_0 + E_N^w + N\frac{\mu_+ - \mu_-}{2} + C(\mu_- - \mu_+ + C_\varepsilon T^{2-\varepsilon})\left(\frac{N}{\sigma_N^2} + Ne^{-N/\sigma_N^2}\right) + C\frac{\sigma_N^2}{N}. \tag{4-34}$$

We now optimize the remainder terms by choosing σ_N^2 as in (4-22). Since we assume $T \sim N^{-\delta}$ for some $\delta > 0$ we have from (A-4)

$$Ne^{-N/\sigma_N^2} \leqslant CN^{-\eta}$$

for any $\eta > 0$, showing that the exponential term in (4-34) is much smaller than N/σ_N^2. Using again (4-22) and (A-4), the two last terms in (4-34) are bounded by $C_\varepsilon T^{1/2-\varepsilon}$ if $0 < \delta < 2$ and by $C_\varepsilon T^{1-\varepsilon}N + CN^{-1} \sim C_\varepsilon N^{-(\delta-1)+\varepsilon\delta} + CN^{-1}$ if $\delta \geqslant 2$. The claimed bounds then follow from

$$\max(T^{1/2-\varepsilon}, N^{-1+\varepsilon\delta}) = \begin{cases} T^{1/2-\varepsilon} & \text{if } 0 < \delta < 2, \\ N^{-1+\varepsilon\delta} & \text{if } \delta \geqslant 2. \end{cases} \qquad \square$$

4C. ***Bose–Hubbard energy and proof of Proposition 2.4.*** The next result of this section will allow us to recover the Bose–Hubbard energy, which is the lowest energy of the Bose–Hubbard Hamiltonian (2-24), in terms of quantities appearing in the bounds for $H_{2\text{-mode}}$.

Proposition 4.6 (Bose–Hubbard energy). *Let E_{BH} be the bottom of the spectrum of the Bose Hubbard Hamiltonian H_{BH} defined in* (2-24) *on the N-body two-mode space $\bigotimes_{\text{sym}}^N(PL^2(\mathbb{R}^d))$. Then*

$$\left|E_{\text{BH}} - \left(\frac{\lambda N^2}{4(N-1)}w_{1111} - \frac{\lambda N}{2(N-1)}w_{1111} + (\mu_+ - \mu_-)\frac{N}{2}\right)\right| \leqslant C_\varepsilon \max(T^{1/2-\varepsilon}, N^{-1+\varepsilon\delta}). \tag{4-35}$$

Proof. Since H_{BH} is defined on $\bigotimes_{\text{sym}}^N(PL^2(\mathbb{R}^d))$ only, we can plug $\mathcal{N}_1 + \mathcal{N}_2 = N$ (i.e., $\mathcal{N}_\perp = 0$) into (2-24). This gives

$$H_{\text{BH}} = \frac{\lambda}{2(N-1)}\left(\frac{N^2}{2} - N\right)w_{1111} + \frac{\mu_+ - \mu_-}{2}(a_1^\dagger a_2 + a_2^\dagger a_1) + \frac{\lambda w_{1111}}{4(N-1)}(\mathcal{N}_1 - \mathcal{N}_2)^2.$$

We then repeat the proof of (4-10) and (4-23) on this simplified Hamiltonian. This gives

$$E_{\mathrm{BH}} \leqslant \langle H_{\mathrm{BH}} \rangle_{\psi_{\mathrm{gauss}}} \leqslant \frac{\lambda N^2}{4(N-1)} w_{1111} - \frac{\lambda N}{2(N-1)} w_{1111} + (\mu_+ - \mu_-)\frac{N}{2} + C_\varepsilon \max(T^{1/2-\varepsilon}, N^{-1+\varepsilon\delta})$$

and

$$H_{\mathrm{BH}} \geqslant \frac{\lambda N^2}{4(N-1)} w_{1111} - \frac{\lambda N}{2(N-1)} w_{1111} + (\mu_+ - \mu_-)\frac{N}{2},$$

which completes the proof. □

Proof of Proposition 2.4. Recall Definition (2-22). We deduce from Proposition 4.4 that

$$E_{2\text{-mode}} \leqslant E_0 + E_N^w + N\frac{\mu_+ - \mu_-}{2} + C_\varepsilon \max(T^{1/2-\varepsilon}, N^{-1+\varepsilon\delta}). \tag{4-36}$$

Since the ground state of $H_{2\text{-mode}}$ entirely lives in the two-mode subspace, for a matching lower bound we may set $\mathcal{N}_\perp = 0$ in (4-10). Thus, recalling that $U \geqslant 0$, we deduce from Proposition 4.2 that

$$E_{2\text{-mode}} \geqslant E_0 + E_N^w + N\frac{\mu_+ - \mu_-}{2}.$$

Let us set

$$\widetilde{E}_0 = E_0 - \frac{\lambda N^2}{4(N-1)}(4w_{1122} - 2w_{1212}) = Nh_{11} - \frac{\lambda N^2}{4(N-1)}(2w_{1122} - w_{1212}).$$

It follows from the two preceding bounds, Proposition 4.6 and the definition (4-3) of E_N^w that

$$\begin{aligned}
&|E_{2\text{-mode}} - \widetilde{E}_0 - E_{\mathrm{BH}}| \\
&\quad\leqslant \left| E_{2\text{-mode}} - E_0 - E_N^w - N\frac{\mu_+ - \mu_-}{2} \right| + \left| -E_{\mathrm{BH}} + N\frac{\mu_+ - \mu_-}{2} + E_N^w + E_0 - \widetilde{E}_0 \right| \\
&\quad\leqslant C_\varepsilon \max(T^{1/2-\varepsilon}, N^{-1+\varepsilon\delta}) + \left| -\frac{\lambda N^2}{4(N-1)} w_{1111} + \frac{\lambda N}{2(N-1)} w_{1111} + E_N^w + \frac{\lambda N^2}{4(N-1)}(4w_{1122} - 2w_{1212}) \right| \\
&\quad\leqslant C_\varepsilon \max(T^{1/2-\varepsilon}, N^{-1+\varepsilon\delta}) + \frac{\lambda N}{2(N-1)} w_{1122}.
\end{aligned}$$

Proposition 2.4 follows by using Lemma 4.1 again. □

5. Derivation of the Bogoliubov Hamiltonian and reduction to right and left modes

The aim of this section is two-fold: we will prove that the Bogoliubov Hamiltonian $\mathbb{H}$ from (3-18) is the leading contribution to $H_N - H_{2\text{-mode}}$, and we will show that $\mathbb{H}$ can be decomposed into the two quadratic Hamiltonians $\mathbb{H}_{\mathrm{right}}$ and $\mathbb{H}_{\mathrm{left}}$ from (3-21) and (3-22). The most delicate part of this program is the fact that there are terms in H_N that contain *exactly one* $a_m^\sharp$ with $m \geqslant 3$, but that are not a priori negligible. We keep track of them in Proposition 5.1, and we will show that they are negligible at a later stage.

Let us state the two main results.

Proposition 5.1 (derivation of the Bogoliubov Hamiltonian). *For any excitation vector $\Phi \in \ell^2(\mathfrak{F}_\perp)$ of the form $\Phi = \mathcal{U}_N \psi$ for some $\psi \in \mathfrak{H}^N$, we have*

$$\begin{aligned}\Bigg| \langle \mathcal{U}_N (H_N - H_{2\text{-mode}}) \mathcal{U}_N^* \rangle_\Phi - \langle \mathbb{H} \rangle_\Phi - \mu_+ \langle \mathcal{N}_\perp \rangle_\Phi \\ - \frac{\lambda}{\sqrt{2(N-1)}} \Big\langle \sum_{m \geqslant 3} w_{+1-m}\, \Theta a_m \mathfrak{D} + \text{h.c.} \Big\rangle_\Phi - \frac{\lambda}{\sqrt{2(N-1)}} \Big\langle \sum_{m \geqslant 3} w_{+2-m}\, \Theta^{-1} a_m \mathfrak{D} + \text{h.c.} \Big\rangle_\Phi \Bigg| \\ \leqslant \frac{C}{N^{1/4}} \left(\langle \mathcal{N}_\perp^2 + 1 \rangle_\Phi + \left\langle \frac{\mathfrak{D}^2}{N} \right\rangle_\Phi \right) + C_\varepsilon \frac{T^{1-\varepsilon}}{N^{1/4}} \langle \mathcal{N}_- \rangle_{\mathcal{U}_N^* \Phi}^{3/4} \langle \mathcal{N}_\perp^2 \rangle_\Phi^{1/4}. \qquad (5\text{-}1)\end{aligned}$$

While proving the decomposition of $\mathbb{H}$ into right and left modes, we will need to project the problem on the eigenmodes of h_{MF} with index smaller than some $M \in \mathbb{N}$. To this end, we define the spectral projections

$$P_{\leqslant M} := \sum_{1 \leqslant \alpha \leqslant M} (|u_{2\alpha+1}\rangle\langle u_{2\alpha+1}| + |u_{2\alpha+2}\rangle\langle u_{2\alpha+2}|) = \sum_{1 \leqslant \alpha \leqslant M} (|u_{r,\alpha}\rangle\langle u_{r,\alpha}| + |u_{\ell,\alpha}\rangle\langle u_{\ell,\alpha}|) \qquad (5\text{-}2)$$

and

$$P_{>M} := \sum_{\alpha > M} (|u_{2\alpha+1}\rangle\langle u_{2\alpha+1}| + |u_{2\alpha+2}\rangle\langle u_{2\alpha+2}|) = \mathbb{1} - P_{\leqslant M} - |u_+\rangle\langle u_+| - |u_-\rangle\langle u_-|.$$

Let us introduce the versions of the Bogoliubov Hamiltonians $\mathbb{H}_{\text{right}}$ and $\mathbb{H}_{\text{left}}$ in the right and left wells with an energy cutoff, obtained by restricting all sums in (3-21) and (3-22) to indices α, β smaller than M,

$$\begin{aligned}\mathbb{H}_{\text{right}}^{(M)} &:= \mathrm{d}\Gamma(P_{\leqslant M}) \mathbb{H}_{\text{right}} \mathrm{d}\Gamma(P_{\leqslant M}) \\ &= \sum_{1 \leqslant \alpha,\beta \leqslant M} \langle u_{r,\alpha}, (h_{\text{MF}} - \mu_+ + \lambda K_{11}) u_{r,\beta} \rangle b_{r,\alpha}^\dagger b_{r,\beta} + \frac{\lambda}{2} \sum_{1 \leqslant \alpha,\beta \leqslant M} \langle u_{r,\alpha}, K_{11} u_{r,\beta} \rangle (b_{r,\alpha}^\dagger b_{r,\beta}^\dagger + b_{r,\alpha} b_{r,\beta}), \quad (5\text{-}3)\end{aligned}$$

$$\begin{aligned}\mathbb{H}_{\text{left}}^{(M)} &:= \mathrm{d}\Gamma(P_{\leqslant M}) \mathbb{H}_{\text{left}} \mathrm{d}\Gamma(P_{\leqslant M}) \\ &= \sum_{1 \leqslant \alpha,\beta \leqslant M} \langle u_{\ell,\alpha}, (h_{\text{MF}} - \mu_+ + \lambda K_{22}) u_{\ell,\beta} \rangle c_{\ell,\alpha}^\dagger c_{\ell,\beta} + \frac{\lambda}{2} \sum_{1 \leqslant \alpha,\beta \leqslant M} \langle u_{\ell,\alpha}, K_{22} u_{\ell,\beta} \rangle (c_{\ell,\alpha}^\dagger c_{\ell,\beta}^\dagger + c_{\ell,\alpha} c_{\ell,\beta}), \quad (5\text{-}4)\end{aligned}$$

where we recall that the operators K_{11}, K_{22} and K_{12} are defined as

$$\langle v, K_{ii} u \rangle = \tfrac{1}{2} \langle v \otimes u_i, w\, u_i \otimes u \rangle, \quad i = 1, 2, \qquad \langle v, K_{12} u \rangle = \langle v \otimes u_1, w\, u_2 \otimes v \rangle.$$

Proposition 5.2 (reduction to right- and left-mode Hamiltonians). *Consider $\Phi \in \ell^2(\mathfrak{F}_\perp)$ such that*

$$\langle \mathrm{d}\Gamma(h_{\text{MF}} - \mu_+) + \mathcal{N}_\perp^2 + \mathrm{d}\Gamma(h_{\text{MF}} - \mu_+) \mathcal{N}_\perp \rangle_\Phi \leqslant C \qquad (5\text{-}5)$$

for a constant C that does not depend on N. For every energy cutoff Λ, let M_Λ be the largest integer such that $\mu_{2M_\Lambda+2} \leqslant \Lambda$, where $\{\mu_m\}_m$ are the eigenvalues of h_{MF} in increasing order. Then,

$$|\langle \mathbb{H} - \mathbb{H}_{\text{right}}^{(M_\Lambda)} - \mathbb{H}_{\text{left}}^{(M_\Lambda)} - \mathrm{d}\Gamma_\perp(P_{\geqslant M_\Lambda} (h_{\text{MF}} - \mu_+) P_{\geqslant M_\Lambda}) \rangle_\Phi| \leqslant C_\Lambda o_N(1) + \frac{C}{(\mu_{2M_\Lambda+2} - \mu_+)^{1/2}}, \qquad (5\text{-}6)$$

where the constant C_Λ does not depend on N.

The results of Propositions 5.1 and 5.2 will enable us to show in the next sections that the expectation value of $H_N - H_{2\text{-mode}}$ in the ground state ψ_{gs} of the N-body Hamiltonian H_N is equal to $\langle \mathbb{H} + \mu_+ \mathcal{N}_\perp \rangle_{\mathcal{U}_N^* \psi_{\text{gs}}}$ up to error terms $o_N(1)$ and, furthermore, that the Bogoliubov Hamiltonian in the last expression can be decomposed as a sum of a "right" and "left" Bogoliubov Hamiltonian up to small errors. Indeed, let us anticipate the following a priori estimates to be proven in Section 6:

$$\langle \mathcal{N}_\perp^2 \rangle_{\psi_{\mathrm{gs}}} \leqslant C, \quad \langle \mathrm{d}\Gamma(h_{\mathrm{MF}} - \mu_+)\mathcal{N}_\perp \rangle_{\psi_{\mathrm{gs}}} \leqslant C, \quad \langle \mathcal{N}_- \rangle_{\psi_{\mathrm{gs}}} \leqslant C_\varepsilon \min\{N, T^{-1-\varepsilon}\},$$

where the constants C and C_ε are independent of N. In particular, taking $\Phi = \mathcal{U}_N \psi_{\mathrm{gs}}$, the second term in the right-hand side of (5-1) is of order $T^{1/2-\varepsilon}$.

To prove Proposition 5.1 we will, in the next three subsections, group the terms in $H_N - H_{2\text{-mode}}$ depending on the number of creation and annihilation operators $a_m^\sharp$ with $m \geqslant 3$ they contain.The proof of Proposition 5.2 is provided in Section 5D.

We first collect a few properties that we will use throughout the section.

Lemma 5.3 (general estimates). (i) *For any functions $f, g, h \in L^2(\mathbb{R}^d)$ we have*

$$\sum_{m \geqslant 3} |\langle f \otimes g, w\, h \otimes u_m \rangle|^2 \leqslant \langle g, |w * (\bar{f}h)|^2 g \rangle \leqslant C \|f\|_2^2 \|g\|_2^2 \|h\|_2^2. \tag{5-7}$$

(ii) *For any two functions $f, g \in L^2(\mathbb{R}^d)$ we have*

$$\sum_{m,n \geqslant 3} |\langle f \otimes g, w\, u_m \otimes u_n \rangle|^2 \leqslant \langle f \otimes g, w^2 f \otimes g \rangle \leqslant C \|f\|_2^2 \|g\|_2^2. \tag{5-8}$$

(iii) *We have the bound*

$$\|w * (u_1 u_2)\|_{L^\infty} = \sup_{x \in \mathbb{R}^d} |w * (u_1 u_2)(x)| \leqslant C_\varepsilon T^{1-\varepsilon}. \tag{5-9}$$

(iv) *The operators K_{11} and K_{22} are positive and trace-class. Moreover*

$$\|K_{12}\|_{\mathrm{op}} \leqslant C_\varepsilon T^{1/2-\varepsilon}. \tag{5-10}$$

Proof. Let us start by proving (5-7). We have

$$\sum_{m \geqslant 3} |\langle f \otimes g, w\, h \otimes u_m \rangle|^2 = \sum_{m \geqslant 3} \langle g, w * (\bar{f}h) |u_m\rangle\langle u_m| w * (\bar{h}f) g \rangle.$$

The first inequality in (5-7) then follows thanks to the operator bound

$$\sum_{m \geqslant 3} |u_m\rangle\langle u_m| \leqslant \mathbb{1}.$$

To pass to the second inequality of (5-7) one uses Young's inequality, recalling that $w \in L^\infty$. A similar argument proves (5-8) as well, using instead the operator bound

$$\sum_{m,n \geqslant 3} |u_m \otimes u_n\rangle\langle u_m \otimes u_n| \leqslant \mathbb{1}.$$

To prove (5-9) we write, recalling that $2u_1u_2 = u_+^2 - u_-^2$ and $w \geqslant 0$,

$$\begin{aligned} \sup_{x \in \mathbb{R}^d} \left| \int_{\mathbb{R}^d} w(x-y) u_1(y) u_2(y)\, dy \right| &\leqslant \frac{1}{2} \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} w(x-y) ||u_+(y)|^2 - |u_-(y)|^2|\, dy \\ &\leqslant C \||u_+|^2 - |u_-|^2\|_{L^1} \leqslant C_\varepsilon T^{1-\varepsilon}, \end{aligned}$$

where the second inequality follows from Young's inequality, while the third one follows from (A-1).

The operators K_{11} and K_{22} are trace-class since they are integral operators with kernels $K_{ii}(x,y) = \frac{1}{2}u_i(x)w(x-y)u_i(y)$ and their trace is equal to

$$\int_{\mathbb{R}^d} K_{ii}(x,x)\,dx = \frac{1}{2}\int_{\mathbb{R}^d} w(0)\,|u_i(x)|^2\,dx = \frac{1}{2}w(0) < \infty \quad \text{for } i,\, j \in \{1,2\}.$$

They are positive because of our assumption that w is of positive type; see (2-1). To prove (5-10) we use the Cauchy–Schwarz inequality to obtain

$$\begin{aligned}\|K_{12}\|_{\mathrm{op}} &= \sup_{u,v\in L^2(\mathbb{R}^d),\ \|u\|=\|v\|=1} |\langle v, K_{12}u\rangle| \leqslant \sup_{\|u\|=\|v\|=1} \iint_{\mathbb{R}^{2d}} |v(x)|u_1(y)w(x-y)u_2(x)|u(y)|\,dx\,dy \\ &\leqslant \sup_{\|u\|=\|v\|=1} \left(\iint_{\mathbb{R}^{2d}} |v(x)|^2 w(x-y)|u(y)|^2\,dx\,dy\right)^{1/2} w_{1212}^{1/2}\end{aligned}$$

and the result then follows from $w \in L^\infty$ and (4-7). □

5A. ***Linear terms.*** The part of the Hamiltonian containing only one $a_m^\sharp$ is, recalling the identities $w_{mnpq} = w_{pnmq} = w_{mqpn} = w_{nmqp}$,

$$\begin{aligned}A_1 = &\sum_{m\geqslant 3}(-\Delta + V_{\mathrm{DW}})_{+m}a_+^\dagger a_m + \text{h.c.} &(5\text{-}11)\\ &+\sum_{m\geqslant 3}(-\Delta + V_{\mathrm{DW}})_{-m}a_-^\dagger a_m + \text{h.c.} &(5\text{-}12)\\ &+\frac{\lambda}{N-1}\sum_{m\geqslant 3} w_{+++m}a_+^\dagger a_+^\dagger a_+ a_m + \text{h.c.} &(5\text{-}13)\\ &+\frac{\lambda}{N-1}\sum_{m\geqslant 3} w_{++-m}a_+^\dagger a_+^\dagger a_- a_m + \text{h.c.} &(5\text{-}14)\\ &+\frac{\lambda}{N-1}\sum_{m\geqslant 3} w_{+-+m}a_+^\dagger a_-^\dagger a_+ a_m + \text{h.c.} &(5\text{-}15)\\ &+\frac{\lambda}{N-1}\sum_{m\geqslant 3} w_{+-m+}a_+^\dagger a_-^\dagger a_m a_+ + \text{h.c.} &(5\text{-}16)\\ &+\frac{\lambda}{N-1}\sum_{m\geqslant 3} w_{+--m}a_+^\dagger a_-^\dagger a_- a_m + \text{h.c.} &(5\text{-}17)\\ &+\frac{\lambda}{N-1}\sum_{m\geqslant 3} w_{+-m-}a_+^\dagger a_-^\dagger a_m a_- + \text{h.c.} &(5\text{-}18)\\ &+\frac{\lambda}{N-1}\sum_{m\geqslant 3} w_{--+m}a_-^\dagger a_-^\dagger a_+ a_m + \text{h.c.} &(5\text{-}19)\\ &+\frac{\lambda}{N-1}\sum_{m\geqslant 3} w_{---m}a_-^\dagger a_-^\dagger a_- a_m + \text{h.c.} &(5\text{-}20)\end{aligned}$$

The main result of this subsection is the following proposition.

Proposition 5.4 (linear terms). *Let $\Phi \in \ell^2(\mathfrak{F}_\perp)$ be such that $\Phi = \mathcal{U}_N\psi$ for some $\psi \in \mathfrak{H}^N$. We have*:

- ***Elimination of subleading terms.***

$$\left|\langle A_1\psi\rangle_\psi - \frac{\lambda}{N-1}\left\langle\left(\sum_{m\geqslant 3}(w_{++-m}a_+^\dagger + w_{+--m}a_-^\dagger)(\mathcal{N}_1-\mathcal{N}_2)a_m + \text{h.c.}\right)\right\rangle_\psi\right| \leqslant \frac{C}{\sqrt{N}}\langle \mathcal{N}_\perp^2+1\rangle_\psi + C_\varepsilon \frac{T^{1-\varepsilon}}{N^{1/4}}\langle\mathcal{N}_-\rangle_\psi^{3/4}\langle\mathcal{N}_\perp^2\rangle_\psi^{1/4}. \tag{5-21}$$

- ***Conjugation with $\mathcal{U}_N$.***

$$\left|\langle \mathcal{U}_N A_1 \mathcal{U}_N^*\rangle_\Phi - \frac{\lambda}{\sqrt{2(N-1)}}\left\langle\sum_{m\geqslant 3} w_{+1-m}\,\Theta a_m\mathfrak{D} + \text{h.c.}\right\rangle_\Phi - \frac{\lambda}{\sqrt{2(N-1)}}\left\langle\sum_{m\geqslant 3} w_{+2-m}\,\Theta^{-1} a_m\mathfrak{D} + \text{h.c.}\right\rangle_\Phi\right| \leqslant \frac{C}{N^{1/4}}\left(\langle\mathcal{N}_\perp^2+1\rangle_\Phi + \left\langle\frac{\mathfrak{D}^2}{N}\right\rangle_\Phi\right) + C_\varepsilon\frac{T^{1-\varepsilon}}{N^{1/4}}\langle\mathcal{N}_-\rangle_{\mathcal{U}_N^*\Phi}^{3/4}\langle\mathcal{N}_\perp^2\rangle_\Phi^{1/4}. \tag{5-22}$$

Some linear terms still appear explicitly in (5-22), of the form

$$\frac{1}{N}a_\pm^\dagger(\mathcal{N}_1-\mathcal{N}_2)a_m, \quad m\geqslant 3.$$

According to the standard prescriptions of Bogoliubov theory ($a_\pm^\sharp \simeq \sqrt{N}$ and $a_m^\sharp \simeq 1$ for $m \geqslant 3$), and using the a priori estimate (6-6), for the variance, this term would not result to be negligible. We will prove that it actually is at a later stage of the proof.

Proof. Let us start with (5-21). The terms (5-11), (5-13), and (5-18) will be considered together (and analogous arguments will hold for (5-12) + (5-15) + (5-20)). Their sum gives

$$\begin{aligned}(5\text{-}11)+(5\text{-}13)+(5\text{-}18) &= \sum_{m\geqslant 3}\left[(-\Delta+V_{\mathrm{DW}})_{+m}a_+^\dagger a_m + \frac{\lambda}{N-1}w_{+++m}a_+^\dagger(\mathcal{N}_+ + \mathcal{N}_-)a_m\right] + \text{h.c.}\\ &\qquad + \frac{\lambda}{N-1}\sum_{m\geqslant 3}[(w_{+-m-}-w_{+++m})a_+^\dagger\mathcal{N}_- a_m] + \text{h.c.}\\ &=: L_1 + L_2.\end{aligned} \tag{5-23}$$

In order to estimate L_1 we write, using $\mathcal{N}_+ + \mathcal{N}_- = N - \mathcal{N}_\perp$ and $w_{+++m} = (w * u_+^2)_{+m}$,

$$L_1 = \sum_{m\geqslant 3}\left[(h_{\mathrm{MF}})_{+m}a_+^\dagger a_m - \frac{\lambda}{N-1}w_{+++m}a_+^\dagger(\mathcal{N}_\perp - 1)a_m\right] + \text{h.c.}$$

But $(h_{\mathrm{MF}})_{+m} = \mu_+\langle u_+, u_m\rangle = 0$ if $m \geqslant 3$ and thus

$$\langle L_1\rangle_\psi = -\frac{\lambda}{N-1}\sum_{m\geqslant 3} w_{+++m}\langle\psi, a_+^\dagger(\mathcal{N}_\perp - 1)a_m\psi\rangle + \text{h.c.}$$

Using the Cauchy–Schwarz inequality twice, inserting (5-7), recalling that $\mathcal{N}_+ \leqslant N$ and $2\mathcal{N}_\perp \leqslant \mathcal{N}_\perp^2 + 1$, we have

$$\begin{aligned}|\langle L_1\rangle_\psi| &\leqslant \frac{C}{N}\biggl[\sum_{m\geqslant 3}|w_{+++m}|^2\biggr]^{1/2}\biggl[\sum_{m\geqslant 3}(\|\mathcal{N}_\perp^{1/2}a_+\psi\|^2\|\mathcal{N}_\perp^{1/2}a_m\psi\|^2 + \|a_+\psi\|^2\|a_m\psi\|^2)\biggr]^{1/2}\\ &\leqslant \frac{C}{N}\biggl[\sum_{m\geqslant 3}(\langle\mathcal{N}_+\mathcal{N}_\perp\rangle_\psi\,\langle\mathcal{N}_\perp\mathcal{N}_m\rangle_\psi + \langle\mathcal{N}_+\rangle_\psi\langle\mathcal{N}_m\rangle_\psi)\biggr]^{1/2} \leqslant \frac{C}{\sqrt{N}}\langle\mathcal{N}_\perp^2+1\rangle_\psi. \qquad (5\text{-}24)\end{aligned}$$

The term L_2 in (5-23) can be rewritten as

$$L_2 = \frac{\lambda}{N-1}\sum_{m\geqslant 3}\langle u_+, w*(|u_-|^2-|u_+|^2)u_m\rangle a_+^\dagger\mathcal{N}_-a_m + \text{h.c.}$$

Hence

$$\begin{aligned}|\langle L_2\rangle_\psi| &\leqslant \frac{C}{N}\biggl[\sum_{m\geqslant 3}|\langle u_+, w*(|u_-|^2-|u_+|^2)u_m\rangle|^2\biggr]^{1/2}\biggl[\sum_{m\geqslant 3}\|\mathcal{N}_-^{1/2}a_+\psi\|^2\|\mathcal{N}_-^{1/2}a_m\psi\|^2\biggr]^{1/2}\\ &\leqslant \frac{C}{N}\langle u_+, (w*(|u_+|^2-|u_-|^2))^2u_+\rangle^{1/2}\langle\mathcal{N}_+\mathcal{N}_-\rangle_\psi^{1/2}\langle\mathcal{N}_\perp\mathcal{N}_-\rangle_\psi^{1/2}\\ &\leqslant C_\varepsilon\frac{T^{1-\varepsilon}}{N^{1/2}}\langle\mathcal{N}_-\rangle_\psi^{1/2}\langle\mathcal{N}_\perp^2\rangle_\psi^{1/4}\langle\mathcal{N}_-^2\rangle_\psi^{1/4}\\ &\leqslant C_\varepsilon\frac{T^{1-\varepsilon}}{N^{1/4}}\langle\mathcal{N}_-\rangle_\psi^{3/4}\langle\mathcal{N}_\perp^2\rangle_\psi^{1/4}.\end{aligned}$$

In the first step we used the Cauchy–Schwarz inequality for the m-sum and for the scalar product. In the second step we used (5-7). In the third one we used Young's inequality, $w \in L^\infty$ and the L^2-bound (A-1), as well as $\mathcal{N}_+ \leqslant N$ and the Cauchy–Schwarz inequality $\langle\mathcal{N}_\perp\mathcal{N}_-\rangle_\psi^2 \leqslant \langle\mathcal{N}_\perp^2\rangle_\psi\langle\mathcal{N}_-^2\rangle_\psi$. In the last step we used $\mathcal{N}_-^2 \leqslant N\mathcal{N}_-$.

Having estimated both L_1 and L_2, we deduce

$$|\langle\psi, ((5\text{-}11)+(5\text{-}13)+(5\text{-}18))\psi\rangle| \leqslant \frac{C}{\sqrt{N}}\langle\mathcal{N}_\perp^2+1\rangle_\psi + C_\varepsilon\frac{T^{1-\varepsilon}}{N^{1/4}}\langle\mathcal{N}_-\rangle_\psi^{3/4}\,\langle\mathcal{N}_\perp^2\rangle_\psi^{1/4}. \qquad (5\text{-}25)$$

Analogous arguments lead to a similar bound for $|\langle\psi, ((5\text{-}12)+(5\text{-}15)+(5\text{-}20))\psi\rangle|$.

The remaining terms in A_1 yield the linear terms in the left-hand side of (5-21). In fact, noticing that $w_{++-m} = w_{-++m} = w_{+-m+}$, and using the identity

$$a_+^\dagger a_- + a_-^\dagger a_+ = \mathcal{N}_1 - \mathcal{N}_2,$$

we find

$$(5\text{-}14)+(5\text{-}16)+(5\text{-}17)+(5\text{-}19) = \frac{\lambda}{N-1}\sum_{m\geqslant 3}(w_{++-m}a_+^\dagger + w_{+--m}a_-^\dagger)(\mathcal{N}_1-\mathcal{N}_2)a_m + \text{h.c.} \qquad (5\text{-}26)$$

The estimate (5-21) is then deduced by merging (5-25) and (5-26).

We now turn to (5-22). Using the definition of u_1 and u_2 in terms of u_+ and u_- (see (2-11)) we can replace $a_+^\sharp$ and $a_-^\sharp$ with linear combinations of $a_1^\sharp$ and $a_2^\sharp$. The action of $\mathcal{U}_N$ on $a_m^\dagger a_n$ is then obtained using (3-17). For example, recalling that $[\mathcal{N}_1, a_m] = [\mathcal{N}_2, a_m] = 0$ for $m \geqslant 3$, and recalling the definition of $\mathfrak{D}$ from (3-13),

$$\begin{aligned}\mathcal{U}_N a_1^\dagger(\mathcal{N}_1 - \mathcal{N}_2) a_m \mathcal{U}_N^* &= \mathcal{U}_N a_1^\dagger a_m \mathcal{U}_N^* \, \mathcal{U}_N (\mathcal{N}_1 - \mathcal{N}_2)\mathcal{U}_N^* \\ &= \Theta \sqrt{\frac{N - \mathcal{N}_\perp + \mathfrak{D} + 1}{2}}\, a_m \mathfrak{D}.\end{aligned}$$

The action of $\mathcal{U}_N$ on the term of (5-22) containing $a_2^\dagger$ is computed analogously, and the same holds for the adjoint operators. Thus, acting with $\mathcal{U}_N$ on the linear terms in the right-hand side of (5-21) and recalling the definition of u_1 and u_2 to re-express the matrix elements of w gives

$$\begin{aligned}&\frac{\lambda}{N-1}\mathcal{U}_N \sum_{m\geqslant 3} (w_{++-m} a_+^\dagger + w_{+--m} a_-^\dagger)(\mathcal{N}_1 - \mathcal{N}_2) a_m \mathcal{U}_N^* + \text{h.c.} \\ &= \frac{\lambda}{\sqrt{2}(N-1)} \sum_{m\geqslant 3} \big[w_{+1-m}\Theta\sqrt{N-\mathcal{N}_\perp+\mathfrak{D}+1} + \text{h.c.} + w_{+2-m}\Theta^{-1}\sqrt{N-\mathcal{N}_\perp-\mathfrak{D}+1}\big]\mathfrak{D} a_m + \text{h.c.} \qquad (5\text{-}27)\end{aligned}$$

The linear terms in (5-22) are obtained by replacing all square roots in the above right-hand side by $\sqrt{N-1}$. We now bound the remainders this operation produces. Consider for example the second line of (5-27), and define

$$R_1 := \frac{\lambda}{\sqrt{2}(N-1)} \sum_{m\geqslant 3} w_{+1-m} \big\langle \Theta\big(\sqrt{N - \mathcal{N}_\perp + \mathfrak{D} + 1} - \sqrt{N-1}\big)\mathfrak{D} a_m \big\rangle_\Phi + \text{h.c.}$$

Proceeding as when estimating $\langle L_1\rangle_\psi$ and $\langle L_2\rangle_\psi$ above, recalling that $[\mathfrak{D}, a_m] = 0$, one obtains

$$|R_1| \leqslant \frac{C}{\sqrt{N}} \bigg(\sum_{m\geqslant 3} |w_{+1-m}|^2\bigg)^{1/2} \langle \mathcal{N}_\perp \mathfrak{D}^2\rangle_\Phi^{1/2} \bigg\langle \Theta\bigg(\sqrt{1 - \frac{\mathcal{N}_\perp}{N-1} + \frac{\mathfrak{D}}{N-1} + \frac{2}{N-1}} - 1\bigg)^2 \Theta^{-1}\bigg\rangle_\Phi^{1/2}.$$

We now use the inequality

$$\bigg(\sqrt{1 + \sum_{j=1}^K X_j} - 1\bigg)^2 \leqslant \bigg(\frac{1}{2}\sum_{j=1}^K X_j\bigg)^2 \leqslant C_K \sum_{j=1}^K X_j^2 \qquad (5\text{-}28)$$

for a collection $X_1, \dots, X_K$ of K mutually commuting self-adjoint operators. Inserting (5-7) and using the Cauchy–Schwarz inequality to get $\langle \mathcal{N}_\perp \mathfrak{D}^2\rangle_\Phi^2 \leqslant \langle \mathcal{N}_\perp^2 \mathfrak{D}^2\rangle_\Phi \langle \mathfrak{D}^2\rangle_\Phi$, we find

$$|R_1| \leqslant \frac{C}{N^{3/4}} \langle \mathcal{N}_\perp^2 \mathfrak{D}^2\rangle_\Phi^{1/4} \bigg\langle \frac{\mathfrak{D}^2}{N}\bigg\rangle_\Phi^{1/4} \bigg(\frac{1}{N}\langle \mathcal{N}_\perp^2 + 1\rangle_\Phi + \bigg\langle \frac{\mathfrak{D}^2}{N}\bigg\rangle_{\Theta^{-1}\Phi}\bigg)^{1/2}.$$

Since $\Phi = \mathcal{U}_N\psi$, we know that

$$\langle \mathcal{N}_\perp^2 \mathfrak{D}^2\rangle_\Phi = \sum_{s,d} s^2 d^2 \|\Phi_{s,d}\|^2 \leqslant N^2 \langle \mathcal{N}_\perp^2\rangle_\Phi.$$

Moreover, the commutation relation $[\mathfrak{D}, \Theta] = \Theta$ implies

$$\langle \mathfrak{D}^2\rangle_{\Theta^{-1}\Phi} = \langle (\Theta\mathfrak{D}\Theta^{-1})^2\rangle_\Phi = \langle (\mathfrak{D} - 1)^2\rangle_\Phi \leqslant 2\langle \mathfrak{D}^2 + 1\rangle_\Phi$$

and we deduce

$$\begin{aligned}|R_1| &\leqslant \frac{C}{N^{1/4}}\langle \mathcal{N}_\perp^2\rangle_\Phi^{1/4}\Big\langle \frac{\mathfrak{D}^2}{N}\Big\rangle_\Phi^{1/4}\bigg(\frac{1}{N}\langle \mathcal{N}_\perp^2+1\rangle_\Phi+\Big\langle \frac{\mathfrak{D}^2}{N}\Big\rangle_\Phi\bigg)^{1/2}\\ &\leqslant \frac{C}{N^{1/4}}\bigg(\langle \mathcal{N}_\perp^2+1\rangle_\Phi+\Big\langle \frac{\mathfrak{D}^2}{N}\Big\rangle_\Phi\bigg).\end{aligned}$$

The remainder for the term in the third line of (5-27) can be treated in the same way, completing the proof of (5-22). □

5B. ***Cubic and quartic terms.*** The part of H_N containing three $a_m^\sharp$ with $m \geqslant 3$ is

$$A_3 := \frac{\lambda}{N-1}\sum_{m,n,p\geqslant 3}[w_{+mnp}a_+^\dagger a_m^\dagger a_n a_p + w_{-mnp}a_-^\dagger a_m^\dagger a_n a_p] + \text{h.c.},$$

while the one containing four is

$$A_4 := \frac{\lambda}{2(N-1)}\sum_{m,n,p,q\geqslant 3} w_{mnpq}a_m^\dagger a_n^\dagger a_p a_q.$$

Proposition 5.5 (cubic and quartic terms). *For any* $\Phi \in \ell^2(\mathfrak{F}_\perp)$ *we have*

$$|\langle \mathcal{U}_N A_3 \mathcal{U}_N^*\rangle_\Phi| \leqslant \frac{C}{\sqrt{N}}\langle \mathcal{N}_\perp^2+1\rangle_\Phi, \tag{5-29}$$

$$|\langle \mathcal{U}_N A_4 \mathcal{U}_N^*\rangle_\Phi| \leqslant \frac{C}{N}\langle \mathcal{N}_\perp^2\rangle_\Phi. \tag{5-30}$$

Proof. To prove (5-30) notice that with the notation (3-6) we have

$$\mathcal{U}_N A_4 \mathcal{U}_N^* = \frac{\lambda}{2(N-1)}\mathrm{d}\Gamma_\perp(w),$$

where w is the operator of multiplication by $w(x-y)$ on $L^2(\mathbb{R}^d)^{\otimes 2}$. Since $w \in L^\infty$, we have

$$\mathcal{U}_N A_4 \mathcal{U}_N^* \leqslant \frac{C}{N}\mathrm{d}\Gamma_\perp(\mathbb{1}\otimes\mathbb{1}) = \frac{C}{N}\mathcal{N}_\perp(\mathcal{N}_\perp-1) \leqslant \frac{C}{N}\mathcal{N}_\perp^2$$

because second quantization preserves operator inequalities. Since $A_4 \geqslant 0$, (5-30) follows.

Let us now prove (5-29). Taking the second quantization of the operator inequality (recall that $w \geqslant 0$)

$$P^\perp\otimes(P^\perp-\varepsilon P_+)wP^\perp\otimes(P^\perp-\varepsilon P_+)+(P^\perp-\varepsilon P_+)\otimes P^\perp w(P^\perp-\varepsilon P_+)\otimes P^\perp \geqslant 0$$

for some $\varepsilon > 0$, we deduce

$$\begin{aligned}\sum_{m,n,p\geqslant 3} w_{+mnp}a_+^\dagger a_m^\dagger a_n a_p + \text{h.c.} &\leqslant \varepsilon\,\mathrm{d}\Gamma_\perp(w*|u_+|^2)\mathcal{N}_+ + \frac{1}{\varepsilon}\sum_{m,n,p,q\geqslant 3} w_{mnpq}a_m^\dagger a_n^\dagger a_p a_q\\ &\leqslant \varepsilon C\mathcal{N}_\perp\mathcal{N}_+ + \frac{1}{\varepsilon}\sum_{m,n,p,q\geqslant 3} w_{mnpq}a_m^\dagger a_n^\dagger a_p a_q.\end{aligned}$$

In the last step we used the inequality $\mathrm{d}\Gamma_\perp(w*|u_+|^2) \leqslant C\mathcal{N}_\perp$, which holds by boundedness of $w*|u_+|^2$. We can use the same arguments for the part of A_3 that contains w_{-mnp}. Adding the two results and multiplying by $\lambda/(N-1)$ we thus obtain

$$A_3 \leqslant \frac{\varepsilon C\lambda}{N-1}\mathcal{N}_\perp(\mathcal{N}_+ + \mathcal{N}_-) + \frac{4}{\varepsilon}A_4.$$

Using the fact that $\mathcal{N}_+ + \mathcal{N}_- \leqslant N$ on $\mathfrak{H}^N$, and then conjugating by $\mathcal{U}_N$, this implies

$$\mathcal{U}_N A_3 \mathcal{U}_N^* \leqslant \varepsilon C\mathcal{N}_\perp + \varepsilon^{-1}C\mathcal{U}_N A_4 \mathcal{U}_N^*$$

and plugging (5-30) in the last term gives

$$\mathcal{U}_N A_3 \mathcal{U}_N^* \leqslant \varepsilon C\mathcal{N}_\perp + \varepsilon^{-1}\frac{C}{N}\mathcal{N}_\perp^2.$$

We optimize this bound by choosing $\varepsilon = N^{-1/2}$. Repeating the same proof with ε replaced by $-\varepsilon$ and with reversed inequalities, this yields

$$-\frac{C}{\sqrt{N}}(\mathcal{N}_\perp + \mathcal{N}_\perp^2) \leqslant \mathcal{U}_N A_3 \mathcal{U}_N^* \leqslant \frac{C}{\sqrt{N}}(\mathcal{N}_\perp + \mathcal{N}_\perp^2).$$

Using also $2\mathcal{N}_\perp \leqslant \mathcal{N}_\perp^2 + 1$, this concludes the proof. □

5C. *Quadratic terms.* The part A_2 of H_N that contains exactly two $a_m^\sharp$ with $m \geqslant 3$ is composed of 24 terms which can be combined together by using the equalities $w_{mnpq} = w_{pnmq} = w_{mqpn} = w_{nmqp}$ and the identities

$$\sum_{m.n\geqslant 3}(w_{imin} + w_{imni})a_m^\dagger a_n = \mathrm{d}\Gamma_\perp(w * |u_i|^2 + 2K_{ii}), \quad i = 1, 2,$$
$$\sum_{m.n\geqslant 3}(w_{1m2n} + w_{1mn2})a_m^\dagger a_n = \mathrm{d}\Gamma_\perp(w * (u_1u_2) + K_{12})$$

to obtain

$$\begin{aligned}A_2 := \sum_{m,n\geqslant 3}(-\Delta + V_{\mathrm{DW}})_{mn}a_m^\dagger a_n + \frac{\lambda}{2(N-1)}\sum_{m,n\geqslant 3}(w_{11mn}a_1^\dagger a_1^\dagger + 2w_{12mn}a_1^\dagger a_2^\dagger + w_{22mn}a_2^\dagger a_2^\dagger)a_m a_n + \mathrm{h.c.}\\ + \frac{\lambda}{N-1}\big(a_1^\dagger a_1 \mathrm{d}\Gamma_\perp(w*|u_1|^2 + 2K_{11}) + a_2^\dagger a_2 \mathrm{d}\Gamma_\perp(w*|u_2|^2 + 2K_{22})\big)\\ + \frac{\lambda}{N-1}\big(a_1^\dagger a_2 \mathrm{d}\Gamma_\perp(w*(u_1u_2) + K_{12}) + \mathrm{h.c.}\big).\end{aligned}$$

The action of $\mathcal{U}_N$ on quadratic terms of the type $a^\dagger a$ was given in Lemma 3.6. To deduce the action of $\mathcal{U}_N$ on terms of the type $a^\dagger a^\dagger aa$ as the ones in A_2, we can always reduce ourselves to terms of type $a^\dagger a$ by commuting operators, as in

$$\mathcal{U}_N a_1^\dagger a_2^\dagger a_m a_n \mathcal{U}_N^* = \mathcal{U}_N a_1^\dagger a_m \mathcal{U}_N^* \mathcal{U}_N a_2^\dagger a_n \mathcal{U}_N^* \quad \text{for } m, n \geqslant 3.$$

This is allowed because for $m, n \geqslant 3$ the operators $a_m^\sharp a_n^\sharp$ commute with $a_1^\sharp$ and $a_2^\sharp$. The same argument holds for terms of the type

$$\mathcal{U}_N a_1^\dagger a_m^\dagger a_2 a_n \mathcal{U}_N^* = \mathcal{U}_N a_1^\dagger a_2 \mathcal{U}_N^* \mathcal{U}_N a_m^\dagger a_n \mathcal{U}_N^*.$$

Arguing in this way to commute operators, one easily deduces the expression

$$\mathcal{U}_N A_2 \mathcal{U}_N^* := \sum_{m,n\geqslant 3} (-\Delta + V_{\mathrm{DW}})_{mn} a_m^\dagger a_n \tag{5-31}$$

$$+\frac{\lambda}{2(N-1)}\Bigg[\sum_{m,n\geqslant 3} w_{11mn}\Theta^2 \sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}+2}{2}}\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}+1}{2}}a_m a_n + \text{h.c.} \tag{5-32}$$

$$+2\sum_{m,n\geqslant 3} w_{12mn}\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}}{2}}\sqrt{\frac{N-\mathcal{N}_\perp-\mathfrak{D}+1}{2}}a_m a_n + \text{h.c.} + \text{h.c.} \tag{5-33}$$

$$+\sum_{m,n\geqslant 3} w_{22mn}\Theta^{-2}\sqrt{\frac{N-\mathcal{N}_\perp-\mathfrak{D}+2}{2}}\sqrt{\frac{N-\mathcal{N}_\perp-\mathfrak{D}+1}{2}}a_m a_n + \text{h.c.} \tag{5-34}$$

$$+(N-\mathcal{N}_\perp+\mathfrak{D})\,d\Gamma_\perp(w*|u_1|^2+2K_{11}) \tag{5-35}$$

$$+(N-\mathcal{N}_\perp-\mathfrak{D})\,d\Gamma_\perp(w*|u_2|^2+2K_{22}) \tag{5-36}$$

$$+2\Theta^2\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}+2}{2}}\sqrt{\frac{N-\mathcal{N}_\perp-\mathfrak{D}}{2}}\,d\Gamma_\perp(w*(u_1u_2)+K_{12}) + \text{h.c.}\Bigg]. \tag{5-37}$$

If we could replace all square roots by $\sqrt{(N-1)/2}$ and $(N-\mathcal{N}_\perp \pm \mathfrak{D})$ by $N-1$, then the expression on the right-hand side would coincide with

$$\begin{aligned}\mathbb{H} + \mu_+\mathcal{N}_\perp := d\Gamma_\perp\Big(-\Delta + V_{\mathrm{DW}} + \frac{\lambda}{2}w*|u_1|^2 + \frac{\lambda}{2}w*|u_2|^2 + \lambda K_{11} + \lambda K_{22}\Big)\\ + \frac{\lambda}{2}\sum_{m,n\geqslant 3}(K_{12}+w*(u_1u_2))_{mn}\Theta^2 a_m^\dagger a_n + \text{h.c.}\\ + \frac{\lambda}{2}\sum_{m,n\geqslant 3}((K_{11})_{mn}\Theta^2 + (K_{22})_{mn}\Theta^{-2} + (K_{12}^*)_{mn})a_m a_n + \text{h.c.};\end{aligned} \tag{5-38}$$

see (3-18). The $\mu_+\mathcal{N}_\perp$ term is there to compensate for a term which we included in the definition of $\mathbb{H}$ but that does not come from $\mathcal{U}_N A_2 \mathcal{U}_N^*$. We will prove the following result, showing that such a replacement can be done at the expense of negligible remainders.

Proposition 5.6 (quadratic terms). *Let $\Phi \in \ell^2(\mathfrak{F}_\perp)$ be such that $\Phi = \mathcal{U}_N\psi$ for some $\psi \in \mathfrak{H}^N$. Then*

$$|\langle \mathcal{U}_N A_2 \mathcal{U}_N^*\rangle_\Phi - \langle \mathbb{H}\rangle_\Phi - \mu_+\langle \mathcal{N}_\perp\rangle_\Phi| \leqslant \frac{C}{\sqrt{N}}\left\langle \frac{\mathcal{N}_\perp^2+\mathfrak{D}^2+1}{N}\right\rangle_\Phi, \tag{5-39}$$

where $\mathbb{H}$ was defined in (3-18).

Proof. The result is proven if we show the following three general estimates:

- **Controlling terms (5-32)–(5-34).** For every $i, k \in \{1, 2\}$, $c_1, c_2 \in \mathbb{Z}$, $j \in \{-2, 0, 2\}$, and $\varepsilon_1, \varepsilon_2 \in \{-1, 1\}$,

$$\begin{aligned}\left|\frac{\lambda}{2(N-1)}\left\langle \sum_{m,n\geqslant 3} w_{ikmn}\Theta^j\left(\sqrt{\frac{N-\mathcal{N}_\perp+\varepsilon_1\mathfrak{D}+c_1}{2}}\sqrt{\frac{N-\mathcal{N}_\perp+\varepsilon_2\mathfrak{D}+c_2}{2}} - \frac{N-1}{2}\right)a_m a_n\right\rangle_\Phi + \text{h.c.}\right|\\ \leqslant \frac{C}{N}\langle \mathcal{N}_\perp^2\rangle_\Phi^{1/2}\left\langle \frac{\mathcal{N}_\perp^4}{N^3} + \frac{\mathfrak{D}^4}{N^3} + \frac{\mathcal{N}_\perp^2}{N} + \frac{\mathfrak{D}^2}{N} + \frac{1}{N}\right\rangle_\Phi^{1/2}.\end{aligned} \tag{5-40}$$

- **Controlling terms (5-35)–(5-36).** For every $i \in \{1, 2\}$,

$$\left|\frac{\lambda}{N-1}\big\langle\big((N-\mathcal{N}_\perp \pm \mathfrak{D})-(N-1)\big)\mathrm{d}\Gamma_\perp(w*|u_i|^2+2K_{ii})\big\rangle_\Phi\right| \leqslant \frac{C}{N}\langle\mathcal{N}_\perp^2+\mathfrak{D}^2+1\rangle_\Phi^{1/2}\langle\mathcal{N}_\perp^2\rangle_\Phi^{1/2}. \quad (5\text{-}41)$$

- **Controlling the last term (5-37).** Finally,

$$\left|\frac{\lambda}{N-1}\left\langle\Theta^2\left(\sqrt{\frac{N-\mathcal{N}_\perp+\mathfrak{D}+2}{2}}\sqrt{\frac{N-\mathcal{N}_\perp-\mathfrak{D}}{2}}-\frac{N-1}{2}\right)\mathrm{d}\Gamma_\perp(w*(u_1u_2)+K_{12})\right\rangle_\Phi+\text{h.c.}\right|$$
$$\leqslant \frac{C}{N}\langle\mathcal{N}_\perp^2\rangle_\Phi^{1/2}\left\langle\frac{\mathcal{N}_\perp^4}{N^3}+\frac{\mathfrak{D}^4}{N^3}+\frac{\mathcal{N}_\perp^2}{N}+\frac{\mathfrak{D}^2}{N}+\frac{1}{N}\right\rangle_\Phi^{1/2}. \quad (5\text{-}42)$$

Let us prove (5-40). We have

$$\left|\frac{\lambda}{2(N-1)}\left\langle\sum_{m,n\geqslant 3} w_{ikmn}\Theta^j\left(\sqrt{\frac{N-\mathcal{N}_\perp+\varepsilon_1\mathfrak{D}+c_1}{2}}\sqrt{\frac{N-\mathcal{N}_\perp+\varepsilon_2\mathfrak{D}+c_2}{2}}-\frac{N-1}{2}\right)a_m a_n\right\rangle_\Phi+\text{h.c.}\right|$$
$$\leqslant \frac{\lambda N}{2(N-1)}\left(\sum_{m,n\geqslant 3}|w_{ikmn}|^2\right)^{1/2}\left(\sum_{m,n\geqslant 3}\|a_m a_n\Phi\|^2\right)^{1/2}$$
$$\times\left\langle\Theta^j\left(\sqrt{1-\frac{\mathcal{N}_\perp}{N}+\varepsilon_1\frac{\mathfrak{D}}{N}+\frac{c_1}{N}}\sqrt{1-\frac{\mathcal{N}_\perp}{N}+\varepsilon_2\frac{\mathfrak{D}}{N}+\frac{c_2}{N}}-1+\frac{1}{N}\right)^2\Theta^{-j}\right\rangle_\Phi^{1/2}$$
$$\leqslant C\langle\mathcal{N}_\perp(\mathcal{N}_\perp-1)\rangle_\Phi^{1/2}\left\langle\Theta^j\left(\frac{\mathcal{N}_\perp^4}{N^4}+\frac{\mathcal{N}_\perp^2}{N^2}+\frac{1}{N^2}+\frac{\mathfrak{D}^2}{N^2}+\frac{\mathfrak{D}^4}{N^4}\right)\Theta^{-j}\right\rangle_\Phi,$$

where in the first step we used the Cauchy–Schwarz inequality for the sum over m, n and for the $\ell^2(\mathfrak{F}_\perp)$ scalar product, and in the second step we used (5-8), the inequality (5-28), the commutation of $\mathcal{N}_\perp$ and $\mathfrak{D}$, and the bound $\mathcal{N}_\perp^2\mathfrak{D}^2 \leqslant N^2\mathfrak{D}^2$. The proof of (5-40) is complete if we show how to get rid of Θ. For the terms containing $\mathcal{N}_\perp^n$ we simply use the fact that $[\Theta, \mathcal{N}_\perp] = 0$ and that Θ is unitary. For the $\mathfrak{D}$ terms we use the identity

$$\Theta\mathfrak{D}\Theta^{-1} = \mathfrak{D}-1,$$

which implies $\Theta\mathfrak{D}^n\Theta^{-1} = (\mathfrak{D}-1)^n$ for each $n \in \mathbb{N}$, and therefore

$$\Theta^2\mathfrak{D}^2\Theta^{-2} = (\mathfrak{D}-2)^2 \leqslant C\mathfrak{D}^2 + C,$$
$$\Theta^2\mathfrak{D}^4\Theta^{-2} \leqslant (\mathfrak{D}-2)^4 \leqslant C(\mathfrak{D}^4+\mathfrak{D}^2+1).$$

This completes the proof of (5-40).

Let us now prove (5-41). We have

$$\left|\frac{\lambda}{N-1}\big\langle\big((N-\mathcal{N}_\perp \pm \mathfrak{D})-(N-1)\big)\mathrm{d}\Gamma_\perp(w*|u_i|^2+K_{ii})\big\rangle_\Phi\right|$$
$$= \left|\frac{\lambda}{N-1}\big\langle\big(-\mathcal{N}_\perp \pm \mathfrak{D}+1\big)\mathrm{d}\Gamma_\perp(w*|u_i|^2+K_{ii})\big\rangle_\Phi\right|$$
$$\leqslant \frac{C}{N}\langle\mathcal{N}_\perp^2+\mathfrak{D}^2+1\rangle_\Phi^{1/2}\langle\mathcal{N}_\perp^2\rangle_\Phi^{1/2},$$

where we used the Cauchy–Schwarz inequality for the $\ell^2(\mathfrak{F}_\perp)$ scalar product, the boundedness of $w*|u_i|^2$ and K_{ii}, and the fact that $|\mathrm{d}\Gamma_\perp(K)| \leqslant \|K\|\mathcal{N}_\perp$ for a bounded one-body operator K.

Finally, one may prove (5-42) in a similar way, using the boundedness of $w*(u_1u_2)$ and K_{12}, inequality (5-28), and commuting Θ with $\mathcal{N}_\perp$ and $\mathfrak{D}$ as done above for (5-40). □

Proposition 5.1 now follows by merging (5-22), (5-29), (5-30), and (5-39), with a rearrangement of the remainder terms.

5D. *Reduction to left and right modes*: *proof of Proposition 5.2.*

Proof of Proposition 5.2. We have the decomposition

$$\mathbb{H} - \mathbb{H}^{(M_\Lambda)}_{\mathrm{right}} - \mathbb{H}^{(M_\Lambda)}_{\mathrm{left}} - \mathrm{d}\Gamma_\perp(P_{>M_\Lambda}(h_{\mathrm{MF}}-\mu_+)P_{>M_\Lambda}) = \mathbb{H}_{12} + \mathbb{K}_{>M_\Lambda} + \sum_{j=1}^{3}\Xi_j, \tag{5-43}$$

where

$$\begin{aligned}\mathbb{H}_{12} := {}&\frac{\lambda}{2}\sum_{m,n\geqslant 3}(w*(u_1u_2))_{mn}(-2+\Theta^2+\Theta^{-2})a^\dagger_m a_n\\ &+\frac{\lambda}{2}\sum_{m,n\geqslant 3}(K_{12})_{mn}\Theta^2 a^\dagger_m a_n + \frac{\lambda}{2}\sum_{m,n\geqslant 3}(K^*_{12})_{mn}\Theta^{-2}a^\dagger_m a_n\\ &\quad+\frac{\lambda}{2}\sum_{m,n\geqslant 3}(K_{12})_{mn}a^\dagger_m a^\dagger_n + \frac{\lambda}{2}\sum_{m,n\geqslant 3}(K^*_{12})_{mn}a_m a_n,\end{aligned} \tag{5-44}$$

$$\mathbb{K}_{>M_\Lambda} := \lambda\sum_{m,n>2M_\Lambda+2}(K_{11}+K_{22})_{mn}a^\dagger_m a_n + \lambda\sum_{\substack{3\leqslant m\leqslant 2M_\Lambda+2\\ n>2M_\Lambda+2}}(K_{11}+K_{22})_{mn}(a^\dagger_m a_n + \mathrm{h.c.}) \tag{5-45}$$

$$\begin{aligned}&+\frac{\lambda}{2}\sum_{m,n>2M_\Lambda+2}[((K_{11})_{mn}\Theta^{-2}+(K_{22})_{mn}\Theta^2)a^\dagger_m a^\dagger_n + \mathrm{h.c.}]\\ &\quad+\lambda\sum_{\substack{3\leqslant m\leqslant 2M_\Lambda+2\\ n>2M_\Lambda+2}}[((K_{11})_{mn}\Theta^{-2}+(K_{22})_{mn}\Theta^2)a^\dagger_m a^\dagger_n + \mathrm{h.c.}]\end{aligned} \tag{5-46}$$

$$\Xi_1 := \sum_{1\leqslant\alpha,\beta\leqslant M_\Lambda}\langle u_{r,\alpha},(h_{\mathrm{MF}}-\mu_+)u_{\ell,\beta}\rangle a^\dagger_{r,\alpha}a_{\ell,\beta} + \mathrm{h.c.}, \tag{5-47}$$

$$\begin{aligned}\Xi_2 := {}&\lambda\sum_{1\leqslant\alpha,\beta\leqslant M_\Lambda}\langle u_{r,\alpha},(K_{11}+K_{22})u_{\ell,\beta}\rangle a^\dagger_{r,\alpha}a_{\ell,\beta} + \mathrm{h.c.}\\ &+\lambda\sum_{1\leqslant\alpha,\beta\leqslant M_\Lambda}[\langle u_{r,\alpha},K_{11}u_{\ell,\beta}\rangle\Theta^{-2} + \langle u_{r,\alpha},K_{22}u_{\ell,\beta}\rangle\Theta^2]a^\dagger_{r,\alpha}a^\dagger_{\ell,\beta} + \mathrm{h.c.},\end{aligned} \tag{5-48}$$

$$\begin{aligned}\Xi_3 := {}&\lambda\sum_{1\leqslant\alpha,\beta\leqslant M_\Lambda}[\langle u_{r,\alpha},K_{22}u_{r,\beta}\rangle a^\dagger_{r,\alpha}a_{r,\beta} + \langle u_{\ell,\alpha},K_{11}u_{\ell,\beta}\rangle a^\dagger_{\ell,\alpha}a_{\ell,\beta}]\\ &+\frac{\lambda}{2}\sum_{1\leqslant\alpha,\beta\leqslant M_\Lambda}[\langle u_{r,\alpha},K_{22}u_{r,\beta}\rangle\Theta^2 a^\dagger_{r,\alpha}a^\dagger_{r,\beta} + \langle u_{\ell,\alpha},K_{11}u_{\ell,\beta}\rangle\Theta^{-2}a^\dagger_{\ell,\alpha}a^\dagger_{\ell,\beta} + \mathrm{h.c.}].\end{aligned} \tag{5-49}$$

Let us briefly explain the rationale behind the above decomposition. First, in view of the definitions of h_{MF} and of the right and left modes $u_{r,\alpha}$ and $u_{\ell,\alpha}$, see (2-6) and (2-26), one has

$$\begin{aligned}\mathrm{d}\Gamma_\perp\Big(-\Delta+V_{\mathrm{DW}}+\frac{\lambda}{2}w*|u_1|^2+\frac{\lambda}{2}w*|u_2|^2-\mu_+\Big)\\ =\sum_{1\leqslant\alpha,\beta\leqslant M_\Lambda}[\langle u_{r,\alpha},(h_{\mathrm{MF}}-\mu_+)u_{r,\beta}\rangle a^\dagger_{r,\alpha}a_{r,\beta}+\langle u_{\ell,m},(h_{\mathrm{MF}}-\mu_+)u_{\ell,n}\rangle a^\dagger_{\ell,m}a_{\ell,n}]\\ +\Xi_1+\mathrm{d}\Gamma_\perp(P_{>M_\Lambda}(h_{\mathrm{MF}}-\mu_+)P_{>M_\Lambda})-\lambda\mathrm{d}\Gamma_\perp(w*(u_1u_2)),\end{aligned}\tag{5-50}$$

where the sum in the first line contains the terms involving $h_{\mathrm{MF}}-\mu_+$ in $\mathbb{H}^{(M_\Lambda)}_{\mathrm{right}}$ and $\mathbb{H}^{(M_\Lambda)}_{\mathrm{left}}$; see (5-3) and (5-4). One can proceed similarly for the terms involving K_{11} and K_{22} in the Bogoliubov Hamiltonian (5-38). Now, we gather in $\mathbb{H}_{12}$ all those terms that involve the operators $w*(u_1u_2)$ and K_{12} (including the last term in (5-50)). For $\mathbb{H}_{12}$ we will prove a cutoff-independent quantitative bound. We then gathered in $\mathrm{d}\Gamma_\perp(P_{>M_\Lambda}(h_{\mathrm{MF}}-\mu_+)P_{>M_\Lambda})$ and $\mathbb{K}_{>M_\Lambda}$ those terms of $\mathbb{H}-\mathbb{H}_{12}$ for which one or two indices m and n are larger than the cutoff M_Λ. We will show that the contribution of $\mathbb{K}_{>M_\Lambda}$ is negligible, while $\mathrm{d}\Gamma_\perp(P_{>M_\Lambda}(h_{\mathrm{MF}}-\mu_+)P_{>M_\Lambda})$, being nonnegative, can be dropped for a lower bound. For the part of $\mathbb{H}-\mathbb{H}_{12}$ in which sums run over modes below the energy cutoff M_Λ, we want to control those terms that contain matrix elements that couple "right" modes with "left" modes. They are of different types, and we collected them in Ξ_1, Ξ_2, and Ξ_3. The remaining terms precisely give $\mathbb{H}^{(M_\Lambda)}_{\mathrm{right}}+\mathbb{H}^{(M_\Lambda)}_{\mathrm{left}}$. We will show that (expectations of) all terms in the right-hand side of (5-43) are controllable in the limit $N\to\infty$ followed by $M\to\infty$.

We first prove that

$$|\langle\mathbb{H}_{12}\rangle_\Phi|\leqslant C_\varepsilon T^{1/2-\varepsilon}\langle\mathcal{N}_\perp^2+1\rangle_\Phi.\tag{5-51}$$

For the first two lines of $\mathbb{H}_{12}$ we write

$$\begin{aligned}I_1=:&\bigg|\Big\langle\frac{\lambda}{2}\sum_{m,n\geqslant3}\big[[(w*(u_1u_2))_{mn}(-2+\Theta^2+\Theta^{-2})+(K_{12})_{mn}\Theta^2+(K^*_{12})_{mn}\Theta^{-2}]a^\dagger_m a_n\big]\Big\rangle_\Phi\bigg|\\ &=\frac{\lambda}{2}|\langle\mathrm{d}\Gamma_\perp(w*(u_1u_2))(-2+\Theta^2+\Theta^{-2})+(\mathrm{d}\Gamma_\perp(K_{12})\Theta^2+\mathrm{h.c.})\rangle_\Phi|\\ &\leqslant\frac{\lambda}{2}\|(-2+\Theta^2+\Theta^{-2})\Phi\|\,\|\mathrm{d}\Gamma_\perp(w*(u_1u_2))\Phi\|+\lambda\|\Theta^2\Phi\|\,\|\mathrm{d}\Gamma_\perp(K^*_{12})\Phi\|.\end{aligned}$$

Recalling that the norms of $w*(u_1u_2)$ and K_{12} were estimated in (5-9) and (5-10), arguing as in Section 5C we find

$$I_1\leqslant C_\varepsilon T^{1/2-\varepsilon}\langle\mathcal{N}_\perp^2\rangle_\Phi.$$

For the other terms of $\mathbb{H}_{12}$ we write

$$I_2=:\bigg|\Big\langle\frac{\lambda}{2}\sum_{m,n\geqslant3}(K_{12})_{mn}a^\dagger_m a^\dagger_n+\mathrm{h.c.}\Big\rangle_\Phi\bigg|\leqslant\lambda\|\Phi\|\,\bigg\|\sum_{m,n\geqslant3}(K_{12})_{mn}a_m a_n\Phi\bigg\|.$$

Since we assumed that all elements of the basis $\{u_m\}_m$ are real-valued functions, we have

$$\langle u_m,K_{12}u_n\rangle\equiv\langle u_m\otimes u_1,w\,u_2\otimes u_n\rangle=\langle u_m\otimes u_n,w\,u_2\otimes u_1\rangle$$

and this gives

$$\Big\|\sum_{m,n\geqslant 3}\langle u_m,\, K_{12}u_n\rangle a_m a_n\Phi\Big\|^2 = \sum_{m,n,p,q\geqslant 3}\langle u_m,\, K_{12}u_n\rangle\langle u_q,\, K_{12}^* u_p\rangle\langle a_p^\dagger a_q^\dagger a_m a_n\rangle_\Phi$$
$$= \sum_{m,n,p,q\geqslant 3}\langle u_m\otimes u_n,\, w\, u_2\otimes u_1\rangle\,\langle u_2\otimes u_1,\, w\, u_p\otimes u_q\rangle\langle a_p^\dagger a_q^\dagger a_m a_n\rangle_\Phi$$
$$= \langle \mathrm{d}\Gamma_\perp(w|u_2\otimes u_1\rangle\langle u_2\otimes u_1|w)\rangle_\Phi.$$

However,

$$\|w|u_1\otimes u_2\rangle\langle u_1\otimes u_2|w\|_{\mathrm{op}}^2 = \sup_{u\in L^2(\mathbb{R}^{2d}),\ \|u\|=1} |\langle u,\, w\, u_1\otimes u_2\rangle|^2\langle u_1\otimes u_2,\ w^2\, u_1\otimes u_2\rangle$$
$$\leqslant \bigg(\int (w(x-y))^2|u_1(x)|^2\,|u_2(y)|^2\,dx\,dy\bigg)^2 \leqslant C_\varepsilon T^{2-\varepsilon},$$

where the last step is due to (4-7). Since the second quantization preserves operator inequalities, we conclude

$$\Big\|\sum_{m,n\geqslant 3}\langle u_m,\, K_{12}u_n\rangle a_m a_n\Phi\Big\|^2 \leqslant C_\varepsilon T^{1-\varepsilon}\langle \mathcal{N}_\perp^2\rangle_\Phi,$$

from which

$$I_2 \leqslant C_\varepsilon T^{1/2-\varepsilon}\langle \mathcal{N}_\perp^2\rangle_\Phi.$$

This completes the proof of (5-51), since the expectation in the right-hand side is uniformly bounded by our assumption (5-5).

We now explain how to bound $\mathbb{K}_{>M_\Lambda}$, focusing, as an example, on the term

$$\mathbb{K}_{>M_\Lambda}^{(1)} := \sum_{\substack{3\leqslant m\leqslant 2M_\Lambda+2\\ n>2M_\Lambda+2}} [(K_{11})_{mn}\Theta^{-2}a_m^\dagger a_n^\dagger + \text{h.c.}].$$

We have

$$|\langle \mathbb{K}_{>M_\Lambda}^{(1)}\rangle_\Phi| \leqslant 2\bigg(\sum_{m,n\geqslant 1}|\langle u_m,\, K_{11}u_n\rangle|^2\bigg)^{1/2}\bigg(\sum_{m\geqslant 3,\ n>2M_\Lambda+2}\|a_n a_m\Phi\|^2\bigg)^{1/2}\|\Theta^{-2}\Phi\|$$
$$\leqslant 2\,\mathrm{Tr}(K_{11}^2)^{1/2}\|\Phi\|\bigg\langle \mathcal{N}_\perp\sum_{n>2M_\Lambda+2}a_n^\dagger a_n\bigg\rangle_\Phi^{1/2}.$$

The first bound follows from the Cauchy–Schwarz inequality both for the sum over m, n and for the $\ell^2(\mathfrak{F}_\perp)$-scalar product. The second one follows from the fact that K_{11} and thus K_{11}^2 are trace-class, as proven in Lemma 5.3, and by commuting $a_n^\dagger a_n$ with a_m and ignoring a negative term coming from the commutator. For the last square root we write

$$\bigg\langle \mathcal{N}_\perp\sum_{n>2M_\Lambda+2}a_n^\dagger a_n\bigg\rangle_\Phi \leqslant \frac{1}{\mu_{2M_\Lambda+2}-\mu_+}\bigg\langle \mathcal{N}_\perp\sum_{n>2M_\Lambda+2}(\mu_n-\mu_+)a_n^\dagger a_n\bigg\rangle_\Phi.$$

We now notice that the sum in the right-hand side satisfies

$$\sum_{n>2M_\Lambda+2}(\mu_n-\mu_+)a_n^\dagger a_n \leqslant \mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+),$$

and since all the operators commute with $\mathcal{N}_\perp$ we can plug this into the expectation value above. We thus find

$$|\langle \mathbb{K}^{(1)}_{>M_\Lambda}\rangle_\Phi| \leqslant C\biggl(\frac{1}{\mu_{2M_\Lambda+2}-\mu_+}\langle \mathcal{N}_\perp \mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+)\rangle_\Phi\biggr)^{1/2}.$$

Since, by the assumptions (5-5) on Φ, the expectation value is bounded uniformly in N, we deduce

$$|\langle \mathbb{K}^{(1)}_{>M_\Lambda}\rangle_\Phi| \leqslant \frac{C}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}}.$$

All the terms in the second and third lines of (5-47) can be estimated in this way. For the terms in the first line the argument is slightly simpler since, arguing as above,

$$\begin{aligned}\biggl|\sum_{\substack{3\leqslant m\leqslant 2M_\Lambda+2\\ n>2M_\Lambda+2}}(K_{11}+K_{22})_{mn}\langle a_m^\dagger a_n+\mathrm{h.c.}\rangle_\Phi\biggr| &\leqslant C\biggl(\sum_{3\leqslant m}\langle a_m^\dagger a_m\rangle_\Phi \sum_{n>2M_\Lambda+2}\langle a_n^\dagger a_n\rangle_\Phi\biggr)^{1/2}\\ &\leqslant C\langle \mathcal{N}_\perp\rangle_\Phi^{1/2}\frac{\langle \mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+)\rangle_\Phi}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}}.\end{aligned}$$

This proves

$$|\langle \mathbb{K}_{>M_\Lambda}\rangle_\Phi| \leqslant \frac{C}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}}. \tag{5-52}$$

We next turn to estimating the Ξ terms in (5-43). Since all sums are finite, it is enough to show that the $L^2(\mathbb{R}^d)$-expectation values multiplying $a^{\sharp_r}_{r,\alpha}a^{\sharp_\ell}_{\ell,\beta}$ in the sums converge to zero as $N\to\infty$ (notice that our assumption (5-5) on Φ ensures that all expectation values in $\ell^2(\mathfrak{F}_\perp)$ are well-defined). For Ξ_1 we notice that

$$\langle u_{r,\alpha}, (h_{\mathrm{MF}}-\mu_+)u_{\ell,\beta}\rangle = \tfrac{1}{2}(\mu_{2\alpha+1}-\mu_{2\alpha+2})\delta_{\alpha,\beta}, \tag{5-53}$$

and therefore, by (A-7), for any $\alpha,\beta\in\{1,\dots,M_\Lambda\}$,

$$\lim_{N\to\infty}\langle u_{r,\alpha}, (h_{\mathrm{MF}}-\mu_+)u_{\ell,\beta}\rangle = 0.$$

The fact that $\langle\Xi_2\rangle_\Phi$ and $\langle\Xi_3\rangle_\Phi$ converge to zero as $N\to\infty$ is a consequence of the localization of $u_{r,\alpha}$ and $u_{\ell,\beta}$ in the right and left wells, respectively. More precisely, for Ξ_2 we notice that, by the definition of K_{11},

$$\begin{aligned}|\langle u_{r,\alpha}, K_{11}u_{\ell,\beta}\rangle| &= \tfrac{1}{2}|\langle u_{r,\alpha}\otimes u_1, w\, u_1\otimes u_{\ell,\beta}\rangle| \leqslant C\langle |u_{\ell,\beta}|, |u_1|\rangle\\ &\leqslant C\biggl(\int_{x_1\geqslant 0}|u_{\ell,\beta}(x)|^2\,dx\biggr)^{1/2} + C\biggl(\int_{x_1\leqslant 0}|u_1(x)|^2\,dx\biggr)^{1/2},\end{aligned} \tag{5-54}$$

and both terms in the right-hand side converge to zero as $N\to\infty$ by (A-8) and (A-9). The expectations of K_{22} in Ξ_2 coincide with those of K_{11} by reflection symmetry, so the same argument applies. For Ξ_3 we argue similarly by noticing that

$$\begin{aligned}|\langle u_{r,\alpha}, K_{22}u_{r,\beta}\rangle| &\leqslant C\langle |u_{r,\alpha}|, |u_2|\rangle\langle |u_{r,\beta}|, |u_2|\rangle,\\ \langle |u_{r,\alpha}|, |u_2|\rangle &\leqslant C\biggl(\int_{x_1\leqslant 0}|u_{r,\beta}(x)|^2\,dx\biggr)^{1/2} + C\biggl(\int_{x_1\geqslant 0}|u_2(x)|^2\,dx\biggr)^{1/2}\end{aligned}$$

and the right-hand side of the second bound converges to zero as $N \to \infty$, once again by (A-8) and (A-9). These arguments prove that, for $i = 1, 2, 3$,

$$|\langle \Xi_i \rangle_\Phi| \leqslant C_{M_\Lambda} o_N(1) \quad \text{as } N \to \infty \tag{5-55}$$

for some constant C_Λ that does not depend on N. Comparing this, (5-51), and (5-52) with (5-43) proves (5-6). □

5E. *Reduction to right and left modes*: *linear terms*. We now prove that the main contribution to the linear terms surviving in the left-hand side of (5-22) actually comes from terms that couple u_1 with the modes $u_{r,\alpha}$ and u_2 with the modes $u_{\ell,\alpha}$. As previously we also show that we can neglect the contribution of modes beyond the energy cutoff M_Λ. First, we remark that using the definition of $b^\sharp$'s and $c^\sharp$'s from (3-19) we can rewrite the linear terms of Proposition 5.1 as

$$\frac{\lambda}{\sqrt{2(N-1)}} \sum_{m \geqslant 3} w_{+1-m}(b_m \mathfrak{D} + \text{h.c.}) + \frac{\lambda}{\sqrt{2(N-1)}} \sum_{m \geqslant 3} w_{+2-m}(c_m \mathfrak{D} + \text{h.c.}).$$

Proposition 5.7 (reduction of linear terms to right and left modes). *Assume $\Phi \in \ell^2(\mathfrak{F}_\perp)$ satisfies*

$$\left\langle \mathcal{N}_\perp + \frac{\mathfrak{D}^2}{N} + \mathrm{d}\Gamma_\perp(h_{\mathrm{MF}} - \mu_+) \right\rangle_\Phi \leqslant C \quad \textit{uniformly in } N. \tag{5-56}$$

For every energy cutoff Λ large, let M_Λ be the largest integer such that $\mu_{2M_\Lambda+2} \leqslant \Lambda$, where $\{\mu_m\}_m$ are the eigenvalues of h_{MF}. We have:

- ***Large cutoff limit.***

$$\left| \frac{\lambda}{\sqrt{2(N-1)}} \sum_{m > 2M_\Lambda+2} (w_{+1-m} \langle b_m \mathfrak{D} \rangle_\Phi + w_{+2-m} \langle c_m \mathfrak{D} \rangle_\Phi + \text{h.c.}) \right| \leqslant \frac{C}{(\mu_{2M_\Lambda+2} - \mu_+)^{1/2}}. \tag{5-57}$$

- ***Reduction to right and left modes.***

$$\begin{aligned}
&\frac{\lambda}{\sqrt{2(N-1)}} \left| \sum_{3 \leqslant m \leqslant 2M_\Lambda+2} w_{+1-m} \langle b_m \mathfrak{D} + \text{h.c.} \rangle_\Phi - \sum_{1 \leqslant \alpha \leqslant M_\Lambda} \langle u_1, w * (u_+ u_-) u_{r,\alpha} \rangle \langle b_{r,\alpha} \mathfrak{D} + \text{h.c.} \rangle_\Phi \right| \leqslant C_{M_\Lambda} o_N(1), \\
&\frac{\lambda}{\sqrt{2(N-1)}} \left| \sum_{3 \leqslant m \leqslant 2M_\Lambda+2} w_{+2-m} \langle c_m \mathfrak{D} + \text{h.c.} \rangle_\Phi - \sum_{1 \leqslant \alpha \leqslant M_\Lambda} \langle u_2, w * (u_+ u_-) u_{\ell,\alpha} \rangle \langle c_{\ell,\alpha} \mathfrak{D} + \text{h.c.} \rangle_\Phi \right| \leqslant C_{M_\Lambda} o_N(1).
\end{aligned} \tag{5-58}$$

Proof. Let us discuss how to prove (5-57), by focusing on the first limit (the second one is treated similarly). We have

$$\begin{aligned}
&\left| \frac{\lambda}{\sqrt{2(N-1)}} \sum_{m > 2M_\Lambda+2} w_{+1-m} \langle b_m \mathfrak{D} + \text{h.c.} \rangle_\Phi \right| \\
&\qquad \leqslant C \bigg(\sum_{m > 2M_\Lambda+2} |w_{+1-m}|^2 \bigg)^{1/2} \frac{\|\mathfrak{D}\Phi\|}{\sqrt{N}} \bigg(\sum_{m > 2M_\Lambda+2} \langle a_m^\dagger a_m \rangle_\Phi \bigg)^{1/2},
\end{aligned} \tag{5-59}$$

where we have used the Cauchy–Schwarz inequality both for the sum and for the $\ell^2(\mathfrak{F}_\perp)$ scalar product and the identities $b_m\mathfrak{D} = (\mathfrak{D}-1)b_m$ and $b_m^\dagger b_m = a_m^\dagger a_m$. The first sum in the right-hand side is bounded by a fixed constant thanks to (5-7). We now multiply and divide by $\mu_{2M_\Lambda+2}-\mu_+$ to get, arguing as in the previous subsection,

$$\sum_{m>2M_\Lambda+2} a_m^\dagger a_m \leqslant \frac{1}{\mu_{2M_\Lambda+2}-\mu_+}\mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+).$$

Plugging this inside (5-59), and using the assumption (5-56), we get

$$\left|\frac{\lambda}{\sqrt{2(N-1)}}\sum_{m>2M_\Lambda+2} w_{+1-m}\langle b_m\mathfrak{D}+\mathrm{h.c.}\rangle_\Phi\right| \leqslant \frac{C}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}},$$

which is the desired bound.

Let us now prove (5-58), again by focusing on the first bound only. By a change of basis we have

$$\sum_{3\leqslant m\leqslant 2M_\Lambda+2} w_{+1-m}\frac{\langle b_m\mathfrak{D}+\mathrm{h.c.}\rangle_\Phi}{\sqrt{2(N-1)}} = \sum_{1\leqslant\alpha\leqslant M_\Lambda}\langle u_1, w*(u_+u_-)u_{r,\alpha}\rangle\frac{\langle b_{r,\alpha}\mathfrak{D}+\mathrm{h.c.}\rangle_\Phi}{\sqrt{2(N-1)}} + \sum_{1\leqslant\alpha\leqslant M_\Lambda}\langle u_1, w*(u_+u_-)u_{\ell,\alpha}\rangle\frac{\langle b_{\ell,\alpha}\mathfrak{D}+\mathrm{h.c.}\rangle_\Phi}{\sqrt{2(N-1)}}. \tag{5-60}$$

The second sum in the right-hand converges to zero in the limit $N\to\infty$ because each summand does, and the sum is finite. Indeed, for instance

$$|\langle u_1, w*(u_+u_-)u_{\ell,\alpha}\rangle| \leqslant C\langle |u_1|, |u_{\ell,\alpha}|\rangle$$

and the right-hand side tends to zero as $N\to\infty$ by (5-54). The expectations on the state Φ in the sum are well-defined thanks to the assumption (5-56). We thus have

$$\left|\sum_{1\leqslant\alpha\leqslant M_\Lambda}\langle u_1, w*(u_+u_-)u_{\ell,\alpha}\rangle\frac{\langle b_{\ell,\alpha}\mathfrak{D}+\mathrm{h.c.}\rangle_\Phi}{\sqrt{2(N-1)}}\right| \leqslant C_{M_\Lambda}o_N(1),$$

which proves (5-58). □

6. A priori estimates on the ground state of H_N

Based on the previous results we can now deduce nontrivial information on the ground state ψ_{gs} of H_N, in particular that $\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}} \leqslant CN$ and $\langle\mathcal{N}_\perp^2\rangle_{\psi_{\mathrm{gs}}} \leqslant C$ with C a constant independent of N.

Proposition 6.1 (number and energy of excitations).

$$\langle\mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}} \leqslant C, \tag{6-1}$$

$$\langle\mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+)\rangle_{\psi_{\mathrm{gs}}} \leqslant C, \tag{6-2}$$

$$\langle\mathcal{N}_-\rangle_{\psi_{\mathrm{gs}}} \leqslant C_\varepsilon\min\{N, T^{-1-\varepsilon}\}. \tag{6-3}$$

Proposition 6.2 (second moment of excitations).

$$\langle \mathcal{N}_\perp^2 \rangle_{\psi_{\mathrm{gs}}} \leqslant \frac{C}{N} \langle (\mathcal{N}_1 - \mathcal{N}_2)^2 \rangle_{\psi_{\mathrm{gs}}} + C, \tag{6-4}$$

$$\langle \mathcal{N}_\perp \mathrm{d}\Gamma_\perp (h_{\mathrm{MF}} - \mu_+) \rangle_{\psi_{\mathrm{gs}}} \leqslant \frac{C}{N} \langle (\mathcal{N}_1 - \mathcal{N}_2)^2 \rangle_{\psi_{\mathrm{gs}}} + C. \tag{6-5}$$

Proposition 6.3 (variance in the two-mode subspace).

$$\langle (\mathcal{N}_1 - \mathcal{N}_2)^2 \rangle_{\psi_{\mathrm{gs}}} \leqslant CN. \tag{6-6}$$

Inserting (6-6) in (6-4) and (6-5) yields

$$\begin{aligned} \langle \mathcal{N}_\perp^2 \rangle_{\psi_{\mathrm{gs}}} &\leqslant C, \\ \langle \mathcal{N}_\perp \mathrm{d}\Gamma_\perp (h_{\mathrm{MF}} - \mu_+) \rangle_{\psi_{\mathrm{gs}}} &\leqslant C. \end{aligned} \tag{6-7}$$

As a consequence of (6-3), (6-6), and (6-7), if one applies Proposition 5.1 to the vector $\Phi = \mathcal{U}_N \psi_{\mathrm{gs}}$, the error terms in the right-hand side of (5-1) are small, being bounded by

$$\frac{C}{N^{1/4}} + C_\varepsilon \frac{T^{-2\varepsilon}}{N^{1/2}}. \tag{6-8}$$

The rest of this section is devoted to the proofs of Propositions 6.1–6.3. The general strategy for the first two results is similar to the single-well case (that is, the case of fixed L) and some arguments are accordingly borrowed from [Grech and Seiringer 2013]. The two-mode nature of our low energy space, however, calls for additional ingredients, in particular as regards the proof of Proposition 6.2. Proposition 6.3 uses as input our results of Sections 4 and 5.

We will use several times Onsager's inequality (see, e.g., [Rougerie 2020, Lemma 2.6]):

$$\begin{aligned} \frac{1}{N} \sum_{i \neq j} w(x_i - x_j) \\ \geqslant -N \iint w(x-y) |u_+(x)|^2 |u_+(y)|^2 \, dx \, dy + 2 \sum_{i=1}^N \int w(x_i - y) |u_+(y)|^2 \, dy - w(0). \end{aligned} \tag{6-9}$$

Proof of Proposition 6.1. Using (6-9) and then the definition of μ_+ from (2-8) we get (since the interaction term in the N-body Hamiltonian (1-2) is nonnegative, we may replace the prefactor $\lambda/(N-1)$ by λ/N)

$$\begin{aligned} \langle H_N \rangle_{\psi_{\mathrm{gs}}} &\geqslant \langle \mathrm{d}\Gamma(h_{\mathrm{MF}}) \rangle_{\psi_{\mathrm{gs}}} - \frac{\lambda N}{2} \iint w(x-y) |u_+(x)|^2 |u_+(y)|^2 \, dx \, dy - C \\ &\geqslant \langle \mathrm{d}\Gamma(h_{\mathrm{MF}} - \mu_+) \rangle_{\psi_{\mathrm{gs}}} + N \mathcal{E}^{\mathrm{H}}[u_+] - C \\ &> \langle \mathrm{d}\Gamma_\perp (h_{\mathrm{MF}} - \mu_+) \rangle_{\psi_{\mathrm{gs}}} + N \mathcal{E}^{\mathrm{H}}[u_+] - C. \end{aligned} \tag{6-10}$$

The last step is due to the identity

$$\mathrm{d}\Gamma(h_{\mathrm{MF}} - \mu_+) = (\mu_- - \mu_+) \mathcal{N}_- + \mathrm{d}\Gamma_\perp (h_{\mathrm{MF}} - \mu_+) \tag{6-11}$$

and to the fact that $\mu_- > \mu_+$. On the other hand, the factorized trial function $u_+^{\otimes N}$ yields the energy upper bound

$$\langle H_N \rangle_{\psi_{\mathrm{gs}}} \leqslant N \mathcal{E}^{\mathrm{H}}[u_+], \tag{6-12}$$

and putting together (6-10) and (6-12) we find

$$\langle \mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+)\rangle_{\psi_{\mathrm{gs}}} \leqslant C, \tag{6-13}$$

which is precisely (6-2). Recalling the spectral decomposition (2-25), and the fact that $\mu_m - \mu_+ \geqslant C$ for $m \geqslant 3$ (by Theorem A.1), we deduce

$$\langle \mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+)\rangle_{\psi_{\mathrm{gs}}} \geqslant C\langle \mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}},$$

which, together with (6-2), proves (6-1).

To prove (6-3) we use (6-11) again and notice that, by the spectral properties of h_{MF} from Theorem A.1,

$$\langle \mathrm{d}\Gamma(h_{\mathrm{MF}}-\mu_+)\rangle_{\psi_{\mathrm{gs}}} \geqslant (\mu_- - \mu_+)\langle \mathcal{N}_-\rangle_{\psi_{\mathrm{gs}}} \geqslant c_\varepsilon T^{1+\varepsilon}\langle \mathcal{N}_-\rangle_{\psi_{\mathrm{gs}}}.$$

This, compared with (6-10) and (6-12), yields (6-3) after recalling that $\langle \mathcal{N}_-\rangle_{\psi_{\mathrm{gs}}} \leqslant N$ also trivially holds. □

Proof of Proposition 6.2. We claim that

$$\langle \mathcal{N}_\perp \mathrm{d}\Gamma(h_{\mathrm{MF}}-\mu_+)\rangle_{\psi_{\mathrm{gs}}} \leqslant \delta\langle \mathcal{N}_\perp^2\rangle_{\psi_{\mathrm{gs}}} + \frac{C}{N}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}} + C_\delta \tag{6-14}$$

for $\delta > 0$ arbitrary and for some constants $C, C_\delta > 0$. This implies the bound (6-4) because

$$\mathrm{d}\Gamma(h_{\mathrm{MF}}-\mu_+) \geqslant c\mathcal{N}_\perp$$

on $L^2(\mathbb{R}^{dN})$ with $c > 0$, and because h_{MF} commutes with $\mathcal{N}_\perp$.

To prove (6-14) we define the operators

$$S := \lambda\sum_{j=1}^N w * |u_+|^2(x_j) - \frac{\lambda}{N-1}\sum_{i<j} w(x_i - x_j) + E(N) - N\mu_+$$

and

$$P_j = |u_+\rangle\langle u_+|_j + |u_-\rangle\langle u_-|_j, \quad P_j^\perp = \mathbb{1} - P_j,$$

with $j = 1, \dots, N$. The latter project a single particle in (or out) the two-mode subspace. We also denote by $h_{\mathrm{MF},j}$ the operator that acts as h_{MF} on the j-th variable and as the identity on all the others. We then have

$$\langle \mathcal{N}_\perp \mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+)\rangle_{\psi_{\mathrm{gs}}} = \left\langle \mathcal{N}_\perp \sum_{j=1}^N (h_{\mathrm{MF},j}-\mu_+)\right\rangle_{\psi_{\mathrm{gs}}} = \langle \mathcal{N}_\perp S\rangle_{\psi_{\mathrm{gs}}} = N\langle P_1^\perp S\rangle_{\psi_{\mathrm{gs}}}, \tag{6-15}$$

where we have used $H_N\psi_{\mathrm{gs}} = E(N)\psi_{\mathrm{gs}}$ in the second equality and the fact that ψ_{gs} is symmetric under permutations of variables in the last one. We split the operator S into the part which commutes with $P_1^\perp$ and the part which does not, according to

$$S = S_a + S_b,$$

where

$$S_a := \lambda\sum_{j=2}^N w * |u_+|^2(x_j) - \frac{\lambda}{N-1}\sum_{2\leqslant i<j\leqslant N} w(x_i - x_j) + E_N - N\mu_+,$$

$$S_b := \lambda w * |u_+|^2(x_1) - \frac{\lambda}{N-1}\sum_{j=2}^N w(x_1 - x_j).$$

We will estimate separately the contributions of the terms containing S_a and S_b inside (6-15). For the contribution of the term containing S_a we use (6-9) for $N-1$ variables, that is,

$$\frac{\lambda}{N-1}\sum_{2\leqslant i<j\leqslant N} w(x_i-x_j) \geqslant -\lambda\frac{N-1}{2}w_{++++}+\lambda\sum_{j=2}^{N} w*|u_+|^2(x_j)-C.$$

We also take advantage of the upper bound

$$\langle H_N\rangle_{\psi_{\mathrm{gs}}}\leqslant N\mu_+-\lambda\frac{N}{2}w_{++++},$$

which follows immediately from (6-12) if we recall the expression (2-8) of μ_+. The two last formulas yield

$$S_a\leqslant C.$$

Since S_a commutes with $P_1^\perp$, we have, using also (6-1),

$$N\langle P_1^\perp S_a\rangle_{\psi_{\mathrm{gs}}}\leqslant C\langle \mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}}\leqslant C. \tag{6-16}$$

To estimate the contribution of S_b, we consider the decomposition

$$\begin{aligned}N\langle P_1^\perp S_b\rangle_{\psi_{\mathrm{gs}}} &= \lambda N\langle P_1^\perp[w*|u_+|^2(x_1)-w(x_1-x_2)]\rangle_{\psi_{\mathrm{gs}}}\\ &= \lambda N\langle P_1^\perp P_2^\perp[w*|u_+|^2(x_1)-w(x_1-x_2)]\rangle_{\psi_{\mathrm{gs}}}\\ &\quad+\lambda N\langle P_1^\perp P_2[w*|u_+|^2(x_1)-w(x_1-x_2)]P_2^\perp\rangle_{\psi_{\mathrm{gs}}}\\ &\quad\;+\lambda N\langle P_1^\perp P_2[w*|u_+|^2(x_1)-w(x_1-x_2)]P_2\rangle_{\psi_{\mathrm{gs}}}\\ &=: \mathrm{Term}_1+\mathrm{Term}_2+\mathrm{Term}_3.\end{aligned}\tag{6-17}$$

We estimate the last three terms separately. For the first one we use the Cauchy–Schwarz inequality and the fact that w and $w*|u_+|^2$ are bounded to get

$$\begin{aligned}|\mathrm{Term}_1|\leqslant CN\langle P_1^\perp P_2^\perp\rangle_{\psi_{\mathrm{gs}}}^{1/2} &= CN\left\langle P_1^\perp\frac{1}{N-1}\sum_{j=2}^{N}P_j^\perp\right\rangle_{\psi_{\mathrm{gs}}}^{1/2}\\ &\leqslant C\langle\mathcal{N}_\perp^2\rangle_{\psi_{\mathrm{gs}}}^{1/2}\leqslant\delta\langle\mathcal{N}_\perp^2\rangle_{\psi_{\mathrm{gs}}}+C_\delta,\end{aligned}$$

with $\delta>0$ arbitrary, where the last bound follows from $\sqrt{x}\leqslant\delta x+1/(4\delta)$ for any $x>0$. For the second term in (6-17) we argue similarly to get

$$|\mathrm{Term}_2|\leqslant CN\langle P_1^\perp\rangle_{\psi_{\mathrm{gs}}}^{1/2}\langle P_2^\perp\rangle_{\psi_{\mathrm{gs}}}^{1/2}=C\langle\mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}}\leqslant C,$$

where the last bound follows from (6-1).

The third term in (6-17) is more delicate, since it contains only one $P_j^\perp$. We write

$$\begin{aligned}\mathrm{Term}_3 &= \lambda N\langle P_1^\perp|u_-\rangle\langle u_-|_2 w*(|u_+|^2-|u_-|^2)(x_1)\rangle_{\psi_{\mathrm{gs}}}\\ &\qquad\qquad\qquad -\lambda N\langle P_1^\perp(|u_+\rangle\langle u_-|_2+|u_-\rangle\langle u_+|_2)w*(u_+u_-)(x_1)\rangle_{\psi_{\mathrm{gs}}}\\ &=: \mathrm{Term}_{3,1}+\mathrm{Term}_{3,2},\end{aligned}\tag{6-18}$$

where we have used several times the operator identity

$$|u\rangle\langle u|_2\, w(x_1-x_2)\, |v\rangle\langle v|_2=|u\rangle\langle v|_2\, w*(\bar{u}v)(x_1).$$

Using the Cauchy–Schwarz and Young inequalities, the L^1-estimate (A-1), and then the a priori estimate (6-3), we find

$$\begin{aligned}|\mathrm{Term}_{3,1}| &\leqslant CN\langle P_1^\perp\rangle_{\psi_{\mathrm{gs}}}^{1/2}\langle|u_-\rangle\langle u_-|_2\rangle_{\psi_{\mathrm{gs}}}^{1/2}\||u_+|^2-|u_-|^2\|_{L^1}\\ &\leqslant C_\varepsilon T^{1-\varepsilon/2}\langle\mathcal{N}_-\rangle_{\psi_{\mathrm{gs}}}^{1/2}\langle\mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}}^{1/2}\\ &\leqslant C_\varepsilon T^{1-\varepsilon/2}\min\left\{N,\frac{1}{T^{1+\varepsilon}}\right\}^{1/2}\langle\mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}}^{1/2}\\ &\leqslant C_\varepsilon T^{1/2-\varepsilon}\langle\mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}}^{1/2}\leqslant C_\varepsilon T^{1/2-\varepsilon}.\end{aligned}$$

Recalling that

$$\sum_{j=1}^N(|u_+\rangle\langle u_-|_j+|u_-\rangle\langle u_+|_j)=a_+^\dagger a_-+a_-^\dagger a_+=\mathcal{N}_1-\mathcal{N}_2,$$

one may write

$$\begin{aligned}-\mathrm{Term}_{3,2}=\frac{N}{N-1}&\langle P_1^\perp w*(u_+u_-)(x_1)(\mathcal{N}_1-\mathcal{N}_2)\rangle_{\psi_{\mathrm{gs}}}\\ &-\frac{N}{N-1}\langle P_1^\perp w*(u_+u_-)(x_1)(|u_+\rangle\langle u_-|_1+|u_-\rangle\langle u_+|_1)\rangle_{\psi_{\mathrm{gs}}}.\end{aligned}$$

The second summand is clearly bounded by a constant and thus we include it in the error. For the first one we write, using the Cauchy–Schwarz inequality and the boundedness of $w*(u_+u_-)$,

$$\frac{N}{N-1}|\langle P_1^\perp w*(u_+u_-)(x_1)(\mathcal{N}_1-\mathcal{N}_2)\rangle_{\psi_{\mathrm{gs}}}|\leqslant C\langle P_1^\perp\rangle_{\psi_{\mathrm{gs}}}^{1/2}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}}^{1/2}.$$

We finally get

$$|\mathrm{Term}_{3,2}|\leqslant C\langle\mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}}^{1/2}\left(\frac{\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}}}{N}\right)^{1/2}\leqslant CN^{-1/2}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}}^{1/2},$$

where we have used (6-1) in the last bound. All in all we proved

$$|\mathrm{Term}_3|\leqslant CN^{-1/2}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}}^{1/2}+C,$$

and therefore

$$N\langle P_1^\perp S_b\rangle_{\psi_{\mathrm{gs}}}\leqslant\delta\langle\mathcal{N}_\perp^2\rangle_{\psi_{\mathrm{gs}}}+\frac{C}{N}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}}+C_\delta. \tag{6-19}$$

The announced bound (6-14) then follows from (6-16) and (6-19). We deduce (6-4) by choosing δ small enough. Plugging (6-4) into (6-14) yields (6-5) as well. □

Proof of Proposition 6.3. We combine Proposition 5.1 with a computation similar to Proposition 4.4 to obtain an energy upper bound. For a corresponding lower bound we use Propositions 4.2 and 4.4 to control the two-mode energy, and argue that the excitation energy must be uniformly bounded with respect to N.

Recall the trial state ψ_{gauss} from (4-20). We apply (5-1) with $\Phi=\mathcal{U}_N\psi_{\mathrm{gauss}}$. Since ψ_{gauss} has no excitation in the subspace $P_\pm^\perp\mathfrak{H}^N$ ($a_m\psi_{\mathrm{gauss}}=0$ for any $m\geqslant 3$), we get

$$\mathcal{N}_\perp\mathcal{U}_N\psi_{\mathrm{gauss}}=\mathbb{H}\,\mathcal{U}_N\psi_{\mathrm{gauss}}=0.$$

The expectations of the linear terms in a_m in the left-hand side of (5-1) also vanish for $\psi = \psi_{\text{gauss}}$. Furthermore, we will use

$$\frac{1}{N}\langle \mathfrak{D}^2\rangle_{\mathcal{U}_N\psi_{\text{gauss}}} = \frac{1}{N}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\text{gauss}}} \leqslant \frac{1}{N}\sigma_N^2 = \sqrt{\mu_- - \mu_+} \leqslant C_\varepsilon T^{1/2-\varepsilon},$$

where the first bound was proven in (4-32). By the variational principle for the ground state problem of H_N we find

$$\begin{aligned} E(N) \leqslant \langle H_N\rangle_{\psi_{\text{gauss}}} &\leqslant \langle H_{2\text{-mode}}\rangle_{\psi_{\text{gauss}}} + \frac{C}{N^{1/4}} \\ &\leqslant E_0 + E_N^w + N\frac{\mu_+-\mu_-}{2} + C_\varepsilon T^{1/2-\varepsilon} + \frac{C}{N^{1/4}} \\ &\leqslant E_0 + E_N^w + N\frac{\mu_+-\mu_-}{2} + C, \end{aligned} \tag{6-20}$$

applying successively (5-1) and (4-23).

For a lower bound we apply (5-1) with $\Phi = \Phi_{\text{gs}} =: \mathcal{U}_N\psi_{\text{gs}}$, obtaining

$$|E(N) - \langle H_{2\text{-mode}} + \mu_+\mathcal{N}_\perp\rangle_{\psi_{\text{gs}}} - \langle\mathbb{H}\rangle_{\Phi_{\text{gs}}} - \langle\text{linear terms}\rangle_{\Phi_{\text{gs}}}| \leqslant \text{error terms}.$$

In this inequality:

(i) The error terms are bounded by using (6-3), the identity $\langle\mathfrak{D}^2\rangle_{\Phi_{\text{gs}}} = \langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\text{gs}}}$, and the inequality $\langle\mathcal{N}_-\rangle_{\psi_{\text{gs}}} \leqslant N$, yielding

$$\text{error terms} \leqslant \left(\frac{C}{N^{1/4}} + C_\varepsilon T^{1-\varepsilon}\right)\left(\frac{\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\text{gs}}}}{N} + 1\right).$$

(ii) The expectation of $H_{2\text{-mode}}+\mu_+\mathcal{N}_\perp$ is bounded from below by using the lower bound of Proposition 4.2,

$$\langle H_{2\text{-mode}} + \mu_+\mathcal{N}_\perp\rangle_{\psi_{\text{gs}}} \geqslant E_0 + E_N^w + N\frac{\mu_+-\mu_-}{2} + \frac{\lambda U}{N-1}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\text{gs}}} - C_\varepsilon T^{1-\varepsilon}\langle\mathcal{N}_\perp\rangle_{\psi_{\text{gs}}}.$$

Thanks to (6-1), the term in the second line can be replaced by $-C$.

(iii) The expectation of $\mathbb{H}$ is bounded from below using the fact that $\mathbb{H}$ is bounded below independently of N (this can easily be seen as in [Lewin et al. 2015, equation (A.6)], keeping in mind that $h_{\text{MF}} - \mu_+$ has a finite gap on the excited subspace).

(iv) The expectation of linear terms can be bounded by using the Cauchy–Schwarz inequality as follows:

$$\begin{aligned} &\left|\frac{\lambda}{\sqrt{2(N-1)}}\sum_{m\geqslant 3}[w_{+1-m}\langle\Theta a_m\mathfrak{D} + \text{h.c.}\rangle_{\Phi_{\text{gs}}} + w_{+2-m}\langle\Theta^{-1}a_m\mathfrak{D} + \text{h.c.}\rangle_{\Phi_{\text{gs}}}]\right| \\ &\quad\leqslant \frac{2\lambda}{\sqrt{2(N-1)}}\left(\sum_{m\geqslant 3}|w_{+1-m}|^2\right)^{1/2}\left(\sum_{m\geqslant 3}\|a_m\Phi_{\text{gs}}\|^2\right)^{1/2}\|\mathfrak{D}\Theta^{-1}\Phi_{\text{gs}}\| \\ &\qquad\qquad + \frac{2\lambda}{\sqrt{2(N-1)}}\left(\sum_{m\geqslant 3}|w_{+2-m}|^2\right)^{1/2}\left(\sum_{m\geqslant 3}\|a_m\Phi_{\text{gs}}\|^2\right)^{1/2}\|\mathfrak{D}\Theta\Phi_{\text{gs}}\|. \end{aligned}$$

The sums of $|w_{+i-m}|^2$ are bounded by constants thanks to (5-7). The other sums equal $\langle \mathcal{N}_\perp \rangle_{\psi_{\mathrm{gs}}}$, for which we use (6-1). Finally, thanks to the commutation relation (3-15) one has

$$\|\mathfrak{D}\Theta^{\pm 1}\Phi_{\mathrm{gs}}\|^2 = \langle (\mathcal{N}_1 - \mathcal{N}_2 \pm 1)^2\rangle_{\psi_{\mathrm{gs}}} \leqslant 2\langle (\mathcal{N}_1 - \mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}} + 2$$

and thus

$$|\langle \text{linear terms}\rangle_{\Phi_{\mathrm{gs}}}| \leqslant \frac{\delta}{N}\langle (\mathcal{N}_1 - \mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}} + C_\delta$$

for any $\delta > 0$ arbitrarily small.

Overall we find

$$E_N \geqslant E_0 + E_N^w + N\frac{\mu_+ - \mu_-}{2} + \frac{c}{N-1}\langle (\mathcal{N}_1 - \mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}} - C$$

for a suitable small enough positive constant c. Notice that we used the fact that the constant U in (4-3) satisfies $U \geqslant C > 0$ independently of N thanks to the estimates of Lemma 4.1, Comparing this with (6-20) gives the desired (6-6). □

7. Shifted Hamiltonians and lower bound

Shifted CCR. Let us introduce the notation

$$\begin{aligned}\mathbb{H}^{(M)}_{\mathrm{right,shift}} &:= \mathbb{H}^{(M)}_{\mathrm{right}} + \frac{\lambda}{\sqrt{2(N-1)}}\sum_{1\leqslant\alpha\leqslant M}\langle u_1, w*(u_+u_-)u_{r,\alpha}\rangle(b_{r,\alpha}\mathfrak{D} + \mathrm{h.c.}),\\ \mathbb{H}^{(M)}_{\mathrm{left,shift}} &:= \mathbb{H}^{(M)}_{\mathrm{left}} + \frac{\lambda}{\sqrt{2(N-1)}}\sum_{1\leqslant\alpha\leqslant M}\langle u_2, w*(u_+u_-)u_{\ell,m}\rangle(c_{\ell,\alpha}\mathfrak{D} + \mathrm{h.c.}).\end{aligned} \tag{7-1}$$

The linear terms are those appearing in (5-1) up to a change of basis from $\{u_m\}_{m\geqslant 3}$ to the right and left mode basis $\{u_{r,\alpha}, u_{\ell,\alpha}\}_{\alpha\geqslant 1}$), where we have ignored the modes beyond the cutoff M and small error terms, as justified in Proposition 5.7.

The estimates of Propositions 5.1, 5.2, and 5.7 yield the lower bound

$$\mathcal{U}_N(H_N - H_{2\text{-mode}})\mathcal{U}_N^* \geqslant \mathbb{H}^{(M_\Lambda)}_{\mathrm{right,shift}} + \mathbb{H}^{(M_\Lambda)}_{\mathrm{left,shift}} + \mu_+\mathcal{N}_\perp - \text{remainders}. \tag{7-2}$$

We will show in this section how to deal with the linear terms in $\mathbb{H}^{(M_\Lambda)}_{\mathrm{right,shift}}$ and $\mathbb{H}^{(M_\Lambda)}_{\mathrm{left,shift}}$. The idea is to define new shifted creation and annihilation operators $\tilde{b}^\sharp_{r,\alpha}$ and $\tilde{c}^\sharp_{\ell,\alpha}$ in such a way that $\mathbb{H}^{(M_\Lambda)}_{\mathrm{right,shift}}$ and $\mathbb{H}^{(M_\Lambda)}_{\mathrm{left,shift}}$ are quadratic in terms of, respectively, $\tilde{b}^\sharp_{r,\alpha}$ and $\tilde{c}^\sharp_{\ell,\alpha}$, up to a constant term. We will do this for each fixed M, not necessarily the M_Λ from Proposition 5.2.

From now on we will use the notation $\{r,\alpha\}$ or $\{\ell,\alpha\}$ to indicate that the mode $u_{r,\alpha}$ or $u_{\ell,\alpha}$ intervenes in an expectation value. For example, for any operator A on $L^2(\mathbb{R}^d)$,

$$A_{\{r,\alpha\}\{\ell,\beta\}} = \langle u_{r,\alpha}, A u_{\ell,\beta}\rangle.$$

Similarly,

$$w_{m\{r,\alpha\}p\{r,\beta\}} = \langle u_m \otimes u_{r,\alpha}, w\, u_p \otimes u_{r,\beta}\rangle,$$

and so on.

Definition 7.1 (shifted creators and annihilators). For any $\alpha \geqslant 1$ we define

$$\begin{aligned}\tilde{b}_{r,\alpha} &:= b_{r,\alpha} + x_\alpha \mathfrak{D},\\ \tilde{c}^\dagger_{\ell,\alpha} &:= c^\dagger_{\ell,\alpha} + y_\alpha \mathfrak{D},\end{aligned} \tag{7-3}$$

where $x_\alpha,\ y_\alpha,\ \alpha = 1, \dots, M$, are real numbers whose values will be given below.

A simple calculation using the commutation relations (3-15), (3-20) yields:

Lemma 7.2 (commutations relations for shifted operators). *One has*

$$\begin{aligned}[\tilde{b}_{r,\alpha}, \tilde{b}^\dagger_{r,\beta}] &= \delta_{\alpha\beta} - x_\beta b_{r,\alpha} - x_\alpha b^\dagger_{r,\beta},\\ [\tilde{b}_{r,\alpha}, \tilde{b}_{r,\beta}] &= -x_\beta b_{r,\alpha} + x_\alpha b_{r,\beta}.\end{aligned} \tag{7-4}$$

Similar commutation relations, with straightforward adaptations, hold for the $\tilde{c}^\sharp_{\ell,\alpha}$.

We define the following quadratic Hamiltonians, obtained from (3-21) and (3-22) by replacing the creation and annihilation operators $b^\sharp$ and $c^\sharp$ by the shifted creators and annihilators (7-3),

$$\widetilde{\mathbb{H}^{(M)}_{\text{right}}} := \frac{1}{2} \sum_{1\leqslant\alpha,\beta\leqslant M} (h_{\text{MF}} - \mu_+ + \lambda K_{11})_{\{r,\alpha\}\{r,\beta\}} (\tilde{b}^\dagger_{r,\alpha}\tilde{b}_{r,\beta} + \tilde{b}_{r,\alpha}\tilde{b}^\dagger_{r,\beta}) \\ + \frac{\lambda}{2} \sum_{1\leqslant\alpha,\beta\leqslant M} (K_{11})_{\{r,\alpha\}\{r,\beta\}} (\tilde{b}^\dagger_{r,\alpha}\tilde{b}^\dagger_{r,\beta} + \tilde{b}_{r,\alpha}\tilde{b}_{r,\beta}), \tag{7-5}$$

$$\widetilde{\mathbb{H}^{(M)}_{\text{left}}} := \frac{1}{2} \sum_{1\leqslant\alpha,\beta\leqslant M} (h_{\text{MF}} - \mu_+ + \lambda K_{22})_{\{\ell,\alpha\}\{\ell,\beta\}} (\tilde{c}^\dagger_{\ell,\alpha}\tilde{c}_{\ell,\beta} + \tilde{c}_{\ell,\alpha}\tilde{c}^\dagger_{\ell,\beta}) \\ + \frac{\lambda}{2} \sum_{1\leqslant\alpha,n\leqslant M} (K_{22})_{\{\ell,\alpha\}\{\ell,\beta\}} (\tilde{c}^\dagger_{\ell,\alpha}\tilde{c}^\dagger_{\ell,\beta} + \tilde{c}_{\ell,\alpha}\tilde{c}_{\ell,\beta}), \tag{7-6}$$

where we have ignored the modes beyond the cutoff M and symmetrized the terms involving one creator and one annihilator.

Let us introduce the orthogonal projections

$$P_{r,\leqslant M} := P_r P_{\leqslant M} = P_{\leqslant M} P_r = \sum_{1\leqslant\alpha\leqslant M} |u_{r,\alpha}\rangle\langle u_{r,\alpha}|, \tag{7-7}$$

$$P_{\ell,\leqslant M} := P_\ell P_{\leqslant M} = P_{\leqslant M} P_\ell = \sum_{1\leqslant\alpha\leqslant M} |u_{\ell,\alpha}\rangle\langle u_{\ell,\alpha}|. \tag{7-8}$$

We will show the following result.

Proposition 7.3 (shifted Hamiltonians). *For any* $\Phi \in \ell^2(\mathfrak{F}_\perp)$ *we have*

$$\Big| \langle \mathbb{H}^{(M)}_{\text{right,shift}} \rangle_\Phi - \langle \widetilde{\mathbb{H}^{(M)}_{\text{right}}} \rangle_\Phi + \tfrac{1}{2}\operatorname{Tr}(P_{r,\leqslant M}(h_{\text{MF}} - \mu_+ + \lambda K_{11})) \\ + \frac{\lambda^2}{2(N-1)} \langle u_1,\ K_{11} W_{r,\leqslant M} K_{11} u_1 \rangle \langle \mathfrak{D}^2 \rangle_\Phi \Big| \leqslant \frac{C}{\sqrt{N}} \langle \mathcal{N}_\perp \rangle_\Phi + \frac{C_\varepsilon T^{1/2-\varepsilon}}{N} \langle \mathfrak{D}^2 \rangle_\Phi, \tag{7-9}$$

where $W_{r,\leqslant M}$ *is defined by*

$$W_{r,\leqslant M} := P_{r,\leqslant M} (P_{r,\leqslant M} (h_{\text{MF}} - \mu_+ + 2\lambda K_{11}) P_{r,\leqslant M})^{-1} P_{r,\leqslant M} \tag{7-10}$$

and we picked

$$x_\alpha = \frac{\lambda}{\sqrt{2(N-1)}} \langle u_{r,\alpha},\, W_{r,\leqslant M}\, w * (u_+ u_-)\, u_1 \rangle. \tag{7-11}$$

A similar bound holds for $\mathbb{H}^{(M)}_{\mathrm{left,shift}}$ *upon replacing* K_{11} *by* K_{22}.

Thus the quadratic Hamiltonian $\mathbb{H}^{(M)}_{\mathrm{right}}$ together with the linear terms coincides, up to remainders, with $\widetilde{\mathbb{H}^{(M)}_{\mathrm{right}}}$ minus a constant term given by the trace in (7-9) and minus a term proportional to $\lambda^2\mathfrak{D}^2$. The latter term will be absorbed using the variance term from $H_{2\text{-mode}}$ which is proportional to λ, and $\widetilde{\mathbb{H}^{(M)}_{\mathrm{right}}}$ minus the constant term will give the correct Bogoliubov energy in the lower bound. Note that the trace in the constant term is finite because we are restricting ourselves to modes $\alpha \leqslant M$.

Proof. Using the commutation relations (7-4) and $[\tilde{b}_{r,\alpha}, \mathfrak{D}] = [\Theta, \mathfrak{D}]a_{r,\alpha} = -b_\alpha$, one finds that $\mathbb{H}^{(M)}_{\mathrm{right,shift}}$ is given in terms of the shifted creators and annihilators $\tilde{b}^\sharp$ by

$$\begin{aligned}
&\mathbb{H}^{(M)}_{\mathrm{right,shift}}\\
&\quad= \frac{1}{2}\sum_{1\leqslant\alpha,\beta\leqslant M} (h_{\mathrm{MF}} - \mu_+ + \lambda K_{11})_{\{r,\alpha\}\{r,\beta\}} (\tilde{b}^\dagger_{r,\alpha}\tilde{b}_{r,\beta} + \tilde{b}_{r,\alpha}\tilde{b}^\dagger_{r,\beta})\\
&\qquad + \frac{\lambda}{2}\sum_{1\leqslant\alpha,\beta\leqslant M} (K_{11})_{\{r,\alpha\}\{r,\beta\}} (\tilde{b}^\dagger_{r,\alpha}\tilde{b}^\dagger_{r,\beta} + \tilde{b}_{r,\alpha}\tilde{b}_{r,\beta}) - \frac{1}{2}\operatorname{Tr}(P_{r,\leqslant M}(h_{\mathrm{MF}} - \mu_+ + \lambda K_{11}))\\
&\qquad - \sum_{1\leqslant\alpha\leqslant M}\Biggl(\sum_{1\leqslant\beta\leqslant M} (h_{\mathrm{MF}} - \mu_+ + 2\lambda K_{11})_{\{r,\alpha\}\{r,\beta\}} x_\beta - \frac{\lambda}{\sqrt{2(N-1)}} w_{+1-\{r,\alpha\}}\Biggr)(\tilde{b}^\dagger_{r,\alpha}\mathfrak{D} + \mathfrak{D}\tilde{b}_{r,\alpha})\\
&\qquad + \sum_{1\leqslant\alpha\leqslant M}\Biggl(\sum_{1\leqslant\beta\leqslant M} (h_{\mathrm{MF}} - \mu_+ + 2\lambda K_{11})_{\{r,\alpha\}\{r,\beta\}} x_\beta - \frac{2\lambda}{\sqrt{2(N-1)}} w_{+1-\{r,\alpha\}}\Biggr) x_\alpha \mathfrak{D}^2\\
&\qquad + \frac{1}{2}\sum_{1\leqslant\alpha\leqslant M}\Biggl(\sum_{1\leqslant\beta M} (h_{\mathrm{MF}} - \mu_+ + 2\lambda K_{11})_{\{r,\alpha\}\{r,\beta\}} x_\beta - \frac{2\lambda}{\sqrt{2(N-1)}} w_{+1-\{r,\alpha\}}\Biggr)(b_\alpha + b^\dagger_\alpha).
\end{aligned}\tag{7-12}$$

The first and second lines in the right-hand side precisely coincide with $\widetilde{\mathbb{H}^{(M)}_{\mathrm{right}}}$ defined in (7-5) minus the constant term $-\operatorname{Tr}(P_{r,\leqslant M}(h_{\mathrm{MF}} - \mu_+ + \lambda K_{11}))/2$. The condition for the vanishing of the linear terms in the third line is

$$\sum_{1\leqslant\beta\leqslant M} (h_{\mathrm{MF}} - \mu_+ + 2\lambda K_{11})_{\{r,\alpha\}\{r,\beta\}} x_\beta = \frac{\lambda}{\sqrt{2(N-1)}} w_{+1-\{r,\alpha\}}, \tag{7-13}$$

which leads to (7-11), using the projection $P_{r,\leqslant M}$ defined in (2-27) and (7-10). With this choice, the expectation in Φ of the last line in (7-12) becomes

$$R_\Phi = -\frac{\lambda}{\sqrt{2(N-1)}} \sum_{1\leqslant\alpha\leqslant M} w_{+1-\{r,\alpha\}} \langle b_{r,\alpha} + b_{r,\alpha^\dagger}\rangle_\Phi.$$

This can be bounded with the help of the Cauchy–Schwarz inequality and the boundedness of $w * (u_+ u_-)$ as in the proofs of Section 5, that is,

$$|R_\Phi| \leqslant \frac{C\lambda}{\sqrt{N}} \biggl\{\sum_{\alpha\geqslant 1} |w_{+1-\{r,\alpha\}}|^2\biggr\}^{1/2} \biggl\{\sum_{\alpha\geqslant 1} \|b_{r,\alpha}\Phi\|^2\biggr\}^{1/2} \leqslant \frac{C}{\sqrt{N}} \langle \mathcal{N}_\perp \rangle_\Phi^{1/2},$$

with C independent of N and M. Plugging (7-13) into (7-12) we only have to compute the contribution of the term proportional to $\mathfrak{D}^2$ in the fifth line, which is given by

$$-\frac{\lambda}{\sqrt{2(N-1)}}\sum_{1\leqslant\alpha\leqslant M} w_{+1-\{r,\alpha\}}x_\alpha\mathfrak{D}^2=-\frac{\lambda^2}{2(N-1)}\langle u_1,\ w*(u_+u_-)W_{r,\leqslant M}w*(u_+u_-)u_1\rangle\mathfrak{D}^2. \tag{7-14}$$

To bring this contribution to the form appearing in (7-9) we have to show that one can replace the multiplication operator $w*(u_+u_-)$ by the integral operator K_{11} up to a small error. To this end we notice that, using (1-6), for any $f\in L^2(\mathbb{R}^d)$,

$$\begin{aligned}|\langle u_1,(w*(u_+u_-)-K_{11})f\rangle|^2&=\left|\left\langle u_1,\left(w*(u_+u_-)-\frac{w*|u_1|^2}{2}\right)f\right\rangle\right|^2\\&=\tfrac{1}{4}|\langle u_1,w*|u_2|^2f\rangle|^2\\&\leqslant\|f\|_2^2\langle u_1,(w*|u_2|^2)^2u_1\rangle\leqslant C\|f\|_2^2w_{1212},\end{aligned}$$

where we have bounded one of the $w*|u_2|^2$ in the square by a constant. Using (4-7) this implies

$$|\langle u_1,(w*(u_+u_-)-K_{11})f\rangle|\leqslant C_\varepsilon T^{1/2-\varepsilon}\|f\|_2.$$

Noting that the operator $W_{r,\leqslant M}$ is bounded (recall that $h_{\mathrm{MF}}-\mu_+$ has a finite gap by (A-5) and $K_{11}\geqslant 0$), this yields

$$\begin{aligned}&|\langle u_1,w*(u_+u_-)W_{r,\leqslant M}w*(u_+u_-)u_1\rangle-\langle u_1,K_{11}W_{r,\leqslant M}K_{11}u_1\rangle|\\&\qquad\leqslant C_\varepsilon T^{1/2-\varepsilon}(\|W_{r,\leqslant M}\,w*(u_+u_-)\,u_1\|_2^2+\|W_{r,\leqslant M}\,K_{11}\,u_1\|_2)\\&\qquad\leqslant C_\varepsilon T^{1/2-\varepsilon}.\end{aligned}$$

This means that we can replace $w*(u_+u_-)$ by K_{11} in (7-14), thus obtaining the term proportional to $\mathfrak{D}^2$ in (7-9), at the expense of a remainder term of the form

$$\frac{C_\varepsilon T^{1/2-\varepsilon}}{N-1}\mathfrak{D}^2. \qquad \square$$

Lower bound on the shifted Hamiltonian. We now discuss how to minimize $\widetilde{\mathbb{H}^{(M)}_{\text{right}}}+\widetilde{\mathbb{H}^{(M)}_{\text{left}}}$.

Proposition 7.4 (lower bound for the full shifted Hamiltonian). *Let E^{Bog} be defined in* (2-29). *Then*

$$\begin{aligned}\widetilde{\mathbb{H}^{(M)}_{\text{right}}}+\widetilde{\mathbb{H}^{(M)}_{\text{left}}}\geqslant E^{\text{Bog}}&+\tfrac{1}{2}\operatorname{Tr}[P_{r,\leqslant M}(h_{\mathrm{MF}}-\mu_++\lambda K_{11})]\\&+\tfrac{1}{2}\operatorname{Tr}[P_{\ell,\leqslant M}(h_{\mathrm{MF}}-\mu_++\lambda K_{22})]-\frac{C_M}{\sqrt{N}}(\mathcal{N}_\perp+1).\end{aligned}\tag{7-15}$$

The lower bound (7-15) is one of the main points in which our proofs significantly deviate from the standard techniques of derivation of Bogoliubov theory. Indeed, the Hamiltonian $\widetilde{\mathbb{H}_{\text{right}}}$ (with or without cutoff) is defined in terms of operators which *do not* satisfy an exact CCR (see Lemma 7.2 above). For this reason, the techniques that are normally used to diagonalize quadratic Hamiltonians (see, e.g., [Lewin et al. 2015, Appendix A]) are not directly applicable here, and we thus need slightly different methods in order to recover the correct energy E^{Bog} in (7-15). We will adopt a method already used in

[Grech and Seiringer 2013], whose main point is to perform a suitable linear symplectic transformation mixing creators and annihilators (Bogoliubov transformation). After such a transformation the original Hamiltonian is brought into a diagonal part in the new creation and annihilation operators $d^\sharp_{r,\alpha}$ and a part containing commutators of these operators. If the $\tilde{b}^\sharp_{r,\alpha}$ satisfied the CCR, then the same would be true for the $d^\sharp_{r,\alpha}$ and after the transformation the Hamiltonian would have the form $\sum_\alpha e_\alpha d^\dagger_{r,\alpha} d_{r,\alpha} + E^{\mathrm{Bog}}$. In our case, however, this is not true, and the commutators will be corrected by terms that need to be controlled. Since we work here with a finite number of modes (due to the energy cutoff), we can simplify the analysis by considering the symmetrized versions of the quadratic Hamiltonians defined in (7-5)–(7-6), instead of the Hamiltonians obtained from (3-21) and (3-22) by replacing the creators and annihilators $b^\sharp_{r,\alpha}$ and $c^\sharp_{\ell,\alpha}$ by $\tilde{b}^\sharp_{r,\alpha}$ and $\tilde{c}^\sharp_{\ell,\alpha}$.

The proof of Proposition 7.4 will occupy the rest of the present section. Define the operators

$$D_r := P_r(h_{\mathrm{MF}} - \mu_+)P_r, \quad D_{r,\leqslant M} := P_{r,\leqslant M}(h_{\mathrm{MF}} - \mu_+)P_{r,\leqslant M}. \tag{7-16}$$

The operators D_ℓ and $D_{\ell,\leqslant M}$ are defined similarly.

Recall from (2-29) that $E^{\mathrm{Bog}} = E^{\mathrm{Bog}}_r + E^{\mathrm{Bog}}_\ell$ with

$$E^{\mathrm{Bog}}_r := -\tfrac{1}{2}\mathrm{Tr}_{\perp,r}\big[D_r + \lambda P_r K_{11} P_r - \sqrt{D_r^2 + 2\lambda D_r^{1/2} P_r K_{11} P_r D_r^{1/2}}\big].$$

The quantity E^{Bog}_r is the ground state energy

$$E^{\mathrm{Bog}}_r = \inf \mathrm{spec}(\mathbb{H}^{\Theta=\mathbb{1}}_{\mathrm{right}}) \tag{7-17}$$

of the quadratic Hamiltonian

$$\mathbb{H}^{\Theta=\mathbb{1}}_{\mathrm{right}} := \sum_{\alpha,\beta\geqslant 1} \langle u_{r,\alpha}, (D_r + \lambda P_r K_{11} P_r) u_{r,\beta}\rangle A^\dagger_\alpha A_\beta + \frac{\lambda}{2} \sum_{\alpha,\beta\geqslant 1} \langle u_{r,\alpha}, P_r K_{11} P_r\, u_{r,\beta}\rangle (A^\dagger_\alpha A^\dagger_\beta + \mathrm{h.c.}), \tag{7-18}$$

where $A^\sharp_\alpha$ are canonical creation and annihilation operators on a Fock space $\mathfrak{F}_{\perp,r}$ whose base space is the span of the right modes $u_{r,\alpha}$, $\alpha \geqslant 1$; that is, the $A^\sharp_\alpha$ are operators on $\mathfrak{F}_{\perp,r}$ satisfying the CCR (the notation $\Theta = \mathbb{1}$ is there to recall that this Hamiltonian can be formally obtained from $\mathbb{H}_{\mathrm{right}}$ by setting Θ equal to the identity inside the $b^\sharp$). Equation (7-17) can be deduced by replicating the arguments of [Grech and Seiringer 2013, Section 4–5] or [Lewin et al. 2015, Appendix A]. The fact that the operator

$$D_r + \lambda P_r K_{11} P_r - \sqrt{D_r^2 + 2\lambda D_r^{1/2} P_r K_{11} P_r D_r^{1/2}}$$

is trace-class on the space $P_r L^2(\mathbb{R}^d)$ is part of the proof; see [Grech and Seiringer 2013, equation (53) and below]. The adaptation to our case is immediate because the method does not depend on the details of D_r.

It follows from the variational principle that E^{Bog}_r is bounded from above by the ground state energy $E^{\mathrm{Bog}}_{r,\leqslant M}$ of a quadratic Hamiltonian obtained from (7-18) by ignoring the modes $u_{r,\alpha}$, $\alpha > M$, i.e.,

$$\begin{aligned}\mathbb{H}^{(M),\Theta=\mathbb{1}}_{\mathrm{right}} := &\sum_{1\leqslant\alpha,\beta\leqslant M} \langle u_{r,\alpha}, (D_{r,\leqslant M} + \lambda P_{r,\leqslant M} K_{11} P_{r,\leqslant M}) u_{r,\beta}\rangle A^\dagger_\alpha A_\beta \\ &+ \frac{\lambda}{2} \sum_{1\leqslant\alpha,\beta\leqslant M} \langle u_{r,\alpha}, P_{r,\leqslant M} K_{11} P_{r,\leqslant M}\, u_{r,\beta}\rangle (A^\dagger_\alpha A^\dagger_\beta + \mathrm{h.c.}).\end{aligned} \tag{7-19}$$

The aforementioned arguments adapted to the finite-dimensional setting ensure that

$$E^{\mathrm{Bog}}_{r,\leqslant M} := -\tfrac{1}{2}\mathrm{Tr}_{\perp,r}\big[D_{r,\leqslant M} + \lambda P_{r,\leqslant M} K_{11} P_{r,\leqslant M} - \sqrt{D^2_{r,\leqslant M} + 2\lambda D^{1/2}_{r,\leqslant M} P_{r,\leqslant M} K_{11} P_{r,\leqslant M} D^{1/2}_{r,\leqslant M}}\big].$$

Notice that E^{Bog}_r is formally obtained from $E^{\mathrm{Bog}}_{r,\leqslant M}$ by replacing $P_{r,\leqslant M}$ by P_r (i.e., $M = \infty$). The ground state energies E^{Bog}_ℓ and $E^{\mathrm{Bog}}_{\ell,\leqslant M}$ of the left Bogoliubov Hamiltonians without and with energy cutoff are given by a similar expressions as in (7-18) and (7-19), with r replaced by ℓ and K_{11} replaced by K_{22}.

Lemma 7.5 (Bogoliubov energies with and without cutoff). *One has*

$$E^{\mathrm{Bog}}_r \leqslant E^{\mathrm{Bog}}_{r,\leqslant M}, \quad E^{\mathrm{Bog}}_\ell \leqslant E^{\mathrm{Bog}}_{\ell\leqslant M}. \tag{7-20}$$

Proof. As we already mentioned, E^{Bog}_r and $E^{\mathrm{Bog}}_{r,\leqslant M}$ are the ground state energies of the quadratic Hamiltonians (7-18) and (7-19). They are reached (see previous references again) by unique (up to a phase) ground states. Let $\Phi^{(M),\Theta=\mathbb{1}}$ be the ground state of $\mathbb{H}^{(M),\Theta=\mathbb{1}}_{\mathrm{right}}$. We have that

$$\langle \mathbb{H}^{\Theta=\mathbb{1}}_{\mathrm{right}} \rangle_{\Phi^{(M),\Theta=\mathbb{1}}} = E^{\mathrm{Bog}}_{r,\leqslant M}$$

because all terms with α, $\beta \geqslant M$ vanish, since $\Phi^{(M),\Theta=\mathbb{1}}$ has no components in the sectors of the Fock space corresponding to those modes. The claimed result thus immediately follows from the variational principle. □

We now prove that $\widetilde{\mathbb{H}^{(M)}_{\mathrm{right}}}$ can be bounded from below by $E^{\mathrm{Bog}}_{r,\leqslant M}$, up to

- a correcting term originating from the symmetrization in the creators and annihilators in the definitions (7-5) and (7-6),
- a controllable error due to operators entering $\widetilde{\mathbb{H}^{(M)}_{\mathrm{right}}}$ do not exactly satisfy the CCR.

Lemma 7.6 (lower bounds for the shifted Hamiltonians). *We have*

$$\widetilde{\mathbb{H}^{(M)}_{\mathrm{right}}} \geqslant \frac{1}{2}\,\mathrm{Tr}[D_{r,\leqslant M} + \lambda P_{r,\leqslant M} K_{11}] + E^{\mathrm{Bog}}_{r,\leqslant M} - \frac{C_M}{\sqrt{N}}(\mathcal{N}_\perp + 1), \tag{7-21}$$

$$\widetilde{\mathbb{H}^{(M)}_{\mathrm{left}}} \geqslant \frac{1}{2}\,\mathrm{Tr}[D_{\ell,\leqslant M} + \lambda P_{\ell,\leqslant M} K_{22}] + E^{\mathrm{Bog}}_{\ell,\leqslant M} - \frac{C_M}{\sqrt{N}}(\mathcal{N}_\perp + 1). \tag{7-22}$$

The bound of Proposition 7.4 immediately follows from (7-21), (7-22), Lemma 7.5, and $E^{\mathrm{Bog}} = E^{\mathrm{Bog}}_r + E^{\mathrm{Bog}}_\ell$. There thus only remains to provide the following proof.

Proof of Lemma 7.6. We discuss (7-21) only, since (7-22) can be obtained by completely analogous arguments. Let us define the $M \times M$ real symmetric matrices

$$\begin{aligned} D &:= (\langle u_{r,\alpha},\ D_{r,\leqslant M}\, u_{r,\beta}\rangle)^M_{\alpha,\beta=1}, \\ V &:= \lambda(\langle u_{r,\alpha},\ P_{r,\leqslant M} K_{11} P_{r,\leqslant M}\, u_{r,\beta}\rangle)^M_{\alpha,\beta=1}, \\ E &:= \sqrt{D^2 + 2D^{1/2} V D^{1/2}}. \end{aligned} \tag{7-23}$$

The notation is chosen to allow direct comparison with the arguments in [Grech and Seiringer 2013, Sections 4–5]. In terms of these matrices, we have

$$\widetilde{\mathbb{H}^{(M)}_{\mathrm{right}}} = \frac{1}{2}\left((\tilde{\boldsymbol{b}}^{\dagger})^{t}, \tilde{\boldsymbol{b}}^{t}\right)\begin{pmatrix} D+V & V \\ V & D+V \end{pmatrix}\begin{pmatrix} \tilde{\boldsymbol{b}} \\ \tilde{\boldsymbol{b}}^{\dagger} \end{pmatrix}, \tag{7-24}$$

where we have used the matrix notation $\tilde{\boldsymbol{b}} = (\tilde{b}_{r,\alpha})_{\alpha=1}^{M}$ and $\tilde{\boldsymbol{b}}^{\dagger} = (\tilde{b}^{\dagger}_{r,\alpha})_{\alpha=1}^{M}$ for the creation and annihilation operators and t denotes the transpose.

Let us introduce new creators and annihilators $d^{\sharp}_{r,\alpha}$ obtained by means of the Bogoliubov transformation

$$\begin{pmatrix} \boldsymbol{d} \\ \boldsymbol{d}^{\dagger} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} A_0^{-1}+B_0^{-1} & A_0^{-1}-B_0^{-1} \\ A_0^{-1}-B_0^{-1} & A_0^{-1}+B_0^{-1} \end{pmatrix}\begin{pmatrix} \tilde{\boldsymbol{b}} \\ \tilde{\boldsymbol{b}}^{\dagger} \end{pmatrix}, \tag{7-25}$$

where A_0 and B_0 are the real $M \times M$ matrices defined by

$$A_0 := D^{1/2}E^{-1/2}U_0, \quad B_0 := (A_0^{-1})^{t} = D^{-1/2}E^{1/2}U_0,$$

with U_0 the orthogonal $M \times M$ matrix diagonalizing E,

$$U_0^{t}EU_0 = \Lambda = \mathrm{diag}(e_{\alpha}).$$

The inverse transformation is

$$\begin{pmatrix} \tilde{\boldsymbol{b}} \\ \tilde{\boldsymbol{b}}^{\dagger} \end{pmatrix} = S\begin{pmatrix} \boldsymbol{d} \\ \boldsymbol{d}^{\dagger} \end{pmatrix} := \frac{1}{2}\begin{pmatrix} A_0+B_0 & A_0-B_0 \\ A_0-B_0 & A_0+B_0 \end{pmatrix}\begin{pmatrix} \boldsymbol{d} \\ \boldsymbol{d}^{\dagger} \end{pmatrix}. \tag{7-26}$$

The matrix S is symplectic and diagonalizes the $2M \times 2M$ symmetric matrix in (7-24),

$$S^{t}\begin{pmatrix} D+V & V \\ V & D+V \end{pmatrix}S = \begin{pmatrix} \Lambda & 0 \\ 0 & \Lambda \end{pmatrix}$$

(this can be checked by an explicit calculation, noting that $A_0^{t}(D+2V)A_0 = B_0^{t}DB_0 = \Lambda$). Thus

$$\widetilde{\mathbb{H}^{(M)}_{\mathrm{right}}} = \frac{1}{2}\left((\boldsymbol{d}^{\dagger})^{t}, \boldsymbol{d}^{t}\right)\begin{pmatrix} \Lambda & 0 \\ 0 & \Lambda \end{pmatrix}\begin{pmatrix} \boldsymbol{d} \\ \boldsymbol{d}^{\dagger} \end{pmatrix} = \sum_{\alpha=1}^{M} e_{\alpha} d^{\dagger}_{r,\alpha} d_{r,\alpha} + \frac{1}{2}\sum_{\alpha=1}^{M} e_{\alpha}[d_{r,\alpha}, d^{\dagger}_{r,\alpha}].$$

If the operators $\tilde{b}^{\sharp}_{r,\alpha}$ satisfied the CCR, the same would be true for the $d^{\sharp}_{r,\alpha}$ and the last sum would be equal to

$$\mathrm{Tr}(E) = \mathrm{Tr}\sqrt{D^{2}_{r,\leqslant M} + 2\lambda D^{1/2}_{r,\leqslant M} P_{r,\leqslant M} K_{11} P_{r,\leqslant M} D^{1/2}_{r,\leqslant M}},$$

which is precisely the sum of the two first terms in the right-hand side of (7-21).

In our case, the sum involving the commutators can be obtained from the following identity: if R is a real $M \times M$ symmetric matrix, then

$$[\boldsymbol{d}^{t},\ R\,\boldsymbol{d}^{\dagger}] := \sum_{1\leqslant\alpha,\beta\leqslant M} R_{\alpha\beta}[d_{r,\alpha},\ d^{\dagger}_{r,\beta}] = \mathrm{Tr}(R) - \boldsymbol{x}^{t}B_0RA_0^{t}(\boldsymbol{b}+\boldsymbol{b}^{\dagger}), \tag{7-27}$$

where $\boldsymbol{x} = (x_{\alpha})_{\alpha=1}^{M}$ is given by (7-13). The identity (7-27) follows by noting that the commutation relations of the $\tilde{b}^{\sharp}_{r,\alpha}$ given in Lemma 7.2 can be rewritten as

$$[\tilde{\boldsymbol{b}}^{t},\ Q\,\tilde{\boldsymbol{b}}] = \boldsymbol{x}^{t}(Q-Q^{t})\boldsymbol{b}, \quad [\tilde{\boldsymbol{b}}^{t},\ Q\,\tilde{\boldsymbol{b}}^{\dagger}] = \mathrm{Tr}(Q) - \boldsymbol{x}^{t}(Q^{t}\,\boldsymbol{b} + Q\,\boldsymbol{b}^{\dagger}) \tag{7-28}$$

for any $M \times M$ matrix Q. One deduces from (7-25) and from $A_0^{-1} = B_0^t$, $B_0^{-1} = A_0^t$ that

$$[\boldsymbol{d}^t,\ R\,\boldsymbol{d}^\dagger] = -\tfrac{1}{4}[\tilde{\boldsymbol{b}}^t,\ (A_0+B_0)R(A_0-B_0)^t\,\tilde{\boldsymbol{b}}] + \text{h.c.} \\ + \tfrac{1}{4}[\tilde{\boldsymbol{b}}^t,\ (A_0+B_0)R(A_0+B_0)^t\,\tilde{\boldsymbol{b}}^\dagger] - \tfrac{1}{4}[\tilde{\boldsymbol{b}}^t,\ (A_0-B_0)R(A_0-B_0)^t\,\tilde{\boldsymbol{b}}^\dagger],$$

from which (7-27) is obtained by relying on (7-28).

Applying (7-27) with $R = \Lambda$ yields

$$\widetilde{\mathbb{H}^{(M)}_{\text{right}}} = \sum_{\alpha=1}^{M} e_\alpha d^\dagger_{r,\alpha} d_{r,\alpha} + \frac{1}{2}\operatorname{Tr}(E) - \frac{\lambda}{2\sqrt{2(N-1)}}\boldsymbol{w}^t_{+1-} D^{1/2}E^{-1}D^{1/2}(\boldsymbol{b}+\boldsymbol{b}^\dagger), \tag{7-29}$$

where $\boldsymbol{w}_{+1-}$ stands for the vector $(w_{+1-\{r,\alpha\}})_{\alpha=1}^{M}$. To deduce the above equation we used

$$(D+2V)^{-1}B_0\Lambda A_0^t = D^{1/2}E^{-1}D^{1/2},$$

which follows thanks to the identities $B_0\Lambda A_0^t = D^{-1/2}ED^{1/2}$ and $D^{-1/2}E^2D^{-1/2} = (D+2V)$. The expectation of the last term in (7-29) on the vector $\Phi \in \ell^2(\mathfrak{F}_\perp)$ can be bounded using the Cauchy–Schwarz inequality, the boundedness of $w*(u_+u_-)$, and the fact that $E^{-1} \leqslant D^{-1}$ by operator monotonicity of the inverse and square root (recall that $E^2 = D^{1/2}(D+2V)D^{1/2} \geqslant D^2$ since $V \geqslant 0$) to write

$$\begin{aligned}\left|\frac{1}{2\sqrt{2(N-1)}}\boldsymbol{w}^t_{+1-}D^{1/2}E^{-1}D^{1/2}\langle \boldsymbol{b}+\boldsymbol{b}^\dagger\rangle_\Phi\right| &\leqslant \frac{C}{\sqrt{N}}\left\{\sum_{\alpha\geqslant 1}|w_{+1-\{r,\alpha\}}|^2\right\}^{1/2}\left\{\sum_{\alpha\geqslant 1}\left\|\sum_{\beta\geqslant 1}(D^{1/2}E^{-1}D^{1/2})_{\alpha\beta}b_{r,\beta}\Phi\right\|^2\right\}^{1/2} \\ &\leqslant \frac{C}{\sqrt{N}}\langle \mathcal{N}_\perp\rangle_\Phi^{1/2}.\end{aligned}$$

The lower bound in the lemma then follows from the fact that the first term in (7-29) is nonnegative (since $E \geqslant 0$ and thus $e_\alpha \geqslant 0$ for all α). □

8. Proof of the main results

Recall that Proposition 2.4 follows from the considerations of Section 4.

8A. ***Energy upper bound.*** We obtain an upper bound on the ground state energy $E(N)$ corresponding to (2-32) by constructing a trial state ψ_{trial} as follows. Recall that by the decomposition (3-8), any wave-function ψ is uniquely identified by the components $\Phi_{s,d}$ of $\mathcal{U}_N\psi$. The d-dependence of the components of $\mathcal{U}_N\psi_{\text{trial}}$ will be encoded in the gaussian coefficients $c_d = e^{-d^2/4\sigma_N^2}/Z_N$ that we already used in Section 4. The s-dependence, in turn, will be chosen so that the expectation of $\mathbb{H}$ on $\mathcal{U}_N\psi_{\text{trial}}$ will coincide (up to remainders) with E^{Bog} defined in (2-29). To evaluate this part of the energy, we need a well-known lemma. Its claims follow, e.g., from arguments[6] in [Grech and Seiringer 2013].

[6] In particular, notice that the transformation in [Grech and Seiringer 2013, equation (26)] is implemented in Fock space by e^{X_a}, where X_a is defined before Lemma 3 in that work.

Lemma 8.1 (minimization of quadratic Hamiltonians). *Let V be a locally bounded external potential such that* $\lim_{|x|\to\infty} V(x) = +\infty$, *and define* $h := -\Delta + V$. *Let k be the integral operator on $L^2(\mathbb{R}^d)$ whose kernel is $u(x)w(x-y)u(y)$ for a real-valued $u \in L^2(\mathbb{R}^d)$ and w as in Assumption 2.1. Given an orthonormal basis $\{u_n\}$ of $L^2(\mathbb{R}^d)$ such that all u_n are real-valued, denote by $h_{mn} = \langle u_m, h\, u_n\rangle$ and $k_{mn} = \langle u_m, k\, u_n\rangle$ the matrix elements of h and k in this basis. Consider the quadratic Hamiltonian*

$$\mathbb{H}_{\text{quad}} = \sum_{m,n} (h+k)_{mn} A_m^\dagger A_n + \frac{1}{2}\sum_{m,n} k_{mn}(A_m^\dagger A_n^\dagger + A_m A_n),$$

where $A_m^\dagger$ and A_n are creation and annihilation operators on the Fock space $\mathcal{G}$ with base $L^2(\mathbb{R}^d)$ satisfying the canonical commutation relations. Then the unique (*up to a phase*) *ground state of* $\mathbb{H}_{\text{quad}}$ *is*

$$\mathbb{U}\,\Omega_{\mathcal{G}},$$

where $\Omega_{\mathcal{G}}$ is the vacuum vector of $\mathcal{G}$ and $\mathbb{U}$ a Bogoliubov transformation, *acting on creation/annihilation operators as*

$$\mathbb{U}^* A_m^\dagger \mathbb{U} = \sum_n (c_{mn} A_n^\dagger + s_{mn} A_n) \tag{8-1}$$

for suitable coefficients c_{mn} and s_{mn}. Moreover, *the ground state energy of* $\mathbb{H}_{\text{quad}}$ *is*

$$\inf \sigma(\mathbb{H}_{\text{quad}}) = -\tfrac{1}{2}\operatorname{Tr}\big(h + k - \sqrt{h^2 + 2h^{1/2}kh^{1/2}}\big). \tag{8-2}$$

We refer to [Lewin et al. 2015; Grech and Seiringer 2013; Nam et al. 2016; Bach and Bru 2016; Bruneau and Dereziński 2007; Dereziński 2017] for more details. It follows from (8-1) that we have

$$\langle \mathbb{U}\,\Omega_{\mathcal{G}} \mid A_m^\dagger \mathbb{U}\,\Omega_{\mathcal{G}}\rangle = 0, \tag{8-3}$$

i.e., particles appear only in pairs in the Bogoliubov ground state. Moreover, by using the fact that $\mathbb{U}\,\Omega_{\mathcal{G}}$ is a quasifree state, one can show that all moments of the number operator $\mathcal{N}_\perp = \sum_n A_n^\dagger A_n$ in this state are finite; i.e., $\langle \mathcal{N}_\perp^k\rangle_{\mathbb{U}\,\Omega_{\mathcal{G}}} < \infty$ for all positive integer k.

Recall the Bogoliubov Hamiltonian $\mathbb{H}_{\text{right}}$ for right modes, defined in (3-21). Let us consider its version in which the d-translation operator Θ is formally set to the identity. This amounts to replacing the $b^\sharp$ with the $a^\sharp$, i.e.,

$$\mathbb{H}_{\text{right}}^{\Theta=\mathbb{1}} := \sum_{\alpha,\beta\geqslant 1} \langle u_{r,\alpha}, (h_{\text{MF}} - \mu_+ + \lambda K_{11})u_{r,\beta}\rangle a_{r,\alpha}^\dagger a_{r,\beta} + \frac{\lambda}{2}\sum_{\alpha,\beta\geqslant 1}\langle u_{r,\alpha}, K_{11}u_{r,\beta}\rangle (a_{r,\alpha}^\dagger a_{r,\beta}^\dagger + a_{r,\alpha}a_{r,\beta}).$$

This operator acts on the right Fock space

$$\mathfrak{F}_\perp^r = \mathfrak{F}(P_{\perp,r}L^2(\mathbb{R}^d)), \quad P_{\perp,r} := \sum_{\alpha\geqslant 1} |u_{r,\alpha}\rangle\langle u_{r,\alpha}|. \tag{8-4}$$

Similarly, we consider the Bogoliubov Hamiltonian $\mathbb{H}_{\text{left}}^{\Theta=\mathbb{1}}$ for the left modes and the left Fock space $\mathfrak{F}_\perp^\ell$, defined by the same formulas with r replaced by ℓ and K_{11} by K_{22}. We extend both operators to the full excited Fock space $\mathfrak{F}_\perp$ by using the unitary equivalence

$$\mathfrak{F}_\perp = \mathfrak{F}\big((P_{\perp,r}L^2(\mathbb{R}^d)) \oplus (P_{\perp,\ell}L^2(\mathbb{R}^d))\big) \simeq \mathfrak{F}_\perp^r \oplus \mathfrak{F}_\perp^\ell$$

and having $\mathbb{H}_{\mathrm{right}}^{\Theta=\mathbb{1}}$ acting as the identity on the left Fock space (respectively $\mathbb{H}_{\mathrm{left}}^{\Theta=\mathbb{1}}$ acting as the identity on the right Fock space). Applying Lemma 8.1, there exist unitary Bogoliubov transformations $\mathbb{U}_{\mathrm{right}}$ and $\mathbb{U}_{\mathrm{left}}$ such that

$$\mathbb{H}_{\mathrm{right}}^{\Theta=\mathbb{1}}\mathbb{U}_{\mathrm{right}}\,\Omega = E_{\mathrm{right}}^{\mathrm{Bog}}\mathbb{U}_{\mathrm{right}}\,\Omega,$$
$$\mathbb{H}_{\mathrm{left}}^{\Theta=\mathbb{1}}\mathbb{U}_{\mathrm{left}}\,\Omega = E_{\mathrm{left}}^{\mathrm{Bog}}\mathbb{U}_{\mathrm{left}}\,\Omega,$$

with Ω the vacuum vector of $\mathfrak{F}_\perp$ and

$$E_{\mathrm{right}}^{\mathrm{Bog}} = -\tfrac{1}{2}\mathrm{Tr}_{\perp,r}\big[D_r + \lambda P_r K_{11} P_r - \sqrt{(D_r)^2 + 2\lambda D_r^{1/2} P_r K_{11} P_r D_r^{1/2}}\big],$$
$$E_{\mathrm{left}}^{\mathrm{Bog}} = -\tfrac{1}{2}\mathrm{Tr}_{\perp,\ell}\big[D_\ell + \lambda P_\ell K_{22} P_\ell - \sqrt{(D_\ell)^2 + 2\lambda D_\ell^{1/2} P_\ell K_{22} P_\ell D_\ell^{1/2}}\big],$$

where D_r, D_ℓ are defined in (7-16).

The latter quantities are those given by adapting (8-2) to our case. Their sum coincides with E^{Bog} defined in (2-29). By construction, $\mathbb{H}_{\mathrm{right}}^{\Theta=\mathbb{1}}$ commutes with $\mathbb{U}_{\mathrm{left}}$, because the latter is defined in terms of left modes only. Similarly, $\mathbb{H}_{\mathrm{left}}^{\Theta=\mathbb{1}}$ commutes with $\mathbb{U}_{\mathrm{right}}$. Thus

$$\begin{aligned}(\mathbb{H}_{\mathrm{right}}^{\Theta=\mathbb{1}} + \mathbb{H}_{\mathrm{left}}^{\Theta=\mathbb{1}})\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\,\Omega &= (E_{\mathrm{right}}^{\mathrm{Bog}} + E_{\mathrm{left}}^{\mathrm{Bog}})\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\,\Omega \\ &= E^{\mathrm{Bog}}\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\,\Omega. \end{aligned} \tag{8-5}$$

We denote by $(\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\,\Omega)_s$ the component of $\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\,\Omega$ in the s-particle sector of $\mathfrak{F}_\perp$.

We are now ready to define our trial state. To control some terms arising from Bogoliubov excitations, our choice of variance differs slightly from that of Section 4.

Definition 8.2 (trial state with fluctuations). We define

$$\psi_{\mathrm{trial}} := \sum_{s=0}^{N}\sum_{|d|\leqslant\sigma_N^2} c_{d,s} u_1^{\otimes(N-s+d)/2} \otimes_{\mathrm{sym}} u_2^{\otimes(N-s-d)/2} \otimes_{\mathrm{sym}} \Phi_{\mathrm{trial},s}, \tag{8-6}$$

where the coefficients $c_{d,s}$ are defined by

$$c_{d,s} = \begin{cases} (1/Z_N)e^{-d^2/4\sigma_N^2} & \text{if } N-s+d \text{ is even and } |d|\leqslant\sigma_N^2, \\ 0 & \text{otherwise,}\end{cases} \tag{8-7}$$

Z_N being a normalization factor such that $\sum_{|d|\leqslant\sigma_N^2} c_{d,s}^2 = 1$ for all s and

$$\sigma_N^2 = \begin{cases} \sqrt{\mu_- - \mu_+ N} & \text{if } \delta < 1 \text{ in the assumption } T\sim N^{-\delta}, \\ N^{1/2} & \text{otherwise.}\end{cases} \tag{8-8}$$

Moreover, let

$$\Phi_{\mathrm{trial},s} := \frac{(\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\,\Omega)_s}{\sqrt{\sum_{s=0}^{N}\|(\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\,\Omega)_s\|^2}}. \tag{8-9}$$

The excitation content of ψ_{trial} is

$$(\mathcal{U}_N\psi_{\mathrm{trial}})_{s,d} = c_{d,s}\Phi_{\mathrm{trial},s}$$

for $0\leqslant s\leqslant N$ and $|d|\leqslant\sigma_N^2$, and zero otherwise. Note that the function of the s-variables $\Phi_{\mathrm{trial},s}$ does not depend on d, and that $c_{d,s} = c_{d,s'}$ for all d if s and s' have the same parity. Note also that ψ_{trial} is normalized to 1. In the rest of this subsection we prove:

Proposition 8.3 (energy upper bound). *Pick a sequence $T(N) \sim N^{-\delta}$ with $0 < \delta$. Then, along this sequence,*

$$\limsup_{N\to\infty}(\langle H_N\rangle_{\psi_{\mathrm{trial}}} - E_{2\text{-mode}} - E^{\mathrm{Bog}}) \leqslant 0. \tag{8-10}$$

Proof. By using Proposition 5.1 with $\Phi = \mathcal{U}_N \psi_{\mathrm{trial}}$ to estimate $\langle H_N\rangle_{\psi_{\mathrm{trial}}}$, one obtains the upper bound

$$\langle H_N\rangle_{\psi_{\mathrm{trial}}} \leqslant \langle H_{2\text{-mode}}\rangle_{\psi_{\mathrm{trial}}} + \langle \mathbb{H}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}} + \mu_+ \langle \mathcal{N}_\perp\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}} + \langle \text{linear terms }\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}} - \text{error terms}. \tag{8-11}$$

We first determine the expectations in the trial state of the two-mode Hamiltonian $H_{2\text{-mode}}$ (Step 1), then that of the Bogoliubov Hamiltonian $\mathbb{H}$ (Steps 2 and 3), before showing that the expectation of the linear terms and the error terms converge to zero as $N \to \infty$.

Step 1: two-mode energy of the trial state. The two-mode Hamiltonian (4-9) does not contain operators that change the number of excitations (i.e., the index s). The only terms in $H_{2\text{-mode}}$ that involve the variable s are those containing $\mathcal{N}_\perp$ or $\mathcal{N}_\perp^2$. For example, we compute

$$\langle \mathcal{N}_\perp^2\rangle_{\psi_{\mathrm{trial}}} = \sum_{s=0}^{N}\sum_{|d|\leqslant\sigma_N^2} |c_{d,s}|^2 s^2 \|\Phi_{\mathrm{trial},s}\|^2 = \sum_{s=0}^{N} s^2\|\Phi_{\mathrm{trial},s}\|^2 = \frac{\langle \mathcal{N}_\perp^2 \mathbb{1}_{\mathcal{N}_\perp\leqslant N}\rangle_{\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\Omega}}{\|\mathbb{1}_{\mathcal{N}_\perp\leqslant N}\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\Omega\|^2}.$$

The denominator in the last line tends to 1 when $N \to \infty$ and it easily follows from the previous definitions that

$$\langle \mathcal{N}_\perp^2\rangle_{\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\Omega} = \langle \mathcal{N}_\perp^2\rangle_{\mathbb{U}_{\mathrm{left}}\Omega} + \langle \mathcal{N}_\perp^2\rangle_{\mathbb{U}_{\mathrm{right}}\Omega}.$$

Since both moments in the right-hand side are finite, it follows that

$$\langle \mathcal{N}_\perp^2\rangle_{\psi_{\mathrm{trial}}} \leqslant C \tag{8-12}$$

for a constant $C > 0$ independent of N. By the Cauchy–Schwarz inequality, this implies $\langle \mathcal{N}_\perp\rangle_{\psi_{\mathrm{trial}}} \leqslant \sqrt{C}$.

For all other terms of $H_{2\text{-mode}}$ in (4-9), i.e., those that only contain $a_1^\sharp$ and $a_2^\sharp$, we will use a general formula of the type

$$\begin{aligned}\langle f(a_1^\sharp, a_2^\sharp)\rangle_{\psi_{\mathrm{trial}}} = \sum_{s=0}^{N}\sum_{|d|\leqslant\sigma_N^2}\sum_{|d'|\leqslant\sigma_N^2} & c_{d,s}c_{d',s}\|\Phi_{\mathrm{trial},s}\|^2 \\ & \times \langle u_1^{\otimes(N-s+d')/2} \otimes_{\mathrm{sym}} u_2^{\otimes(N-s-d')/2}, f(a_1^\sharp, a_2^\sharp)\, u_1^{\otimes(N-s+d)/2} \otimes_{\mathrm{sym}} u_2^{\otimes(N-s-d)/2}\rangle.\end{aligned}$$

To compute the expectations in the second line, we can repeat the calculations performed in the proof of the upper bound (4-23) for the two-mode Hamiltonian, keeping track of the fact that the total number of particles is now $N - s$ for a generic $0 \leqslant s \leqslant N$. Let

$$\psi_{\mathrm{trial},s} := \sum_{|d|\leqslant\sigma_N^2} c_{d,s} u_1^{\otimes(N-s+d)/2} \otimes_{\mathrm{sym}} u_2^{\otimes(N-s-d)/2} \otimes_{\mathrm{sym}} \Phi_{\mathrm{trial},s}$$

be the component of ψ_{trial} with exactly s excitations. One finds

$$\langle (\mathcal{N}_1 + \mathcal{N}_2)^n\rangle_{\psi_{\mathrm{trial},s}} = (N-s)^n\|\Phi_{\mathrm{trial},s}\|^2,$$

$$\begin{aligned}\langle \mathcal{N}_-\rangle_{\psi_{\mathrm{trial},s}} = \tfrac{1}{2}\langle \mathcal{N}_1+\mathcal{N}_2 - a_1^\dagger a_2 - a_2^\dagger a_1\rangle_{\psi_{\mathrm{trial},s}} &\leqslant C\biggl(1+\frac{N-s}{\sigma_N^2}+(N-s)e^{-(N-s)/\sigma_N^2}\biggr)\|\Phi_{\mathrm{trial},s}\|^2,\\ \langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\mathrm{trial},s}} &\leqslant C_\varepsilon (N-s)T^{1/2-\varepsilon}\|\Phi_{\mathrm{trial},s}\|^2.\end{aligned}$$

Using $\sum_{s=0}^{N}\psi_{\mathrm{trial},s}=\psi_{\mathrm{trial}}$ and splitting the sum into two parts for $0\leqslant s<N/2$ and for $N/2\leqslant s\leqslant N$, one has for example ($C$ is a generic constant which may change from line to line)

$$\begin{aligned}\langle \mathcal{N}_-\rangle_{\psi_{\mathrm{trial}}} &\leqslant C\sum_{0\leqslant s<N/2}\biggl(1+\frac{N-s}{\sigma_N^2}+(N-s)e^{-(N-s)/\sigma_N^2}\biggr)\|\Phi_{\mathrm{trial},s}\|^2+CN\sum_{N/2\leqslant s\leqslant N}\|\Phi_{\mathrm{trial},s}\|^2\\ &\leqslant C\biggl(1+\frac{N}{\sigma_N^2}+Ne^{-N/2\sigma_N^2}\biggr)+\frac{C}{N}\\ &\leqslant C\biggl(1+\frac{N}{\sigma_N^2}\biggr)\leqslant C(1+\max(C_\varepsilon T^{-1/2-\varepsilon},N^{1/2})),\end{aligned}\tag{8-13}$$

where in the second line we have used $\sum_{s=0}^{N}\|\Phi_{\mathrm{trial},s}\|^2=1$ and the bound

$$\frac{N^2}{4}\sum_{N/2\leqslant s\leqslant N}\|\Phi_{\mathrm{trial},s}\|^2\leqslant\sum_{N/2\leqslant s\leqslant N}s^2\|\Phi_{\mathrm{trial},s}\|^2\leqslant\langle\mathcal{N}_\perp^2\rangle_{\psi_{\mathrm{trial}}}\leqslant C,$$

and in the third line we have used (A-4), the assumption $T\sim N^{-\delta}$, and the fact that $Ne^{-N/2\sigma_N^2}$ can be bounded by a constant times $N(\sigma_N^2/N)^{2\delta^{-1}(1-\varepsilon)^{-1}}$. Similarly, we find

$$\begin{aligned}\langle(\mathcal{N}_1+\mathcal{N}_2)^n\rangle_{\psi_{\mathrm{trial}}}=\langle(N-\mathcal{N}_\perp)^n\rangle_{\psi_{\mathrm{trial}}}&\leqslant CN^n,\\ 0\leqslant\langle N-\mathcal{N}_\perp-a_1^\dagger a_2-a_2^\dagger a_1\rangle_{\psi_{\mathrm{trial}}}=2\langle\mathcal{N}_-\rangle_{\psi_{\mathrm{trial}}}&\leqslant C(1+\max(C_\varepsilon T^{-1/2-\varepsilon},N^{1/2})),\\ \langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\mathrm{trial}}}&\leqslant\max(C_\varepsilon NT^{1/2-\varepsilon},N^{1/2}),\\ \langle\mathcal{N}_-^2\rangle_{\psi_{\mathrm{trial}}}\leqslant N\langle\mathcal{N}_-\rangle_{\psi_{\mathrm{trial}}}&\leqslant CN(1+\max(C_\varepsilon T^{-1/2-\varepsilon},N^{1/2})).\end{aligned}\tag{8-14}$$

According to the identity (4-9) of Proposition 4.2, one has

$$\begin{aligned}\langle H_{2\text{-mode}}\rangle_{\psi_{\mathrm{trial}}}=E_0+E_N^w+N\frac{\mu_+-\mu_-}{2}&+\frac{\mu_--\mu_+}{2}\langle N-a_1^\dagger a_2-a_2^\dagger a_1\rangle_{\psi_{\mathrm{trial}}}\\ &-\frac{\lambda N}{N-1}((w_{1112}+w_{1122})\langle\mathcal{N}_\perp\rangle_{\psi_{\mathrm{trial}}}-w_{1122})\\ &\quad+\frac{\lambda}{N-1}\langle((w_{1112}+w_{1122})\mathcal{N}_\perp-w_{1122})(N-a_1^\dagger a_2-a_2^\dagger a_1)\rangle_{\psi_{\mathrm{trial}}}\\ &\quad\quad-\mu\langle\mathcal{N}_\perp\rangle_{\psi_{\mathrm{trial}}}+\frac{\lambda U}{N-1}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\mathrm{trial}}}\\ &\quad\quad\quad+\frac{2\lambda}{N-1}w_{1122}\langle\mathcal{N}_-^2\rangle_{\psi_{\mathrm{trial}}}+\frac{\lambda}{4(N-1)}(w_{1111}-2w_{1122}+w_{1212})\langle\mathcal{N}_\perp^2\rangle_{\psi_{\mathrm{trial}}}.\end{aligned}$$

Plugging (8-12) and (8-14) into this identity, bounding the expectation in the third line by

$$(|w_{1112}|+w_{1122})N\langle N-a_1^\dagger a_2-a_2^\dagger a_1\rangle_{\psi_{\mathrm{trial}}},$$

and recalling the estimates for the various w-coefficients and for $\mu - \mu_+$ from Lemma 4.1, we deduce that

$$\langle H_{2\text{-mode}}\rangle_{\psi_{\mathrm{trial}}} \leqslant E_0 + E_N^w + N\frac{\mu_+ - \mu_-}{2} - \mu_+\langle \mathcal{N}_\perp\rangle_{\psi_{\mathrm{trial}}} + o_N(1)$$

in both cases of (8-8). Arguing as in Section 4C, we conclude

$$\langle H_{2\text{-mode}}\rangle_{\psi_{\mathrm{trial}}} \leqslant E_{2-\mathrm{mode}} - \mu_+\langle \mathcal{N}_\perp\rangle_{\psi_{\mathrm{trial}}} + o_N(1). \tag{8-15}$$

Step 2: Bogoliubov energy of the trial state. We want to compute $\langle \mathbb{H}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}$. We consider the decomposition analogously to (5-43):

$$\mathbb{H} = \mathbb{H}_{\mathrm{right}} + \mathbb{H}_{\mathrm{left}} + \mathbb{H}_{12} + \sum_{j=1}^{3}\xi_j, \tag{8-16}$$

with $\mathbb{H}_{\mathrm{right}}$, $\mathbb{H}_{\mathrm{left}}$ given by (3-21)–(3-22), $\mathbb{H}_{12}$ given by (5-44), and

$$\begin{aligned}
\xi_1 &= \sum_{\alpha,\beta\geqslant 1}\langle u_{r,\alpha}, (h_{\mathrm{MF}} - \mu_+)u_{\ell,\beta}\rangle a^\dagger_{r,\alpha}a_{\ell,\beta} + \text{h.c.},\\
\xi_2 &= \lambda\sum_{\alpha,\beta\geqslant 1}\langle u_{r,\alpha}, (K_{11}+K_{22})u_{\ell,\beta}\rangle a^\dagger_{r,\alpha}a_{\ell,\beta} + \text{h.c.}\\
&\qquad\qquad + \lambda\sum_{\alpha,\beta\geqslant 1}[\langle u_{r,\alpha}, K_{11}u_{\ell,\beta}\rangle\Theta^{-2} + \langle u_{r,\alpha}, K_{22}u_{\ell,\beta}\rangle\Theta^{2}]a^\dagger_{r,\alpha}a^\dagger_{\ell,\beta} + \text{h.c.},\\
\xi_3 &= \sum_{\alpha,\beta\geqslant 1}\langle u_{r,\alpha}, K_{22}u_{r,\beta}\rangle a^\dagger_{r,\alpha}a_{r,\alpha} + \sum_{\alpha,\beta\geqslant 1}\langle u_{\ell,\alpha}, K_{11}u_{\ell,\beta}\rangle a^\dagger_{\ell,\alpha}a_{\ell,\alpha}\\
&\qquad + \frac{\lambda}{2}\sum_{\alpha,\beta\geqslant 1}(\langle u_{r,\alpha}, K_{22}u_{r,\beta}\rangle\Theta^{2}a^\dagger_{r,\alpha}a^\dagger_{r,\beta} + \langle u_{\ell,\alpha}, K_{11}u_{\ell,\beta}\rangle\Theta^{-2}a^\dagger_{\ell,\alpha}a^\dagger_{\ell,\beta} + \text{h.c.}).
\end{aligned} \tag{8-17}$$

We will show below (see Step 3) that the main part of the energy in the trial state comes from the expectation of $\mathbb{H}_{\mathrm{right}} + \mathbb{H}_{\mathrm{left}}$. We now prove that the latter expectation is equal to E^{Bog} up to errors of order $N^{-1}T^{-1/2-\varepsilon}$. Each term of $\mathbb{H}_{\mathrm{right}} + \mathbb{H}_{\mathrm{left}}$ contains Θ elevated to a certain power, either -2, 0, or $+2$ (this power is zero for the $b^\dagger b$ and $c^\dagger c$ parts). We know that the excitation content of ψ_{trial} is

$$\{\mathcal{U}_N\psi_{\mathrm{trial}}\}_{s,d} = c_{d,s}\Phi_{\mathrm{trial},s};$$

thus the operator Θ acts on $\mathcal{U}_N\psi_{\mathrm{trial}}$ by simply translating the $c_{d,s}$-coefficient as $c_{d,s}\to c_{d-1,s}$. Taking one term of $\mathbb{H}_{\mathrm{right}}$ as an example, we have

$$\begin{aligned}
\sum_{\alpha,\beta\geqslant 1}(K_{11})_{\alpha\beta}\langle\Theta^2 a_{r,\alpha}a_{r,\beta}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}} &= \sum_{\alpha,\beta\geqslant 1}(K_{11})_{\alpha\beta}\sum_{s=0}^{N}\biggl(\sum_{|d|\leqslant\sigma_N^2}\langle(\mathcal{U}_N\psi_{\mathrm{trial}})_{s,d}, (a_{r,\alpha}a_{r,\beta}\mathcal{U}_N\psi_{\mathrm{trial}})_{s,d-2}\rangle\biggr)\\
&= \sum_{\alpha,\beta\geqslant 1}(K_{11})_{\alpha\beta}\sum_{s=0}^{N}\biggl(\sum_{|d|\leqslant\sigma_N^2}c_{d,s}c_{d-2,s}\biggr)\langle\Phi_{\mathrm{trial},s}, a_{r,\alpha}a_{r,\beta}\Phi_{\mathrm{trial},s+2}\rangle,
\end{aligned}$$

where we have used that $c_{d,s}$ only depends of the parity of s. For the sum over d, we know that, by (4-25), for all $\kappa\in 2\mathbb{Z}$,

$$\biggl|\sum_{|d|\leqslant\sigma_N^2}c_{d,s}c_{d\pm\kappa,s} - 1\biggr| \leqslant \frac{C}{\sigma_N^2} \leqslant \begin{cases}1/(c_\varepsilon N T^{1/2+\varepsilon}) & \text{if } \delta < 1,\\ 1/N^{1/2} & \text{otherwise,}\end{cases} \tag{8-18}$$

having used the lower bound (A-4) on the gap for the second inequality and recalled the choice (8-8). This proves that

$$\left|\sum_{\alpha,\beta\geqslant 1}(K_{11})_{\alpha\beta}\langle\Theta^2 a_{r,\alpha}a_{r,\beta}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}-\sum_{\alpha,\beta\geqslant 1}(K_{11})_{\alpha\beta}\langle a_{r,\alpha}a_{r,\beta}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}\right|=\left|\sum_{s=0}^{N}g(s)\langle\Phi_{\mathrm{trial},s},\widetilde{K}\Phi_{\mathrm{trial},s+2}\rangle\right|\leqslant o_N(1),$$

where

$$g(s)=1-\sum_{d}c_{d,s}c_{d-2,s}\ \text{and}\ \widetilde{K}=\sum_{\alpha,\beta}(K_{11})_{\alpha\beta}a_{r,\alpha}a_{r,\beta}.$$

We used the Cauchy–Schwarz inequality, (8-18) and the fact that, since K_{11} is trace-class, $\widetilde{K}$ is controlled by $\mathcal{N}_\perp^2$, whose expectation in Φ_{trial} is uniformly bounded. All terms in $\mathbb{H}_{\mathrm{right}}$ and $\mathbb{H}_{\mathrm{left}}$ that contain $\Theta^{\pm 2}$ can be treated similarly. This shows that, up to a remainder, $\mathbb{H}_{\mathrm{right}}+\mathbb{H}_{\mathrm{left}}$ acts on $\mathcal{U}_N\psi_{\mathrm{trial}}$ as if Θ were set to the identity, and therefore

$$|\langle\mathbb{H}_{\mathrm{right}}+\mathbb{H}_{\mathrm{left}}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}-E^{\mathrm{Bog}}|\leqslant|\langle\mathbb{H}_{\mathrm{right}}^{\Theta=\mathbb{1}}+\mathbb{H}_{\mathrm{left}}^{\Theta=\mathbb{1}}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}-E^{\mathrm{Bog}}|+o_N(1).$$

On the other hand, recalling the definition of $\mathcal{U}_N\psi_{\mathrm{trial}}$, the normalization of c_d, and (8-5), we see that

$$\begin{aligned}\langle\mathbb{H}_{\mathrm{right}}^{\Theta=\mathbb{1}}+\mathbb{H}_{\mathrm{left}}^{\Theta=\mathbb{1}}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}&=\frac{\sum_{s=0}^{N}\langle(\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\Omega)_s,\,((\mathbb{H}_{\mathrm{right}}^{\Theta=\mathbb{1}}+\mathbb{H}_{\mathrm{left}}^{\Theta=\mathbb{1}})(\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\Omega))_s\rangle}{\sum_{s=0}^{N}\|(\mathbb{U}_{\mathrm{left}}\mathbb{U}_{\mathrm{right}}\Omega)_s\|^2}\\&=E^{\mathrm{Bog}}+o_N(1),\end{aligned}$$

where the error is due to sum reaching only to $N<\infty$. Hence

$$|\langle\mathbb{H}_{\mathrm{right}}+\mathbb{H}_{\mathrm{left}}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}-E^{\mathrm{Bog}}|\leqslant o_N(1). \tag{8-19}$$

Step 3: remainder terms in $\mathbb{H}$. We now have to compute the contributions of $\mathbb{H}_{12}$ and of the ξ_j in (8-17). For $\mathbb{H}_{12}$ we have the a priori estimate (5-51), which implies

$$|\langle\mathbb{H}_{12}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}|\leqslant C_\varepsilon T^{1/2-\varepsilon}. \tag{8-20}$$

The terms inside ξ_1 and ξ_2 each contain exactly one operator $a_{r,\alpha}^\sharp$ and one $a_{\ell,\beta}^\sharp$. Using (8-3) and the fact that all the $a_{r,\alpha}^\sharp$'s commute with $a_{\ell,\beta}^\sharp$ and with $\mathbb{U}_{\mathrm{left}}$, we obtain

$$\langle\xi_1\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}=\langle\xi_2\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}=0. \tag{8-21}$$

We now consider ξ_3, focusing on its second line. As in Proposition 5.2, we introduce an energy cutoff Λ and an integer M_Λ which is the largest integer such that $\mu_{2M_\Lambda+2}\leqslant\Lambda$, where $\{\mu_m\}_m$ are the eigenvalues of h_{MF}. We have

$$\begin{aligned}&\left|\sum_{\alpha,\beta\geqslant 1}\langle u_{r,\alpha},\,K_{22}u_{r,\beta}\rangle\langle\Theta^{-2}a_{r,\alpha}a_{r,\beta}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}\right|\\&\quad\leqslant\left|\sum_{1\leqslant\alpha,\beta\leqslant M_\Lambda}\langle u_{r,\alpha},\,K_{22}u_{r,\beta}\rangle\langle\Theta^{-2}a_{r,\alpha}a_{r,\beta}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}\right|+2\left|\sum_{\alpha\geqslant 1,\ \beta>M_\Lambda}\langle u_{r,\alpha},\,K_{22}u_{r,\beta}\rangle\langle\Theta^{-2}a_{r,\alpha}a_{r,\beta}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}\right|\\&\quad=:\xi_3^{\leqslant M_\Lambda}+2\xi_3^{>M_\Lambda}.\end{aligned}$$

For each fixed α and β, the matrix element $\langle u_{r,\alpha}, K_{22}u_{r,\beta}\rangle$ tends to zero as $N\to\infty$ by the argument presented in the proof of Proposition 5.2; see Section 5D. Consequently, $\xi_3^{\leqslant M_\Lambda}$ vanishes as $N\to\infty$ for each fixed M_Λ. For $\xi_3^{>M_\Lambda}$ we argue as in the estimate of $\mathbb{K}_{>M_\Lambda}$ in the proof of Proposition 5.2. By repeated use of the Cauchy–Schwarz inequality, we have

$$\begin{aligned}\xi_3^{>M_\Lambda} &\leqslant \Bigg(\sum_{\alpha\geqslant 1,\,\beta>M_\Lambda} |\langle u_{r,\alpha}, K_{22}u_{r,\beta}\rangle|^2\Bigg)^{1/2}\Bigg(\sum_{\alpha\geqslant 1,\,\beta>M_\Lambda} \|a_{r,\alpha}a_{r,\beta}\mathcal{U}_N\Psi_{\mathrm{trial}}\|^2\Bigg)^{1/2}\\ &\leqslant \Bigg(\sum_{\alpha,\beta\geqslant 1}\langle u_{r,\alpha}, K_{22}u_{r,\beta}\rangle\langle u_{r,\beta}, K_{22}u_{r,\alpha}\rangle\Bigg)^{1/2}\Bigg\langle \mathcal{N}_\perp\sum_{\beta>M_\Lambda} a^\dagger_{r,\beta}a_{r,\beta}\Bigg\rangle^{1/2}_{\mathcal{U}_N\psi_{\mathrm{trial}}}.\end{aligned}$$

The square root that contains K_{22} in the right-hand side is equal to $\mathrm{Tr}K_{22}$, recalling that K_{22} is trace-class as proven in Lemma 5.3. For the other square root we notice that

$$\sum_{\beta>M_\Lambda} a^\dagger_{r,\beta}a_{r,\beta}\leqslant\sum_{\beta>M_\Lambda}(a^\dagger_{r,\beta}a_{r,\beta}+a^\dagger_{\ell,\beta}a_{\ell,\beta})=\sum_{n>2M_\Lambda+2}a^\dagger_n a_n,$$

having passed to the basis (2-25) in the second step. Since all operators commute with $\mathcal{N}_\perp$, we deduce using the same arguments as in the proof of Proposition 5.2 that

$$\Bigg\langle\mathcal{N}_\perp\sum_{\beta>M_\Lambda}a^\dagger_{r,\beta}a_{r,\beta}\Bigg\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}\leqslant\frac{1}{\mu_{2M_\Lambda+2}-\mu_+}\Bigg\langle\mathcal{N}_\perp\sum_{n>2M_\Lambda+2}(\mu_n-\mu_+)a^\dagger_n a_n\Bigg\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}.$$

The operators $a^\dagger_n a_n$ commute with $\mathcal{N}_\perp$ and we can bound the sum in the right-hand side by $\mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+)$. Hence

$$\xi_3^{>M_\Lambda}\leqslant C\bigg(\frac{1}{\mu_{2M_\Lambda+2}-\mu_+}\langle\mathcal{N}_\perp\mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+)\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}\bigg)^{1/2}.$$

The matrix element in the right-hand side is bounded by an N-independent constant. Indeed, since $\mathcal{U}_N\psi_{\mathrm{trial}}$ is a quasifree state, Wick's theorem gives the expectation of a quartic operator such as $\mathcal{N}_\perp\mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+)$ in terms of the expectations of $\mathcal{N}_\perp$ and $\mathrm{d}\Gamma_\perp(h_{\mathrm{MF}}-\mu_+)$, which are uniformly bounded in N. This proves

$$\xi_3^{>M_\Lambda}\leqslant\frac{C}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}}. \tag{8-22}$$

Plugging (8-19), (8-20), (8-21) and (8-22) into (8-16) gives the final bound

$$|\langle\mathbb{H}\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}-E^{\mathrm{Bog}}|\leqslant\frac{C}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}}+C_\Lambda o_N(1). \tag{8-23}$$

Step 4: error and linear terms. Note that, with the choice (8-8),

$$\bigg\langle\frac{\mathfrak{D}^2}{N}\bigg\rangle_{\mathcal{U}_N\psi_{\mathrm{trial}}}=\frac{1}{N}\sum_{s=0}^{N}\sum_{|d|\leqslant\sigma_N^2}d^2c_{d,s}^2\|\Phi_{\mathrm{trial},s}\|^2\leqslant C\frac{\sigma_N^2}{N}\leqslant o_N(1),$$

where the second bound follows from Lemma 4.5. In view also of (8-12), the first error term in (5-1) when $\Phi=\mathcal{U}_N\psi_{\mathrm{trial}}$ is bounded by $CN^{-1/4}$. The second error terms, in turn, can be bounded by an $o_N(1)$, relying on (8-12) and (8-13). Let us now show that the expectations in ψ_{trial} of the linear terms in (5-1) are also negligible. Using the Cauchy–Schwarz inequality we find

$$\begin{aligned}\Bigg|\frac{\lambda}{\sqrt{2(N-1)}}\Big\langle\sum_{m\geqslant 3} w_{+1-m}\, b_m\mathfrak{D}+\text{h.c.}\Big\rangle_{\mathcal{U}_N\psi_{\text{trial}}}+\frac{\lambda}{\sqrt{2(N-1)}}\Big\langle\sum_{m\geqslant 3} w_{+2-m}\, c_m\mathfrak{D}+\text{h.c.}\Big\rangle_{\mathcal{U}_N\psi_{\text{trial}}}\Bigg|&\\ \leqslant \frac{\lambda}{\sqrt{2(N-1)}}\Bigg[\Big(\sum_{m\geqslant 3}|w_{+1-m}|^2\Big)^{1/2}+\Big(\sum_{m\geqslant 3}|w_{+2-m}|^2\Big)^{1/2}\Bigg]\langle\mathcal{N}_\perp\rangle^{1/2}_{\psi_{\text{trial}}}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle^{1/2}_{\psi_{\text{trial}}}&\\ \leqslant o_N(1),&\end{aligned}$$

where the last inequality follows from (5-7) and (8-14). Hence we deduce from (8-11) that

$$\langle H_N\rangle_{\psi_{\text{trial}}}\leqslant\langle H_{2\text{-mode}}\rangle_{\psi_{\text{trial}}}+\langle\mathbb{H}\rangle_{\mathcal{U}_N\psi_{\text{trial}}}+\mu_+\langle\mathcal{N}_\perp\rangle_{\psi_{\text{trial}}}+o_N(1).$$

Plugging (8-15) and (8-23) into this inequality gives precisely (8-10) by passing to the limit $N\to\infty$ and then $\Lambda\to\infty$. □

8B. *Energy lower bound.*

We now prove the following:

Proposition 8.4 (energy lower bound). *Assume $T\ll 1$. For every large enough energy cutoff Λ, let M_Λ be the largest integer such that $\mu_{2M_\Lambda+2}\leqslant\Lambda$, where $\{\mu_m\}_m$ are the eigenvalues of h_{MF} in increasing order (this implies that $M_\Lambda\to\infty$ as $\Lambda\to\infty$). Then there exists $\lambda_0>0$ such that, for all $0\leqslant\lambda\leqslant\lambda_0$,*

$$\begin{aligned}\langle H_N\rangle_{\psi_{\text{gs}}}\geqslant E_0+E_N^w+N\frac{\mu_+-\mu_-}{2}+E^{\text{Bog}}+\frac{c\lambda}{N-1}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\text{gs}}}&\\ -C_\Lambda o_N(1)-C_\varepsilon\frac{T^{-\varepsilon}}{N^{1/2}}-C_\varepsilon T^{1/2-\varepsilon}-\frac{C}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}},&\end{aligned}\tag{8-24}$$

where c is a positive constant.

We first need to prove that the (negative) coefficients multiplying the variance $\langle\mathfrak{D}^2\rangle_\Phi$ in (7-9), and its analog for $\mathbb{H}^{(M)}_{\text{left,shift}}$, can be absorbed by the variance term of the two-mode Hamiltonian. Recall that

$$W_{r,\leqslant M_\Lambda}=P_{r,\leqslant M_\Lambda}(P_{r,\leqslant M_\Lambda}(h_{\text{MF}}-\mu_++2\lambda K_{11})P_{r,\leqslant M_\Lambda})^{-1}P_{r,\leqslant M_\Lambda},$$

with a similar formula for $W_{\ell,\leqslant M_\Lambda}$ (replacing K_{11} by K_{22}).

Lemma 8.5 (variance coefficients). *Let U be the coefficient from* (4-3). *We have*

$$\langle u_1,K_{11}W_{r,\leqslant M_\Lambda}K_{11}u_1\rangle\leqslant C,\quad\langle u_2,K_{22}W_{\ell,\leqslant M_\Lambda}K_{22}u_2\rangle\leqslant C\tag{8-25}$$

for some constant C that does not depend on λ and Λ. Consequently, if $0<\lambda\leqslant\lambda_0$ with λ_0 small enough, then

$$\lambda U-\frac{\lambda^2}{2}\langle u_1,K_{11}W_{r,\leqslant M_\Lambda}K_{11}u_1\rangle-\frac{\lambda^2}{2}\langle u_2,K_{22}W_{\ell,\leqslant M_\Lambda}K_{22}u_2\rangle\geqslant c\lambda\tag{8-26}$$

for some $c>0$.

Proof. Using the positivity of K_{11} and the finite energy gap (A-5), one has

$$P_{r,\leqslant M_\Lambda}(h_{\text{MF}}-\mu_++2\lambda K_{11})P_{r,\leqslant M_\Lambda}\geqslant P_{r,\leqslant M_\Lambda}(h_{\text{MF}}-\mu_+)P_{r,\leqslant M_\Lambda}>C^{-1}P_{r,\leqslant M_\Lambda}$$

for some $C > 0$. Hence

$$W_{r,\leqslant M_\Lambda} \leqslant C P_{r,\leqslant M_\Lambda}$$

because the inverse power is operator monotone [Bhatia 1997] and we are restricting everything to the range of $P_{r,\leqslant M_\Lambda}$. Since K_{11} is also bounded, the first inequality in (8-25) follows, and the second one is proven in the same way. The estimate (8-26) is a consequence of (8-25). Actually, the right-hand side in this estimate is bounded from below by $\lambda(U - C\lambda)$ and $U - C\lambda > 0$ for λ smaller than some λ_0 that depends on C, because $U > 0$ by (4-4). □

The rest of the subsection is devoted to the proof of Proposition 8.4. We use the a priori estimates of Section 6 systematically, without further mention. We also use the fact that $\mathcal{N}_\perp^4 \leqslant N^2 \mathcal{N}_\perp^2$ when evaluated on ψ_{gs}, and similarly for $\mathfrak{D}^4$.

Proof of Proposition 8.4. We first use Proposition 5.1 with $\Phi = \mathcal{U}_N \psi_{\mathrm{gs}}$. For such Φ, the error terms are bounded as in (6-8) and one gets

$$\begin{aligned}\langle H_N\rangle_{\psi_{\mathrm{gs}}} \geqslant \langle H_{2\text{-mode}}\rangle_{\psi_{\mathrm{gs}}} + \mu_+ \langle \mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}} + \langle \mathbb{H}\rangle_{\mathcal{U}_N\psi_{\mathrm{gs}}} \\ + \frac{\lambda}{\sqrt{2(N-1)}}\Big\langle \sum_{m\geqslant 3}(w_{+1-m}\, b_m \mathfrak{D} + w_{+2-m}\, c_m \mathfrak{D} + \text{h.c.})\Big\rangle_{\psi_{\mathrm{gs}}} - \frac{C}{N^{1/4}} - C_\varepsilon \frac{T^{-\varepsilon}}{N^{1/2}}.\end{aligned} \tag{8-27}$$

Next we use Proposition 5.2 to separate the full excitation energy into the excitation energy of right and left modes, at the expense of the appearance of the cutoff Λ. For a lower bound, we ignore the positive $d\Gamma_\perp(P_{\geqslant M_\Lambda}(h_{\mathrm{MF}} - \mu_+)P_{\geqslant M_\Lambda})$. We also use Proposition 5.7 to reduce the linear terms to modes below the cutoff without coupling between right and left modes. We thus obtain for any $\Lambda > 0$ large enough

$$\begin{aligned}\langle H_N\rangle_{\psi_{\mathrm{gs}}} \geqslant \langle H_{2\text{-mode}}\rangle_{\psi_{\mathrm{gs}}} + \mu_+ \langle \mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}} + \langle \mathbb{H}^{(M_\Lambda)}_{\text{right}} + \mathbb{H}^{(M_\Lambda)}_{\text{left}}\rangle_{\mathcal{U}_N\psi_{\mathrm{gs}}} \\ + \frac{\lambda}{\sqrt{2(N-1)}}\Big\langle \sum_{1\leqslant\alpha\leqslant M_\Lambda}(w_{+1-\{r,\alpha\}}\, b_{r,\alpha} \mathfrak{D} + w_{+2-\{\ell,\alpha\}}\, c_{\ell,\alpha} \mathfrak{D} + \text{h.c.})\Big\rangle_{\mathcal{U}_N\psi_{\mathrm{gs}}} \\ - C_\Lambda o_N(1) - C_\varepsilon \frac{T^{-\varepsilon}}{N^{1/2}} - \frac{C}{(\mu_{2M_\Lambda+2} - \mu_+)^{1/2}}.\end{aligned}$$

Let us now plug into the above estimate the lower bound on $H_{2\text{-mode}}$ from Proposition 4.2; see (4-10). This produces, among other terms, a term $-\mu_+\langle \mathcal{N}_\perp\rangle_{\psi_{\mathrm{gs}}}$ that cancels the one above. The expectation in the ground state of the last term in (4-10) is bounded from below by $-C_\varepsilon T^{1-\varepsilon}$ due to (6-1). We also recognize that $\mathbb{H}^{(M_\Lambda)}_{\text{right}} + \mathbb{H}^{(M_\Lambda)}_{\text{left}}$ together with the linear terms coincide with $\mathbb{H}^{(M_\Lambda)}_{\text{right,shift}} + \mathbb{H}^{(M_\Lambda)}_{\text{left,shift}}$ from (7-1). Thus

$$\begin{aligned}\langle H_N\rangle_{\psi_{\mathrm{gs}}} \geqslant E_0 + E_N^w + N\frac{\mu_+ - \mu_-}{2} + \frac{\lambda U}{N-1}\langle(\mathcal{N}_1 - \mathcal{N}_2)^2\rangle_{\psi_{\mathrm{gs}}} + \langle \mathbb{H}^{(M_\Lambda)}_{\text{right,shift}} + \mathbb{H}^{(M_\Lambda)}_{\text{left,shift}}\rangle_{\mathcal{U}_N\psi_{\mathrm{gs}}} \\ - C_\Lambda o_N(1) - C_\varepsilon \frac{T^{-\varepsilon}}{N^{1/2}} - C_\varepsilon T^{1/2-\varepsilon} - \frac{C}{(\mu_{2M_\Lambda+2} - \mu_+)^{1/2}}.\end{aligned}$$

We now use Proposition 7.3 to bound the term containing the shift Bogoliubov Hamiltonians, which enable the absorption of the linear terms at the expense of passing to $\tilde{b}^\sharp$ and $\tilde{c}^\sharp$ operators and of the appearance of a negative variance term. According to the a priori bound (6-6) on $\langle \mathfrak{D}^2\rangle_{\mathcal{U}_N\psi_{\text{gs}}} = \langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\text{gs}}}$, the error terms in Proposition 7.3 are bounded by $C/\sqrt{N}+C_\varepsilon T^{1/2-\varepsilon}$. The new lower bound looks like

$$\begin{aligned}\langle H_N\rangle_{\psi_{\text{gs}}} \geqslant{}& E_0+E_N^w+N\frac{\mu_+-\mu_-}{2}+\langle\widetilde{\mathbb{H}^{(M_\Lambda)}_{\text{right}}}+\widetilde{\mathbb{H}^{(M_\Lambda)}_{\text{left}}}\rangle_{\mathcal{U}_N\psi_{\text{gs}}}\\ &-\frac{1}{2}\operatorname{Tr}(P_{r,\leqslant M_\Lambda}(h_{\text{MF}}-\mu_++\lambda K_{11}))-\frac{1}{2}\operatorname{Tr}(P_{\ell,\leqslant M_\Lambda}(h_{\text{MF}}-\mu_++\lambda K_{22}))\\ &+\frac{1}{N-1}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\text{gs}}}\biggl[\lambda U-\frac{\lambda^2}{2}\langle u_1,\ K_{11}W_{r,\geqslant M_\Lambda}K_{11}u_1\rangle-\frac{\lambda^2}{2}\langle u_2,\ K_{22}W_{\ell,\geqslant M_\Lambda}K_{22}u_2\rangle\biggr]\\ &-C_\Lambda o_N(1)-C_\varepsilon\frac{T^{-\varepsilon}}{N^{1/2}}-C_\varepsilon T^{1/2-\varepsilon}-\frac{C}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}}.\end{aligned}$$

By relying on Proposition 7.4, we bound the difference of the expectation of $\widetilde{\mathbb{H}^{(M_\Lambda)}_{\text{right}}}+\widetilde{\mathbb{H}^{(M_\Lambda)}_{\text{left}}}$ and the terms in the second line by E^{Bog}, up to remainders $C_\Lambda o_N(1)$. Finally, the terms in the square brackets can be bounded from below by using the Lemma 8.5 above; see (8-26). This yields the desired result (8-24). □

8C. ***Proof of Theorem 2.3.*** Putting together Propositions 8.3 and 8.4, we can now conclude the proof of Theorem 2.3. Taking the limit $N\to\infty$ in (8-24) yields

$$\liminf_{N\to\infty}\biggl(\langle H_N\rangle_{\psi_{\text{gs}}}-E_0-E_N^w-N\frac{\mu_+-\mu_-}{2}-E^{\text{Bog}}\biggr)\geqslant\limsup_{N\to\infty}\biggl(\frac{c\lambda}{N}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\text{gs}}}-\frac{C}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}}\biggr).$$

On the other hand, combining (8-10) and the estimate (4-36) on $E_{\text{2-mode}}$, which follows from Proposition 2.4, we have

$$\limsup_{N\to\infty}\biggl(\langle H_N\rangle_{\psi_{\text{gs}}}-E_0-E_N^w-N\frac{\mu_+-\mu_-}{2}-E^{\text{Bog}}\biggr)\leqslant 0.$$

This gives

$$\limsup_{N\to\infty}\frac{c\lambda}{N}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\text{gs}}}\leqslant\limsup_{N\to\infty}\frac{C}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}}.$$

As argued below Proposition 5.2, the limit of the eigenvalue $\mu_{2M_\Lambda+2}$ as $N\to\infty$ is the M_Λ-th eigenvalue of a fixed one-well Hamiltonian with compact resolvent. Hence, letting $\Lambda\to\infty$,

$$\limsup_{N\to\infty}\frac{c\lambda}{N}\langle(\mathcal{N}_1-\mathcal{N}_2)^2\rangle_{\psi_{\text{gs}}}=0,\tag{8-28}$$

thus proving (2-31). Inserting (8-28) in the energy upper and lower bounds (8-10) and (8-24), we find by using (4-36) again

$$o_N(1)-\frac{C}{(\mu_{2M_\Lambda+2}-\mu_+)^{1/2}}+E_{\text{2-mode}}+E^{\text{Bog}}\leqslant E(N)\leqslant E_{\text{2-mode}}+E^{\text{Bog}}+o_N(1).$$

Thus we may let first $N\to\infty$ and then $\Lambda\to\infty$ to conclude the proof of (2-32).

Appendix A: The one-body Hartree problem

We recall here a number of results that were proved in our companion paper [Olgiati and Rougerie 2021], i.e., properties of the eigenvectors and eigenfunctions of the one-body Hamiltonian h_{MF}.

In Section 1 we defined u_+ and u_- as the first and second eigenfunctions of h_{MF}, corresponding to the eigenvalues μ_+ and μ_-, and the full spectral decomposition of h_{MF} is

$$h_{\mathrm{MF}} = \mu_+ |u_+\rangle\langle u_+| + \mu_- |u_-\rangle\langle u_-| + \sum_{m\geqslant 3} \mu_m |u_m\rangle\langle u_m|.$$

Moreover, we defined right and left modes as

$$u_{r,\alpha} := \frac{u_{2\alpha+1} + u_{2\alpha+2}}{\sqrt{2}} \quad \text{and} \quad u_{\ell,\alpha} := \frac{u_{2\alpha+1} - u_{2\alpha+2}}{\sqrt{2}}$$

for any $\alpha \geqslant 1$.

We have the following result.

Theorem A.1 (one-body Hartree problem). (i) ***Lower eigenvectors convergence.***

$$\| |u_+|^2 - |u_-|^2 \|_{L^1} \leqslant C_\varepsilon T^{1-\varepsilon}, \tag{A-1}$$

$$\| |u_+| - |u_-| \|_{L^2} \leqslant C_\varepsilon T^{1/2-\varepsilon}, \tag{A-2}$$

$$\| |u_+| - |u_-| \|_{L^\infty} \leqslant C_\varepsilon T^{1/2-\varepsilon}. \tag{A-3}$$

(ii) ***Bounds on the fist spectral gap.***

$$c_\varepsilon T^{1+\varepsilon} \leqslant \mu_- - \mu_+ \leqslant C_\varepsilon T^{1-\varepsilon}. \tag{A-4}$$

(iii) ***Second gap.***

$$\mu_{\mathrm{m}} - \mu_- \geqslant C \quad \textit{for all } m \geqslant 3 \tag{A-5}$$

independently of L.

(iv) ***Properties of u_+.*** *The function u_+ is smooth, strictly positive (up to a phase), and even under reflections across the $\{x_1 = 0\}$ hyperplane.*

(v) ***Properties of u_-.*** *The function u_- is smooth and odd under reflections across the $\{x_1 = 0\}$ hyperplane. Moreover, up to a phase,*

$$u_1(x) > 0 \quad \textit{for } x_1 \geqslant 0. \tag{A-6}$$

(vi) ***Higher spectrum.*** *For any $\alpha \geqslant 1$ we have*

$$\lim_{T\to 0} (\mu_{2\alpha+2} - \mu_{2\alpha+1}) = 0, \tag{A-7}$$

and, for an appropriate phase choice of the u_m

$$\lim_{T\to 0} \int_{x_1\leqslant 0} |u_{r,\alpha}|^2\, dx = \lim_{T\to 0} \int_{x_1 \geqslant 0} |u_{\ell,\alpha}|^2\, dx = 0. \tag{A-8}$$

Items (i), (ii), and (iii) follow from [Olgiati and Rougerie 2021, Theorem 2.1]. The fact that u_+ can be chosen as positive is a standard fact already recalled in Section 2. Since the h_{MF} commutes with reflection across $\{x_1 = 0\}$, we can choose its eigenvectors to be either odd or even under such a permutation. Since

u_+ is positive, it must be even. The fact that u_- is odd and its sign follow from [Olgiati and Rougerie 2021, Lemma 4.2]. Notice that, for u_1 defined in (2-11), as a consequence of (iv) and (v) we have

$$\int_{x_1\leqslant 0} |u_1(x)|^2\,dx = \frac{1}{\sqrt{2}}\int_{x_1\leqslant 0} |u_+(x)+u_-(x)|^2\,dx = \frac{1}{\sqrt{2}}\int_{x_1\leqslant 0} ||u_+(x)|-|u_-(x)||^2\,dx.$$

Hence, by (A-2),

$$\int_{x_1\leqslant 0} |u_1(x)|^2\,dx = \int_{x_1\geqslant 0} |u_2(x)|^2\,dx \leqslant C_\varepsilon T^{1-\varepsilon}, \tag{A-9}$$

which is the analog of (A-8) for the low energy modes.

Appendix B: Estimates and identities in the two-mode space

We prove here some results that were stated in Section 4.

Proof of Lemma 4.1. The upper bound on w_{1111} follows immediately from Young's inequality (recall that $w \in L^\infty$). To prove the lower bound, we use the pointwise lower bound on u_+ (see [Olgiati and Rougerie 2021, Proposition 3.1])

$$u_+(x) \geqslant c_\varepsilon e^{-(1+\varepsilon)A_{\mathrm{DW}}(x)}, \tag{B-1}$$

where

$$A_{\mathrm{DW}} = \begin{cases} A(|x-x_L|), & x_1 \geqslant 0,\\ A(|x+x_L|), & x_1 \leqslant 0,\end{cases}$$

and A is the Agmon distance (2-12). Let us notice that, using the definition (2-11) of u_1 and u_2,

$$w_{1111} \geqslant \iint_{\substack{x_1\geqslant 0\\ y_1\geqslant 0}} w(x-y)|u_1(x)|^2|u_1(y)|^2\,dx\,dy \geqslant \frac{1}{4}\iint_{\substack{x_1\geqslant 0\\ y_1\geqslant 0}} w(x-y)|u_+(x)|^2|u_+(y)|^2\,dx\,dy,$$

having used in the second inequality the fact that $u_+(x) > 0$ and $u_-(x) \geqslant 0$ for $x_1 \geqslant 0$, as granted by Theorem A.1. Using the lower bound (B-1) we deduce

$$\begin{aligned} w_{1111} &\geqslant c_\varepsilon \iint_{\substack{x_1\geqslant 0\\ y_1\geqslant 0}} w(x-y)e^{-2(1+\varepsilon)A(|x-x_L|)}e^{-2(1+\varepsilon)A(|y-x_L|)}\,dx\,dy\\ &= c_\varepsilon \iint_{\substack{x_1\geqslant -L/2\\ y_1\geqslant -L/2}} w(x-y)e^{-2(1+\varepsilon)A(|x|)}e^{-2(1+\varepsilon)A(|y|)}\,dx\,dy\\ &\geqslant c_\varepsilon \iint_{\substack{x_1\geqslant 0\\ y_1\geqslant 0}} w(x-y)e^{-2(1+\varepsilon)A(|x|)}e^{-2(1+\varepsilon)A(|y|)}\,dx\,dy =: c > 0,\end{aligned}$$

where all the steps are justified since the functions in the integral are manifestly positive and summable.

To prove (4-5) we use the definition of u_1 and u_2 in terms of u_+ and u_- from (2-11), then Young's inequality and (A-1), to get

$$|w_{1112}| \leqslant \frac{1}{2}\int_{\mathbb{R}^d} w*|u_1|^2\,||u_+|^2-|u_-|^2| \leqslant \|w*|u_1|^2\|_{L^\infty}\,\||u_+|^2-|u_-|^2\|_{L^1} \leqslant C_\varepsilon T^{1-\varepsilon}.$$

Similarly, for (4-6) we write

$$w_{1122} \leqslant \frac{1}{4}\int_{\mathbb{R}^d} w * \big||u_+|^2 - |u_-|^2\big|\, \big||u_+|^2 - |u_-|^2\big| \leqslant C\big\||u_+|^2 - |u_-|^2\big\|_{L^1}^2 \leqslant C_\varepsilon T^{2-\varepsilon}.$$

On the other hand, the positivity of w_{1122} is deduced by noticing that

$$w_{1122} = \int \hat{w}(k)|\widehat{u_1 u_2}(k)|^2\, dk \geqslant 0, \tag{B-2}$$

since $\hat{w}(k) \geqslant 0$ by assumption.

To estimate w_{1212} we use the fact that w has compact support and is bounded by a constant to write

$$w_{1212} \leqslant C \iint_{\substack{x_1 \leqslant 0\\ y_1 \leqslant C}} |u_1(x)|^2 |u_2(y)|^2\, dx\, dy + C \iint_{\substack{x_1 \geqslant 0\\ y_1 \geqslant -C}} |u_1(x)|^2 |u_2(y)|^2\, dx\, dy.$$

In the first integral we recognize that $\sqrt{2}u_1(x) = u_+(x) + u_-(x) = |u_+(x)| - |u_-(x)|$ for $x_1 \leqslant 0$ (recall that Theorem A.1 ensures that u_- is negative for negative x_1's), and, using (A-2),

$$C\int_{x_1 \leqslant 0,\ y_1 \leqslant C} |u_1(x)|^2\, |u_2(y)|^2\, dx\, dy \leqslant C\big\||u_+| - |u_-|\big\|_{L^2}^2\, \|u_2\|_{L^2}^2 \leqslant C_\varepsilon T^{1-\varepsilon}.$$

In the second integral we can ignore the region in which $-C \leqslant y_1 \leqslant 0$, since both u_1 and u_2 are exponentially small there, because u_+ and u_- are; see [Olgiati and Rougerie 2021, Proposition 3.1]. For the region in which $y_1 \geqslant 0$ we argue as in the integral above by recognizing that $\sqrt{2}u_2(y) = u_+(y) - u_-(y) = |u_+(y)| - |u_-(y)|$ for $y_1 \geqslant 0$. This proves (4-7).

To prove (4-8) we only have to notice that

$$\mu - \mu_+ = \frac{\lambda}{2(N-1)}(w_{1212} - w_{1112} + (1-2N)w_{1122}),$$

and the result follows from the estimates above. □

Proof of Lemma 4.3. Since $\mathcal{N}_- = (\mathcal{N}_1 + \mathcal{N}_2 - a_1^\dagger a_2 - a_2^\dagger a_1)/2$ and $[\mathcal{N}_1 + \mathcal{N}_2, a_1^\dagger a_2 + a_2^\dagger a_1] = 0$, one has

$$4\mathcal{N}_-^2 = (\mathcal{N}_1 + \mathcal{N}_2)^2 - 2(\mathcal{N}_1 + \mathcal{N}_2)(a_1^\dagger a_2 + a_2^\dagger a_1) + (a_1^\dagger a_2 + a_2^\dagger a_1)^2$$

and thus

$$\begin{aligned} 2(\mathcal{N}_1 + \mathcal{N}_1)(a_1^\dagger a_2 + a_2^\dagger a_1) - (\mathcal{N}_1 + \mathcal{N}_1)^2 + 4\mathcal{N}_-^2 - \mathcal{N}_1 - \mathcal{N}_2 \\ &= (a_1^\dagger a_2)^2 + (a_2^\dagger a_1)^2 + a_1^\dagger a_2 a_2^\dagger a_1 + a_2^\dagger a_1 a_1^\dagger a_2 - \mathcal{N}_1 - \mathcal{N}_2 \\ &= (a_1^\dagger a_2)^2 + (a_2^\dagger a_1)^2 + 2\mathcal{N}_1\mathcal{N}_2, \end{aligned}$$

where the last equality follows from the commutation relations of a_1, $a_1^\dagger$, a_2 and $a_2^\dagger$. □

Acknowledgments

Funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement CORFRONMAT no. 758620) is gratefully acknowledged. Spehner acknowledges support from the Fondecyt project no. 1190134.

References

[Agmon 1982] S. Agmon, *Lectures on exponential decay of solutions of second-order elliptic equations*: *bounds on eigenfunctions of N-body Schrödinger operators*, Math. Notes **29**, Princeton Univ. Press, 1982. MR Zbl

[Ammari 2013] Z. Ammari, *Systèmes hamiltoniens en théorie quantique des champs*: *dynamique asymptotique et limite classique*, habilitation à diriger des recherches, Université de Rennes 1, 2013.

[Anapolitanos et al. 2017] I. Anapolitanos, M. Hott, and D. Hundertmark, "Derivation of the Hartree equation for compound Bose gases in the mean field limit", *Rev. Math. Phys.* **29**:7 (2017), art. id. 1750022. MR Zbl

[Bach and Bru 2016] V. Bach and J.-B. Bru, *Diagonalizing quadratic bosonic operators by non-autonomous flow equations*, Mem. Amer. Math. Soc. **1138**, Amer. Math. Soc., Providence, RI, 2016. MR Zbl

[Benedikter et al. 2016] N. Benedikter, M. Porta, and B. Schlein, *Effective evolution equations from quantum dynamics*, SpringerBriefs Math. Phys. **7**, Springer, 2016. MR Zbl

[Bhatia 1997] R. Bhatia, *Matrix analysis*, Grad. Texts in Math. **169**, Springer, 1997. MR Zbl

[Boccato et al. 2019] C. Boccato, C. Brennecke, S. Cenatiempo, and B. Schlein, "Bogoliubov theory in the Gross–Pitaevskii limit", *Acta Math.* **222**:2 (2019), 219–335. MR Zbl

[Boccato et al. 2020] C. Boccato, C. Brennecke, S. Cenatiempo, and B. Schlein, "The excitation spectrum of Bose gases interacting through singular potentials", *J. Eur. Math. Soc.* **22**:7 (2020), 2331–2403. MR Zbl

[Bruneau and Dereziński 2007] L. Bruneau and J. Dereziński, "Bogoliubov Hamiltonians and one-parameter groups of Bogoliubov transformations", *J. Math. Phys.* **48**:2 (2007), art. id. 022101. MR Zbl

[Buchholz et al. 2014] S. Buchholz, C. Saffirio, and B. Schlein, "Multivariate central limit theorem in quantum dynamics", *J. Stat. Phys.* **154**:1-2 (2014), 113–152. MR Zbl

[Dereziński 2017] J. Dereziński, "Bosonic quadratic Hamiltonians", *J. Math. Phys.* **58**:12 (2017), art. id. 121101. MR Zbl

[Dereziński and Napiórkowski 2014] J. Dereziński and M. Napiórkowski, "Excitation spectrum of interacting bosons in the mean-field infinite-volume limit", *Ann. Henri Poincaré* **15**:12 (2014), 2409–2439. MR Zbl

[Dimassi and Sjöstrand 1999] M. Dimassi and J. Sjöstrand, *Spectral asymptotics in the semi-classical limit*, Lond. Math. Soc. Lect. Note Ser. **268**, Cambridge Univ. Press, 1999. MR Zbl

[Dimonte et al. 2021] D. Dimonte, M. Falconi, and A. Olgiati, "On some rigorous aspects of fragmented condensation", *Nonlinearity* **34**:1 (2021), 1–32. MR Zbl

[Gallagher et al. 2013] I. Gallagher, L. Saint-Raymond, and B. Texier, *From Newton to Boltzmann*: *hard spheres and short-range potentials*, Eur. Math. Soc., Zurich, 2013. MR Zbl

[Golse 2016] F. Golse, "On the dynamics of large particle systems in the mean field limit", pp. 1–144 in *Macroscopic and large scale phenomena*: *coarse graining*, *mean field limits and ergodicity* (Enschede, Netherlands, 2012), edited by A. Muntean et al., Lect. Notes Appl. Math. Mech. **3**, Springer, 2016. MR

[Grech and Seiringer 2013] P. Grech and R. Seiringer, "The excitation spectrum for weakly interacting bosons in a trap", *Comm. Math. Phys.* **322**:2 (2013), 559–591. MR Zbl

[Gustafson and Sigal 2011] S. J. Gustafson and I. M. Sigal, *Mathematical concepts of quantum mechanics*, 2nd ed., Springer, 2011. MR Zbl

[Helffer 1988] B. Helffer, *Semi-classical analysis for the Schrödinger operator and applications*, Lecture Notes in Math. **1336**, Springer, 1988. MR Zbl

[Jabin 2014] P.-E. Jabin, "A review of the mean field limits for Vlasov equations", *Kinet. Relat. Models* **7**:4 (2014), 661–711. MR Zbl

[Lewin et al. 2015] M. Lewin, P. T. Nam, S. Serfaty, and J. P. Solovej, "Bogoliubov spectrum of interacting Bose gases", *Comm. Pure Appl. Math.* **68**:3 (2015), 413–471. MR Zbl

[Lieb and Loss 1997] E. H. Lieb and M. Loss, *Analysis*, Grad. Stud. in Math. **14**, Amer. Math. Soc., Providence, RI, 1997. MR Zbl

[Lieb et al. 2005] E. H. Lieb, R. Seiringer, J. P. Solovej, and J. Yngvason, *The mathematics of the Bose gas and its condensation*, Oberwolfach Seminars **34**, Birkhäuser, Basel, 2005. MR Zbl

[Michelangeli and Olgiati 2017] A. Michelangeli and A. Olgiati, "Mean-field quantum dynamics for a mixture of Bose–Einstein condensates", *Anal. Math. Phys.* **7**:4 (2017), 377–416. MR Zbl

[Michelangeli et al. 2019] A. Michelangeli, P. T. Nam, and A. Olgiati, "Ground state energy of mixture of Bose gases", *Rev. Math. Phys.* **31**:2 (2019), art. id. 1950005. MR Zbl

[Mischler 2011] S. Mischler, "Estimation quantitative et uniforme en temps de la propagation du chaos et introduction aux limites de champ moyen pour des systèmes de particules", course notes, 2011, available at https://www.ceremade.dauphine.fr/~mischler/Enseignements/MiMaM2/EDVersionProvisoire.pdf.

[Nam and Seiringer 2015] P. T. Nam and R. Seiringer, "Collective excitations of Bose gases in the mean-field regime", *Arch. Ration. Mech. Anal.* **215**:2 (2015), 381–417. MR Zbl

[Nam et al. 2016] P. T. Nam, M. Napiórkowski, and J. P. Solovej, "Diagonalization of bosonic quadratic Hamiltonians by Bogoliubov transformations", *J. Funct. Anal.* **270**:11 (2016), 4340–4368. MR Zbl

[Olgiati and Rougerie 2021] A. Olgiati and N. Rougerie, "The Hartree functional in a double well", *J. Spectr. Theory* **11**:4 (2021), 1727–1778. MR Zbl

[Pulvirenti and Simonella 2016] M. Pulvirenti and S. Simonella, "Propagation of chaos and effective equations in kinetic theory: a brief survey", *Math. Mech. Complex Syst.* **4**:3-4 (2016), 255–274. MR Zbl

[Rademacher and Schlein 2019] S. Rademacher and B. Schlein, "Central limit theorem for Bose–Einstein condensates", *J. Math. Phys.* **60**:7 (2019), art. id. 071902. MR Zbl

[Rougerie 2014] N. Rougerie, "Théorèmes de de Finetti, limites de champ moyen et condensation de Bose–Einstein", lecture notes, Collège de France, 2014. arXiv 1409.1182

[Rougerie 2015] N. Rougerie, "de Finetti theorems, mean-field limits and Bose–Einstein condensation", lecture notes, Ludwig-Maximilians-Universität München, 2015. arXiv 1506.05263

[Rougerie 2020] N. Rougerie, "Scaling limits of bosonic ground states, from many-body to non-linear Schrödinger", *EMS Surv. Math. Sci.* **7**:2 (2020), 253–408. MR Zbl

[Rougerie and Spehner 2018] N. Rougerie and D. Spehner, "Interacting bosons in a double-well potential: localization regime", *Comm. Math. Phys.* **361**:2 (2018), 737–786. MR Zbl

[Schlein 2013] B. Schlein, "Derivation of effective evolution equations from microscopic quantum dynamics", pp. 511–572 in *Evolution equations* (Zurich, 2008), edited by D. Ellwood et al., Clay Math. Proc. **17**, Amer. Math. Soc., Providence, RI, 2013. MR Zbl

[Seiringer 2011] R. Seiringer, "The excitation spectrum for weakly interacting bosons", *Comm. Math. Phys.* **306**:2 (2011), 565–578. MR Zbl

[Spohn 1980] H. Spohn, "Kinetic equations from Hamiltonian dynamics: Markovian limits", *Rev. Modern Phys.* **52**:3 (1980), 569–615. MR Zbl

[Spohn 1991] H. Spohn, *Large scale dynamics of interacting particles*, Springer, 1991. Zbl

Received 31 May 2021. Revised 28 Jan 2022. Accepted 16 Mar 2022.

ALESSANDRO OLGIATI: alessandro.olgiati@math.uzh.ch
Institut fur Mathematik, Universitat Zurich, Zurich, Switzerland

NICOLAS ROUGERIE: nicolas.rougerie@ens-lyon.fr
École Normale Supérieure de Lyon & CNRS, UMPA (UMR 5669), Lyon, France

DOMINIQUE SPEHNER: dspehner@ing-mat.udec.cl
Departamento de Ingeniería Matemática, Universidad de Concepción, Concepción, Chile

and

Université Grenoble Alpes & CNRS, Institut Fourier & LPMMC, Grenoble, France

ANALYSIS AND PDE
Vol. 16 (2023), No. 8, pp. 1955–1988
DOI: 10.2140/apde.2023.16.1955

A GENERAL NOTION OF UNIFORM ELLIPTICITY AND THE REGULARITY OF THE STRESS FIELD FOR ELLIPTIC EQUATIONS IN DIVERGENCE FORM

UMBERTO GUARNOTTA AND SUNRA MOSCONI

For solutions of $\operatorname{Div}(DF(Du)) = f$ we show that the quasiconformality of $z \mapsto DF(z)$ is the key property leading to the Sobolev regularity of the stress field $DF(Du)$, in relation with the summability of f. This class of nonlinearities encodes in a general way the notion of uniform ellipticity and encompasses all known instances where the stress field is known to be Sobolev regular. We provide examples showing the optimality of this assumption and present two applications: a nonlinear Cordes condition for equations in divergence form and some partial results on the $C^{p'}$ conjecture.

1. Introduction

In this work we are interested in Sobolev regularity results for the often-called "stress field" $DF(Du)$ corresponding to solutions of

$$\operatorname{div}(DF(Du)) = f, \tag{1-1}$$

seen as the Euler–Lagrange equation for the energy functional

$$J(w, \Omega) = \int_\Omega F(Dw) + f\, w\, dx, \tag{1-2}$$

where $\Omega \subset \mathbb{R}^N,\ N \geqslant 2,\ f \in L^m(\Omega)$ for some $m > 1$, and $F \in C^1(\mathbb{R}^N)$ is a strictly convex function obeying a suitable local form of uniform ellipticity condition. Questions regarding regularity of the stress field recently gained increasing interest as a basic tool to attack further regularity and locality properties of solutions to divergence form equations; see, e.g., [Avelin et al. 2018; Balci et al. 2020; Breit et al. 2018; 2022; Ciraolo et al. 2020; Colombo and Figalli 2014; Kuusi and Mingione 2013]. Despite its usefulness, however, most of the results are constrained to special kinds of nonlinearities of either p-Laplacian-type or having Uhlenbeck structure. Simple nonlinearities, such as

$$F(z) = |z - z_0|^p + |z + z_0|^p, \quad 1 < p < 2,\ z_0 \neq 0, \tag{1-3}$$

do not fall within most of the currently available regularity theory. Since we focus on the stress field instead of Du itself, we briefly justify this point of view, recalling the general situation for functionals of the calculus of variations and their local minimizers.

Let u solve (1-1). If $F \in C^2(\mathbb{R}^N)$ is uniformly elliptic, i.e.,

$$\lambda|\xi|^2 \leqslant (D^2F(z)\,\xi, \xi) \leqslant \Lambda|\xi|^2 \quad \text{for all } z, \xi \in \mathbb{R}^N,$$

MSC2020: primary 30C65, 35B65, 35J62; secondary 49K20.
Keywords: elliptic equations in divergence form, regularity, quasiconformal maps, Calderón–Zygmund estimates, Cordes condition.

it is a classical result that $f \in L^2_{\mathrm{loc}}(\Omega)$ if and only if $u \in W^{2,2}_{\mathrm{loc}}(\Omega)$, and in this case the regularity of Du and $V = DF(Du)$ coincide, since $Du = DF^{-1}(V)$ and DF is bi-Lipschitz. Regarding $W^{2,m}$-regularity for $m \neq 2$, it is clear that if $u \in W^{2,m}_{\mathrm{loc}}(\Omega)$, then $DF(Du) \in W^{1,m}_{\mathrm{loc}}(\Omega)$, and thus $f \in L^m_{\mathrm{loc}}(\Omega)$. Conversely, suppose $f \in L^m_{\mathrm{loc}}(\Omega)$; differentiating (1-1) gives

$$\operatorname{div}(D^2F(Du)Dv_k) = \partial_k f, \quad v_k = \partial_k u, \ k = 1, \dots, N.$$

If $f \in L^m_{\mathrm{loc}}(\Omega)$ for $m > N$, then Du is Hölder continuous by the De Giorgi–Nash theorem; so, freezing the (now continuous) coefficients of the matrix $D^2F(Du)$, we can apply the Calderón–Zygmund theorem to obtain that $u \in W^{2,m}_{\mathrm{loc}}(\Omega)$, i.e., $DF(Du) \in W^{1,m}_{\mathrm{loc}}(\Omega)$.

Next, consider the p-Poisson equation

$$\Delta_p u = f, \quad p > 1, \tag{1-4}$$

corresponding to the integrand $F(z) = |z|^p/p$. In this case F satisfies

$$\lambda\, |z|^{p-2}\, |\xi|^2 \leqslant (D^2F(z)\,\xi, \xi) \leqslant \Lambda\, |z|^{p-2}\, |\xi|^2,$$

and the Sobolev regularity of Du is much more involved. We are aware of only one result giving second-order Sobolev regularity of u from an L^m assumption on f: in the nondegenerate case $p \in (1, 2]$, that $u \in W^{2,p}(\mathbb{R}^N)$ if $f \in L^{p'}(\mathbb{R}^N)$ ($1/p + 1/p' = 1$, as usual) was proved in [Simon 1978] for global solutions and in [de Thélin 1982] for local ones.

Through difference quotients and Caccioppoli inequalities one can usually infer Sobolev regularity of Du from Sobolev regularity of f. In this framework, [Cellina 2017; Mercuri et al. 2016] treat the case when $p > 2$ is near uniform ellipticity, proving, respectively, $u \in W^{2,2}_{\mathrm{loc}}(\Omega)$ for $2 \leqslant p < 3$ and $u \in W^{2,m}_{\mathrm{loc}}(\Omega)$ for any m, as long as $p - 2$ is sufficiently small. The postulated regularity for f is $f \in W^{1,2}_{\mathrm{loc}}(\Omega)$ in [Cellina 2017] and $f \in W^{1,m}_{\mathrm{loc}}(\Omega)$ for $m > N$ in [Mercuri et al. 2016].

For $p > 2$ and only assuming L^m regularity of f in (1-4), the best results available prove *fractional* differentiability of Du; see [Mingione 2007; 2010; Miśkiewicz 2018; Savaré 1998; Simon 1978]. The main idea of [Kuusi and Mingione 2013; Mingione 2007; 2010], which goes back to [Uhlenbeck 1977], is to study the regularity properties of the field

$$\mathcal{V} = |Du|^{(p-2)/2} Du \tag{1-5}$$

and deduce from the latter suitable regularity of Du. This approach is nowadays widespread, but still failed to produce estimates in terms of the Lebesgue norm of f paralleling the second-order Calderón–Zygmund theory depicted above in the nondegenerate case.

An alternative route is to consider the regularity of the stress field

$$V = |Du|^{p-2}\, Du,$$

which arises as an interesting object per se in a variety of situations, e.g., in the framework of nonconvex variational problems [Carstensen and Müller 2002] and in the dual formulation of traffic congestion problems; see [Brasco et al. 2010]. In particular, the applicability of DiPerna–Lions theory in the latter is tied to the Sobolev regularity of the stress field of the dual problem, which was the main concern of [Brasco et al. 2010] for a very degenerate functional of the form (1-2). When f is Sobolev regular, variants of Caccioppoli inequalities and difference quotient methods have been used in the cited works to obtain Sobolev regularity of V; see also [Damascelli and Sciunzi 2004].

For less regular f the seminal paper is [Lou 2008], treating the case $f \in L^m$ with $m \geqslant \max\{2, N/p\}$. In more recent years, the regularity of the stress field has been the object of fruitful investigations, also thanks to the fact that it seems to provide more natural estimates than (1-5). Starting from [Diening et al. 2012], a first-order nonlinear Calderón–Zygmund theory for the p-Laplacian problem with right-hand side in divergence form

$$\operatorname{div}(|Du|^{p-2}\, Du) = \operatorname{div} G$$

is nowadays well-developed, showing the principle that the divergence operator can be "canceled out" to get estimates for $|Du|^{p-2}\, Du$ in terms of G in the same space. We refer to [Balci et al. 2020; Breit et al. 2018; 2022] for this line of research, but let us remark that the order of differentiability for V considered in these works is always less than 1.

Indeed, regarding the second-order Calderón–Zygmund theory (i.e., full Sobolev regularity for V), much less is known. A natural conjecture for solutions of (1-4) via the same principle would be

$$f \in L^m_{\mathrm{loc}}(\Omega) \quad \Longleftrightarrow \quad V \in W^{1,m}_{\mathrm{loc}}(\Omega), \tag{1-6}$$

which is actually false for $p \simeq 1$. The case $m = +\infty$, $p > 2$, corresponds to the well-known $C^{p'}$ conjecture, which will be discussed later, proved in the plane in [Araújo et al. 2017]. The endpoint $m = 1$ of (1-6) is considered by [Avelin et al. 2018], where, e.g., for $f \in L^1_{\mathrm{loc}}(\Omega)$, it is proved that $V \in W^{1-\varepsilon,1}$ for all $\varepsilon > 0$ (actually, f can be a general Radon measure in [Avelin et al. 2018]).

The only positive result of the type (1-6) (beyond the close one in [Lou 2008]) concerns the Hilbertian case $m = 2$, which has been recently proved in [Cianchi and Mazya 2018] for equations having *Uhlenbeck structure*, i.e., of the form

$$\operatorname{div}(a(|Du|)\, Du) = f \quad (\text{or } \operatorname{div} G). \tag{1-7}$$

The role of this structural assumption is in fact the main motivation of this work: indeed, (with the exception of [Mingione 2007; Avelin et al. 2018]), the higher-order Calderón–Zygmund theorem exposed so far is restricted to equations of the form (1-7) and in this case it can be actually extended to systems; see [Breit et al. 2018; 2022; Cianchi and Mazya 2019; Balci et al. 2022]. As L^2-theory seems to be the basic step to deal with the general problem (1-6), it is worth investigating *to what extent the Uhlenbeck structure is necessary to develop such a theory* and whether the general nonlinear problem (1-1) enjoys the same Sobolev regularity for its stress field $V = DF(Du)$.

It turns out that such regularity holds true when the map $z \mapsto DF(z)$ is *quasiconformal.* A quasiconformal map $G : \mathbb{R}^N \to \mathbb{R}^N$ is a homeomorphism belonging to $W^{1,N}_{\mathrm{loc}}(\mathbb{R}^N)$ such that, for some finite K,

$$|\lambda_{\max}(DG(z))|^N \leqslant K\, |\det DG(z)| \tag{1-8}$$

almost everywhere, with $\lambda_{\max}(DG)$ being the maximum singular value of DG; we refer to [Martin 2014] for a short modern survey on quasiconformal mappings. The main outcome of our results is that quasiconformality of DF in (1-1) is a natural and robust notion of uniform ellipticity, flexible enough to encompass anisotropic examples such as (1-3), still allowing a reasonable regularity theory.

Quasiconformal maps which are gradients of convex functions (or being, more generally, monotone) have been systematically studied in [Kovalev and Maldonado 2005; Kovalev 2007]; convex potentials of quasiconformal mappings are called *quasiuniformly convex* functions.

Definition 1.1. A convex $F \in C^1(\mathbb{R}^N)$ is called *K-quasiuniformly convex* for some $K \in [1, +\infty)$ if it is not affine, $DF \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N)$, and

$$\lambda_{\max}(D^2F(z)) \leqslant K\, \lambda_{\min}(D^2F(z)) \quad \text{for a.e. } z \in \mathbb{R}^N, \tag{1-9}$$

where $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ denote the minimum and maximum eigenvalues of M.

For the most part, this will be the main assumption on F for the study of (1-1), and we will use the acronym “q.u.c.” for “quasiuniformly convex”. Clearly, (1-9) and (1-8) for $G = DF$ are equivalent (but (1-9) implies (1-8) with a constant K^{N-1}); in Proposition 2.3 we will show that any q.u.c. function is strictly convex and of (p, q)-growth. In Section 3C we will discuss concrete examples but, in the meantime, we note that $z \mapsto |z|^p$ is q.u.c. for any $p > 1$, and the sum of q.u.c. functions is q.u.c.; hence (1-3) is q.u.c.

1A. *Main results.*

We now present our main results, referring to the appropriate theorems in the following sections for more complete statements. The first one concerns local minimizers $u \in W^{1,p}_{\mathrm{loc}}(\Omega)$ for the functional J in (1-2), i.e., those u obeying $J(u) < +\infty$ and $J(u, B) \leqslant J(u + w, B)$ for all $B \Subset \Omega$ and $w \in W^{1,p}_0(B)$.

Theorem 1.2 (Theorem 3.3). *Let $\Omega \subseteq \mathbb{R}^N$ be open, $F \in C^1(\mathbb{R}^N)$ be K-q.u.c., and $f \in L^2(\Omega) \cap W^{-1,p'}(\Omega)$ for $p = 1 + 1/K$. Then any local minimizer $u \in W^{1,p}_{\mathrm{loc}}(\Omega)$ of J satisfies*

$$\|DF(Du)\|_{W^{1,2}(B_R)} \leqslant C(1 + \|f\|_{L^2(B_{2R})} + \|F(Du)\|^K_{L^1(B_{2R})})$$

for some $C = C(K, N, R)$ and all $B_{4R} \subseteq \Omega$.

Remark 1.3. Let us briefly discuss the main assumptions. Other comments may be found in Remark 3.4.

(1) *Assumptions on F.* The Sobolev regularity assumption $DF \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N)$ is necessary, as shown in Example 3.5, where a strictly convex, radial $F \in C^1(\mathbb{R}^N)$ obeying (1-9) is constructed in such a way that the stress field of a suitable solution to (1-1) with $f = 0$ is not absolutely continuous.

The q.u. convexity condition fails in simple examples such as the orthotropic p-Laplacian related to the integrand

$$F(z) = \sum_{i=1}^{N} |z_i|^p.$$

We remark that Giaquinta’s example [1987] in $\mathbb{R}^6$ is a local minimizer of an analytic functional whose integrand is not q.u.c. and such that the stress field is in $W^{1,s}_{\mathrm{loc}}(\mathbb{R}^N)$ only for $s < \frac{5}{4}$. Playing around with examples of similar structure suggests intricate interplays between the maximal Sobolev regularity of the stress field and the possibly nonstandard structure of DF; hence, it is not clear what to expect from functionals with non-quasiconformal gradient mapping.

In Example 3.6 we discuss integrands of the form $F(z) = F(|z|)$, while Examples 3.7 and 3.8 investigate more general anisotropic functionals.

- *Assumptions on f.* The hypothesis $f \in W^{-1,p'}(\Omega)$ has been made for expository reasons and the relation between p and K will be derived in Proposition 2.3. On one hand, our results are mostly local in nature and therefore it suffices to require $f \in L^2_{\mathrm{loc}}(\Omega) \cap W^{-1,p'}_{\mathrm{loc}}(\Omega)$, meaning with the latter the intersection of the spaces $W^{-1,p'}(\Omega_n)$ on an exhausting sequence of open relatively compact $\Omega_n \uparrow \Omega$. For $p \geqslant 2N/(N+2)$, the condition $f \in W^{-1,p'}_{\mathrm{loc}}(\Omega)$ automatically follows from Sobolev embedding and $f \in L^2_{\mathrm{loc}}(\Omega)$; for $p < 2\,N/(N+2)$, $L^2(\Omega)$ is not embedded in $W^{-1,p'}(\Omega)$, destroying the variational framework we chose to be in. One should then resort to the notion of *approximable solutions* (briefly described in the last section), in order to deal with those cases. We refer to [Alberico et al. 2019] for a comprehensive theory of approximable solutions in the anisotropic framework. Anyway, in terms of summability, $f \in W^{-1,p'}(\Omega)$ is implied by $f \in L^{(p^*)'}(\Omega)$, which is a weaker summability than the one in [Lou 2008].

- *The exponent p.* As will be clear from the proof, the exponent $p = 1 + 1/K$ is not the only possible choice. Any p satisfying
$$|z|^p \leqslant C(F(z)+1), \quad z \in \mathbb{R}^N,$$
will serve the purpose of proving Theorem 1.2, and the previous estimate holds true for $p = 1 + 1/K$ and F K-q.u.c., thanks to Proposition 2.3 below. In Example 3.6 we will see that higher choices of p are sometimes feasible. Clearly, the variational condition $f \in W^{-1,p'}(\Omega)$ is weaker for higher p's.

We will give two applications of Theorem 1.2. The first one deals with nonlinear Cordes conditions for variational problems. Cordes conditions usually refer to L^m-theory for elliptic equations in nondivergence form with measurable coefficients, namely, to solutions of
$$\sum_{ij=1}^{N} a_{ij} D_{ij} u = f, \tag{1-10}$$
with measurable coefficients $a_{ij} : \Omega \subseteq \mathbb{R}^N \to \mathbb{R}$ satisfying
$$\lambda\,|\xi|^2 \leqslant \sum_{ij=1}^{N} a_{ij}(x)\,\xi_i\,\xi_i \leqslant \Lambda\,|\xi|^2$$
for all $\xi \in \mathbb{R}^N$ and a.e. $x \in \Omega$. Under these measurability assumptions alone, there is no hope in general for the Calderón–Zygmund inequality
$$\|u\|_{W^{2,m}(\Omega)} \leqslant C_m \|f\|_{L^m(\Omega)} \quad \text{for all } m > 1. \tag{1-11}$$
Some regularity has to be assumed on a_{ij} in order to obtain (1-11) for all $m > 1$ (VMO regularity suffices; see [Chiarenza et al. 1991]). Roughly stated, a Cordes condition for (1-10) with discontinuous a_{ij} says that (1-11) holds if either m is sufficiently near 2, or Λ/λ is sufficiently near 1. A similar situation takes place for nonlinear equations in divergence form.

Theorem 1.4 (Theorem 4.3). *Let F obey the assumptions of Theorem 1.2 and $f \in L^m(\Omega) \cap W^{-1,p'}(\Omega)$ for some $m > 1$. Then any local minimizer $u \in W^{1,p}(\Omega)$ for J in* (1-2) *is such that $DF(Du) \in W^{1,m}_{\mathrm{loc}}(\Omega)$ and*
$$\|DF(Du)\|_{W^{1,m}(B_R)} \leqslant C(\|f\|_{L^m(B_{2R})} + \|DF(Du)\|_{L^m(B_{2R})})$$

for $C = C(K, N, m, R)$ and all $B_{4R} \subseteq \Omega$, in either of the following cases:

(i) $K \leqslant K_0$, *with* $K_0 = K_0(N, m) > 1$.

(ii) $|m - 2| \leqslant \delta_0$ *for* $\delta_0 = \delta_0(K, N) > 0$.

As a second application we will derive some partial results pertaining the $C^{p'}$ conjecture, which states that any solution of the p-Poisson equation with bounded right-hand side is $C^{p'}$ regular if $p > 2$ (hereafter we use the notation $C^\gamma = C^{[\gamma],\gamma-[\gamma]}$, where $[\gamma]$ is the integer part of γ).

Theorem 1.5 (Corollary 4.5, Theorem 4.7, and Corollary 4.8). *Let $u : \Omega \to \mathbb{R}$ be an approximable solution of* (1-4).

(i) *If $f \in L^\infty(\Omega)$, then for all sufficiently small $|p-2|$ it holds $u \in C^{2-\alpha_p}(\Omega)$, where $\alpha_p = c(N)|p-2| > 0$.*

(ii) *For any $m, p > 1$, if u and Ω have cylindrical symmetry, then $|Du|^{p-2}\, Du \in W^{1,m}_{\mathrm{loc}}(\Omega)$ whenever $f \in L^m_{\mathrm{loc}}(\Omega)$. In particular, any cylindrical approximable solution of* (1-4) *belongs to $C^{\min\{p',2\}-\varepsilon}(\Omega)$ for any $\varepsilon > 0$.*

Remark 1.6. • Item (i) confirms the validity of the $C^{p'}$ conjecture near uniform ellipticity, and is inspired by [Mercuri et al. 2016]. Here, however, we obtain an explicit rate of the Hölder exponent and, more substantially, we do not require Sobolev regularity on f. When $f \in L^\infty(\Omega)$, the usual notion of weak solution suffices.

• Point (ii) gives a weak form of the $C^{p'}$ conjecture (namely, $u \in C^{p'-}$) in the class of cylindrical solutions for any $p \geqslant 2$, with a different approach than the one of [Araújo et al. 2018]. We need the notion of approximable solutions as in general f may fail to belong to $W^{-1,p'}(\Omega)$ for small $m > 1$, destroying the variational setting. For details on such a notion we refer to Section 4B.

By a cylindrical solution we mean a function of the form $u(x) = v(|x'|)$, $x = (x', x'') \in \mathbb{R}^k \times \mathbb{R}^{N-k}$, $k \leqslant N$. It is worth noticing that the domain Ω may not contain the origin, in which case the approach of [Araújo et al. 2018] cannot be easily applied.

1B. ***Outline of the proofs.*** Consider, as a first step, a smooth compactly supported solution u of

$$\operatorname{div} DF(Du) = f \quad \text{in } \mathbb{R}^N,$$

with F and f smooth. Our starting point is the well-known identity

$$\|DV\|^2_{L^2(\mathbb{R}^N)} = \|\operatorname{div} V\|^2_{L^2(\mathbb{R}^N)} + \tfrac{1}{2}\|\operatorname{curl} V\|^2_{L^2(\mathbb{R}^N)} \quad \text{for all } V \in C^\infty_c(\mathbb{R}^N; \mathbb{R}^N), \tag{1-12}$$

where $\operatorname{curl} V = DV - DV^t$. Applying (1-12) to the stress field $V = DF(Du)$, we are reduced to estimate $\operatorname{curl} V$.

The main observation is that, in the smooth setting, DV is of special type, namely

$$DV = D^2F(Du)\, D^2u,$$

where $D^2F(Du)$ is a symmetric positive definite matrix and D^2u is symmetric. An elementary lemma shows that any matrix of the form

$$X = PS, \quad P \text{ symmetric positive definite, } S \text{ symmetric,}$$

satisfies

$$|X - X^t|_2^2 \leqslant 2\left(1 - \frac{\lambda_{\min}}{\lambda_{\max}}\right)^2 |X|_2^2, \tag{1-13}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are, respectively, the minimum and maximum eigenvalues of P. Thus the curl term in (1-12) can be reabsorbed to the left if $\lambda_{\max} \leqslant K\,\lambda_{\min}$ holds a.e. for the matrix D^2F, giving

$$\|DV\|_{L^2(\mathbb{R}^N)}^2 \leqslant K^2\,\|f\|_{L^2(\mathbb{R}^N)}^2$$

for $V = DF(Du)$ in the smooth, global setting.

In order to prove Theorem 1.2 we have to localize the estimate and to suitably build smooth approximating problems. We regularize the integrand, the source, and the boundary data through convolution but, in order to have strongly elliptic problems, we would like to add a small multiple of a strongly elliptic functional. Since we do not want to alter the q.u. convexity constant, the only viable choice is to add small multiples of $|z|^2$ to the regularized integrands. This is a quite unnatural choice if F is not of standard quadratic growth, and it forces some interplay between the regularization parameters. Here, the explicit a priori Lipschitz estimate for the corresponding solutions taken from [Bousquet and Brasco 2016] plays a key role.

The main step of the proof of Theorem 1.4 is to represent the solutions of

$$\begin{cases} \operatorname{div} V = f, \\ \operatorname{curl} V = G \end{cases}$$

in $\mathbb{R}^N$ through Riesz transforms and generalize (1-12) to the L^m-setting as

$$\|DV\|_{L^m(\mathbb{R}^N)} \leqslant C(m, N)(\|\operatorname{div} V\|_{L^m(\mathbb{R}^N)} + \|\operatorname{curl} V\|_{L^m(\mathbb{R}^N)}).$$

Thus, inequality (1-13) does the trick in case (i) of Theorem 1.4, allowing reabsorption of the curl term for K sufficiently near 1. A standard Riesz interpolation argument, together with a careful choice of the norms involved, allows proving case (ii).

Finally, point (i) of Theorem 1.5 is an immediate consequence of point (i) of Theorem 1.4, while point (ii) stems from this observation: if u exhibits cylindrical symmetry, then the stress field $a(|Du|)\,Du$ is irrotational; by the Helmoltz decomposition, it can be locally represented as the gradient of a solution of $\Delta v = f \in L^m$, so the standard Calderón–Zygmund theorem applies.

1C. ***Structure of the paper.*** In Section 2 we recall some functional inequalities and properties about quasiuniform convexity. In Section 3A we develop our basic estimate in the smooth setting; Section 3B is devoted to the proof of Theorem 1.2, where the main approximation procedure, used also later, is described; Section 3C contains the relevant examples depicted above. In Section 4 we focus on the applications: first we treat the Cordes conditions, and finally we collect the partial results pertaining the $C^{p'}$ conjecture.

Notations. • The euclidean norm of a vector $v \in \mathbb{R}^N$ is denoted by $|v|$ and (v, w) denotes the scalar product. By Ω we denote a bounded open subset of $\mathbb{R}^N$, while B_r denotes a ball of radius r not necessarily centered at the origin. Similarly, B denotes a ball with unspecified center and radius; if $B = B_r(x_0)$, then we set $\lambda B = B_{\lambda r}(x_0)$.

- For a $N \times N$ matrix $A = (a_{ij})$, its transpose is denoted by A^t; on such matrices we consider the Frobenius norm

$$|A|_2 = \Bigl(\sum_{i,j=1}^{N} |a_{ij}|^2 \Bigr)^{1/2},$$

arising from the scalar product $(A, B)_2 = \mathrm{Tr}(A\, B^t)$. For $v, w \in \mathbb{R}^N$, $v \otimes w$ is the matrix with entries $(v_i\, w_j)$, while $v \wedge w = v \otimes w - w \otimes v$; I denotes the identity matrix, while Id the identity function of $\mathbb{R}^N$; O_N is the orthogonal group of $N \times N$ matrices such that $A\, A^t = A^t\, A = I$; if A is symmetric, $\lambda_{\min}(A)$, $\lambda_{\max}(A)$ are its minimum and its maximum eigenvalues.

- If a domain of integration is missing, it means that we are integrating on $\mathbb{R}^N$. We also set for brevity $\|f\|_m = \|f\|_{L^m(\mathbb{R}^N)}$. Finally, we sometimes will use the L^m "norm" also for $m \in (0, 1)$, when it is only positively 1-homogeneous.

- As is customary, we indicate with $W^{-1,p'}(\Omega)$ the dual space of $W_0^{1,p}(\Omega)$.

2. Preliminaries

2A. ***Functional inequalities.*** For $0 < \theta < m$, $m > 1$, starting from the inequality

$$\int_{B_1} |v|^m \, dx \leqslant \int_{B_1} |Dv|^m \, dx + C(N, m, \theta) \Bigl(\int_{B_1} |v|^\theta \, dx \Bigr)^{m/\theta} \tag{2-1}$$

(obtained by a standard compactness argument), and replicating, with the obvious modifications, the proof of [Cianchi and Mazya 2018, equation (5.4)], we get the following functional inequality.

Lemma 2.1. *Let $m > 1$, $0 < \theta < m$, and $R \leqslant r < s < 2R$. There exists $C = C(N, m, \theta)$ such that for any $v \in W^{1,m}(B_s \setminus B_r)$ and any $\delta \in (0, 1)$ it holds*

$$\int_{B_s \setminus B_r} |v|^m \, dx \leqslant \delta^m R^m \int_{B_s \setminus B_r} |Dv|^m \, dx + \frac{C}{(\delta^N\, (s - r)\, R^{N-1})^{(m-\theta)/\theta}} \Bigl(\int_{B_s \setminus B_r} |v|^\theta \, dx \Bigr)^{m/\theta}.$$

The following lemma is a straightforward generalization of the well-known identity

$$\int |DV|_2^2 \, dx = \frac{1}{2} \int |\operatorname{curl} V|_2^2 \, dx + \int (\operatorname{div} V)^2 \, dx, \tag{2-2}$$

valid for $V \in C_c^1(\mathbb{R}^N; \mathbb{R}^N)$.

Lemma 2.2. *If $V \in C^2(\mathbb{R}^N, \mathbb{R}^N)$, then for any $\varphi \in C_c^2(\mathbb{R}^N)$ it holds*

$$\int \varphi^2 \, |DV|_2^2 \, dx = \frac{1}{2} \int \varphi^2 \, |\operatorname{curl} V|_2^2 \, dx + \int \varphi^2 \, (\operatorname{div} V)^2 \, dx + \int [2(D\varphi^2, V) \operatorname{div} V + (D^2\varphi^2, V \otimes V)_2] \, dx. \tag{2-3}$$

Proof. Write, through parallelogram identity,

$$|DV|_2^2 = \tfrac{1}{4} |DV - DV^t|_2^2 + \tfrac{1}{4} |DV + DV^t|_2^2,$$

then multiply by φ^2 and integrate to obtain

$$\begin{aligned}\int \varphi^2\,|DV|_2^2\,dx &= \int \frac{\varphi^2}{4}\,|\operatorname{curl} V|_2^2\,dx + \int \frac{\varphi^2}{4}\textstyle\sum_{ij}(D_j V^i + D_i V^j)^2\,dx \\ &= \int \frac{\varphi^2}{4}|\operatorname{curl} V|_2^2\,dx + \int \frac{\varphi^2}{2}\,|DV|_2^2\,dx + \int \frac{\varphi^2}{2}\textstyle\sum_{ij} D_i V^j D_j V^i\,dx. \qquad (2\text{-}4)\end{aligned}$$

The last term is computed integrating by parts twice: for any $i, j = 1, \dots, N$

$$\begin{aligned}&\int \varphi^2\, D_i V^j D_j V^i\,dx \\ &\quad = -\int D_i\varphi^2\, V^j D_j V^i\,dx - \int \varphi^2\, V^j D_{ij} V^i\,dx \\ &\quad = \int D_{ij}\varphi^2\, V^j V^i\,dx + \int D_i\varphi^2\, V^i D_j V^j\,dx + \int D_j\varphi^2\, V^j D_i V^i\,dx + \int \varphi^2\, D_j V^j D_i V^i\,dx,\end{aligned}$$

so that summing over i, j gives

$$\int \varphi^2 \textstyle\sum_{ij} D_i V^j D_j V^i\,dx = \int [\varphi^2(\operatorname{div} V)^2 + 2\,(D\varphi^2, V)\operatorname{div} V + (D^2\varphi^2, V\otimes V)_2]\,dx.$$

Inserting this formula into (2-4) yields (2-3). □

2B. ***Quasiuniform convexity.*** In this section we show that the q.u. convexity condition provides, in a unified way, many of the properties that the usual integrands of the calculus of variations satisfy.

Let us begin by observing that the gradient of a convex F is defined a.e. and belongs to $BV_{\mathrm{loc}}(\mathbb{R}^N)$; see [Alberti and Ambrosio 1999]. Accordingly, the second derivative of F can be decomposed into an absolutely continuous part, a jump part, and a Cantor part. If $F \in C^1(\mathbb{R}^N)$ the jump part vanishes; hence by requiring that $F \in C^1(\mathbb{R}^N) \cap W^{2,1}_{\mathrm{loc}}(\mathbb{R}^N)$ we are actually excluding that $D^2 F$ has a Cantor part.

Now we discuss in detail some consequences of the q.u. convexity condition; although condition (iv) of Proposition 2.3 will be not used in the sequel, we prove it for the sake of completeness.

Proposition 2.3 (properties of K-quasiuniformly convex functions). *Let F be a K-quasiuniformly convex function. Then*:

(i) *DF is K^{N-1}-quasiconformal*; *hence $C^{1/K}(\mathbb{R}^N)$.*

(ii) *F is strictly convex and of (p, q)-growth, i.e., there exists $C = C(N, K, F) > 0$ and $1 < p < q < +\infty$ such that*

$$C^{-1}|z|^p - C \leqslant F(z) \leqslant C(|z|^q + 1), \qquad (2\text{-}5)$$

$$C^{-1}|z|^{p-1} - C \leqslant |DF(z)| \leqslant C(|z|^{q-1} + 1) \qquad (2\text{-}6)$$

for all $z \in \mathbb{R}^N$. More precisely, one can take $p = 1 + 1/K$ and $q = 1 + K$.

(iii) *If $\varphi \in C_c^\infty(\mathbb{R}^N; [0, +\infty))$, then $F * \varphi$ is K-q.u.c.*

(iv) *Its Moreau–Yoshida regularization*

$$F_\delta(z) = \inf_{y\in\mathbb{R}^N}\left\{F(y) + \frac{1}{2\,\delta}|y-z|^2\right\} \tag{2-7}$$

is K-q.u.c.

Proof. (i) By the Alexandrov theorem DF is differentiable a.e., and [Väisälä 1971, Theorem 32.3] ensures that $DF \in W^{1,N}_{\mathrm{loc}}(\mathbb{R}^N)$. Since (1-8) and (1-9) are equivalent up to changing the constants, u is quasiuniformly convex in the sense of [Kovalev and Maldonado 2005]. In particular Theorem 3.1 of that work shows that DF is K^{N-1}-quasiconformal. The regularity statement holds for any quasiconformal mapping; see [Martin 2014, Theorem 2.14].

(ii) The strict convexity of F follows from [Kovalev and Maldonado 2005, Lemma 3.2]. If z_0 is the unique minimum point for F we can consider $F(z_0 + \cdot\,) - F(z_0)$, so there is no loss in generality assuming $DF(0) = 0,\ F(0) = 0$. Let $G = DF$, which then is K^{N-1}-quasiconformal. By [Martin 2014, Theorem 2.14]

$$|G(z)| \leqslant C(N, K) \sup_{y\in B_1} |G(y)|\,|z|^{1/K}, \quad z \in B_1.$$

Since G^{-1} is still K^{N-1} quasiconformal, it obeys a similar estimate, proving the lower bound

$$|G(z)| \geqslant C(N, K, G)|z|^K, \quad z \in B_1.$$

Finally, the inversion

$$G^*(x) = \frac{G(x/|x|^2)}{|G(x/|x|^2)|^2}$$

is again K^{N-1} quasiconformal on B_1, so that the previous estimates are transferred to the outside of B_1 as

$$C^{-1}|z|^{1/K} \leqslant |G(z)| \leqslant C|z|^K, \quad |z| \geqslant 1, \tag{2-8}$$

where $C = C(N, K, G)$. For $G = DF,\ p = 1 + 1/K,\ q = 1 + K$, we thus obtained (2-6). Moreover, by [Kovalev 2007, Lemma 18], DF is δ-convex for some $\delta = \delta(N, K) > 0$, meaning that

$$(DF(z) - DF(y), z - y) \geqslant \delta|DF(z) - DF(y)|\,|z - y| \quad \text{for all } z, y \in \mathbb{R}^N. \tag{2-9}$$

Using (2-9) and (2-6), we get

$$F(z) = \int_0^1 (DF(tz), z)\,dt \geqslant \delta \int_0^1 |DF(tz)|\,|z|\,dt \geqslant \frac{\delta}{p\,C}\,|z|^p - C\,|z| \geqslant \frac{\delta}{p\,C}|z|^p - C,$$

by sufficiently increasing C in the last inequality. This produces the lower bound in (2-5), while the upper bound follows from (2-6) alone through a similar calculation.

(iii) Let $\lambda_{\min}(z) = \lambda_{\min}(D^2F(z))$, where z is a second-order differentiability point for F. From the representation

$$\lambda_{\min}(z) := \inf\{(D^2F(z)\,\xi, \xi) : \xi \in D\},$$

where D is a fixed countable dense subset of $\mathbb{S}^{N-1}$, we infer that $\lambda_{\min}$ is measurable and in $L^1_{\mathrm{loc}}(\mathbb{R}^N)$. Then, for any $z \in \mathbb{R}^N$ and $\xi \in D$,

$$\begin{aligned}(D^2F * \varphi(z)\,\xi, \xi) &= \int \varphi(z-y)\,(D^2F(y)\,\xi, \xi)\,dy \\ &\geqslant \int \varphi(z-y)\,\lambda_{\min}(y)\,|\xi|^2\,dx =: \tilde{\lambda}_{\min}(z)\,|\xi|^2,\end{aligned}$$

while

$$\int \varphi(z-y)\,(D^2F(y)\,\xi, \xi)\,dy \leqslant \int \varphi(z-y)\,K\,\lambda_{\min}(y)\,|\xi|^2\,dy = K\,\tilde{\lambda}_{\min}(z)\,|\xi|^2,$$

implying the claim.

(iv) Recall that the minimum in (2-7) is attained at a unique point $P_\delta(z)$ satisfying

$$P_\delta(z) + \delta\,DF(P_\delta(z)) = z, \quad DF_\delta(z) = DF(P_\delta(z)), \tag{2-10}$$

and the so-defined function $P_\delta = (\mathrm{Id} + \delta\,DF)^{-1}$ is 1-Lipschitz and a homeomorphism of $\mathbb{R}^N$, since $F \in C^1(\mathbb{R}^N)$. Let E be the set of points where DF fails to be differentiable. The map $\mathrm{Id} + \delta\,DF$ is the gradient of a K-quasiuniformly convex function, and hence by point (i) is quasiconformal and satisfies the Lusin (N) property, i.e., it sends null-measure sets to null-measure sets; see [Väisälä 1971]. Thus, $P_\delta^{-1}(E)$ has zero measure. Moreover, since P_δ is Lipschitz continuous, Rademacher's theorem ensures that the set M where P_δ is not differentiable has zero measure. We will prove (1-9) at any point

$$z \notin M \cup P_\delta^{-1}(E),$$

the latter set having zero measure. Indeed, at any such point z we have that DF is differentiable at $P_\delta(z)$ and P_δ is differentiable at z. The chain rule applied to (2-10) then gives

$$(I + \delta D^2F(P_\delta(z)))\,DP_\delta(z) = I, \quad D^2F_\delta(z) = D^2F(P_\delta(z))\,DP_\delta(z),$$

which yields

$$D^2F_\delta(z) = D^2F(P_\delta(z))(I + \delta\,D^2F(P_\delta(z)))^{-1}. \tag{2-11}$$

Let the eigenvalues of $D^2F(P_\delta(z))$ be $\lambda_{\min} = \lambda_1 \leqslant \ldots \leqslant \lambda_N = \lambda_{\max}$. The matrices $D^2F(P_\delta(z))$ and $(I+\delta\,D^2F(P_\delta(z)))^{-1}$ have the same basis of eigenvectors, with eigenvalues λ_i and $(1+\delta\,\lambda_i)^{-1}$ respectively. Hence (2-11) implies that $D^2F_\delta(z)$ has eigenvalues $\lambda_i/(1+\delta\,\lambda_i)$. As $t \mapsto t/(1+\delta\,t)$ is increasing, its minimum and maximum eigenvalues are

$$\lambda_{\delta,\min} := \frac{\lambda_{\min}}{1+\delta\lambda_{\min}}, \quad \lambda_{\delta,\max} := \frac{\lambda_{\max}}{1+\delta\lambda_{\max}},$$

which obey $\lambda_{\delta,\max} \leqslant K\lambda_{\delta,\min}$ as long as $\lambda_{\max} \leqslant K\lambda_{\min}$. □

Due to the previous proposition, we will denote henceforth by p and q the powers of the lower and upper bounds, respectively, for a given K-q.u.c. function F.

2C. *Extensions.*

We conclude with a couple of tools which will be occasionally used in the following.

Lemma 2.4. *Let $F \in C^1(B_R)$ be a nonnegative, strictly convex function, and $\sigma \in (0, 1)$.*

(i) *There exists a strictly convex $\widetilde{F} \in C^1(\mathbb{R}^N)$ such that $\widetilde{F}|_{B_{\sigma R}} = F$ and, for some C, α depending on F, R, N, σ, it holds*

$$|z|^2 \leqslant \alpha(\widetilde{F}(z)+1), \quad C^{-1}|z| - C \leqslant |D\widetilde{F}(z)| \leqslant C(|z|+1). \tag{2-12}$$

(ii) *If F is K-q.u.c. in B_R as per Definition 1.1, and for some $\varepsilon > 0$ it holds $\lambda_{\min}(z) \geqslant \varepsilon$ in $B_R \setminus B_{\sigma R}$, in addition to* (2-12) *$\widetilde{F}$ can be chosen to be $\widetilde{K}$-q.u.c., with $\widetilde{K} = \widetilde{K}(F, R, N, \sigma, \varepsilon)$.*

Proof. To prove (i), let $\tau = (1+\sigma)/2$, choose a radial cut-off function $\eta \in C^\infty_c(B_R; [0, 1])$ such that $\eta \equiv 1$ in $B_{\tau R}$, and define

$$\widetilde{F}(z) = \eta(z)\, F(z) + (1-\eta(z))\, \frac{|z|^2}{2} + C\, (|z| - \sigma\, R)^2_+,$$

where $C > 0$ is a constant to be chosen. Clearly $\widetilde{F} \in C^1(\mathbb{R}^N)$ and obeys (2-12), so it remains to show that F is strictly convex for a suitable C. To this aim, set

$$A(z) := \frac{1}{2}D^2(|z| - \sigma R)^2 = \frac{\sigma R}{|z|}\frac{z}{|z|} \otimes \frac{z}{|z|} + \left(1 - \frac{\sigma R}{|z|}\right) I,$$

whose eigenvalues are 1 and $1 - \sigma R/|z|$. In particular, A is nonnegative definite in $B_R \setminus B_{\sigma R}$, and its eigenvalues are uniformly bounded below in $\mathbb{R}^N \setminus B_{\tau R}$ by a positive constant; since $\widetilde{F}$ agrees with F in $B_{\sigma R}$, it follows that $\widetilde{F}$ is strictly convex in $B_{\tau R}$ and in $\mathbb{R}^N \setminus B_R$. A straightforward computation yields

$$D^2\widetilde{F} = \eta\, D^2 F + M + 2\, C A,$$

$$M := (1-\eta)\, I + D\eta \otimes DF + DF \otimes D\eta - 2\, D\eta \otimes z + \left(F - \frac{|z|^2}{2}\right) D^2\eta$$

a.e. outside $B_{\sigma R}$, and we can choose C so that

$$\left(1 - \frac{\sigma}{\tau}\right) C = \max_{z \in B_R} |M(z)|_2,$$

ensuring

$$\lambda_{\min}(D^2\widetilde{F}) \geqslant \eta\, \lambda_{\min}(D^2 F) + \left(1 - \frac{\sigma}{\tau}\right) C \quad \text{in } B_R \setminus B_{\tau R}.$$

Summing up, $\widetilde{F}$ is globally strictly convex by an elementary argument.

To prove (ii), let $\tilde{\lambda}_{\min}(z)$ and $\tilde{\lambda}_{\max}(z)$ denote the minimum and maximum eigenvalues of $D^2\widetilde{F}(z)$, and $\lambda_{\min}$, $\lambda_{\max}$ those for $D^2 F$. It holds

$$\varepsilon \leqslant \lambda_{\min} \leqslant \tilde{\lambda}_{\min} \leqslant \tilde{\lambda}_{\max} \leqslant 3\, C + \lambda_{\max}$$

in $B_{\tau R} \setminus B_{\sigma R}$, so that from (1-9) we get

$$\frac{\tilde{\lambda}_{\max}}{\tilde{\lambda}_{\min}} \leqslant \frac{\tilde{\lambda}_{\max}}{\lambda_{\min}} \leqslant \frac{3\, C}{\lambda_{\min}} + K \leqslant \frac{3\, C}{\varepsilon} + K.$$

In $\mathbb{R}^N \setminus B_{\tau R}$ it holds

$$(1-\sigma/\tau)\, C + \eta\, \lambda_{\min} \leqslant \tilde{\lambda}_{\min} \leqslant \tilde{\lambda}_{\max} \leqslant 3\, C + \eta\, \lambda_{\max},$$

so that

$$\frac{\tilde{\lambda}_{\max}}{\tilde{\lambda}_{\min}} \leqslant \frac{3\, C + K\, \eta\, \lambda_{\min}}{(1-\sigma/\tau)\, C + \eta\, \lambda_{\min}} \leqslant \frac{3\, \tau}{\tau - \sigma} + K.$$

Since $\widetilde{F} = F$ on $B_{\sigma R}$ and (1-9) holds for F there, the claim is proved. □

3. Divergence form, quasiconformal equations

3A. ***The smooth setting.*** The core of our approach lies in the following elementary observation.

Lemma 3.1. *Let $X = P S$, where P and S are symmetric $N \times N$ matrices and P is positive definite with minimum and maximum eigenvalues $\lambda_{\min}$ and $\lambda_{\max}$. Then*

$$|X - X^t|_2^2 \leqslant 2\, \frac{(1-\lambda_{\min}/\lambda_{\max})^2}{1+(\lambda_{\min}/\lambda_{\max})^2}\, |X|_2^2. \tag{3-1}$$

Proof. Inequality (3-1) is invariant under rotations; thus, without loss of generality, we can suppose $P_{ij} = \lambda_i \delta_{ij}$ with $0 < \lambda_1 \leqslant \ldots \leqslant \lambda_N$. Then from $X = P\, S$ we get

$$X_{ij} = \lambda_i\, S_{ij},$$

so that

$$|X - X^t|_2^2 = \sum_{ij} |X_{ji} - X_{ij}|^2 = 2 \sum_{i<j} |X_{ji} - X_{ij}|^2 = 2 \sum_{i<j} |\lambda_j\, S_{ji} - \lambda_i\, S_{ij}|^2,$$

and from the symmetry of S we conclude

$$|X - X^t|_2^2 = 2 \sum_{i<j} S_{ij}^2\, (\lambda_j - \lambda_i)^2. \tag{3-2}$$

Similarly, we have

$$|X|_2^2 \geqslant \sum_{i<j} (X_{ij}^2 + X_{ji}^2) = \sum_{i<j} S_{ij}^2 (\lambda_i^2 + \lambda_j^2). \tag{3-3}$$

Let

$$\varphi(t) = \frac{(1-t)^2}{1+t^2},$$

which is decreasing in [0, 1], and observe that for $j > i$ we have $\lambda_i/\lambda_j \in [0, 1]$. Therefore

$$(\lambda_j - \lambda_i)^2 = \frac{(\lambda_j - \lambda_i)^2}{\lambda_j^2 + \lambda_i^2}\, (\lambda_j^2 + \lambda_i^2) = \varphi\!\left(\frac{\lambda_i}{\lambda_j}\right)(\lambda_j^2 + \lambda_i^2) \leqslant \varphi\!\left(\frac{\lambda_{\min}}{\lambda_{\max}}\right)(\lambda_j^2 + \lambda_i^2).$$

Inserting this estimate in (3-3) and recalling (3-2) we get

$$|X - X^t|_2^2 \leqslant 2\, \varphi\!\left(\frac{\lambda_{\min}}{\lambda_{\max}}\right) \sum_{i<j} S_{ij}^2 (\lambda_i^2 + \lambda_j^2) = 2\, \varphi\!\left(\frac{\lambda_{\min}}{\lambda_{\max}}\right) |X|_2^2.$$ □

Theorem 3.2. *Let $u \in C^2(B_{2R})$ solve*

$$\operatorname{div}(DF(Du)) = f \quad \textit{in } B_{2R}$$

for a K-q.u.c. $F \in C^2(\mathbb{R}^N)$, and let

$$V(x) = DF(Du(x)).$$

Then for any $\theta \in (0, 2]$ there exist $C = C(N, K, \theta)$ and $C_R = C(N, K, \theta, R)$ such that

$$\|V\|_{W^{1,2}(B_R)} \leqslant C\, \|f\|_{L^2(B_{2R})} + C_R\, \|V\|_{L^\theta(B_{2R})}. \tag{3-4}$$

Proof. For any $\varepsilon > 0$ let

$$F_\varepsilon(z) = F(z) + \varepsilon \frac{|z|^2}{2}, \quad f_\varepsilon = f + \varepsilon\, \Delta u,$$

so that $D^2 F_\varepsilon$ is symmetric and positive definite. It holds

$$\lambda_{\min}(D^2 F_\varepsilon(z)) = \lambda_{\min}(D^2 F(z)) + \varepsilon, \quad \lambda_{\max}(D^2 F_\varepsilon(z)) = \lambda_{\max}(D^2 F(z)) + \varepsilon,$$

so that if (1-9) holds for F, it does so for $D^2 F_\varepsilon$ as well, with the same constant K.

Clearly u solves

$$\operatorname{div}(DF_\varepsilon(Du)) = f_\varepsilon$$

in B_{2R}. Letting

$$V_\varepsilon = DF_\varepsilon(Du),$$

it holds

$$DV_\varepsilon = D^2 F_\varepsilon(Du)\, D^2 u,$$

where the first matrix is symmetric positive definite and the second one is symmetric. We thus apply Lemma 3.1 to the matrix

$$P = D^2 F_\varepsilon(Du), \quad S = D^2 u, \quad X = DV_\varepsilon = PS.$$

Recall that $D^2 F_\varepsilon(Du)$ satisfies (1-9) with constant K, whence

$$\frac{\big(\lambda_{\max}(D^2 F_\varepsilon(Du)) - \lambda_{\min}(D^2 F_\varepsilon(Du))\big)^2}{\lambda^2_{\max}(D^2 F_\varepsilon(Du)) + \lambda^2_{\min}(D^2 F_\varepsilon(Du))} \leqslant \frac{(K-1)^2}{K^2+1}.$$

From (3-1) we get

$$|\operatorname{curl} V_\varepsilon|_2^2 \leqslant 2\, \frac{(K-1)^2}{K^2+1}\, |DV_\varepsilon|_2^2. \tag{3-5}$$

For any r, s with $R \leqslant r < s \leqslant 2\,R$, fix $\varphi \in C_c^\infty(B_s, [0, 1])$ such that

$$\varphi|_{B_r} \equiv 1, \quad |D\varphi| \leqslant \frac{C}{s-r}, \quad |D^2\varphi| \leqslant \frac{C}{(s-r)^2}. \tag{3-6}$$

This will allow us to consider $\varphi\, V_\varepsilon$ as defined on the whole $\mathbb{R}^N$, so that (2-3) holds true. The stipulated properties of φ ensure that

$$\int 2(D\varphi^2, V_\varepsilon)\, f_\varepsilon + (D^2\varphi^2, V_\varepsilon \otimes V_\varepsilon)_2\, dx \leqslant \frac{C}{(s-r)^2} \int_{B_s \setminus B_r} |V_\varepsilon|^2\, dx + C \int_{B_{2R}} f_\varepsilon^2\, dx,$$

where we used the Schwartz inequality on the first term and $s \leqslant 2R$. Using also (3-5) to control the curl term of Lemma 2.2 yields

$$\int \varphi^2 |DV_\varepsilon|_2^2\,dx \leqslant \left(1-\frac{1}{K}\right)^2 \int \varphi^2\,|DV_\varepsilon|_2^2\,dx + \frac{C}{(s-r)^2}\int_{B_s\setminus B_r} |V_\varepsilon|^2\,dx + C\int_{B_{2R}} f_\varepsilon^2\,dx.$$

We let $\varepsilon \to 0$ and bring the first term on the right to the left-hand side; recalling that $\varphi \equiv 1$ on B_r, we obtain

$$\int_{B_r} |DV|_2^2\,dx \leqslant \frac{C_K}{(s-r)^2}\int_{B_s\setminus B_r} |V|^2 + C_K \int_{B_{2R}} f^2\,dx \tag{3-7}$$

for any $R \leqslant r < s \leqslant 2R$. We next proceed as in [Cianchi and Mazya 2018]: by Lemma 2.1 with $m = 2$ and

$$\delta = \frac{s-r}{2\sqrt{C_K}\,R},$$

we get

$$\frac{C_K}{(s-r)^2}\int_{B_s\setminus B_r} |V|^2\,dx \leqslant \frac{1}{4}\int_{B_s\setminus B_r} |DV|_2^2\,dx + \frac{C_K R^{(2-\theta)/\theta}}{(s-r)^{2+(2-\theta)/\theta(N+1)}}\left(\int_{B_s\setminus B_r} |V|^\theta\,dx\right)^{2/\theta},$$

which inserted into (3-7) gives, for all $R \leqslant r < s \leqslant 2R$,

$$\int_{B_r} |DV|_2^2\,dx \leqslant \frac{1}{4}\int_{B_s} |DV|_2^2\,dx + C_K\int_{B_{2R}} f^2\,dx + \frac{C_K R^{(2-\theta)/\theta}}{(s-r)^{2+(2-\theta)/\theta(N+1)}}\left(\int_{B_s\setminus B_r} |V|^\theta\,dx\right)^{2/\theta}.$$

A standard iteration lemma (see [Giaquinta 1983, Lemma 3.1, Chapter 5]) improves the latter to

$$\int_{B_R} |DV|_2^2\,dx \leqslant C_K\int_{B_{2R}} f^2\,dx + \frac{C_K}{R^{2+(2-\theta)/\theta N}}\left(\int_{B_{2R}} |V|^\theta\,dx\right)^{2/\theta},$$

which is the desired estimate on the derivative of V. In order to control $\|V\|_{L^2(B_R)}^2$, we invoke the rescaled form of (2-1) which, in conjunction with the previous estimate, completes the proof of (3-4). □

3B. *Local minimizers.*

For a bounded $\Omega \subseteq \mathbb{R}^N$ we let

$$J(w,\Omega) = \int_\Omega F(Dw)\,dx + \int_\Omega f w\,dx$$

whenever the two integrands are in $L^1(\Omega)$, sometimes omitting the dependence on Ω when this causes no confusion. We will consider J under p-coercivity assumptions on F and for $f \in W^{-1,p'}(\Omega)$, so that it is well-defined on $W^{1,p}(\Omega)$.

Recall that $u \in W^{1,p}_{\mathrm{loc}}(\Omega)$ is a local minimizer for J in $W^{1,p}(\Omega)$ if, for any $B \Subset \Omega$,

$$J(u,B) = \inf\{J(w,B) : w \in u + W_0^{1,p}(B)\}. \tag{3-8}$$

Theorem 3.3. *Let $F \in C^1(\mathbb{R}^N)$ be a q.u.c. function and $q > p > 1$ be given in Proposition 2.3*(ii). *For $f \in L^2(\Omega) \cap W^{-1,p'}(\Omega)$, let u be a local minimizer u of J in Ω. Then, for any ball B such that $4B \subseteq \Omega$ it holds $DF(Du) \in W^{1,2}(B)$,*

$$\|DF(Du)\|_{L^2(B)} \leqslant C(1 + \|f\|_{L^2(2B)} + \|F(Du)\|_{L^1(2B)}^{(q-1)/p}) \tag{3-9}$$

for $C = C(K, N, B) > 0$, *and for any* $\theta \in (0, 2]$

$$\|DF(Du)\|_{W^{1,2}(B)} \leqslant C(\|f\|_{L^2(2B)} + \|DF(Du)\|_{L^\theta(2B)}) \tag{3-10}$$

for $C = C(K, N, B, \theta) > 0$. *Moreover*, u *satisfies the Euler–Lagrange equation*

$$\int_\Omega (DF(Du), D\varphi)\, dx = \int_\Omega f\varphi\, dx \quad \textit{for all } \varphi \in C_c^\infty(\Omega). \tag{3-11}$$

Proof. Let $\operatorname{Argmin}(F) = \{z_0\}$. By considering $\widetilde{F}(z) = F(z + z_0) - F(z_0)$ and $\tilde{u}(x) = u(x) - (z_0, x)$ and noting that $f(\,\cdot\,)\,(z_0, \cdot\,) \in L^1_{\text{loc}}(\Omega)$, it is readily checked that $\tilde{u}$ turns out to be a local minimizer of

$$\widetilde{J}(w, \Omega) = \int_\Omega \widetilde{F}(Dw) + fw\, dx.$$

Hence, hereafter we suppose $F(z) \geqslant F(0) = 0$. Finally, recalling Proposition 2.3(ii), we know that F is strictly convex and

$$C^{-1}|z|^p - C \leqslant F(z) \leqslant C(|z|^q + 1), \quad |DF(z)| \leqslant C(|z|^{q-1} + 1), \tag{3-12}$$

so that u is the unique minimizer locally, with respect to its own boundary values. We split the proof into several steps.

Step 1: approximating problems. Fix $\varphi \in C_c^\infty(B_1, [0, +\infty))$ such that $\|\varphi\|_1 = 1$ and let $\varphi_\varepsilon(x) = \varepsilon^{-N}\varphi(x/\varepsilon)$. For $B \Subset \Omega$ and $n \in \mathbb{N}$, let $f_n = f * \varphi_{1/n}$ and, for $\varepsilon_n, \mu_n \to 0^+$ to be chosen,

$$J_n(w) = \int_B F * \varphi_{\varepsilon_n}(Dw) + \frac{\mu_n}{2}|Dw|^2 + f_n\, w\, dx.$$

Set

$$\operatorname{Lip}_\psi(B) = \{w \in \operatorname{Lip}(\bar{B}) : w = \psi \text{ on } \partial B\}.$$

According to [Stampacchia 1963, Theorem 9.2], there is a solution $v_n \in \operatorname{Lip}_{u*\varphi_{1/n}}(B)$ of

$$J_n(v_n) = \inf\{J_n(w) : w \in \operatorname{Lip}_{u*\varphi_{1/n}}(B)\},$$

since $u * \varphi_{1/n}$ is smooth on ∂B (thus satisfying the bounded slope condition) and also f_n is smooth. Moreover, for any *fixed n* there is no Lavrentiev gap for J_n; see [Bousquet et al. 2014, p. 5923]. Hence v_n also solves

$$J_n(v_n) = \inf\{J_n(w) : w \in u * \varphi_{1/n} + W_0^{1,p}(B)\}.$$

Step 2: determining the parameters. For any choice of ε_n, μ_n, the integrand

$$F_n(z) := F * \varphi_{\varepsilon_n}(z) + \frac{\mu_n}{2}\,|z|^2$$

is μ_n-uniformly convex; hence [Bousquet and Brasco 2016, Theorem 4.1] ensures the existence of constants A_n (depending only on B and $\|f_n\|_\infty$, as well as on the regularity of $u * \varphi_{1/n}$, but not on ε_n, μ_n), such that

$$\operatorname{Lip}(v_n) \leqslant \frac{A_n}{\mu_n} =: L_n. \tag{3-13}$$

Without loss of generality, we can assume $L_n \geqslant 1$. We first choose $\mu_n \downarrow 0$ so that

$$\lim_n \mu_n \int_B |Du * \varphi_{1/n}|^2\, dx = 0, \quad \lim_n \mu_n^{p-1} A_n^{2-p} = 0, \tag{3-14}$$

and observe that L_n is independent of ε_n. Then we choose ε_n: define the numbers

$$M_n = 1 + \sup_B |Du * \varphi_{1/n}| + L_n.$$

Since $\varphi_\varepsilon * F \to F$ in $C^1_{\mathrm{loc}}(\mathbb{R}^N)$ as $\varepsilon \downarrow 0$, we can pick $(\varepsilon_n) \subseteq (0, 1)$, $\varepsilon_n \downarrow 0$, so that

$$\|F * \varphi_{\varepsilon_n} - F\|_{C^1(B_{M_n})} \leqslant \frac{1}{n}. \tag{3-15}$$

Clearly, it still holds

$$F_n \to F \quad \text{in } C^1_{\mathrm{loc}}(\mathbb{R}^N). \tag{3-16}$$

Step 3: the limsup inequality. Testing the minimality of v_n against the admissible function $u * \varphi_{1/n}$ gives

$$J_n(v_n) \leqslant J_n(u * \varphi_{1/n}).$$

Owing to $u * \varphi_{1/n} \to u$ in $W^{1,p}(B)$ and (3-14), one has

$$\overline{\lim_n}\, J_n(u * \varphi_{1/n}) = \int_B f u\, dx + \overline{\lim_n} \int_B F * \varphi_{\varepsilon_n}(Du * \varphi_{1/n})\, dx. \tag{3-17}$$

To estimate the last integral, we use (3-15) to get

$$\int_B F * \varphi_{\varepsilon_n}(Du * \varphi_{1/n})\, dx \leqslant \frac{|B|}{n} + \int_B F(Du * \varphi_{1/n})\, dx.$$

The vector-valued Jensen inequality then leads to

$$\begin{aligned} \overline{\lim_n} \int_B F * \varphi_{\varepsilon_n}(Du * \varphi_{1/n})\, dx &\leqslant \overline{\lim_n} \int_B F(Du * \varphi_{1/n})\, dx \\ &\leqslant \overline{\lim_n} \int_{(1+1/n)B} \varphi_{1/n} * F(Du)\, dx = \int_B F(Du)\, dx. \end{aligned}$$

Inserting the latter into (3-17) we conclude

$$\overline{\lim_n}\, J_n(v_n) \leqslant J(u). \tag{3-18}$$

Step 4: convergence of (v_n) to u. From (3-18) we have that $(J_n(v_n))$ is bounded. By Jensen's inequality $F \leqslant F * \varphi_{\varepsilon_n}$ so that, through (3-12), for some constant $C = C(N, F, B) > 0$ we have

$$\begin{aligned} J_n(v_n) &\geqslant \int_B F(Dv_n)\, dx + \int_B f_n\, v_n\, dx \\ &\geqslant \frac{\|Dv_n\|^p_{L^p(B)}}{C} - C - \|f_n\|_{W^{-1,p'}(B)} \|D(v_n - u * \varphi_{1/n})\|_{L^p(B)} + \int_B f_n\, u * \varphi_{1/n}\, dx \\ &\geqslant \frac{\|Dv_n\|^p_{L^p(B)}}{C} - C - \|f_n\|_{W^{-1,p'}(B)} (\|Dv_n\|_{L^p(B)} + \|Du * \varphi_{1/n}\|_{L^p(B)}) + \int_B f_n\, u * \varphi_{1/n}\, dx. \end{aligned}$$

Since $f_n \to f$ in $W^{-1,p'}(B)$ and $u * \varphi_{1/n} \to u$ in $W^{1,p}(B)$, by (3-18) we deduce that (Dv_n) is bounded in $L^p(B)$. Moreover, by Poincaré's inequality,

$$\begin{aligned}\|v_n\|_{L^p(B)} &\leqslant \|v_n - u * \varphi_{1/n}\|_{L^p(B)} + \|u * \varphi_{1/n}\|_{L^p(B)} \\ &\leqslant C(\|D(v_n - u * \varphi_{1/n})\|_{L^p(B)} + \|u * \varphi_{1/n}\|_{L^p(B)}) \\ &\leqslant C(\|Dv_n\|_{L^p(B)} + \|u * \varphi_{1/n}\|_{W^{1,p}(B)}),\end{aligned}$$

so that (v_n) is bounded in $L^p(B)$ as well. Therefore the sequence (v_n) is bounded in $W^{1,p}(B)$, and hence possesses a (not relabeled) subsequence weakly convergent to some $v \in W^{1,p}(B)$; actually, it is readily checked that $v \in u + W_0^{1,p}(B)$. We claim that

$$\lim_n \mu_n \int_B |Dv_n|^2 \, dx = 0.$$

Indeed, this is obvious by Hölder's inequality when $p \geqslant 2$, while if $p < 2$ we use (3-13) and (3-14) to infer

$$\mu_n \int_B |Dv_n|^2 \, dx \leqslant \mu_n L_n^{2-p} \int_B |Dv_n|^p \, dx \leqslant \mu_n^{p-1} A_n^{2-p} \int_B |Dv_n|^p \, dx \to 0,$$

where we used the boundedness of (Dv_n) in $L^p(B)$. Thus

$$\lim_n \int_B \frac{\mu_n}{2} |Dv_n|^2 + f_n \, v_n \, dx = \int_B f \, v \, dx. \tag{3-19}$$

The functional

$$w \mapsto \int_B F(Dw) \, dx$$

is weakly lower semicontinuous in $W^{1,p}(B)$, whence (again by the Jensen inequality)

$$\begin{aligned} J(v) &\leqslant \varliminf_n \left[\int_B F(Dv_n) \, dx + \int_B f \, v_n \, dx \right] \\ &\leqslant \varliminf_n \left[\int_B F * \varphi_{\varepsilon_n}(Dv_n) \, dx + \frac{\mu_n}{2} \int_B |Dv_n|^2 + f_n \, v_n \, dx \right] = \varliminf_n J_n(v_n). \end{aligned} \tag{3-20}$$

Coupling the latter with (3-18) we get $J(v) \leqslant J(u)$, implying $v = u$ by the strict convexity of J. In particular we obtain, up to subsequences,

$$Dv_n \rightharpoonup Du \quad \text{in } L^p(B), \tag{3-21}$$

and from (3-20), (3-18) we infer $J_n(v_n) \to J(u)$. Subtracting (3-19) we get

$$\int_B F * \varphi_{\varepsilon_n}(Dv_n) \, dx \to \int_B F(Du) \, dx,$$

which, thanks to (3-15), implies

$$\int_B F(Dv_n) \, dx \to \int_B F(Du) \, dx. \tag{3-22}$$

<u>Step 5</u>: uniform Sobolev bound on $DF(Dv_n)$. By Proposition 2.3(iii), and the beginning of the proof of Theorem 3.2, F_n satisfies (1-9) with the same constant K; since $F_n \in C^3(\mathbb{R}^N)$, (1-9) actually holds

everywhere. Moreover, standard regularity theory ensures that $v_n \in C^2(B)$, so we can apply Theorem 3.2, and in particular (3-4), to obtain, for any $\theta \in (0, 2]$ and $r = \frac{1}{2}, 1$,

$$\|DF_n(Dv_n)\|_{W^{1,2}(rB/2)} \leqslant C\,(\|f_n\|_{L^2(rB)} + \|DF_n(Dv_n)\|_{L^\theta(rB)}). \tag{3-23}$$

The first term on the right is clearly bounded by a multiple of $\|f\|_{L^2(B)}$. For the second one, we let p, q be given in (3-12) and choose $\bar\theta = \min\{p/(q-1), 1\}$. By (3-12) we get

$$|DF(z)|^{\bar\theta} \leqslant C\,(|z|^{\bar\theta\,(q-1)} + 1) \leqslant C\,(|z|^p + 1)^{\bar\theta\,(q-1)/p} \leqslant C(F(z) + 1)^{\bar\theta\,(q-1)/p}. \tag{3-24}$$

Using (3-15) and (3-24) we obtain

$$\begin{aligned}
&\int_B |DF_n(Dv_n)|^{\bar\theta}\,dx \\
&\qquad\leqslant \int_B |DF(Dv_n)|^{\bar\theta}\,dx + n^{-\bar\theta}\,|B| + \mu_n^{\bar\theta}\int_B |Dv_n|^{\bar\theta}\,dx \\
&\qquad\leqslant C\int_B (F(Dv_n)+1)^{\bar\theta\,(q-1)/p}\,dx + n^{-\bar\theta}\,|B| + \mu_n^{\bar\theta}\,\|Dv_n\|_{L^p(B)}^{\bar\theta}|B|^{1-\bar\theta/p} \\
&\qquad\leqslant C\,|B|^{1-\bar\theta\,(q-1)/p}\biggl(\int_B (F(Dv_n)+1)\,dx\biggr)^{\bar\theta\,(q-1)/p} + n^{-\bar\theta}\,|B| + \mu_n^{\bar\theta}\,\|Dv_n\|_{L^p(B)}^{\bar\theta}|B|^{1-\bar\theta/p}.
\end{aligned} \tag{3-25}$$

The first integral is bounded by (3-22) and the remaining terms vanish when $n \to \infty$, so

$$\overline{\lim_n}\,\|DF_n(Dv_n)\|_{L^{\bar\theta}(B)} \leqslant C\biggl(\int_B (F(Du)+1)\,dx\biggr)^{(q-1)/p}. \tag{3-26}$$

Thanks to (3-23) for $r = 1$, (3-26) implies the Sobolev bound

$$\overline{\lim_n}\,\|DF_n(Dv_n)\|_{W^{1,2}(B/2)} \leqslant C. \tag{3-27}$$

Step 6: passage to the limit. Let $B' = \frac{1}{2}B$ and

$$V_n = DF_n(Dv_n).$$

Thanks to (3-27), (V_n) is bounded in $W^{1,2}(B')$; hence we can pick a subsequence satisfying

$$V_n \to V \quad \text{weakly in } W^{1,2}(B'), \text{ strongly in } L^2(B'), \text{ and pointwise a.e. in } B', \tag{3-28}$$

for a suitable $V \in W^{1,2}(B')$.

Each F_n is strictly convex and superlinear by construction; hence DF_n is a homeomorphism of $\mathbb{R}^N$. Moreover, by (3-16) we know that $DF_n \to DF$ locally uniformly. According to a theorem by Arens (see [Dijkstra 2005] for a modern exposition), this implies that $DF_n^{-1} \to DF^{-1}$ locally uniformly. Since $V_n \to V$ pointwise a. e., we infer that

$$Dv_n = DF_n^{-1}(V_n) \to DF^{-1}(V) \quad \text{pointwise a.e.},$$

which, coupled with (3-21), allows the identification $Du = DF^{-1}(V)$. Therefore, $V_n \to DF(Du)$ in B' in all the senses prescribed in (3-28).

By considering $4B$ instead of B, estimate (3-9) follows from (3-23) and (3-26), as long as $4B \subseteq \Omega$. Let $B'' = \frac{1}{4}B$. By Lebesgue's dominated convergence theorem $\|V_n\|_{L^\theta(B')} \to \|V\|_{L^\theta(B')}$; hence, exploiting also the lower semicontinuity of the $W^{1,2}(B'')$ norm, we can pass to the limit in (3-23) with $r = \frac{1}{2}$. Again considering $4B$ instead of B yields (3-10). Finally, the validity of (3-11) can be checked only on balls B such that $4B \subseteq \Omega$, by a standard partition of unity argument. If B is such a ball, we can pass to the limit in the Euler Lagrange equations for the approximating problems constructed as before in $4B$, and since $DF_n(Dv_n) \to DF(Du)$, $f_n \to f$ strongly in $L^2(B)$, we get (3-11). □

Remark 3.4. The previous theorem has an immediate consequence. The class of q.u.c. functionals is a (proper) subclass of the so-called *functionals with* (p, q)*-growth*, i.e., those obeying (2-5), (2-6). For example, the integrand

$$F(z) = |z_1|^p + |z_2|^q, \quad z = (z_1, z_2) \in \mathbb{R}^2, \quad p, q > 1,$$

is of (p, q)-growth but not q.u.c., even if $p = q$. Given a local minimizer of a convex functional of *general* (p, q)-growth, the a priori regularity on Du is just $Du \in L^p_{\mathrm{loc}}(\Omega)$, and the first step towards higher regularity is showing that actually $Du \in L^q_{\mathrm{loc}}(\Omega)$. For $2 \leqslant p \leqslant N$, this is to be expected only when $q < p\,(N+2)/N$; see [Giaquinta 1987; Esposito et al. 1999]. For the subclass of q.u.c. integrands, from $DF(Du) \in W^{1,2}_{\mathrm{loc}}(\Omega)$, we infer by Sobolev's embedding that $DF(Du) \in L^{2^*}_{\mathrm{loc}}(\Omega)$. Since $|DF(z)| \gtrsim |z|^{p-1}$, it holds $Du \in L^{2^*(p-1)}_{\mathrm{loc}}(\Omega)$; hence for such class of integrands we obtain the condition $q \leqslant 2^*\,(p-1)$, which gives a larger range if $p \geqslant 2$. This range may not be optimal in the q.u.c. class, but on one hand it shows the advantages of considering the stress field instead of (1-5), while on the other hand it holds for any $f \in L^2_{\mathrm{loc}}(\Omega) \cap W^{-1,p'}(\Omega)$. It is quite possible, in light of the results of [Beck and Mingione 2020], that, for f having a sufficiently high degree of summability, minimizers for q.u.c. integrands are automatically Lipschitz continuous, regardless of the largeness of the ratio q/p, a fact that, if true, would bypass completely the higher integrability issue for the gradient in the q.u.c. class. This is actually the case for functionals with Uhlenbeck structure; see [Cianchi and Mazya 2011].

3C. *Examples.*

Example 3.5 (on the assumption $F \in W^{2,1}$). In this example we show that, in order to obtain Sobolev regularity of $DF(Du)$, it is not sufficient to require that condition (1-9) holds at almost every point, but that Sobolev regularity of DF is a necessary assumption.

Let $N = 2$. For any ball $B \Subset \{(x, y) \in \mathbb{R}^2 : x > 0\}$, consider the smooth function

$$u(x, y) = \arctan(y/x).$$

We claim that, for any (not necessarily convex) C^2 function $F : \mathbb{R} \to \mathbb{R}$, u solves

$$\operatorname{div} DF(Du) = 0 \quad \text{in } B, \tag{3-29}$$

where here and in what follows we make the identification $F(z) = F(|z|)$. Letting $z = (x, y)$, $z^\perp = (-y, x)$, it holds

$$Du(z) = \frac{z^\perp}{|z|^2}, \quad D^2u(z) = \frac{1}{|z|^4}\begin{pmatrix} 2xy & y^2-x^2 \\ y^2-x^2 & -2xy \end{pmatrix},$$

while

$$DF(w) = F'(|w|)\frac{w}{|w|}, \quad D^2F(w) = F''(|w|)\frac{w}{|w|}\otimes\frac{w}{|w|} + \frac{F'(|w|)}{|w|}\Big(I - \frac{w}{|w|}\otimes\frac{w}{|w|}\Big).$$

An elementary computation then yields

$$D^2F(Du)\,D^2u(z) = \frac{1}{|z|^4}\begin{pmatrix} xy(F''(1/|z|)+F'(1/|z|)|z|) & F''(1/|z|)y^2-F'(1/|z|)|z|x^2 \\ F'(1/|z|)|z|y^2-F''(1/|z|)x^2 & -xy(F''(1/|z|)+F'(1/|z|)|z|)\end{pmatrix},$$

which has zero trace, proving the claim.

Now, let $h : [0, 1] \to [0, 1]$ denote the Cantor function. By abuse of notation, we still denote by h its extension to the whole $\mathbb{R}$ defined as

$$h(t) = k + h(t-k) \quad \text{if } k \leqslant t < k+1,\ k \in \mathbb{Z},$$

and we also denote by $\mathcal{C}$ the periodic extension of the Cantor set to the whole $\mathbb{R}$. Consider

$$F(|w|) := \frac{|w|^2}{2} + H(|w|), \qquad H(t) := \int_0^t h(\tau)\,d\tau,$$

which is a strictly convex C^1 function with quadratic growth. Clearly, F can be approximated in C^1 by a sequence $\{F_n\}$ of smooth radial functions, so that we can pass to the limit into the corresponding weak formulations of (3-29) to obtain that u solves

$$\operatorname{div}(DF(Du)) = 0 \quad \text{weakly in } B.$$

However,

$$DF(Du(z)) = \frac{z^\perp}{|z|^2} + h(|z|^{-1})\,\frac{z^\perp}{|z|}$$

is not even absolutely continuous in B, since its distributional derivative has a Cantor part concentrated on $\{z : 1/|z| \in \mathcal{C}\}$, which has zero measure.

Since $h'(t) = 0$ in the classical sense for a.e. $t \in \mathbb{R}$, it is readily verified that F obeys (1-9) with $K = 1$ at every point of

$$\mathbb{R}^N \setminus \mathcal{C}_{\text{rad}}, \qquad \mathcal{C}_{\text{rad}} = \{z \in \mathbb{R}^N : |z| \in \mathcal{C}\},$$

thus almost everywhere. Notice that DF is of bounded variation but does not belong to $W^{1,1}_{\text{loc}}(\mathbb{R}^N)$, since its derivative has a Cantor part concentrated on $\mathcal{C}_{\text{rad}}$.

Example 3.6 (Uhlenbeck structure). For divergence form equations having the Uhlenbeck structure

$$\operatorname{div}(a(|Du|)\,Du) = f \tag{3-30}$$

we recover the local regularity result of [Cianchi and Mazya 2018, Theorem 2.1], under the additional assumption $f \in W^{-1,p'}(\Omega)$ (see the second point in Remark 1.3 in this respect). Here, the exponent p is related to the function a as follows. Define F by

$$F(z) = \int_0^{|z|} t\,a(t)\,dt$$

to get $DF(z) = a(|z|)z$. If $a \in C^1(0, +\infty)$, it holds

$$D^2F(z) = a(|z|)\, I + |z|\, a'(|z|)\, \frac{z}{|z|} \otimes \frac{z}{|z|},$$

possessing the eigenvector $z/|z|$ with eigenvalue $a(|z|) + |z|\, a'(|z|)$, while its orthogonal eigenspace is relative to the unique eigenvalue $a(|z|)$. The equation is elliptic if and only if both eigenvalues are nonnegative, and in order to bound the ratio between them we look at

$$K = \sup_{t>0} \max\left\{ \frac{a(t)}{a(t) + ta'(t)}, \frac{a(t) + ta'(t)}{a(t)} \right\}.$$

It is readily checked that if

$$i_a = \inf_{t>0} \frac{t\, a'(t)}{a(t)}, \qquad s_a = \sup_{t>0} \frac{t\, a'(t)}{a(t)},$$

then

$$K = \max\left\{ \frac{1}{1 + i_a}, 1 + s_a \right\},$$

so that the q.u. convexity condition (1-9) is equivalent to the common requirement

$$-1 < i_a \leqslant s_a < \infty, \tag{3-31}$$

which is the one used, e.g., in [Cianchi and Mazya 2018]. In our framework, the exponent p is given by Proposition 2.3 and turns out to be

$$p = 1 + \frac{1}{K} = \min\left\{ 2 + i_a, \frac{s_a + 2}{s_a + 1} \right\}.$$

In the model case $a(z) = |z|^{p-2}$, which corresponds to the p-Poisson equation, we directly have $i_a = s_a = p - 2$, so that the previous exponent is actually p. In the general case $i_a < i_s$, the exponent p can be improved, since it holds (see [Cianchi and Mazya 2011, Proposition 2.15])

$$F(z) \geqslant a(1) \frac{|z|^{2+i_a}}{2 + i_a}, \qquad |z| \geqslant 1,$$

so that one can take $p = 2 + i_a > 1$ by (3-31) (see the third point in Remark 1.3). Indeed, for $p = 2 + i_a$ and $f \in W^{-1,p'}(\Omega)$, J is also coercive on $W^{1,p}(\Omega)$ when supplemented with reasonable boundary conditions. The variational treatment of (3-30) in standard Sobolev spaces is thus justified if one is not looking for optimal rearrangement invariant estimates.

Example 3.7 (anisotropic examples). In [Ciraolo et al. 2020; Antonini et al. 2022; Cozzi et al. 2014; 2016] anisotropic equations whose principal part arises as the Euler–Lagrange equation of

$$\int_\Omega G(H(Du))\, dx$$

are considered, where $H \in C^2(\mathbb{R}^N \setminus \{0\}, \mathbb{R}_+)$ is a convex, positively 1-homogeneous function and $G \in C^2(\mathbb{R}_+, \mathbb{R}_+)$ is an increasing, strictly convex function of p-growth. Clearly, H is fully determined by the unit “ball”

$$B_H = \{z \in \mathbb{R}^N : H(z) < 1\} \ni 0,$$

and any open, bounded, convex B with $0 \in B$ will uniquely determine such an H through its Minkowski functional. Notice that in general H may not even be a norm, due to the possible lack of symmetry of B_H.

The more general ellipticity assumption in the cited works reads as follows: H is said to be *uniformly convex* if the principal curvatures of ∂B_H are bounded from below by a positive constant. From [Cozzi et al. 2014, Appendix A], the uniform ellipticity of H amounts to

$$(H(z)\, D^2 H(z)\, \eta, \eta) \geqslant \delta |\eta|^2 \quad \text{for all } z \in \mathbb{R}^N \setminus \{0\},\ \ \eta \in DH(z)^{\perp},$$

for some $\delta > 0$ (which is actually equivalent to the same type of inequality for $\eta \in z^{\perp}$). Under this assumption various results can be proved, and in particular the Sobolev regularity of the associated stress field is treated in [Ciraolo et al. 2020; Antonini et al. 2022] as a stepping-stone to more general results. Here we show that anisotropic functionals of this kind fall within our general framework of q.u.c. functionals.

To this end, set $F(z) = G(H(z))$ so that

$$D^2 F(z) = G''(H(z))\, DH(z) \otimes DH(z) + \frac{G'(H(z))}{H(z)}\, H(z)\, D^2 H(z)$$

for $z \neq 0$, and notice that both

$$DH(z) \otimes DH(z) \quad \text{and} \quad H(z)\, D^2 H(z)$$

are 0-homogeneous. Inspecting the proof of [Cozzi et al. 2014, Appendix A], we see that for any $z, \xi \in \mathbb{R}^N \setminus \{0\}$ it either holds

$$(DH(z) \otimes DH(z)\, \xi, \xi) \geqslant \lambda_1\, |\xi|^2$$

or

$$(H(z)\, D^2 H(z)\, \xi, \xi) \geqslant \lambda_2\, |\xi|^2$$

(with $\lambda_i = \lambda_i(H) > 0$), while altogether

$$(H(z)\, D^2 H(z)\, \xi, \xi) \leqslant \Lambda\, |\xi|^2, \quad (DH(z) \otimes DH(z)\, \xi, \xi) \leqslant \Lambda\, |\xi|^2$$

for some $\Lambda = \Lambda(H)$. It follows that the minimum eigenvalue of $D^2 F(z)$ is bounded from below by

$$\min\left\{\lambda_1\, G''(H(z)),\, \lambda_2\, \frac{G'(H(z))}{H(z)}\right\},$$

while its maximum eigenvalue is bounded from above by

$$\Lambda\left(G''(H(z)) + \frac{G'(H(z))}{H(z)}\right).$$

Therefore F is K-q.u.c. for

$$K = \frac{\Lambda}{\min\{\lambda_1, \lambda_2\}} \sup_{t \in \mathbb{R}_+} \max\left\{1 + \frac{G'(t)}{G''(t)\, t},\, 1 + \frac{G''(t)\, t}{G'(t)}\right\}.$$

The finiteness of the latter is compatible with the standard Uhlenbeck example: for $a(t) = G'(t)/t$, a straightforward calculation shows that

$$\sup_{t\in\mathbb{R}_+} \max\left\{ \frac{G'(t)}{G''(t)\,t}, \frac{G''(t)\,t}{G'(t)} \right\} < +\infty \quad \Longleftrightarrow \quad -1 < i_a \leqslant s_a < +\infty.$$

Example 3.8 (nonstandard anisotropic growth). It is straightforward to check that if F_1 and F_2 are q.u.c. C^1 functions satisfying (1-9) with constants K_1 and K_2 respectively, then $F_1 + F_2$ is $\max\{K_1, K_2\}$-q.u.c. This observation allows to consider anisotropic Euler–Lagrange equations arising from integrands of the form

$$F(z) = \sum_{i=1}^{M} |A_i(z - z_i)|^{p_i}, \qquad z_i \in \mathbb{R}^N,\ A_i \geqslant I,\ p_i > 1 \ \text{ for } i = 1, \dots, M.$$

As a more common example of anisotropic functionals, consider the weak solution of the Dirichlet problem

$$\begin{cases} \Delta_p u + D_1(|D_1 u|^{q-2} D_1 u) = f & \text{in } B, \\ u = 0 & \text{on } \partial B \end{cases}$$

for a ball B, $f \in L^\infty(B)$ and $q > p > 2$. This corresponds to the unique minimizer $u \in W_0^{1,p}(B)$ of

$$J(w) = \int_B \frac{1}{p}|Dw|^p + \frac{1}{q}|D_1 w|^q + f\,w\,dx,$$

which is globally Lipschitz continuous, since

$$F(z) = \frac{1}{p}|z|^p + \frac{1}{q}|z_1|^q$$

is uniformly p-convex outside B_1 (see [Bousquet and Brasco 2016]), i.e.,

$$(D^2F(z)\,\xi, \xi) \geqslant c_p(1 + |z|^2)^{(p-2)/2}|\xi|^2, \quad |z| \geqslant 1.$$

Moreover,

$$D^2F(z) = |z|^{p-2}\left(I + (p-2)\frac{z}{|z|} \otimes \frac{z}{|z|}\right) + (q-1)|z_1|^{q-2}\, e_1 \otimes e_1,$$

so that the minimum and maximum eigenvalues of $D^2F(z)$ satisfy

$$\lambda_{\min}(z) \geqslant |z|^{p-2}, \quad \lambda_{\max}(z) \leqslant (p-1)\,|z|^{p-2} + (q-1)\,|z|^{q-2}.$$

Notice that the integrand F is *not* q.u.c. globally on $\mathbb{R}^N$, but on the range of Du we have $|z|^{q-2} \leqslant C\,|z|^{p-2}$ for C depending on $\operatorname{Lip} u$, leading to

$$\lambda_{\max}(z)/\lambda_{\min}(z) \leqslant p - 1 + C\,(q-1).$$

Hence Theorem 3.3 applies thanks to the local nature of (1-9) described in Lemma 2.4. Notice the role of the assumption $f \in L^\infty(B)$ (which could be weakened, but not down to $L^2(B)$, see [Beck and Mingione 2020]) and of the smooth boundary condition $u \in W_0^{1,p}(B)$: they provide the Lipschitz regularity of u, which in turn allows to employ Lemma 2.4 and to consider u as a minimizer of a functional with q.u.c. integrand on the whole $\mathbb{R}^N$.

4. Applications

4A. ***Cordes-type conditions.*** We start with a generalization of (2-2). In this section, given a matrix field $M : \Omega \to \mathbb{R}^N \otimes \mathbb{R}^N$, its L^m-norm will be computed with respect to the Frobenius norm of $M = (m_{ij})$, i.e.,

$$\|M\|_m = \left(\int |M(x)|_2^m \, dx \right)^{1/m}, \quad |M(x)|_2^2 = \sum_{i,j=1}^{N} |m_{ij}(x)|^2.$$

The proofs of this section make use of some tools from harmonic analysis; for an introductory exposition on this topic, the reader can consult [Duoandikoetxea 2001].

Lemma 4.1. *Let $V \in C_c^1(\mathbb{R}^N; \mathbb{R}^N)$ and, for $m > 1$, set $\hat{m} = \max\{m, m/(m-1)\} \geqslant 2$. Then*

$$\|DV\|_m \leqslant N^2(\hat{m}-1)(\|\operatorname{div} V\|_m + \|\operatorname{curl} V\|_m). \tag{4-1}$$

Proof. Let R_j be the Riesz transform, defined as the Fourier multiplier with symbol $-i\,\xi_j/|\xi|$; see [Duoandikoetxea 2001, p. 76]. Then it holds

$$D_h V_k = -R_h\, R_k \operatorname{div} V - \sum_{j=1}^{N} R_h\, R_j \operatorname{curl}_{kj} V, \tag{4-2}$$

which follows, taking the Fourier transform (denoted by $g \mapsto \hat{g}$), from the identity

$$i\,\xi_h\,\widehat{V_k} = \frac{\xi_h\,\xi_k}{|\xi|^2} \sum_{j=1}^{N} i\,\xi_j\,\widehat{V_j} + \sum_{j=1}^{N} \frac{\xi_h\,\xi_j}{|\xi|^2}(i\,\xi_j\,\widehat{V_k} - i\,\xi_k\,\widehat{V_j}).$$

The second-order Riesz transform has $L^m - L^m$ norm $\hat{m}-1$ on diagonal terms and $(\hat{m}-1)/2$ on off-diagonal terms, i.e.,

$$\|R_h\, R_k\, g\|_m \leqslant \frac{\hat{m}-1}{2}\|g\|_m, \quad h \neq k,$$
$$\|R_h^2\, g\|_m \leqslant (\hat{m}-1)\|g\|_m;$$

see [Bañuelos and Méndez-Hernández 2003, Theorem 2.4]. Thus from (4-2) we get

$$\|D_h V_k\|_m \leqslant (\hat{m}-1)\left(\|\operatorname{div} V\|_m + \sum_{j=1}^{N} \|\operatorname{curl}_{kj} V\|_m\right).$$

We sum over $k = 1, \dots, N$ and use the Hölder inequality to get

$$\sum_{k,j=1}^{N} \|\operatorname{curl}_{kj} V\|_m \leqslant N^{2(1-1/m)} \left(\int \sum_{k,j=1}^{N} |\operatorname{curl}_{kj} V|^m \right)^{1/m}.$$

Since

$$\sum_{k,j=1}^{N} |\operatorname{curl}_{kj} V|^m \leqslant N^{2-m} \left(\sum_{k,j=1}^{N} |\operatorname{curl}_{kj} V|^2 \right)^{m/2},$$

we obtain

$$\sum_{k,j=1}^{N} \|\operatorname{curl}_{kj} V\|_m \leqslant N\, \|\operatorname{curl} V\|_m.$$

Summing also over $h = 1, \dots, N$ we finally get

$$\|DV\|_m \leqslant \sum_{h,k=1}^{N} \|D_h V_k\|_m \leqslant (\hat{m}-1)\, N^2(\|\operatorname{div} V\|_m + \|\operatorname{curl} V\|_m). \qquad \square$$

Remark 4.2. Estimate (4-1) is very rough in its dependence on N. It is a common feature of L^m-bounds on Riesz transform that they do not depend on the dimension of the euclidean space. Up to our knowledge, optimal L^m estimates for the operator $D\operatorname{curl}^{-1}$ (let alone for the resolvent operator of the div-curl system) are not known. It is also not optimal as $m \to 2$; compare it with (2-2) in the case $m = 2$.

Theorem 4.3. *Let F obey the assumptions of Theorem 3.3, in particular* (1-9) *with a constant $K \geqslant 1$ and p, q given in Proposition 2.3. Let furthermore $m > 1$ and $f \in L^m(\Omega) \cap W^{-1,p'}(\Omega)$. Then any local minimizer $u \in W^{1,p}_{\mathrm{loc}}(\Omega)$ for J given in* (3-8) *is such that $DF(Du) \in W^{1,m}_{\mathrm{loc}}(\Omega)$ and satisfies the estimates*

$$\|DF(Du)\|_{L^m(B)} \leqslant C(1 + \|f\|_{L^m(2B)} + \|F(Du)\|_{L^1(2B)}^{(q-1)/p}), \tag{4-3}$$

$$\|DF(Du)\|_{W^{1,m}(B)} \leqslant C(\|f\|_{L^m(2B)} + \|DF(Du)\|_{L^m(2B)}) \tag{4-4}$$

in each of the following cases:

(1) $K \leqslant K_0$, *with $K_0 > 1$ depending on N and m, with $C = C(N, m, B)$.*

(2) $|m-2| \leqslant \delta_0$ *for $\delta_0 > 0$ depending on N and K, with $C = C(N, K, B)$.*

Proof. Given B such that $4B \Subset \Omega$, we follow the first four steps of the proof of Theorem 3.3 to find $u_n \in C^\infty(B)$, $f_n \in C^\infty(B)$, $F_n \in C^\infty(\mathbb{R}^N)$ and $v_n \in C^2(B)$ such that

(i) $u_n \to u$ in $W^{1,p}(B)$, $f_n \to f$ in $L^m(B) \cap W^{-1,p'}(B)$,

(ii) F_n is K-q.u.c.,

(iii) $v_n \rightharpoonup u$ in $W^{1,p}(B)$, $v_n \to u$ in $L^p(B)$, and v_n solves

$$\operatorname{div}(DF_n(Dv_n)) = f_n.$$

(iv) $\|F_n(Dv_n)\|_{L^1(B)} \to \|F(Du)\|_{L^1(B)}$.

We then proceed as in Theorem 3.2, in order to find a uniform bound for $DF_n(Dv_n)$ in $W^{1,m}(B')$, $B' = \frac{1}{2}B$. To simplify the notation, we omit for the moment the dependence of v, F, and f on n.

Let $V = DF(Dv)$ and observe that the assumptions of Lemma 3.1 hold true for the matrix $X = DV = D^2F(Dv)\, D^2 v$; hence, pointwise,

$$|\operatorname{curl} V|_2 \leqslant \sqrt{2}\,\mathsf{e}(K)\,|DV|_2, \qquad \mathsf{e}(K) = 1 - \frac{1}{K}. \tag{4-5}$$

Suppose $B = B_{2R}$ and, for any $R \leqslant r < s \leqslant 2R$, fix $\varphi \in C^\infty_c(B_s; [0, 1])$ as in (3-6).

We split the proof in two cases, according to the situations under consideration in the two statements. For the first assertion, we apply (4-1) to the field $W := \varphi V$ to get

$$\begin{aligned}\|\varphi\, DV\|_m &\leqslant \|DW\|_m + \|D\varphi \otimes V\|_m \\ &\leqslant N^2\,(\hat{m}-1)(\|\operatorname{div} W\|_m + \|\operatorname{curl} W\|_m) + \frac{C}{s-r}\|V\|_{L^m(B_s \setminus B_r)}\end{aligned}$$

for any $m > 1$. From

$$\operatorname{div} W = \varphi f + (V, D\varphi), \qquad \operatorname{curl} W = \varphi \operatorname{curl} V + V \wedge D\varphi$$

we thus infer

$$\|\varphi\, DV\|_m \leqslant N^2\,(\hat{m}-1)(\|\varphi \operatorname{curl} V\|_m + \|\varphi\, f\|_m) + \frac{C}{s-r}\|V\|_{L^m(B_s\setminus B_r)}. \tag{4-6}$$

We then let $K_0 = K_0(N, m) > 1$ satisfying

$$\sqrt{2}N^2(\hat{m}-1)\,\mathrm{e}(K_0) < 1 \tag{4-7}$$

so that, if $K \leqslant K_0$, the curl term in (4-6) can be reabsorbed on the left. Thanks to the properties (3-6) of φ we deduce

$$\|DV\|_{L^m(B_r)} \leqslant C\,\|f\|_{L^m(B_R)} + \frac{C}{s-r}\|V\|_{L^m(B_s\setminus B_r)}, \qquad R \leqslant r < s \leqslant 2R, \tag{4-8}$$

for a constant $C = C(N, m)$.

Regarding the second assertion, we consider the linear operator $T(f, G) = DV$, where V solves

$$\begin{cases}\operatorname{div} V = f,\\ \operatorname{curl} V = \sqrt{2}\,G,\end{cases} \qquad f \in C^\infty_c(\mathbb{R}^N),\ \ G \in C^\infty_c(\mathbb{R}^N; \mathbb{R}^N \wedge \mathbb{R}^N),$$

which, as already noted, is represented in terms of Riesz transforms as

$$T(f, G)_{kh} = -R_h\, R_k\, f - \sqrt{2}\sum_{j=1}^{N} R_h\, R_j\, G_{kj}.$$

Estimate (4-1) implies that T has an extension $T : X_m \to Y_m$, where

$$X_m := L^m(\mathbb{R}^N) \times L^m(\mathbb{R}^N; \mathbb{R}^N \wedge \mathbb{R}^N), \qquad Y_m := L^m(\mathbb{R}^N; \mathbb{R}^N \otimes \mathbb{R}^N),$$

with the norms

$$\|(f, G)\|_m = \left(\int |f|^m + |G|_2^m\, dx\right)^{1/m}, \qquad \|M\|_m = \left(\int |M|_2^m\, dx\right)^{1/m}$$

on X_m and Y_m, respectively. For the complex interpolation spaces it holds

$$[X_{m_1}, X_{m_2}]_\theta = X_m, \qquad \frac{1}{m} = \frac{1-\theta}{m_1} + \frac{\theta}{m_2}, \qquad \theta \in [0, 1],$$

with equality of norms, and the same holds for the Y_m. On the other hand, Lemma 2.2 ensures that, with respect to these norms,

$$\|T\|_{\mathcal{L}(X_2, Y_2)} = 1.$$

Fix $\bar{m}' < 2 < \bar{m}$. The Riesz–Thorin interpolation theorem [Duoandikoetxea 2001, Theorem 1.19] yields

$$\|T\|_{\mathcal{L}(X_m, Y_m)} \leqslant \|T\|^\theta_{\mathcal{L}(X_{\bar{m}}, Y_{\bar{m}})}, \qquad \frac{1}{m} = \frac{1-\theta}{2} + \frac{\theta}{\bar{m}},$$

for any $2 \leqslant m \leqslant \bar{m}$, and a similar estimate holds also for $\bar{m}' \leqslant m \leqslant 2$. We infer that there exist $\eta : [\bar{m}', \bar{m}] \to [0, +\infty)$ such that

$$\lim_{m\to 2} \eta(m) = 0 \tag{4-9}$$

and for all $m \in [\bar{m}', \bar{m}]$ it holds

$$\|T\|_{\mathcal{L}(X_m,Y_m)} \leqslant 1+\eta(m).$$

To complete the choice of δ_0 in the second case, we proceed as in the proof of (4-6), setting $W=\varphi V$ and using the previous operator norm estimate, to get, for $\eta=\eta(m)$,

$$\begin{aligned}
\|\varphi\, DV\|_m &\leqslant \|DW\|_m + \|V\otimes D\varphi\|_m \\
&\leqslant \|T(\operatorname{div} W, 2^{-1/2}\operatorname{curl} W)\|_m + \|V\otimes D\varphi\|_m \\
&\leqslant (1+\eta)\, \|(\operatorname{div} W, 2^{-1/2}\operatorname{curl} W)\|_m + \|V\otimes D\varphi\|_m \\
&\leqslant (1+\eta)\big(\|(\varphi f, \varphi\, 2^{-1/2}\operatorname{curl} V)\|_m + \|((V, D\varphi), V\wedge D\varphi)\|_m\big) + \|V\otimes D\varphi\|_m \\
&\leqslant \frac{1+\eta}{\sqrt{2}}\, \|\varphi \operatorname{curl} V\|_m + C\, \|\varphi f\|_m + \frac{C}{s-r}\, \|V\|_{L^m(B_s\setminus B_r)} \\
&\underset{(4\text{-}5)}{\leqslant} (1+\eta)\,\mathsf{e}(K)\, \|\varphi\, DV\|_m + C\, \|f\|_{L^m(B_R)} + \frac{C}{s-r}\, \|V\|_{L^m(B_s\setminus B_r)}.
\end{aligned}$$

Since $\mathsf{e}(K)<1$, thanks to (4-9) we can choose $\delta_0=\delta_0(K,N)$ in such a way that

$$(1+\eta(m))\,\mathsf{e}(K)<1 \quad \text{for all } m\in[2-\delta_0, 2+\delta_0],$$

which again gives estimate (4-8) with a constant $C=C(N,K)$.

Proceeding as in the proof of Theorem 3.2, we see that estimate (4-8) improves to

$$\|V\|_{W^{1,m}(B_R)} \leqslant C\, \|f\|_{L^m(B_{2R})} + C_{R,\theta}\|V\|_{L^\theta(B_{2R})}, \quad \theta\in(0,m], \tag{4-10}$$

which therefore holds uniformly for all $V_n = DF_n(Dv_n)$ constructed at the beginning. If p, q are given in Proposition 2.3, point (ii), we set $\bar{\theta}=\min\{p/(q-1), m\}$ and proceed as in (3-25) to get a uniform bound on $\|V_n\|_{L^{\bar{\theta}}(B_{2R})}$. Thanks to (4-10), the latter in turn implies the compactness of the V_n in $L^m(B_R)$. The rest of the proof of Theorem 3.3 follows verbatim, providing estimates (4-3) and (4-4) in both the stated cases. We omit the details. □

Remark 4.4. Closely inspecting the previous proof yields the following asymptotic estimates. The constant K_0 goes to 1 as $m\to\infty$ or $m\to 1$. Similarly, (4-3), (4-4) hold true for $m\in(a(K), b(K))$ with $a(K)\to 1$ and $b(K)\to+\infty$ as $K\to 1$.

We made no attempt to obtain optimal estimates for K_0 and δ_0 as $m\to 2$, mainly due to the roughness of estimate (4-1) outlined in Remark 4.2. Thus, we do not recover Theorem 3.3 by simply letting $m\to 2$ in the previous statement.

4B. ***On the $C^{p'}$ conjecture.*** An immediate corollary of the Cordes condition proved in the previous section is the following one.

Corollary 4.5. *Any weak solution $u\in W^{1,p}_{\mathrm{loc}}(\Omega)$ of $\Delta_p u = f\in L^\infty_{\mathrm{loc}}(\Omega)$ belongs to $C^{2-\alpha}(\Omega)$, where $\alpha=\alpha(N,p)\leqslant C(N)\,|p-2|$, provided $|p-2|<1/(2N^3)$.*

Proof. Recall that $z\mapsto |z|^p$ is K-q.u.c. with constant

$$K_p=\max\{p-1, 1/(p-1)\}=\begin{cases} p-1 & \text{if } p\geqslant 2,\\ 1/(p-1) & \text{if } p\in(1,2).\end{cases}$$

We apply the Cordes estimates of the previous theorem, so that $|Du|^{p-2}\,Du \in W^{1,m}_{\mathrm{loc}}(\Omega)$ for any $m \geqslant 2$ such that (see (4-7))

$$\sqrt{2}N^2\,(m-1)(1-1/K_p) < 1. \tag{4-11}$$

Given p, we let

$$m_p = \frac{1}{2N^2\,|p-2|}.$$

It is readily checked that, for all p such that $|p-2| < 1/(2N^3)$, inequality (4-11) holds for m_p and moreover $m_p > N$. By the Morrey embedding we thus have that $|Du|^{p-2}\,Du \in C^{1-N/m_p}(B)$.

The map $\Psi : \mathbb{R}^N \to \mathbb{R}^N$, defined as

$$\Psi(y) = \begin{cases} |y|^{(2-p)/(p-1)}\,y & \text{if } y \neq 0, \\ 0 & \text{if } y = 0 \end{cases} \tag{4-12}$$

is the inverse of the similarly defined map $z \mapsto |z|^{p-2}z$, for which the well-known inequalities (see the last chapter of [Lindqvist 2019])

$$(|z_1|^{p-2}z_1 - |z_2|^{p-2}z_2, z_1 - z_2) \geqslant \begin{cases} 2^{2-p}\,|z_1 - z_2|^p & \text{if } p \geqslant 2, \\ (p-1)|z_1 - z_2|^2(1 + |z_1|^2 + |z_2|^2)^{(p-2)/2} & \text{if } 1 < p < 2 \end{cases}$$

hold true. By the Schwartz inequality we deduce

$$||z_1|^{p-2}z_1 - |z_2|^{p-2}z_2| \geqslant \begin{cases} 2^{2-p}\,|z_1 - z_2|^{p-1} & \text{if } p \geqslant 2, \\ c_M\,|z_1 - z_2| & \text{if } 1 < p < 2 \text{ and } |z_1| + |z_2| \leqslant M, \end{cases}$$

which means that Ψ is globally $1/(p-1)$-Hölder continuous if $p \geqslant 2$ and locally Lipschitz continuous if $1 < p < 2$. In the second case, we observe that $f \in L^\infty_{\mathrm{loc}}(\Omega)$ implies that $Du \in L^\infty_{\mathrm{loc}}(\Omega)$; hence Ψ is Lipschitz continuous on the range of Du. In both cases we thus have

$$Du \in C^{\alpha_p}(\Omega), \quad \alpha_p = \begin{cases} \dfrac{1}{p-1}\left(1 - \dfrac{N}{m_p}\right) = \dfrac{1 - 2N^3(p-2)}{p-1} & \text{if } p \geqslant 2, \\ 1 - 2N^3(2-p) & \text{if } 1 < p < 2, \end{cases}$$

giving the claim. □

Remark 4.6. A similar conclusion can be drawn for $W^{1,p}_{\mathrm{loc}}(\Omega)$ local minimizers of $J(\,\cdot\,,\Omega)$ when F is K-q.u.c. and $f \in L^\infty_{\mathrm{loc}}(\Omega)$. Indeed, DF is K^{N-1} quasiconformal; hence so is DF^{-1}. In particular, DF^{-1} is $1/K$-Hölder continuous and the α-Hölder regularity of $DF(Du)$ translates to α/K-Hölder regularity for Du. The dependence of the Hölder exponent of Du from K turns out to be $1 - c_N\,(K-1)$ for K sufficiently near 1.

Consider now a solution u of the inhomogeneous elliptic equation with Uhlenbeck structure

$$\operatorname{div}(a(|Du|)\,Du) = f \in L^m_{\mathrm{loc}}(\Omega), \quad m > 1, \tag{4-13}$$

where $a : (0, +\infty) \to (0, +\infty)$ is $C^1(0, +\infty)$ and satisfies the ellipticity condition

$$-1 < i_a \leqslant s_a < +\infty.$$

Solutions of (4-13) with f merely in $L^2_{\rm loc}(\Omega)$ are to be meant in a generalized sense and, as in [Cianchi and Mazya 2018], we will use the notion of *approximable solutions* (or SOLA, solution obtained as limit of approximation) for (4-13): u is an approximable solution for (4-13) in Ω if $a(|Du|)\,Du \in L^1_{\rm loc}(\Omega)$, (4-13) holds in the distributional sense in Ω, and there exists a sequence $(f_n) \subseteq C^\infty_c(\Omega)$ and corresponding weak solutions u_n of (4-13) in Ω with right-hand side f_n such that

$$f_n \to f \quad \text{in } L^m_{\rm loc}(\Omega), \qquad u_n \to u \text{ and } Du_n \to Du \text{ a.e. in } \Omega,$$

and

$$\lim_n \int_{\Omega'} a(|Du_n|)\,|Du_n|\,dx = \int_{\Omega'} a(|Du|)\,|Du|\,dx$$

for all $\Omega' \Subset \Omega$. Notice that u may fail to belong to $W^{1,1}_{\rm loc}(\Omega)$, but rather falls into the larger space

$$\mathcal{T}^{1,1}_{\rm loc}(\Omega) = \{v : T_k v \in W^{1,1}_{\rm loc}(\Omega) \text{ for all } k > 0\}, \qquad T_k v = \max\{-k, \min\{v, k\}\},$$

for which a pointwise notion of Du is well-defined almost everywhere. Whenever $f \in W^{-1,p'}(\Omega)$, with $p = 2 + i_a$ (see the discussion at the end of Example 3.6), weak and SOLA solutions coincide, and in particular u belongs to the relevant Orlicz–Sobolev space. For the existence and uniqueness theory of SOLA we refer to [Cianchi and Mazya 2017].

We say that $u \in \mathcal{T}^{1,1}_{\rm loc}(\Omega)$ is a cylindrical solution of (4-13) if there exists a point $x_0 \in \mathbb{R}^N$ and a k-dimensional vector subspace $V \subseteq \mathbb{R}^N$ with corresponding orthogonal projection $\pi_V : \mathbb{R}^N \to V$ such that

$$u(x) = v(|\pi_V(x - x_0)|)$$

for some $v : I \to \mathbb{R}$ with $I \subseteq [0, +\infty)$ open and $\Omega \subseteq \{x \in V : |x| \in I\} \times V^\perp$. In other terms, a cylindrical function *only* depends on the distance from some vector subspace.

In order to study the regularity properties of a cylindrical solution of (4-13), we first perform some straightforward reductions. It is clear that we can assume

$$V = \{x \in \mathbb{R}^N : x_i \equiv 0 \text{ for all } i = k+1, \dots, N\}, \qquad \Omega = A \times \mathbb{R}^{N-k},$$

with $A \subseteq \mathbb{R}^k$ invariant by the action of the orthogonal group O_k on $\mathbb{R}^k$. Actually, by the structure of (4-13), we can directly suppose that $k = N$, reducing to the case of radial solutions on a radial (meaning, invariant by O_N) domain Ω.

Theorem 4.7. *Let $a \in C^1((0, +\infty); (0, +\infty))$ satisfy (3-31) and $f \in L^m(\Omega)$ for some $m > 1$, where Ω is a radial domain. If u is a radial approximable solution of*

$$\operatorname{div}(a(|Du|)\,Du) = f$$

in Ω then, for any $m > 1$ and B_R such that $B_{2R} \Subset \Omega$, it holds

$$\|a(|Du|)\,Du\|_{W^{1,m}(B_{R/2})} \leqslant C_{m,R}\big(\|f\|_{L^m(B_{2R})} + \|a(|Du|)\,Du\|_{L^1(B_{2R})}\big). \tag{4-14}$$

Proof. The field $V = a(|Du|)\,Du \in L^1_{\rm loc}(\Omega)$ is the pointwise limit of the fields $V_k = a(|DT_k u|)\,DT_k u$ for $k \to +\infty$, which satisfy

$$T \circ V_k = V_k \circ T \quad \text{for all } T \in O_N, \qquad |(V_k(x), x)| = |V(x)|\,|x|; \tag{4-15}$$

hence it satisfies these as well. We extend V and f as zero outside Ω (thus keeping the previous properties for V) and let $V_\varepsilon = V * \varphi_\varepsilon$, $f_\varepsilon = f * \varphi_\varepsilon$, where φ_ε is a standard radial convolution kernel supported in B_ε. We claim that V_ε obeys (4-15). By changing variables and using the radiality of φ_ε, it is readily checked that V_ε satisfies $T \circ V_\varepsilon = V_\varepsilon \circ T$ for all $T \in O_N$. Thus, in order to check the second condition in (4-15) it suffices to prove it at a point $x = r\, e_1$, where

$$V_\varepsilon(r\, e_1) = \int V(y)\, \varphi_\varepsilon(|r\, e_1 - y|)\, dy.$$

The integrand above is odd with respect to the reflections $y_k \mapsto -y_k$, $k = 2, \dots, N$, which implies that $V_\varepsilon(r\, e_1)$ is parallel to e_1, and this concludes the proof of (4-15) for V_ε. It follows that for

$$h_\varepsilon(x) := (V_\varepsilon(x), x/|x|^2) \in C^\infty(\mathbb{R}^N \setminus \{0\})$$

it holds, with a slight abuse of notation, $h_\varepsilon(x) = h_\varepsilon(|x|)$ and

$$V_\varepsilon(x) = h_\varepsilon(|x|)\, x. \tag{4-16}$$

Given a radial subdomain $\Omega' \Subset \Omega$ and using Fubini's theorem, we have

$$\begin{aligned}\int_{\Omega'} (V_\varepsilon, D\psi)\, dx &= \int_{\Omega'} (V, \varphi_\varepsilon * D\psi)\, dx = \int_{\Omega'} (V, D(\psi * \varphi_\varepsilon))\, dx \\ &= -\int_{\Omega'} f\, \psi * \varphi_\varepsilon\, dx = -\int_{\Omega'} f_\varepsilon\, \psi\, dx;\end{aligned}$$

thus V_ε satisfies $\operatorname{div} V_\varepsilon = f_\varepsilon$ weakly (and thus strongly) in Ω' for all sufficiently small $\varepsilon > 0$. From (4-16) we compute, for $x \neq 0$,

$$DV_\varepsilon(x) = h_\varepsilon(|x|)\, I + |x|\, h'_\varepsilon(|x|)\, \frac{x}{|x|} \otimes \frac{x}{|x|},$$

which is a symmetric matrix, so that

$$\operatorname{curl} V_\varepsilon = 0 \quad \text{in } \Omega'.$$

By the Poincaré lemma, for any $B_{2R} \subseteq \Omega'$ and sufficiently small $\varepsilon > 0$, we thus have $V_\varepsilon = Dv_\varepsilon$ for some $v_\varepsilon \in C^2(B_{2R})$, satisfying weakly $\Delta v_\varepsilon = f_\varepsilon$. For $R < r < s < 2\,R$ we choose cut-off functions as in (3-6) and suppose, without loss of generality, that v_ε has zero mean in B_s. By the standard Calderón–Zygmund estimates (see [Duoandikoetxea 2001, Theorem 5.1]) and Poincaré's inequality, it holds

$$\begin{aligned}\|D^2 v_\varepsilon\|^m_{L^m(B_r)} &\leqslant C_m \left(\|f_\varepsilon\|^m_{L^m(B_{2R})} + \frac{1}{(s-r)^m} \|Dv_\varepsilon\|^m_{L^m(B_s)} + \frac{1}{(s-r)^{2m}} \|v_\varepsilon\|^m_{L^m(B_s)} \right) \\ &\leqslant C_m\, \|f_\varepsilon\|^m_{L^m(B_{2R})} + C_m \left(\frac{1}{(s-r)^m} + \frac{R^m}{(s-r)^{2m}} \right) \|Dv_\varepsilon\|^m_{L^m(B_s)}.\end{aligned}$$

Hence, we can proceed as in the final part of the proof of Theorem 3.2 to improve the latter to

$$\|V_\varepsilon\|_{W^{1,m}(B_{R/2})} \leqslant C_m\, \|f_\varepsilon\|_{L^m(B_{2R})} + C_{m,R}\, \|V_\varepsilon\|_{L^1(B_{2R})}.$$

Since $V_\varepsilon \to V$ in $L^1(B_R)$, we obtain the claimed estimate (4-14) by lower semicontinuity. □

Corollary 4.8. *Let $u \in W^{1,p}(B_R)$ be a cylindrical weak solution of*

$$\Delta_p u = f \in L^\infty(B_R).$$

Then

(1) *If $p \geqslant 2$, $u \in C^{p'-\varepsilon}_{\mathrm{loc}}(B_R)$ for all $\varepsilon > 0$, and also for $\varepsilon = 0$ if furthermore $f \in C^0_{\mathrm{Dini}}(B_R)$.*

(2) *If $p \leqslant 2$, $u \in C^{2-\varepsilon}_{\mathrm{loc}}(B_R)$ for all $\varepsilon > 0$, and also for $\varepsilon = 0$ if furthermore $f \in C^0_{\mathrm{Dini}}(B_R)$.*

Proof. According to Theorem 4.7, the field $V = |Du|^{p-2}\, Du$ belongs to $W^{1,m}_{\mathrm{loc}}(B_R)$ for any $m > 1$; hence, by the Morrey embedding, it lies in $C^{1-\varepsilon}(B_R)$ for any $\varepsilon > 0$. Similarly, $V \in \mathrm{Lip}_{\mathrm{loc}}(B_R)$ if f is Dini continuous, being the gradient of a solution of $\Delta v = f \in C^0_{\mathrm{Dini}}(B_R)$. We conclude through the properties of the map Ψ in (4-12), as in the proof of Corollary 4.5. □

Acknowledgements

We warmly thank Lorenzo Brasco, Giulio Ciraolo and Alberto Farina for fruitful conversations on some of the themes of this research. We also thank an anonymous referee for several suggestions which improved the quality of the manuscript.

Mosconi is supported by grant PRIN no. 2017AYM8XW of the Ministero dell'Istruzione, Università e Ricerca and grant PIACERI 20-22, Linea 2 and Linea 3, of the University of Catania.

References

[Alberico et al. 2019] A. Alberico, I. Chlebicka, A. Cianchi, and A. Zatorska-Goldstein, "Fully anisotropic elliptic problems with minimally integrable data", *Calc. Var. Partial Differential Equations* **58**:6 (2019), art. id. 186. MR Zbl

[Alberti and Ambrosio 1999] G. Alberti and L. Ambrosio, "A geometrical approach to monotone functions in $\mathbb{R}^n$", *Math. Z.* **230**:2 (1999), 259–316. MR Zbl

[Antonini et al. 2022] C. A. Antonini, G. Ciraolo, and A. Farina, "Interior regularity results for inhomogeneous anisotropic quasilinear equations", *Math. Ann.* (online publication November 2022).

[Araújo et al. 2017] D. J. Araújo, E. V. Teixeira, and J. M. Urbano, "A proof of the $C^{p'}$-regularity conjecture in the plane", *Adv. Math.* **316** (2017), 541–553. MR Zbl

[Araújo et al. 2018] D. J. Araújo, E. V. Teixeira, and J. M. Urbano, "Towards the $C^{p'}$-regularity conjecture in higher dimensions", *Int. Math. Res. Not.* **2018**:20 (2018), 6481–6495. MR Zbl

[Avelin et al. 2018] B. Avelin, T. Kuusi, and G. Mingione, "Nonlinear Calderón–Zygmund theory in the limiting case", *Arch. Ration. Mech. Anal.* **227**:2 (2018), 663–714. MR Zbl

[Bañuelos and Méndez-Hernández 2003] R. Bañuelos and P. J. Méndez-Hernández, "Space-time Brownian motion and the Beurling–Ahlfors transform", *Indiana Univ. Math. J.* **52**:4 (2003), 981–990. MR Zbl

[Balci et al. 2020] A. K. Balci, L. Diening, and M. Weimar, "Higher order Calderón–Zygmund estimates for the p-Laplace equation", *J. Differential Equations* **268**:2 (2020), 590–635. MR Zbl

[Balci et al. 2022] A. K. Balci, A. Cianchi, L. Diening, and V. Mazya, "A pointwise differential inequality and second-order regularity for nonlinear elliptic systems", *Math. Ann.* **383**:3-4 (2022), 1775–1824. MR Zbl

[Beck and Mingione 2020] L. Beck and G. Mingione, "Lipschitz bounds and nonuniform ellipticity", *Comm. Pure Appl. Math.* **73**:5 (2020), 944–1034. MR Zbl

[Bousquet and Brasco 2016] P. Bousquet and L. Brasco, "Global Lipschitz continuity for minima of degenerate problems", *Math. Ann.* **366**:3-4 (2016), 1403–1450. MR Zbl

[Bousquet et al. 2014] P. Bousquet, C. Mariconda, and G. Treu, “On the Lavrentiev phenomenon for multiple integral scalar variational problems”, *J. Funct. Anal.* **266**:9 (2014), 5921–5954. MR Zbl

[Brasco et al. 2010] L. Brasco, G. Carlier, and F. Santambrogio, “Congested traffic dynamics, weak flows and very degenerate elliptic equations”, *J. Math. Pures Appl.* (9) **93**:6 (2010), 652–671. MR Zbl

[Breit et al. 2018] D. Breit, A. Cianchi, L. Diening, T. Kuusi, and S. Schwarzacher, “Pointwise Calderón–Zygmund gradient estimates for the p-Laplace system”, *J. Math. Pures Appl.* (9) **114** (2018), 146–190. MR Zbl

[Breit et al. 2022] D. Breit, A. Cianchi, L. Diening, and S. Schwarzacher, “Global Schauder estimates for the p-Laplace system”, *Arch. Ration. Mech. Anal.* **243**:1 (2022), 201–255. MR Zbl

[Carstensen and Müller 2002] C. Carstensen and S. Müller, “Local stress regularity in scalar nonconvex variational problems”, *SIAM J. Math. Anal.* **34**:2 (2002), 495–509. MR Zbl

[Cellina 2017] A. Cellina, “The regularity of solutions to some variational problems, including the p-Laplace equation for $2 \leq p < 3$”, *ESAIM Control Optim. Calc. Var.* **23**:4 (2017), 1543–1553. MR Zbl

[Chiarenza et al. 1991] F. Chiarenza, M. Frasca, and P. Longo, “Interior $W^{2,p}$ estimates for nondivergence elliptic equations with discontinuous coefficients”, *Ric. Mat.* **40**:1 (1991), 149–168. MR Zbl

[Cianchi and Mazya 2011] A. Cianchi and V. G. Mazya, “Global Lipschitz regularity for a class of quasilinear elliptic equations”, *Comm. Partial Differential Equations* **36**:1 (2011), 100–133. MR Zbl

[Cianchi and Mazya 2017] A. Cianchi and V. Mazya, “Quasilinear elliptic problems with general growth and merely integrable, or measure, data”, *Nonlinear Anal.* **164** (2017), 189–215. MR Zbl

[Cianchi and Mazya 2018] A. Cianchi and V. G. Mazya, “Second-order two-sided estimates in nonlinear elliptic problems”, *Arch. Ration. Mech. Anal.* **229**:2 (2018), 569–599. MR Zbl

[Cianchi and Mazya 2019] A. Cianchi and V. G. Mazya, “Optimal second-order regularity for the p-Laplace system”, *J. Math. Pures Appl.* (9) **132** (2019), 41–78. MR Zbl

[Ciraolo et al. 2020] G. Ciraolo, A. Figalli, and A. Roncoroni, “Symmetry results for critical anisotropic p-Laplacian equations in convex cones”, *Geom. Funct. Anal.* **30**:3 (2020), 770–803. MR Zbl

[Colombo and Figalli 2014] M. Colombo and A. Figalli, “Regularity results for very degenerate elliptic equations”, *J. Math. Pures Appl.* (9) **101**:1 (2014), 94–117. MR Zbl

[Cozzi et al. 2014] M. Cozzi, A. Farina, and E. Valdinoci, “Gradient bounds and rigidity results for singular, degenerate, anisotropic partial differential equations”, *Comm. Math. Phys.* **331**:1 (2014), 189–214. MR Zbl

[Cozzi et al. 2016] M. Cozzi, A. Farina, and E. Valdinoci, “Monotonicity formulae and classification results for singular, degenerate, anisotropic PDEs”, *Adv. Math.* **293** (2016), 343–381. MR Zbl

[Damascelli and Sciunzi 2004] L. Damascelli and B. Sciunzi, “Regularity, monotonicity and symmetry of positive solutions of m-Laplace equations”, *J. Differential Equations* **206**:2 (2004), 483–515. MR Zbl

[Diening et al. 2012] L. Diening, P. Kaplický, and S. Schwarzacher, “BMO estimates for the p-Laplacian”, *Nonlinear Anal.* **75**:2 (2012), 637–650. MR Zbl

[Dijkstra 2005] J. J. Dijkstra, “On homeomorphism groups and the compact-open topology”, *Amer. Math. Monthly* **112**:10 (2005), 910–912. MR Zbl

[Duoandikoetxea 2001] J. Duoandikoetxea, *Fourier analysis*, Grad. Stud. in Math. **29**, Amer. Math. Soc., Providence, RI, 2001. MR Zbl

[Esposito et al. 1999] L. Esposito, F. Leonetti, and G. Mingione, “Higher integrability for minimizers of integral functionals with (p, q) growth”, *J. Differential Equations* **157**:2 (1999), 414–438. MR Zbl

[Giaquinta 1983] M. Giaquinta, *Multiple integrals in the calculus of variations and nonlinear elliptic systems*, Ann. of Math. Stud. **105**, Princeton Univ. Press, 1983. MR Zbl

[Giaquinta 1987] M. Giaquinta, “Growth conditions and regularity: a counterexample”, *Manuscripta Math.* **59**:2 (1987), 245–248. MR Zbl

[Kovalev 2007] L. V. Kovalev, “Quasiconformal geometry of monotone mappings”, *J. Lond. Math. Soc.* (2) **75**:2 (2007), 391–408. MR Zbl

[Kovalev and Maldonado 2005] L. V. Kovalev and D. Maldonado, "Mappings with convex potentials and the quasiconformal Jacobian problem", *Illinois J. Math.* **49**:4 (2005), 1039–1060. MR Zbl

[Kuusi and Mingione 2013] T. Kuusi and G. Mingione, "Linear potentials in nonlinear potential theory", *Arch. Ration. Mech. Anal.* **207**:1 (2013), 215–246. MR Zbl

[Lindqvist 2019] P. Lindqvist, *Notes on the stationary p-Laplace equation*, Springer, 2019. MR Zbl

[Lou 2008] H. Lou, "On singular sets of local solutions to p-Laplace equations", *Chinese Ann. Math. Ser. B* **29**:5 (2008), 521–530. MR Zbl

[Martin 2014] G. J. Martin, "The theory of quasiconformal mappings in higher dimensions, I", pp. 619–677 in *Handbook of Teichmüller theory*, *IV*, edited by A. Papadopoulos, IRMA Lect. Math. Theor. Phys. **19**, Eur. Math. Soc., Zürich, 2014. MR Zbl

[Mercuri et al. 2016] C. Mercuri, G. Riey, and B. Sciunzi, "A regularity result for the p-Laplacian near uniform ellipticity", *SIAM J. Math. Anal.* **48**:3 (2016), 2059–2075. MR Zbl

[Mingione 2007] G. Mingione, "The Calderón–Zygmund theory for elliptic problems with measure data", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **6**:2 (2007), 195–261. MR Zbl

[Mingione 2010] G. Mingione, "Gradient estimates below the duality exponent", *Math. Ann.* **346**:3 (2010), 571–627. MR Zbl

[Miśkiewicz 2018] M. Miśkiewicz, "Fractional differentiability for solutions of the inhomogeneous p-Laplace system", *Proc. Amer. Math. Soc.* **146**:7 (2018), 3009–3017. MR Zbl

[Savaré 1998] G. Savaré, "Regularity results for elliptic equations in Lipschitz domains", *J. Funct. Anal.* **152**:1 (1998), 176–201. MR Zbl

[Simon 1978] J. Simon, "Régularité de la solution d'une équation non linéaire dans $\mathbb{R}^N$", pp. 205–227 in *Journées d'analyse non linéaire* (Besançon, France, 1977), edited by P. Bénilan and J. Robert, Lecture Notes in Math. **665**, Springer, 1978. MR Zbl

[Stampacchia 1963] G. Stampacchia, "On some regular multiple integral problems in the calculus of variations", *Comm. Pure Appl. Math.* **16** (1963), 383–421. MR Zbl

[de Thélin 1982] F. de Thélin, "Local regularity properties for the solutions of a nonlinear partial differential equation", *Nonlinear Anal.* **6**:8 (1982), 839–844. MR Zbl

[Uhlenbeck 1977] K. Uhlenbeck, "Regularity for a class of non-linear elliptic systems", *Acta Math.* **138**:3-4 (1977), 219–240. MR Zbl

[Väisälä 1971] J. Väisälä, *Lectures on n-dimensional quasiconformal mappings*, Lecture Notes in Math. **229**, Springer, 1971. MR Zbl

Received 21 Jun 2021. Revised 18 Jan 2022. Accepted 7 Mar 2022.

UMBERTO GUARNOTTA: umberto.guarnotta@phd.unict.it
Dipartimento di Matematica e Informatica, Università degli Studi di Catania, Catania, Italy

SUNRA MOSCONI: sunra.mosconi@unict.it
Dipartimento di Matematica e Informatica, Università degli Studi di Catania, Catania, Italy

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at msp.org/apde.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in APDE are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ANALYSIS & PDE

Volume 16 No. 8 2023