Tunisian Journal of Mathematics

an international publication organized by the Tunisian Mathematical Society

2019 vol. 1 no. 3

msp

Tunisian Journal of Mathematics

msp.org/tunis

The Tunisian Journal of Mathematics is an international publication organized by the Tunisian Mathematical Society (http://www.tms.rnu.tn) and published in electronic and print formats by MSP in Berkeley.

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

The subscription price for 2019 is US $315/year for the electronic version, and $370/year (+$20, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Tunisian Journal of Mathematics (ISSN 2576-7666 electronic, 2576-7658 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

TUNISIAN JOURNAL OF MATHEMATICS
Vol. 1, No. 3, 2019
dx.doi.org/10.2140/tunis.2019.1.295

Rigid local systems and alternating groups

Robert M. Guralnick, Nicholas M. Katz and Pham Huu Tiep

We show that some very simple to write one parameter families of exponential sums on the affine line in characteristic p have alternating groups as their geometric monodromy groups.

1. Introduction

In earlier work Katz [2018] exhibited some very simple one parameter families of exponential sums which gave rigid local systems on the affine line in characteristic p whose geometric (and usually, arithmetic) monodromy groups were $\mathrm{SL}_2(q)$, and he exhibited other such very simple families giving $\mathrm{SU}_3(q)$. (Here q is a power of the characteristic p, and p is odd.) In this paper, we exhibit equally simple families whose geometric monodromy groups are the alternating groups $\mathrm{Alt}(2q)$. We also determine their arithmetic monodromy groups. See Theorem 3.1 (Of course from the resolution [Raynaud 1994] of the Abhyankar conjecture, any finite simple group whose order is divisible by p will occur as the geometric monodromy group of some local system on $\mathbb{A}^1/\overline{\mathbb{F}}_p$; the interest here is that it occurs in our particularly simple local systems.)

Guralnick was partially supported by NSF grant DMS-1600056 and Tiep was partially supported by NSF grant DMS-1840702. Guralnick would also like to thank the Institute for Advanced Study, Princeton for its support.
The authors are grateful to the referees for careful reading and helpful comments on the paper.
MSC2010: 11T23, 20D05.
Keywords: rigid local system, monodromy, alternating group.

In the earlier work of Katz, he used a theorem to Kubert to know that the monodromy groups in question were finite, then work of Gross [2010] to determine which finite groups they were. Here we do not have, at present, any direct way of showing this finiteness. Rather, the situation is more complicated and more interesting. Using some basic information about these local systems (see Theorem 6.1), the first and third authors prove a fundamental dichotomy: the geometric monodromy group is either $\mathrm{Alt}(2q)$ or it is the special orthogonal group $\mathrm{SO}(2q-1)$. The second author uses an elementary polynomial identity to compute the third moment as being 1 (see Theorem 7.1), which rules out the $\mathrm{SO}(2q-1)$ case. This roundabout method establishes the theorem. It would be interesting to find a "direct" proof that these local systems have integer (rather than rational) traces; this integrality is in fact equivalent to their monodromy groups being finite, see [Katz 1990, 8.14.6]. But even if one had such a direct proof, it would still require serious group theory to show that their geometric monodromy groups are the alternating groups.

2. The local systems in general

Throughout this paper, p is an odd prime, q is a power of p, k is a finite field of characteristic p, ℓ is a prime $\neq p$,

$$\psi = \psi_k : (k, +) \to \mu_p \subset \overline{\mathbb{Q}}_\ell^\times$$

is a nontrivial additive character of k, and

$$\chi_2 = \chi_{2,k} : k^\times \to \pm 1 \subset \overline{\mathbb{Q}}_\ell^\times$$

is the quadratic character, extended to k by $\chi_2(0) := 0$. For L/k a finite extension, we have the nontrivial additive character

$$\psi_{L/k} := \psi_k \circ \mathrm{Trace}_{L/k}$$

of L, and the quadratic character $\chi_{2,L} = \chi_{2,k} \circ \mathrm{Norm}_{L/k}$ of $L^\times$, extended to L by $\chi_{2,L}(0) = 0$.

On the affine line $\mathbb{A}^1/k$, we have the Artin–Schreier sheaf $\mathcal{L}_{\psi(x)}$. On $\mathbb{G}_m/k$ we have the Kummer sheaf $\mathcal{L}_{\chi_2(x)}$ and its extension by zero $j_!\mathcal{L}_{\chi_2(x)}$ (for $j : \mathbb{G}_m \subset \mathbb{A}^1$ the inclusion) on $\mathbb{A}^1/k$.

For an odd integer $n = 2d + 1$ which is prime to p, we have the rigid local system (rigid by [Katz 1996, 3.0.2 and 3.2.4])

$$\mathcal{F}(k, n, \psi) := FT_\psi(\mathcal{L}_{\psi(x^n)} \otimes j_!\mathcal{L}_{\chi_2(x)})$$

on $\mathbb{A}^1/k$. Let us recall the basic facts about it, see [Katz 2004, 1.3 and 1.4].

It is lisse of rank n, pure of weight one, and orthogonally self-dual, with its geometric monodromy group

$$G_{\mathrm{geom}} \subset \mathrm{SO}(n, \overline{\mathbb{Q}}_\ell).$$

Recall that G_{geom} is the Zariski closure in $\mathrm{SO}(n, \overline{\mathbb{Q}}_\ell)$ of the image of the geometric fundamental group $\pi_1(\mathbb{A}^1/\bar{k})$ in the representation which "is" the local system $\mathcal{F}(k, n, \psi)$. For ease of later reference, we recall the following fundamental fact.

Lemma 2.1. *For any lisse local system $\mathcal{H}$ on $\mathbb{A}^1/\bar{k}$, the subgroup Γ_p of its G_{geom} generated by elements of p-power order is Zariski dense.*

Proof. Denote by N the Zariski closure of Γ_p in G_{geom}. Then N is a normal subgroup of G_{geom}. We must show that the quotient $M := G_{\mathrm{geom}}/N$ is trivial.

To see this, we argue as follows. The local system $\mathcal{H}$ gives us a group homomorphism

$$\pi_1(\mathbb{A}^1/\bar{k}) \to G_{\mathrm{geom}} \subset \mathrm{GL}(\mathrm{rank}(\mathcal{H}), \overline{\mathbb{Q}}_\ell)$$

with Zariski dense image. Under this homomorphism, the wild inertia group P_∞ has finite image in G_{geom} (because $\ell \neq p$). This image being a finite p group in G_{geom}, it lies in N, and hence dies in $M := G_{\mathrm{geom}}/N$. Therefore M/M^0 is a finite quotient of $\pi_1(\mathbb{A}^1/\bar{k})$ in which P_∞ dies. So any irreducible representation of M/M^0 gives an irreducible local system on $\mathbb{A}^1/\bar{k}$ which is tame at ∞, hence trivial. Thus $M = M^0$ is connected. We next show that $M^{\mathrm{red}} := M/\mathcal{R}_u$, the quotient of M by its unipotent radical, is trivial. For this, it suffices to show that M has no nontrivial irreducible representations. But any such representation is a local system on $\mathbb{A}^1/\bar{k}$ which is tamely ramified at ∞ (again because P_∞ dies in M), so is trivial. Thus M is unipotent. But $H^1(\mathbb{A}^1/\bar{k}, \overline{\mathbb{Q}}_\ell)$ vanishes, so any unipotent local system on $\mathbb{A}^1/\bar{k}$ is trivial, and hence M is trivial. □

Let us denote by $A(k, n, \psi)$ the Gauss sum

$$A(k, n, \psi) := -\chi_2(n(-1)^d) \sum_{x \in k^\times} \psi(x)\chi_2(x).$$

By the Hasse–Davenport relation, for L/k an extension of degree d, we have

$$A(L, n, \psi_{L/k}) = (A(k, n, \psi))^d.$$

The twisted local system

$$\mathcal{G}(k, n, \psi) := \mathcal{F}(k, n, \psi) \otimes A(n, k, \psi)^{-\deg}$$

is pure of weight zero and has

$$G_{\mathrm{geom}} \subset G_{\mathrm{arith}} \subset \mathrm{SO}(n, \overline{\mathbb{Q}}_\ell).$$

Concretely, for L/k a finite extension, and $t \in L$, the trace at time t of $\mathcal{G}(k, n, \psi)$ is

$$\begin{aligned}\operatorname{Trace}(\operatorname{Frob}_{t,L} | \mathcal{G}(k, n, \psi)) &= -(1/A(L, n, \psi_{L/k})) \sum_{x \in L^\times} \psi_{L/k}(x^n + tx)\chi_{2,L}(x) \\ &= -(1/A(L, n, \psi_{L/k})) \sum_{x \in L} \psi_{L/k}(x^n + tx)\chi_{2,L}(x),\end{aligned}$$

the last equality because the χ_2 factor kills the $x = 0$ term.

Let us recall also [Katz 2004, 3.4] that the geometric monodromy group of $\mathcal{F}(k, n, \psi)$, or equivalently of $\mathcal{G}(k, n, \psi)$, is independent of the choice of the pair (k, ψ).

To end this section, let us recall the relation of the local system $\mathcal{F}(k, n, \psi)$ to the hypergeometric sheaf

$$\mathcal{H}_n := \mathcal{H}(!, \psi; \text{all characters of order dividing } n; \chi_2).$$

According to [Katz 1990, 9.2.2], $\mathcal{F}(k, n, \psi)|\mathbb{G}_m$ is geometrically isomorphic to a multiplicative translate of the Kummer pullback $[n]^\star\mathcal{H}_n$. (An explicit descent of $\mathcal{F}(k, n, \psi)|\mathbb{G}_m$ through the n-th power map is given by the lisse sheaf on $\mathbb{G}_m$ whose trace function at time $t \in L^\times$, for L/k a finite extension, is

$$t \mapsto -\sum_{x \in L^\times} \psi_{L/k}(x^n/t + x)\chi_{2,L}(x/t).$$

The structure theory of hypergeometric sheaves shows that this descent is, geometrically, a multiplicative translate of the asserted $\mathcal{H}_n$.)

3. The candidate local systems for Alt(2q)

In this section, we specialize the n of the previous section to

$$n = 2q - 1 = 2(q - 1) + 1.$$

The target theorem is this:

Theorem 3.1. *Let p be an odd prime, q a power of p, k a finite field of characteristic p, ℓ a prime $\neq p$, and ψ a nontrivial additive character of k. For the ℓ-adic local system $\mathcal{G}(k, 2q - 1, \psi)$ on $\mathbb{A}^1/k$, its geometric and arithmetic monodromy groups are given as follows*:

(1) $G_{\text{geom}} = \operatorname{Alt}(2q)$ *in its unique irreducible representation of dimension* $2q - 1$.

(2) (a) *If -1 is a square in k, then* $G_{\text{geom}} = G_{\text{arith}} = \operatorname{Alt}(2q)$.

(b) *If -1 is not a square in k, then* $G_{\text{arith}} = \operatorname{Sym}(2q)$, *the symmetric group, in its irreducible representation labeled by the partition* $(2, 1^{2q-2})$, *i.e.*,

$$(\textit{the deleted permutation representation of } \operatorname{Sym}(2q)) \otimes \operatorname{sgn}.$$

Remark 3.2. The traces of elements of $\mathrm{Alt}(n)$ (respectively of $\mathrm{Sym}(n)$) in its deleted permutation representation (respectively in every irreducible representation) are integers. One sees easily (look at the action of $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$) that the local system $\mathcal{G}(k, 2q-1, \psi)$ has traces which all lie in $\mathbb{Q}$, but as mentioned in the introduction, we do not know a direct proof that these traces all lie in $\mathbb{Z}$.

4. Basic facts about $\mathcal{H}_n$

In this section, we assume that $n \geq 3$ is odd and that $n(n-1)$ is prime to p. The geometric local monodromy at 0 is tame, and a topological generator of the tame inertia group $I(0)^{\mathrm{tame}}$, acting on $\mathcal{H}_n$, has as eigenvalues all the roots of unity of order dividing n.

The geometric local monodromy at ∞ is the direct sum

$$\mathcal{L}_{\chi_2} \oplus W, \quad W \text{ has rank } n-1, \text{ and all slopes } 1/(n-1).$$

Because n is odd, the local system $\mathcal{H}_n$ is (geometrically) orthogonally self-dual, and $\det(\mathcal{H}_n)$ is geometrically trivial (because trivial at 0, lisse on $\mathbb{G}_m$, and all ∞ slopes are $\leq 1/(n-1) < 1$). Therefore $\det(W)$ is geometrically $\mathcal{L}_{\chi_2}$. From [Katz 1990, 8.6.4 and 8.7.2], we see that up to multiplicative translation, the geometric isomorphism class is determined entirely by its rank $n-1$ and its determinant $\mathcal{L}_{\chi_2}$. Because $n-1$ is even and prime to p, it follows that up to multiplicative translation, the geometric isomorphism class of W is that of the $I(\infty)$-representation of the Kloosterman sheaf

$$\mathrm{Kl}_{n-1} := \mathrm{Kl}(\psi; \text{all characters of order dividing } n-1).$$

By [Katz 1988, 5.6.1], we have a global Kummer direct image geometric isomorphism

$$\mathrm{Kl}_{n-1} \cong [n-1]_\star \mathcal{L}_{\psi_{n-1}},$$

where we write ψ_{n-1} for the additive character $x \mapsto \psi((n-1)x)$. Therefore, up to multiplicative translation, the geometric isomorphism class of W is that of $[n-1]_\star \mathcal{L}_\psi$. Pulling back by $[n-1]$, which does not change the restriction of W to the wild inertia group $P(\infty)$, we get

$$[n-1]^\star W \cong \bigoplus_{\zeta \in \mu_{n-1}} \mathcal{L}_{\psi(\zeta x)}.$$

A further pullback by n-th power, which also does not change the restriction of W to $P(\infty)$, gives

$$[n-1]^\star [n]^\star W \cong \bigoplus_{\zeta \in \mu_{n-1}} \mathcal{L}_{\psi(\zeta x^n)}.$$

Thus we find that the $I(\infty)$ representation attached to a multiplicative translate[1] of $[n-1]^\star\mathcal{F}(k,n,\psi)$ is the direct sum

$$\mathbb{1}\bigoplus_{\zeta\in\mu_{n-1}}\mathcal{L}_{\psi(\zeta x^n)}=\bigoplus_{\alpha\in\mu_{n-1}\cup\{0\}}\mathcal{L}_{\psi(\alpha x^n)}.$$

This description shows that the image of $P(\infty)$ in the $I(\infty)$-representation attached to $\mathcal{F}(k,n,\psi)$ is an abelian group killed by p.

Lemma 4.1. *Let $L/\mathbb{F}_p$ be a finite extension which contains the $(n-1)$-st roots of unity. Denote by $V\subset L$ the additive subgroup of L spanned by the $(n-1)$-st roots of unity. Denote by $V^\star$ the Pontryagin dual of V:*

$$V^\star:=\mathrm{Hom}_{\mathbb{F}_p}(V,\mu_p(\overline{\mathbb{Q}}_\ell)).$$

Then the image of $P(\infty)$ in the $I(\infty)$-representation attached to $\mathcal{F}(k,n,\psi)$ is $V^\star$, and the representation restricted to $V^\star$ is the direct sum

$$\mathbb{1}\bigoplus_{\zeta\in\mu_{n-1}(L)}(\textit{evaluation at }\zeta)=\bigoplus_{\alpha\in\mu_{n-1}(L)\cup\{0\}}(\textit{evaluation at }\alpha).$$

Proof. Each of the characters $\mathcal{L}_{\psi(\alpha x^n)}$ of $I(\infty)$ has order dividing p. Given an n-tuple of elements $(a_\alpha)_{\alpha\in\mu_{n-1}(L)\cup\{0\}}$, consider the character

$$\Lambda:=\bigotimes_{\alpha\in\mu_{n-1}(L)\cup\{0\}}(\mathcal{L}_{\psi(\alpha x^n)})^{\otimes a_\alpha}=\mathcal{L}_{\psi((\sum_{\alpha\in\mu_{n-1}(L)\cup\{0\}}a_\alpha\alpha)x^n)}.$$

The following conditions are equivalent:

(a) $\sum_{\alpha\in\mu_{n-1}(L)\cup\{0\}}a_\alpha\alpha=0$.

(b) The character Λ is trivial on $I(\infty)$.

(c) The character Λ is trivial on $P(\infty)$.

Indeed, it is obvious that (a) $\Longrightarrow$ (b) $\Longrightarrow$ (c). If (c) holds, then for

$$A:=\sum_{\alpha\in\mu_{n-1}(L)\cup\{0\}}a_\alpha\alpha,$$

we have that $\mathcal{L}_{\psi(Ax)}$ is trivial on $P(\infty)$, so is a character of $I(\infty)/P(\infty)=I(\infty)^{\mathrm{tame}}$, a group of order prime to p. But $\mathcal{L}_{\psi(Ax)}$ has order dividing p, so is trivial on $I(\infty)$, hence $A=0$.

This equivalence shows that the character group of the image of $P(\infty)$ is indeed the $\mathbb{F}_p$ span of the α's, i.e., it is V. The rest is just Pontryagin duality of finite abelian groups. □

[1]The referee has kindly explained to us that the results of [Fu 2010, Proposition 0.7, 0.8] allow one to make precise the multiplicative translates in the above paragraphs.

5. Basic facts about $\mathcal{H}_{2q-1}$

Taking $n = 2q - 1$, the geometric local monodromy at 0 of $\mathcal{H}_{2q-1}$ is tame, and a topological generator of the tame inertia group $I(0)^{\text{tame}}$, acting on $\mathcal{H}_n$, has as eigenvalues all the roots of unity of order dividing $2q - 1$.

Turning now to the action of $P(\infty)$, we have:

Lemma 5.1. *Denote by $\zeta_{2q-2} \in \mathbb{F}_{q^2}$ a primitive $(2q-2)$-th root of unity. In the $I(\infty)$-representation attached to $\mathcal{F}(k, 2q-1, \psi)$, the character group V of the image of $P(\infty)$ is the $\mathbb{F}_p$-space*

$$V = \mathbb{F}_q \oplus \zeta_{2q-2}\mathbb{F}_q.$$

Fix a nontrivial additive character ψ_0 of $\mathbb{F}_q$, and denote by ψ_1 the nontrivial additive character of $\mathbb{F}_{q^2}$ given by

$$\psi_1 := \psi_0 \circ \mathrm{Trace}_{\mathbb{F}_{q^2}/\mathbb{F}_q}.$$

Then the image $V^\star$ of $P(\infty)$ is itself isomorphic to V, and the representation of $P(\infty)$ is the direct sum of the characters

$$\bigoplus_{\alpha \in \mathbb{F}_q} \psi_1(\alpha x) \oplus \bigoplus_{\beta \in \mathbb{F}_q^\times} \psi_1(\zeta_{2q-2}\beta x).$$

Proof. When $n = 2q - 1$, then $n - 1 = 2(q - 1)$. The field $\mathbb{F}_{q^2}$ contains the $2(q-1)$-th roots of unity. The group $\mu_{2(q-1)}(\mathbb{F}_{q^2})$ contains the subgroup $\mu_{q-1}(\mathbb{F}_{q^2}) = \mathbb{F}_q^\times$ with index 2, the other coset being $\zeta_{2(q-1)}\mathbb{F}_q^\times$. Thus the $\mathbb{F}_p$ span of $\mu_{2(q-1)}(\mathbb{F}_{q^2})$ inside the additive group of $\mathbb{F}_{q^2}$ is indeed the asserted V. The characters $\psi_1(\alpha x)$, as α varies over $\mathbb{F}_q$, are each trivial on $\zeta_{2q-2}\mathbb{F}_q$ (because $\mathrm{Trace}_{\mathbb{F}_{q^2}/\mathbb{F}_q}(\zeta_{2q-2}) = 0$) and give all the additive characters of $\mathbb{F}_q$ (on which $\mathrm{Trace}_{\mathbb{F}_{q^2}/\mathbb{F}_q}$ is simply the map $x \mapsto 2x$). The characters $\psi_1(\zeta_{2q-2}\beta x)$, as β varies over $\mathbb{F}_q$, are trivial on $\mathbb{F}_q$ (because $\mathrm{Trace}_{\mathbb{F}_{q^2}/\mathbb{F}_q}(\zeta_{2q-2}) = 0$) and give all the characters of $\zeta_{2q-2}\mathbb{F}_q$ (because $\zeta_{2(q-1)}^2$ lies in $\mathbb{F}_q^\times$). □

Corollary 5.2. *The image of $P(\infty)$ in the $I(\infty)$-representation attached to*

$$\mathcal{F}(k, 2q-1, \psi) \oplus \mathbb{1}$$

is the direct sum

$$V = \mathbb{F}_q \oplus \zeta_{2q-2}\mathbb{F}_q$$

acting through the representation

$$\mathrm{Reg}_{\mathbb{F}_q} \oplus \mathrm{Reg}_{\zeta_{2q-2}\mathbb{F}_q}.$$

6. Basic facts about the group G_{geom} for $\mathcal{F}(k, 2q-1, \psi)$

Recall that G_{geom} is the Zariski closure in $\mathrm{SO}(2q-1, \overline{\mathbb{Q}}_\ell)$ of the image of $\pi_1^{\text{geom}} := \pi_1(\mathbb{A}^1/\overline{\mathbb{F}}_p)$ in the representation attached to $\mathcal{F}(k, 2q-1, \psi)$. Thus G_{geom} is an irreducible subgroup of $\mathrm{SO}(2q-1, \overline{\mathbb{Q}}_\ell)$.

Theorem 6.1. *We have the following two results*:

(i) G_{geom} *is normalized by an element of* $\mathrm{SO}(2q-1, \overline{\mathbb{Q}}_\ell)$ *whose eigenvalues are all the roots of unity of order dividing* $2q-1$ *in* $\overline{\mathbb{Q}}_\ell$.

(ii) G_{geom} *contains a subgroup isomorphic to* $\mathbb{F}_q \oplus \mathbb{F}_q$, *acting through the virtual representation*

$$\mathrm{Reg}_{\text{first}} \oplus \mathrm{Reg}_{\text{second}} - \mathbb{1}.$$

Proof. The local system $\mathcal{F}(k, 2q-1, \psi)$ is, geometrically, a multiplicative translate of the Kummer pullback $[2q-1]^\star \mathcal{H}_{2q-1}$. Therefore G_{geom} is a normal subgroup of the group G_{geom} for $\mathcal{H}_{2q-1}$, so is normalized by any element of this possibly larger group. As already noted, local monodromy at 0 for $\mathcal{H}_{2q-1}$ is an element of the asserted type. This proves (i). Statement (ii) is just a repeating of what was proved in the previous lemma. □

7. The third moment of $\mathcal{F}(k, 2q-1, \psi)$ and of $\mathcal{G}(k, 2q-1, \psi)$

Let us recall the general set up. We are given a lisse $\mathcal{G}$ on a lisse, geometrically connected curve C/k. We suppose that $\mathcal{G}$ is ι-pure of weight zero, for an embedding ι of $\overline{\mathbb{Q}}_\ell$ into $\mathbb{C}$. We denote by V the $\overline{\mathbb{Q}}_\ell$-representation given by $\mathcal{G}$, and by G_{geom} the Zariski closure in $\mathrm{GL}(V)$ of the image of $\pi_1^{\text{geom}}(C/k)$. For an integer $n \geq 1$, the n-th moment of $\mathcal{G}$ is the dimension of the space of invariants

$$M_n(\mathcal{G}) := \dim((V^{\otimes n})^{G_{\text{geom}}}).$$

Recall [Katz 2005, 1.17.4] that we have an archimedean limit formula for $M_n(\mathcal{G})$ as the lim sup over finite extensions L/k of the sums

$$(1/\#L) \sum_{t \in C(L)} (\mathrm{Trace}(\mathrm{Frob}_{t,L} \mid \mathcal{G}))^n,$$

which we call the empirical moments.

Theorem 7.1. *For the lisse sheaf* $\mathcal{G}(k, 2q-1, \psi)$ *on* $\mathbb{A}^1/k$, *we have*

$$M_3(\mathcal{G}(k, 2q-1, \psi)) = 1.$$

Proof. Fix a finite extension L/k. For $t \in L$, we have

$$\mathrm{Trace}(\mathrm{Frob}_{t,L} \,|\mathcal{G}(k, 2q-1, \psi)) \\ = (-1/A(L, 2q-1, \psi_{L/k})) \sum_{x \in L} \psi_{L/k}(x^{2q-1} + tx)\chi_{2,L}(x),$$

with the twisting factor given explicitly as

$$A(L, 2q-1, \psi_{L/k}) = -\chi_{2,L}(-1) \sum_{x \in L^\times} \psi_{L/k}(x)\chi_{2,L}(x).$$

Write g_L for the Gauss sum

$$g_L := \sum_{x \in L^\times} \psi_{L/k}(x)\chi_{2,L}(x).$$

Then the empirical M_3 is the sum

$$(1/\#L)(\chi_{2,L}(-1)/g_L)^3 \sum_{t \in L} \sum_{x,y,z \in L} \psi_{L/k}\big(x^{2q-1} + y^{2q-1} + z^{2q-1} + t(x+y+z)\big) \\ \cdot \chi_{2,L}(xyz)$$
$$= (\chi_{2,L}(-1)/g_L)^3 \sum_{x,y,z \in L, x+y+z=0} \psi_{L/k}\big(x^{2q-1} + y^{2q-1} + z^{2q-1}\big)\chi_{2,L}(xyz)$$
$$= (\chi_{2,L}(-1)/g_L)^3 \sum_{x,y \in L} \psi_{L/k}\big(x^{2q-1} + y^{2q-1} + (-x-y)^{2q-1}\big)\chi_{2,L}(xy(-x-y)).$$

The key is now the following identity.

Lemma 7.2. *In $\mathbb{F}_q[x, y]$, we have the identity*

$$x^{2q-1} + y^{2q-1} + (-x-y)^{2q-1} = xy(x+y) \prod_{\alpha \in \mathbb{F}_q \setminus \{0,-1\}} (x - \alpha y)^2.$$

If we write $q = p^f$, then collecting Galois-conjugate terms this is

$$xy(x+y) \prod_{h \in \mathcal{P}_f} h(x, y)^2,$$

where $\mathcal{P}_f$ is the set of irreducible $h(x, y) \in \mathbb{F}_p[x, y]$ which are homogeneous of degree dividing f, monic in x, other than x or $x + y$.

Proof. Because $x^{2q-1} + y^{2q-1} + (-x-y)^{2q-1}$ is homogeneous of odd degree $2q-1$ and visibly divisible by y, it suffices to prove the inhomogeneous identity, that in $\mathbb{F}_q[x]$ we have

$$x^{2q-1} + 1 - (x+1)^{2q-1} = x(x+1) \prod_{\alpha \in \mathbb{F}_q \setminus \{0,-1\}} (x - \alpha)^2.$$

The left side

$$P(x) := x^{2q-1} + 1 - (x+1)^{2q-1}$$

has degree $2q-2$, and visibly vanishes at $x = 0$ and at $x = -1$.

So it suffices to show that for each $\alpha \in \mathbb{F}_q \setminus \{0, -1\}$, $P(x)$ is divisible by $(x-\alpha)^2$. The key point is that for $\beta \in \mathbb{F}_q$, we have

$$\beta^{2q-1} = \beta,$$

and for $\alpha \in \mathbb{F}_q^\times$ we have

$$\alpha^{2q-2} = 1.$$

Thus for any $\beta \in \mathbb{F}_q$, we trivially have $P(\beta) = 0$. The derivative $P'(x)$ is equal to

$$P'(x) = -x^{2q-2} + (x+1)^{2q-2}.$$

So if both α and $\alpha + 1$ lie in $\mathbb{F}_q^\times$, then $P'(\alpha) = -1 + 1 = 0$. □

With this identity in hand, we now return to the calculation of the empirical moment, which is now

$$(\chi_{2,L}(-1)/g_L)^3 \sum_{x,y \in L} \psi_{L/k}(xy(x+y) \prod_{h \in \mathcal{P}_f} h(x,y)^2)\chi_{2,L}(xy(-x-y)).$$

The set of $(x, y) \in \mathbb{A}^2(L)$ with $xy \neq 0$ and at which $\prod_{h \in \mathcal{P}_f} h(x, y) = 0$ has cardinality $(q-2)(\#L-1)$. So the empirical sum differs from the modified empirical sum

$$(\chi_{2,L}(-1)/g_L)^3 \sum_{x,y \in L} \psi_{L/k}(xy(x+y) \prod_{h \in \mathcal{P}_f} h(x,y)^2)\chi_{2,L}(xy(-x-y) \prod_{h \in \mathcal{P}_f} h(x,y)^2)$$

by a difference which is

$$(\chi_{2,L}(-1)/g_L)^3 \quad \big(\text{a sum of at most } (q-2)(\#L-1) \text{ terms, each of absolute value } 1\big).$$

So the difference in absolute value is at most $q/\sqrt{\#L}$, which tends to zero as L grows (remember q is fixed). The modified empirical sum we now rewrite as

$$(\chi_{2,L}(-1)/g_L)^3 \sum_{t \in L^\times} \psi_{L/k}(t)\chi_{2,L}(-t)N_L(t),$$

with $N_L(t)$ the number of L-points on the curve $\mathcal{C}_t$ given by

$$\mathcal{C}_t : xy(x+y) \prod_{h \in \mathcal{P}_f} h(x,y)^2 = t.$$

Because $xy(x+y) \prod_{h \in \mathcal{P}_f} h(x, y)^2$ is homogeneous of degree $2q-1$ prime to p and is not a d-th power for any $d \geq 2$, the curves $\mathcal{C}_t$ are smooth and geometrically irreducible for all $t \neq 0$, see [Katz 1989, proof of 6.5]. Moreover, by the homogeneity, these curves are each geometrically isomorphic to $\mathcal{C}_1$, indeed the family become constant after the tame Kummer pullback $[2q-1]^\star$. Thus for the structural map $\pi : \mathcal{C} \to \mathbb{G}_m/\mathbb{F}_p$, $R^2\pi_!(\mathbb{Q}_\ell) = \mathbb{Q}_\ell(-1)$, $R^1\pi_!\mathbb{Q}_\ell$ is lisse of some rank r, tame at both 0 and ∞, and mixed of weight ≤ 1, and all other $R^i\pi_!(\mathbb{Q}_\ell) = 0$.

So our modified empirical moment is

$$(\chi_{2,L}(-1)/g_L)^3 \sum_{t\in L^\times} \psi_{L/k}(t)\chi_{2,L}(-t)(\#L - \mathrm{Trace}(\mathrm{Frob}_{t,L} \mid R^1\pi_!\mathbb{Q}_\ell)$$
$$= (\chi_{2,L}(-1)/g_L)^3 \sum_{t\in L^\times} \psi_{L/k}(t)\chi_{2,L}(-t)(\#L)$$
$$- (\chi_{2,L}(-1)/g_L)^3 \sum_{t\in L^\times} \mathrm{Trace}\big(\mathrm{Frob}_{t,L} \mid \mathcal{L}_{\psi(t)} \otimes \mathcal{L}_{\chi_2(t)} \otimes R^1\pi_!\mathbb{Q}_\ell\big).$$

Remembering that $g_L^2 = \chi_{2,L}(-1)\#L$, we see that the first sum is $\chi_{2,L}(-1)$. We now show that the second sum is $O(1/\sqrt{\#L})$, or equivalently that the sum

$$\sum_{t\in L^\times} \mathrm{Trace}\big(\mathrm{Frob}_{t,L} \mid \mathcal{L}_\psi \otimes \mathcal{L}_{\chi_2} \otimes R^1\pi_!\mathbb{Q}_\ell\big)$$

is $O(\#L)$. By the Lefschetz trace formula [Grothendieck 1968], the second sum is

$$\mathrm{Trace}\big(\mathrm{Frob}_L \mid H^2_c(\mathbb{G}_m/\overline{\mathbb{F}}_p, \mathcal{L}_\psi \otimes \mathcal{L}_{\chi_2} \otimes R^1\pi_!\mathbb{Q}_\ell)\big)$$
$$- \mathrm{Trace}\big(\mathrm{Frob}_L \mid H^1_c(\mathbb{G}_m/\overline{\mathbb{F}}_p, \mathcal{L}_\psi \otimes \mathcal{L}_{\chi_2} \otimes R^1\pi_!\mathbb{Q}_\ell)\big).$$

The H^2_c group vanishes, because the coefficient sheaf is totally wild at ∞ (this because it is $\mathcal{L}_\psi$ tensored with a lisse sheaf which is tame at ∞). The second sum is $O(\#L)$, by Deligne's fundamental estimate [Deligne 1980, 3.3.1] (because the coefficient sheaf is mixed of weight ≤ 1, its H^1_c is mixed of weight ≤ 2).

Thus the empirical moment is $\chi_{2,L}(-1)$ plus an error term which, as L grows, is $O(1/\sqrt{\#L})$. So the lim sup is 1, as asserted. □

8. Exact determination of G_{arith}

Theorem 8.1. *Suppose known that $\mathcal{G}(k, 2q-1, \psi)$ has $G_{\mathrm{geom}} = \mathrm{Alt}(2q)$. Then its G_{arith} is as asserted in Theorem 3.1, namely it is $\mathrm{Alt}(2q)$ if -1 is a square in k, and is $\mathrm{Sym}(2q)$ if -1 is not a square in k.*

Proof. For $q > 3$, the outer automorphism group of Alt$(2q)$ has order 2, induced by the conjugation action of Sym$(2q)$. Therefore the normalizer of Alt$(2q)$ in SO$(2q-1)$ (viewed there by its deleted permutation representation) is the group Sym$(2q)$ (viewed in SO$(2q-1)$ by (deleted permutation representation) $\otimes$ sgn). If $q = 3$, the automorphism group is slightly bigger but the stabilizer of the character of the deleted permutation module is just Sym$(2q)$. (Indeed, either of the exotic automorphisms of Alt(6) maps the cycle (123) to an element which in Sym(6) is conjugate to (123)(456). The element (123) has trace 2, whereas (123)(456) has trace -1 (both viewed in SO(5) by the deleted permutation representation)). Since we have a priori inclusions

$$G_{\text{geom}} = \text{Alt}(2q) \lhd G_{\text{arith}} \subset \text{SO}(2q-1),$$

the only choices for G_{arith} are $\text{Alt}(2q)$ or $\text{Sym}(2q)$.

Denoting by V the representation of G_{arith} given by $\mathcal{G}(k, 2q-1, \psi)$, the action of G_{arith} on the line

$$\mathbb{L} := (V^{\otimes 3})^{G_{\text{geom}}}$$

is a character of $G_{\text{arith}}/G_{\text{geom}}$. We claim that this character is the sign character sgn of $G_{\text{arith}} \subset \text{Sym}(2q)$. To see this, we argue as follows.

For any $n \geq 3$, denoting by V_n the deleted permutation representation of $\text{Sym}(n+1)$, one knows that

$$(V_n^{\otimes 3})^{\text{Sym}(n+1)} = (V_n^{\otimes 3})^{\text{Alt}(n+1)}$$

is one dimensional. (Indeed, if S^λ denotes the complex irreducible representation of $\text{Sym}(n+1)$ labeled by the partition λ of $n+1$, then $V_n = S^{(n,1)}$ and $\text{sgn} = S^{(1^{n+1})}$. An application of the Littlewood–Richardson rule to

$$S^\lambda \otimes \text{Ind}_{\text{Sym}(n)}^{\text{Sym}(n+1)}(S^{(n)}) = S^\lambda \oplus (S^\lambda \otimes V_n)$$

yields

$$V_n \otimes V_n = S^{(n+1)} \oplus S^{(n,1)} \oplus S^{(n-1,2)} \oplus S^{(n-1,1^2)}$$

see [Fulton and Harris 1991, Exercise 4.19]. Further similar applications of the Littlewood–Richardson rule then show that $V_n \otimes V_n \otimes V_n$ contains the trivial representation $S^{(n+1)}$ once but does not contain sgn.) Hence that the action of $\text{Sym}(n+1)$ on

$$((V_n \otimes \text{sgn})^{\otimes 3})^{\text{Alt}(n+1)}$$

is $\text{sgn}^3 = \text{sgn}$. Taking $n = 2q-1$, we get the claim.

Now apply Deligne's equidistribution theorem, in the form [Katz and Sarnak 1999, 9.7.10]. It tells us that if $G_{\text{arith}}/G_{\text{geom}}$ has order 2 instead of 1, then the Frobenii $\text{Frob}_{t,L}$ as L runs over larger and larger extensions of k of even (respectively odd) degree become equidistributed in the conjugacy classes of G_{arith} lying in G_{geom} (respectively lying in the other coset $G_{\text{arith}} \setminus G_{\text{geom}}$). If -1 is not a square in k, then $\chi_L(-1) = -1$ for all odd degree extensions L/k, and the empirical third moment over all odd degree extensions will be $-1 + O(1/\sqrt{\#L})$, by the proof of Theorem 7.1, whereas the empirical moment will be $1 + O(1/\sqrt{\#L})$ over even degree extensions. So if -1 is not a square in k, then $G_{\text{arith}} = \text{Sym}(2q)$. If -1 is a square in k, then every empirical moment will be $1 + O(1/\sqrt{\#L})$, and hence $G_{\text{arith}} = \text{Alt}(2q) = G_{\text{geom}}$.

9. Identifying the group

In this section, we use the information obtained earlier to identify the group. We choose a field embedding $\overline{\mathbb{Q}}_\ell \subset \mathbb{C}$, so that we may view $G := G_{\text{geom}}$ as an algebraic group over $\mathbb{C}$.

So let p be an odd prime with q a power of p. We start by assuming that G is an irreducible, Zariski closed subgroup of $\mathrm{SO}(2q-1, \mathbb{C}) = \mathrm{SO}(V)$ such that G contains Q, an elementary abelian subgroup of order q^2. Moreover, we assume that we may write $Q = Q_1 \times Q_2$ with $|Q_1| = |Q_2| = q$ so that $V = V_0 \oplus V_1 \oplus V_2$, where V_0 is a trivial Q-module, $V_0 \oplus V_i$ is the regular representation for Q_i and Q_i acts trivially on the other summand. Moreover, we assume that G is a quasi-p group (in the sense that the subgroup generated by its p-elements is Zariski dense), see Lemma 2.1.

Lemma 9.1. *V is tensor indecomposable for Q_1. More precisely, $V \neq X_1 \otimes X_2$, where the X_i are Q_1-modules each of dimension ≥ 2.*

Proof. We argue by contradiction. Suppose $V = X_1 \otimes X_2$ with each X_i of (necessarily odd) dimensional ≥ 2. Let χ_{X_i} be the character of Q_1 on X_i. So $\chi_{X_1} = a_0\mathbb{1} + \sum a_\chi \chi$ and $\chi_{X_2} = b_0\mathbb{1} + \sum b_\chi \chi$, where the χ are the nontrivial characters of Q_1.

We first reduce to the case when both a_0, b_0 are nonzero. The multiplicity of the trivial character of Q_1 in V is q, so we have

$$q = a_0 b_0 + \sum_\chi a_\chi b_{\overline{\chi}}.$$

So either $a_0 b_0$ is nonzero, and we are done, or for some nontrivial χ we have $a_\chi b_{\overline{\chi}}$ nonzero. In this latter case, replace X_1 by $X_1 \otimes \overline{\chi}$ and X_2 by $X_2 \otimes \chi$.

Since each nontrivial character χ of Q_1 occur exactly once in V, for each such χ we have

$$1 = a_0 b_\chi + a_\chi b_0 + \sum_{\rho \neq \chi} a_\rho b_{\chi \bar{\rho}}.$$

In particular we have the inequalities

$$a_0 b_\chi \leq 1, \quad a_\chi b_0 \leq 1.$$

Because a_0, b_0 are both nonzero, we infer that if $a_\chi \neq 0$, then $a_\chi = b_0 = 1$ (respectively that if $b_\chi \neq 0$, then $a_0 = b_\chi = 1$). It cannot be the case that all a_χ vanish, otherwise X_1 is the trivial module of dimension > 1. This is impossible so long as X_2 is nontrivial, as each nontrivial character of Q_1 occurs in V exactly once. But if all a_χ and all b_χ vanish, then V is the trivial Q_1 module, which it is not. Therefore $a_0 = 1$ and, similarly, $b_0 = 1$, and all a_χ, b_χ are either 0 or 1. Now use again that the multiplicity of the trivial character of Q_1 in V is q, so we have

$$q = a_0 b_0 + \sum_{\chi} a_\chi b_{\bar{\chi}}.$$

This is possible only if all a_χ and all b_χ are 1. But then each X_i has dimension q, which is impossible, as the product of their dimensions is $2q-1$. □

Lemma 9.2. *The following statements hold for G:*

(i) *G preserves no nontrivial orthogonal decomposition of V.*

(ii) *V is not tensor induced for G.*

Proof. We first prove (i). We argue by contradiction. Suppose that

$$V = W_1 \perp \cdots \perp W_r \quad \text{with } r > 1.$$

Because G acts irreducibly, G transitively permutes the W_i, and all the W_i have the same odd dimension d (because $2q-1 = rd$). Since r divides $2q-1$, $\gcd(r, p) = 1$, so the p-group Q fixes at least one of the W_i, say W_1. Because $r > 1$, there are other orbits of Q on the set of blocks. Any of these has cardinality some power of p, so the corresponding direct sum of W_i's has odd dimension. As $2q-1$ is odd, there must be evenly many other orbits, so at least three orbits in total. In each Q-stable odd-dimensional orthogonal space, Q lies in a maximal torus of the corresponding SO group, so has a fixed line. Hence $\dim V^Q \geq 3$, contradiction.

We next show that V is not tensor induced. We argue by contradiction. If V is tensor induced, write $V = W \otimes \cdots \otimes W$ (with $f \geq 2$ tensor factors, $\dim W < \dim V$). Then Q_1 must act transitively on the set of tensor factors (otherwise the representation for Q_1 is tensor decomposable and the previous lemma gives a contradiction).

So by Jordan's theorem [1872] (see also [Serre 2003, Theorem 4]), there exists an element $y \in Q_1$ that acts fixed point freely on the set of the f tensor factors. All such elements are conjugate in the wreath product $\mathrm{GL}(W) \wr \mathrm{Sym}(f)$ and we have

$$\chi_V(y) = (\dim W)^{f/p}.$$

(Indeed, after replacing y by a $\mathrm{GL}(W) \wr \mathrm{Sym}(f)$-conjugate, the situation is this. Each orbit of $\langle y \rangle$ on the set of tensor factors has length p, and y acts on each corresponding p-fold self-product of W, indexed by $\mathbb{F}_p$, by mapping $\bigotimes_i w_i$ to $\bigotimes_i w_{i+1}$. In terms of a basis $B := \{e_j\}_{j=1,\ldots,\dim W}$ of W, the only diagonal entries of the matrix of y on this $W^{\otimes p}$ are given by the $\dim W$ vectors $e \otimes e \otimes \cdots \otimes e$ with $e \in B$.) On the other hand, we have $\chi_V(y) = q-1$ for any nonzero element y of Q_1. Thus, if $d = \dim W$, we have $d^{f/p} = q-1$. Thus, $\dim V = d^f = (q-1)^p > 2q-1$, a contradiction. □

Corollary 9.3. *Let $L \leq \mathrm{SO}(V)$ be any subgroup containing G and let $1 \neq N \lhd L$. Then N acts irreducibly on V.*

Proof. We argue by contradiction. Note that the conclusions of Lemmas 9.1 and 9.2 also hold for L.

(i) Because N is normal in L, V is completely reducible for N. Let $V_1, \dots, V_r$ be the distinct N-isomorphism classes of N-irreducible submodules of V. Because V is L-self-dual, it is a fortiori N-self-dual. Therefore the set of V_i is stable by passage to the N-dual, $V_i \mapsto V_i^\star$. The group L acts transitively on the set of the V_i. Either every V_i is N-self-dual, or none is (the L-conjugates of an N-self-dual representation are N-self-dual).

When we write V as the direct sum of its N-isotypic ("homogeneous" in the terminology of [Curtis and Reiner 1962, 49.5]) components,

$$V = W_1 \oplus \cdots \oplus W_r,$$

then for some integer $e \geq 1$ we have N-isomorphisms

$$W_i \cong eV_i := \text{ the direct sum of } e \text{ copies of } V_i.$$

If $r > 1$ and all the W_i are self-dual, then this is an orthogonal decomposition (because for $i \neq j$, the inner product pairing of (any) V_i with (any) V_j is an N-homomorphism from V_i to $V_j^\star \cong V_j$, so vanishes). This contradicts Lemma 9.2.

Suppose $r > 1$ and no V_i is self-dual. Then the V_i occur in pairs of duals. Therefore both r and $\dim V$ are even, again a contradiction.

(ii) We have shown that $r = 1$ and $e > 1$, i.e., $V \cong eV_1$. Now we apply Clifford's theorem, see [Curtis and Reiner 1962, Theorem 51.7]. Thus L preserves the N-isomorphism class of V_1, and so we get an irreducible projective representation $L \mapsto \mathrm{PGL}(V_1) = \mathrm{PSL}(V_1)$, and V as a projective representation of L is $V_1 \otimes X$ with L acting (projectively) irreducibly on X through L/N, and X of dimension e. Furthermore, the two factor sets associated to these two projective representations can be chosen to be inverses to each other (as functions $L \times L \to \mathbb{C}^\times$), because the tensor product $V_1 \otimes X = V$ is a linear representation of Q_1. Since $e \dim V_1 = \dim V = 2q - 1$, both e and $n = \dim V_1$ are coprime to p.

We now claim that, restricted to Q_1, each of the tensor factors V_1 and X lifts to a genuine linear representation. Indeed, using the fact that $\mathrm{PGL}(n, \mathbb{C}) = \mathrm{PSL}(n, \mathbb{C})$ and the short exact sequence

$$1 \to \mu_n \to \mathrm{SL}(n, \mathbb{C}) \to \mathrm{PSL}(n, \mathbb{C}) \to 1,$$

the obstruction for $(V_1)|_{Q_1}$, which is given by the first factor set restricted to Q_1, lies in the cohomology group $H^2(Q_1, \mu_n)$. As $p \nmid n$ while Q_1 is a p-group, this cohomology group vanishes; and so the first factor set restricted to Q_1 is cohomologically trivial. As the second set is the inverse of the first set, it is also cohomologically trivial. Thus the Q_1-module V is tensor decomposable, contradicting Lemma 9.1. □

We next show that G is finite. It is convenient to use one more fact about G. There is a subgroup A (namely the group G_{geom} for the hypergeometric sheaf $\mathcal{H}_{2q-1}$) of $\mathrm{SO}(V)$ such that G is normal in A, A/G is cyclic of order dividing $2q-1$ and A contains an element x of order $2q-1$ with distinct eigenvalues on V.

We also use the fact that G has a nontrivial fixed space on $V\otimes V\otimes V$ (Theorem 7.1).

Theorem 9.4. *G is finite.*

Proof. Suppose not. Let N be any nontrivial normal (closed) subgroup of G. By Corollary 9.3, N is irreducible on V.

(i) Let G^0 be the identity component of G. We now show that G^0 is a simple algebraic group. Taking $N = G^0$, we have that G^0 acts irreducibly and hence is semisimple (as it lies in $\mathrm{SO}(V)$). Moreover, the center of G^0 is trivial (because it consists of scalars in $\mathrm{SO}(V)$). Therefore if G^0 is not simple, it is the product of adjoint groups L_j, $1 \le j \le t$ (namely the adjoint forms of the factors of its universal cover), and V is the (outer) tensor product $V = \bigotimes_{j=1}^{t} V_j$ of nontrivial irreducible L_j-modules V_j. By [Guralnick and Tiep 2008, Corollary 2.7], G permutes these tensor factors V_j. This action is transitive, otherwise we contradict Lemma 9.1. But this implies that V is tensor induced for G, contradicting Lemma 9.2. Thus G^0 is a simple algebraic group.

(ii) Because the subgroup of G generated by its p-elements is Zariski dense, the finite group G/G^0 is generated by its p-elements. As p is odd, it follows that either $G = G^0$ is a simple algebraic group or $p = 3$ and $G^0 = D_4(\mathbb{C})$. (In all other cases, the outer automorphism group, i.e., the automorphism group of the Dynkin diagram of G^0, has order at most 2.) Since A/G has odd order, it follows that $A \le G^0$ as well, unless $G^0 = D_4(\mathbb{C})$ and 3 divides $2q-1$.

Suppose first that A is connected and so a simple algebraic group. Then it contains a semisimple element x acting with distinct eigenvalues. This implies that a maximal torus has all weight spaces of dimension at most 1. Moreover, the module is in the root lattice (since it is odd dimensional and orthogonal). By a result of Howe [1990] (see also [Panyushev 2004, Table]), it follows if $G \ne \mathrm{SO}(V)$, then either $G = G_2(\mathbb{C})$ with $\dim V = 7$ or $G = \mathrm{PGL}_2(\mathbb{C})$. If $\dim V = 7$, then $q = 4$, a contradiction. If $G = \mathrm{PGL}_2(\mathbb{C})$, then any finite abelian subgroup of odd order is cyclic and so Q does not embed in G.

So $G = \mathrm{SO}(V)$. However, $\mathrm{SO}(V)$ has no nonzero fixed points on $V \otimes V \otimes V$ and this contradicts Theorem 7.1.

Thus, it follows that A is disconnected. So the connected component is $D_4(\mathbb{C})$ and this acts irreducibly on V. If $D_4(\mathbb{C})$ contains the element of order $2q-1$, then a maximal torus has all weight space of dimension 1 and again using [Howe 1990], we obtain a contradiction. If not, then 3 divides $2q-1$, whence $p \geq 5$ and $Q \leq D_4(\mathbb{C})$. Any elementary abelian p-subgroup of $D_4(\mathbb{C})$ is contained in a torus and so again we see that the connected component has all weight spaces of dimension at most 1 and we obtain the final contradiction using [Howe 1990]. □

Let $F^*(X)$ denote the generalized Fitting subgroup of a finite group X (so X is almost simple precisely when $F^*(X)$ is a nonabelian simple group).

Corollary 9.5. *A and G are almost simple and $F^*(A) = F^*(G)$ acts irreducibly on V.*

Proof. Let N be a minimal normal subgroup of G. By Corollary 9.3, N acts irreducibly, and so by Schur's lemma $\boldsymbol{C}_A(N) = \boldsymbol{Z}(N) = 1$ as $A < \mathrm{SO}(V)$ with $\dim V$ odd. So N is nonabelian, and so, being a minimal normal subgroup, it is a direct product of nonabelian simple groups. Arguing as in part (i) of the proof of Theorem 9.4, we see that N is nonabelian simple (otherwise the module V would be tensor induced). As $\boldsymbol{C}_G(N) = 1$, we see that $N \lhd G \leq \mathrm{Aut}(N)$, and so G is almost simple and $F^*(G) = N$.

Now, as $G \lhd A$, A normalizes N. Again since $\boldsymbol{C}_A(N) = 1$ we have that $N \lhd A \leq \mathrm{Aut}(N)$, and so A is almost simple and $F^*(A) = N$. □

We next observe:

Lemma 9.6. *$F^*(G)$ is not a sporadic simple group.*

Proof. Notice that both G and A are generated by elements of odd order (p-elements for G, these and elements of order $2q-1$ for A). On the other hand, we have $S \leq G \leq A \leq \mathrm{Aut}(S)$ for $S = F^*(G)$. One knows [Conway et al. 1985] that if S is sporadic, then $|\mathrm{Out}(S)| \leq 2$. Therefore, if S is a sporadic simple group, then $G = A = S$. The result now follows easily from information in [Conway et al. 1985]. Namely, we observe that if q is an odd prime power with q^2 dividing $|G|$, then G has no irreducible representation of dimension $2q-1$. □

We next consider the case $F^*(G) = \mathrm{Alt}(n)$. First note $\mathrm{Alt}(5)$ contains no noncyclic elementary abelian groups of odd order and so is ruled out. Since $2q-1$ is odd, we see that if $G = \mathrm{Alt}(n)$, then $A = G = \mathrm{Alt}(n)$ (as the outer automorphism group of $\mathrm{Alt}(n)$ is a 2-group).

Theorem 9.7. *Let $\Gamma = \mathrm{Alt}(n)$ with $n \geq 6$. Suppose that $x \in \Gamma$ has odd order and V is an irreducible $\mathbb{C}[\Gamma]$-module such that x acts as a semisimple regular element on V. Then one of the following holds*:

(i) *V is the deleted permutation module of dimension $n-1$* (i.e., the nontrivial irreducible constituent of $\mathbb{C}^{\mathrm{Alt}(n)}_{\mathrm{Alt}(n-1)}$), *and x is either an n-cycle* (*for n odd*) *or a product of two disjoint cycles of coprime lengths* (*for n even*); *or*

(ii) $n = 8$, *x has order* 15 *and* $\dim V = 14$.

Proof. First note that if V is the deleted permutation module of dimension $n-1$, an element with 3 or more disjoint cycles has at least a two-dimensional fixed space on V. Next assume that x has two disjoint cycles of lengths a and b which are not coprime. Then x affords a 2-dimensional eigenspace on $\mathbb{C}^n$ for an eigenvalue λ, a primitive $\gcd(a, b)$-th root of unity in $\mathbb{C}$. As $\lambda \neq 1$ and V is obtained from $\mathbb{C}^n$ by modding out the trivial eigenspace of $\mathrm{Sym}(n)$, it follows that x has a two-dimensional eigenspace on V as well.

Next we observe that if x is semisimple regular on V, then the order of x is at least $\dim V$. This proves the result for $6 \le n \le 14$ by inspection of the odd order elements and dimensions of the irreducible modules, aside from the case $n = 8$ and $\dim V = 14$ (note that Alt(8) contains an element of order 15). Recall that $\mathrm{Alt}(8) \cong \mathrm{GL}_4(2)$ and it acts 2-transitively on the nonzero vectors. The only irreducible module of dimension 14 is the irreducible summand of the permutation module of dimension 15. In this case x has a single orbit in the permutation representation and so x is semisimple regular on V.

Now assume that $n \ge 15$.

Suppose first that x has at most three nontrivial cycles. Then the order of x is less than $(n/3)^3 = n^3/27$ and so $\dim V < n^3/27$. Let W be a complex irreducible $\mathrm{Sym}(n)$-module whose restriction to $\mathrm{Alt}(n)$ contains $V|_{\mathrm{Alt}(n)}$. Since $2 \le \dim W < 2n^3/27$, it follows by [Rasala 1977, Result 3] that $W \cong S^\lambda$ or $S^\lambda \otimes \mathrm{sgn}$, where S^λ is the Specht module labeled by the partition λ of n, with $\lambda = (n-1, 1)$, $(n-2, 2)$, or $(n-2, 1, 1)$. Restricting back to $\mathrm{Alt}(n)$, we see that $V|_{\mathrm{Alt}(n)} = S^\lambda|_{\mathrm{Alt}(n)}$.

Note that

$$\dim S^{(n-2,1,1)} = (n-1)(n-2)/2, \quad \dim S^{(n-2,2)} = n(n-3)/2.$$

It is straightforward to see that the dimension of the fixed space of x on either of these modules is at least two dimensional, a contradiction. Hence $\lambda = (n-1, 1)$ and $V|_{\mathrm{Alt}(n)}$ is the deleted permutation module of dimension $n-1$.

We now induct on n. The base case $n \le 14$ has already done. We may assume that x has at least four nontrivial cycles (each of odd length, as x has odd order). View $x \in J := \mathrm{Alt}(a) \times \mathrm{Alt}(b)$. where the projection into $\mathrm{Alt}(b)$ is a b-cycle and so the projection into $\mathrm{Alt}(a)$ is a product of at least three disjoint cycles. Thus, $a \ge 9$. Let W be an irreducible J-submodule of V with $\mathrm{Alt}(a)$ acting nontrivially. So $W = W_1 \otimes W_2$ with W_1 an irreducible $\mathrm{Alt}(a)$-module. Then x must be multiplicity free on each W_i and by induction x can have at most two cycles in $\mathrm{Alt}(a)$, a contradiction. □

Note that the previous result does fail for $n = 5$. $\mathrm{Alt}(5)$ has a 5-dimensional representation in which an element of order 5 has all eigenvalues occurring once. Thus if $G = \mathrm{Alt}(n)$, we see that $n = 2q$ and V is the deleted permutation module.

Corollary 9.8. *If $G = G_{\mathrm{geom}}$ is an alternating group $\mathrm{Alt}(n)$ for some n, then $n = 2q$.*

Finally, we consider the case where G is an almost simple finite group of Lie type, defined over $\mathbb{F}_s$, where $s = s_0^f$ is a power of a prime s_0. Let us denote

$$S := F^*(G) = F^*(A).$$

Recall that S is simple, irreducible on V, and $\mathbf{Z}(S) = 1$ by Corollary 9.5. We will freely use information on character tables of simple groups available in [Conway et al. 1985; GAP 2004], as well as degrees of complex irreducible characters of various quasisimple groups of Lie type available in [Lübeck 2007]. Finally, we will also use bounds on the smallest degree $d(S)$ of nontrivial complex irreducible representations of S as listed in [Tiep 2003, Table 1].

Theorem 9.9. *Suppose $s_0 \neq p$. Then $S \cong \mathrm{Alt}(m)$ with $m \in \{5, 6, 8\}$.*

Proof. (i) Assume the contrary. We will exploit the existence of the subgroup $Q \leq G$. Recall that the *p-rank* $m_p(G)$ is the largest rank of elementary abelian p-subgroups of G. Furthermore,

$$\mathrm{Aut}(S) \cong \mathrm{Inndiag}(S) \rtimes \Phi_S\Gamma_S, \tag{9.9.1}$$

where $\mathrm{Inndiag}(S)$ is the subgroup of inner-diagonal automorphisms of S, Φ_S is a subgroup of field automorphisms of S and Γ_S is a subgroup of graph automorphisms of S, as defined in [Gorenstein et al. 1998, Theorem 2.5.12]. As $F^*(G) = S$, we can embed G in $\mathrm{Aut}(S)$. Now, given an elementary abelian p-subgroup $P < G$ of rank $m_p(G)$, we can define a normal series

$$1 \leq P_1 \leq P_2 \leq P,$$

where $P_1 = P \cap \mathrm{Inndiag}(S)$ and $P_2 = P \cap (\mathrm{Inndiag}(S) \rtimes \Phi_S)$. As Φ_S is cyclic and P is elementary abelian, P_2/P_1 has order 1 or p. Set $e = 1$ if $S \cong P\Omega_8^+(s)$ and $p = 3$, and $e = 0$ otherwise. Then $|P/P_2| \leq p^e$.

Next we bound $|P_1|$ when S is not a Suzuki–Ree group. Let $\Phi_j(t)$ denote the j-th cyclotomic polynomial in the variable t, and let m denote the multiplicative order of s modulo p, so that $p \mid \Phi_m(s)$. Note that we can find a simple algebraic group $\mathcal{G}$ of adjoint type defined over $\overline{\mathbb{F}}_s$ and a Frobenius endomorphism $F : \mathcal{G} \to \mathcal{G}$

such that $\mathrm{Inndiag}(S) \cong \mathcal{G}^F$. Letting r denote the rank of $\mathcal{G}$, then one can find r positive integers $k_1, \ldots, k_r$ and $\epsilon_1, \ldots, \epsilon_r = \pm 1$ such that

$$|\,\mathrm{Inndiag}(S)| = s^N \prod_{j \geq 1} \Phi_j(s)^{n_j} = s^N \prod_{i=1}^{r} (s^{k_i} - \epsilon_i)$$

for suitable integers N, n_j. Then, according to [Gorenstein et al. 1998, Theorem 4.10.3(b)], $|P_1| \leq p^{n_m}$. Let $\varphi(\cdot)$ denote the Euler function, so $\deg(\Phi_m) = \varphi(m)$. Inspecting the integers $k_1, \ldots, k_r$, one sees that $n_m \leq r/\varphi(m)$. It follows that

$$|P_1| \leq \Phi_m(s)^{n_m} \leq ((s+1)^{\varphi(m)})^{r/\varphi(m)} \leq (s+1)^r.$$

In fact, one can verify that this bound on $|P_1|$ also holds for Suzuki–Ree groups. Putting all the above estimates together, we obtain that

$$q^2 = |Q| \leq |P| \leq (s+1)^{r+1+e}. \tag{9.9.2}$$

We will show that this upper bound on q contradicts the lower bound

$$2q - 1 = \dim V \geq d(S) \tag{9.9.3}$$

in most of the cases. Let f^* denote the odd part of the integer f.

(ii) First we handle the case when S is of type D_4 or ${}^3D_4$. Here, $q \leq (s+1)^3$ by (9.9.2). On the other hand, $d(S) \geq s(s^4 - s^2 + 1)$, contradicting (9.9.3) if $s \geq 3$. If $s = 2$, then $\Phi_S \Gamma_S = C_3$, and so instead of (9.9.2) we now have that $q^2 \leq 3^5$, whence $q \leq 13$, $2q - 1 \leq 25 < d(S)$, again a contradiction.

From now on we may assume $e = 0$.

Next we consider the case $S = \mathrm{PSL}_2(s)$. Then $\mathrm{Out}(S) = C_{\gcd(2,s-1)} \times C_f$, and $m_p(S) \leq 1$. It follows that Q is not contained in S but in $S \rtimes C_f$ and $3 \leq p \mid f^*$, and furthermore $q^2 = |Q| \leq (s+1)f^*$. As $d(S) \geq (s-1)/2$, (9.9.3) now implies that

$$s + 1 = s_0^f + 1 \leq 16 f^*,$$

a contradiction if $s_0 \geq 5$, or $s_0 = 3$ and $f \geq 5$, or $s_0 = 2$ and $f \geq 7$. If $s_0 = 3$ and $f \leq 4$, then $f^* = 3 = f = p$, forcing $p = s_0$, a contraction. Suppose $s_0 = 2$ and $f \leq 6$. If $p = 5$, then $f^* = 5$ and $m_p(G) = 1$, ruling out the existence of Q. If $p = 3$, then $f = 3, 6$, whence $q^2 \leq 9$ and $2q - 1 \leq 5 < d(S)$.

Suppose that $S = {}^2B_2(s)$ or ${}^2G_2(s)$ with $s \geq 8$. Since $m_p(S) \leq 1$, we see that $q^2 \leq (s+1)f$, contradicting (9.9.3) as $d(S) \geq (s-1)\sqrt{s/2}$.

Now we consider the remaining cases with $r = 2$. Then $q \leq (s+1)^{\frac{3}{2}}$ by (9.9.2). This contradicts (9.9.3) if $S = G_2(s)$ (and $s \geq 3$), as $d(S) \geq s^3 - 1$. Similarly, $S \not\cong \mathrm{PSL}_3(s)$ with $s \geq 5$ and $S \not\cong \mathrm{PSU}_3(s)$ with $s \geq 8$. If $S = \mathrm{PSp}_4(s)$, then the case $2 \nmid s \geq 19$ is ruled out since $d(S) \geq (s^2-1)/2$, and similarly the case $2 | s \geq 8$ is ruled out since $d(S) = s(s-1)^2/2$. In the remaining cases, $\Phi_S \Gamma_S$ is a 2-group, and so $Q \le S$, $q^2 \le (s+1)^2$, $q \le s+1$. Now $\mathrm{PSL}_3(s)$ and $\mathrm{PSU}_3(s)$ with $s \ge 4$ are ruled out by (9.9.3), and the same for $\mathrm{PSp}_4(s)$ with $s \ge 4$. Note that when $s = 3$, $q \ge 4$ and so $\gcd(q, 2s) \ne 1$, a contradiction. If $S = \mathrm{SL}_3(2)$, then $q = 3$ and S has no irreducible character of degree $2q-1$. Finally, $\mathrm{Sp}_4(2)' \cong \mathrm{Alt}(6)$.

Next we handle the groups with $r = 3$. Here $q \le (s+1)^2$ by (9.9.2). Then (9.9.3) implies that $s \le 5$. In this case, $\mathrm{Out}(S)$ is a 2-group, and so $Q \le S$ and $q \le (s+1)^{\frac{3}{2}}$ by (9.9.2). Using (9.9.3), we see that $s \le 3$. The remaining groups S cannot occur, since S does not have a real-valued complex irreducible character of degree $2q-1$.

(iii) From now we may assume that $r \ge 4$ (and S is not of type D_4 or ${}^3D_4$). First we consider the case $s = 2$. If $S = \mathrm{SL}_n(2)$ with $n \ge 5$, then since $\mathrm{Out}(S) = C_2$, the arguments in (i) show that $q^2 \le 3^{n-1}$. This contradicts (9.9.3), since $d(S) = 2^n - 2$. Suppose $S = \mathrm{SU}_n(2)$ with $n \ge 7$. Note by [Tiep and Zalesskii 1996, Theorem 4.1] that the first three nontrivial irreducible characters of S are Weil characters and either non-real-valued or of even degree, and the next characters have degree at least $(2^n-1)(2^{n-1}-4)/9$. Hence (9.9.3) can be improved to

$$2q-1 \ge (2^n-1)(2^{n-1}-4)/9,$$

which is impossible since $q^2 \le 3^n$ by (9.9.2). If $S = \mathrm{PSU}_n(2)$ with $n = 5, 6$, then $q^2 \le 3^6$, and S has no nontrivial real-valued irreducible character of odd degree $\le 2q-1 \le 53$. If $S = {}^2F_4(2)'$ or $F_4(2)$, then $q^2 \le 3^5$, $q \le 13$, and S has no nontrivial real-valued irreducible character of odd degree $\le 2q-1 \le 25$.

Suppose $S = \mathrm{Sp}_{2n}(2)$ or $\Omega^{\pm}_{2n}(2)$. Then $\mathrm{Out}(S)$ is a 2-group (recall S is not of type D_4), and so $q^2 \le 3^n$. On the other hand, $d(S) \ge (2^n-1)(2^{n-1}-2)/3$, contradicting (9.9.3). Finally, if S is of type E_8, E_7, E_6, or ${}^2E_6$, then $q^2 \le 3^8$ whereas $d(S) > 2^{10}$, again contradicting (9.9.3).

(iv) Suppose that $S = \mathrm{PSp}_{2n}(s)$ with $n \ge 4$ and $2 \nmid s \ge 3$. Then by (9.9.2) and (9.9.3) we have

$$(s^n-1)/2 \le 2q-1 \le 2(s+1)^{(n+1)/2} - 1,$$

implying $n \le 5$ and $s = 3$. In this case, inspecting the order of $\mathrm{PSp}_{10}(3)$ we see that $q^2 \le 121$, and so $2q-1 \le 21 < d(S)$, a contradiction.

Next suppose that $S = \mathrm{PSU}_n(3)$ with $n \ge 5$. Then $q \le 2^n$ and $d(S) \ge (3^n-3)/4$, and so (9.9.3) implies that $n = 5$. In this case, inspecting the order of $\mathrm{SU}_5(3)$ we see that $q^2 \le 61$, and so $2q-1 \le 13 < d(S)$, again a contradiction.

Now we may assume that $r \ge 4$, $s \ge 3$, $S \not\cong \mathrm{PSp}_{2n}(s)$ if $2 \nmid s$, and moreover $s \ge 4$ if $S \cong \mathrm{PSU}_n(s)$. Then one can check that $d(S) \ge s^r \cdot (51/64)$ (with equality attained exactly when $S \cong \mathrm{PSU}_5(4)$). Hence (9.9.2) and (9.9.3) imply that

$$(51/64)^2 \cdot s^{2r} \le d(S)^2 \le (2q-1)^2 < 4(s+1)^{r+1} \le 4 \cdot (4s/3)^{r+1},$$

and so

$$(3s/4)^{r-1} < 4 \cdot (64/51)^2 \cdot (4/3)^2,$$

which is impossible for $r \geq 4$. □

Theorem 9.10. *Suppose $s_0 = p$. Then $S \cong \mathrm{Alt}(6)$.*

Proof. (i) Assume the contrary. We now exploit the existence of the element $x \in A$ of order $2q-1$ which has simple spectrum on V. As before, we can embed A in $\mathrm{Aut}(S)$ and again use the decomposition (9.9.1). Let $\langle y \rangle = \langle x \rangle \cap \mathrm{Inndiag}(S)$. We also view $S = \mathcal{G}^F$ for some Frobenius endomorphism $F : \mathcal{G} \to \mathcal{G}$ of a simple algebraic group $\mathcal{G}$ of adjoint type, defined over $\overline{\mathbb{F}}_p$. Note that y is an F-stable semisimple element in $\mathcal{G}$, hence it is contained in an F-stable maximal torus $\mathcal{T}$ of $\mathcal{G}$ by [Digne and Michel 1991, Corollary 3.16]. It follows that $|y| \leq |\mathcal{T}^F| \leq (s+1)^r$, if r is the rank of $\mathcal{G}$. Set $e = 3$ if S is of type D_4 or ${}^3D_4$, and $e = 1$ otherwise. Then the decomposition (9.9.1) shows that

$$|x|/|y| \leq ef^*,$$

where f^* denotes the odd part of f as before (and $s = p^f$). We have thus shown that

$$2q - 1 = |x| \leq (s+1)^r ef^*. \tag{9.10.1}$$

We will frequently use the following remark:

$$\text{either} \quad f = 1 \text{ and } s \geq 3f^*, \quad \text{or} \quad s \geq 9f^*. \tag{9.10.2}$$

We will show that in most of the cases (9.10.1) contradicts (9.9.3). First we handle the case S is of type D_4 or ${}^3D_4$, whence $d(S) \geq s(s^4 - s^2 + 1)$. Hence (9.10.1) and (9.10.2) imply that

$$s(s^4 - s^2 + 1) \leq 2q - 1 \leq s(s+1)^4/3$$

if $f > 1$, a contradiction. If $f = 1$, then since $2q - 1 = \dim V$ is coprime to $2s$, we see by [Lübeck 2007] that

$$2q - 1 > s^7/2 > 3(s+1)^4,$$

contradicting (9.10.1).

(ii) From now on we may assume that $e = 1$. Next we rule out the case where $V|_S$ is a Weil module of $S \in \{\mathrm{PSL}_n(s), \mathrm{PSU}_n(s)\}$ with $n \geq 3$, or $S = \mathrm{PSp}_{2n}(s)$ with $n \geq 2$. Indeed, in this case, if $S = \mathrm{PSL}_n(s)$ then

$$\dim V = (s^n - s)/(s-1),\ (s^n - 1)/(s-1)$$

is congruent to 0 or 1 modulo p and so cannot be equal to $2q-1$. Similarly, if $S = \mathrm{PSU}_n(s)$, then $V|_S$ can be a Weil module of dimension $2q-1$ only when $2 \mid n$ and $\dim V = (s^n-1)/(s+1)$. In this case,

$$q = (2q-1)_p = ((s^n+s)/(s+1))_p = s$$

(where N_p denotes the p-part of the integer N), and so $2s-1 = (s^n+s)/(s+1)$, a contradiction. Likewise, if $S = \mathrm{PSp}_{2n}(s)$, then $V|_S$ can be a Weil module of dimension $2q-1$ only when $p=3$ and $\dim V = (s^n+1)/2$. In this case,

$$s^n = (2\dim V - 1)_p = (4q-3)_p,$$

and so $q=3$ and $n=2$. One can show that $\mathrm{PSp}_4(3)$ does possess a complex irreducible module of dimension $2q-1=5$, with an element x of order 5 with simple spectrum on V and a subgroup $Q \cong C_3^2$ with desired prescribed action on V; however, any such module is not self-dual. Henceforth, for the aforementioned possibilities for S we may assume that $\dim V \geq d_2(S)$, the next degree after the degree of Weil characters. Note that $d_2(S)$ for these simple groups S is determined in Theorems 3.1, 4.1, and 5.2 of [Tiep and Zalesskii 1996].

(iii) Suppose $S = \mathrm{PSL}_2(s)$; in particular, $s \neq 9$. Assume $f \geq 4$. As $\mathrm{Out}(S) = C_{2,s-1} \times C_f$, we see that $q^2 \leq sf_p < s^2/20$, whereas $2q-1 \geq d(S) \geq (s-1)/2$, a contradiction. If $f \leq 3$ but $f_p > 1$, then $f = p = 3$, $s = 3^3$, $q^2 \leq sf = 3^4$, forcing $q = 9$. But then $S = \mathrm{PSL}_2(27)$ has no irreducible character of degree $2q-1 = 17$. Thus $f_p = 1$, $q^2 \leq s$, and so (9.9.3) implies that $s \leq 17$. As $s \neq 9$, we see that $m_p(G) = m_p(S) = 1$, contradicting the existence of Q.

Next we consider the case $S = \mathrm{PSL}_3(s)$ or $\mathrm{PSU}_3(s)$. If $f > 1$, then (9.9.3)–(9.10.2) imply

$$(s-1)(s^2-s+1)/3 \leq d_2(S) \leq 2q-1 \leq (s+1)^2 s/9,$$

which is impossible. Thus $f = 1$, whence

$$(s-1)(s^2-s+1)/3 \leq d_2(S) \leq 2q-1 \leq (s+1)^2,$$

yielding $s \leq 5$. But if $s = 3$ or 5, then any nontrivial $\chi \in \mathrm{Irr}(S)$ of odd degree coprime to s is a Weil character, which has been ruled out in (ii).

Suppose now that $S = \mathrm{PSL}_4(s)$ or $\mathrm{PSU}_4(s)$. For $s \geq 5$ we have

$$(s-1)(s^3-1)/2 \leq d_2(S) \leq 2q-1 \leq (s+1)^3 s/3,$$

which is possible only when $s \leq 11$. Thus $3 \leq s \leq 11$, whence $f^* = 1$, and so

$$(s-1)(s^3-1)/2 \leq d_2(S) \leq 2q-1 \leq (s+1)^3,$$

leading to $s = 3$. If $s = 3$, then any odd-order element in G has order ≤ 13, whereas $d(S) = 21$, contradicting (9.9.3).

To finish off type A, assume now that $S = \mathrm{PSL}_n(s)$ or $\mathrm{PSU}_n(s)$ with $n \geq 5$. Then (9.9.3)–(9.10.2) imply

$$\frac{(s^n+1)(s^{n-1}-s^2)}{(s+1)(s^2-1)} \leq d_2(S) \leq 2q-1 \leq (s+1)^{n-1}s/3,$$

whence

$$s^{2n-3} < (s+1)^n s/3 < s^{51n/40}$$

(because $(s+1)/s \leq 4/3 < 3^{11/40}$), a contradiction as $n \geq 5$.

(iv) Suppose $S = P\Omega^{\pm}_{2n}(s)$ with $n \geq 4$. For $n \geq 5$ we get that

$$\frac{(s^n-1)(s^{n-1}-s)}{s^2-1} \leq d(S) \leq 2q-1 \leq (s+1)^n f \leq (s+1)^n s/3,$$

whence

$$s^{2n-3.1} < (s+1)^n s/3 < s^{51n/40},$$

a contradiction. If $n = 4$, then $S = P\Omega^-_8(s)$. In this case, since $2q-1$ is coprime to $2s$, [Lübeck 2007] implies that

$$2q-1 \geq (s^4+1)(s^2-s+1)/2 > (s+1)^4 s/3,$$

again a contradiction.

Suppose $S = \mathrm{PSp}_{2n}(s)$ with $n \geq 2$ or $\Omega_{2n+1}(s)$ with $n \geq 3$. Using the bound $2q-1 \geq d_2(S)$ for $S = \mathrm{PSp}_{2n}(s)$ and $2q-1 \geq d(S)$ otherwise, we get for $n \geq 3$ that

$$\frac{(s^n-1)(s^n-s)}{s^2-1} \leq 2q-1 \leq (s+1)^n f \leq (s+1)^n s/3,$$

whence

$$s^{2n-2.1} < (s+1)^n s/3 < s^{51n/40},$$

a contradiction. If $n = 2$, then $S = \mathrm{PSp}_4(s)$, and we have

$$s(s-1)^2 \leq 2q-1 \leq (s+1)^2 s/3,$$

forcing $q \leq 9$. If $5 \leq q \leq 9$, then since the degree $2q-1 = \dim V$ is coprime to $2s$, we again get $2q-1 > 300 \geq (s+1)^2 s/3$. Finally, $\mathrm{PSp}_4(3)$ has no nontrivial non-Weil character of degree coprime to 6.

(v) If S is of type E_6, ${}^2E_6$, E_7, or E_8, then

$$(s^5+s)(s^6-s^3+1) \leq d(S) \leq 2q-1 \leq (s+1)^8 f \leq (s+1)^8 s/3,$$

a contradiction. Similarly, if $S = F_4(s)$, then

$$s^8 - s^4 + 1 = d(S) \leq 2q-1 \leq (s+1)^4 s/3,$$

which is impossible. Likewise, if $S = G_2(s)$ with $s \geq 5$, then

$$s^3 - 1 \leq d(S) \leq 2q - 1 \leq (s+1)^2 s/3,$$

again a contradiction. Next, if $S = G_2(3)$, then $2q - 1 \leq 16$ cannot be a degree of an irreducible character of S. Finally, if $S = {}^2G_2(s)$, then

$$s^2 - s + 1 = d(S) \leq 2q - 1 \leq (s+1)f \leq (s+1)s/3,$$

again a contradiction since $s \geq 27$. □

Our proof is now concluded by applying Theorem 9.7. □

References

[Conway et al. 1985] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *Atlas of finite groups: maximal subgroups and ordinary characters for simple groups*, Oxford University Press, Eynsham, 1985. MR Zbl

[Curtis and Reiner 1962] C. W. Curtis and I. Reiner, *Representation theory of finite groups and associative algebras*, Pure and Applied Mathematics **XI**, Interscience, New York, 1962. MR Zbl

[Deligne 1980] P. Deligne, "La conjecture de Weil, II", *Inst. Hautes Études Sci. Publ. Math.* **52** (1980), 137–252. MR Zbl

[Digne and Michel 1991] F. Digne and J. Michel, *Representations of finite groups of Lie type*, London Mathematical Society Student Texts **21**, Cambridge University Press, 1991. MR Zbl

[Fu 2010] L. Fu, "Calculation of ℓ-adic local Fourier transformations", *Manuscripta Math.* **133**:3-4 (2010), 409–464. MR Zbl

[Fulton and Harris 1991] W. Fulton and J. Harris, *Representation theory: a first course*, Graduate Texts in Mathematics **129**, Springer, 1991. MR Zbl

[GAP 2004] *GAP — Groups, Algorithms, and Programming, Version 4.4*, The GAP Group, 2004, available at https://www.gap-system.org.

[Gorenstein et al. 1998] D. Gorenstein, R. Lyons, and R. Solomon, *The classification of the finite simple groups*, vol. 3, Mathematical Surveys and Monographs **40**, American Mathematical Society, Providence, RI, 1998. MR Zbl

[Gross 2010] B. H. Gross, "Rigid local systems on $\mathbb{G}_m$ with finite monodromy", *Adv. Math.* **224**:6 (2010), 2531–2543. MR Zbl

[Grothendieck 1968] A. Grothendieck, "Formule de Lefschetz et rationalité des fonctions L", pp. 31–45 in *Dix exposés sur la cohomologie des schémas*, Adv. Stud. Pure Math. **3**, North-Holland, Amsterdam, 1968. MR Zbl

[Guralnick and Tiep 2008] R. M. Guralnick and P. H. Tiep, "Symmetric powers and a problem of Kollár and Larsen", *Invent. Math.* **174**:3 (2008), 505–554. MR Zbl

[Howe 1990] R. Howe, "Another look at the local θ-correspondence for an unramified dual pair", pp. 93–124 in *Festschrift in honor of I. I. Piatetski-Shapiro on the occasion of his sixtieth birthday, I* (Ramat Aviv, Israel, 1989), Israel Math. Conf. Proc. **2**, Weizmann, Jerusalem, 1990. MR Zbl

[Jordan 1872] C. Jordan, "Recherches sur les substitutions", *Journal de Mathématiques Pures et Appliquées* **17** (1872), 351–367. Zbl

[Katz 1988] N. M. Katz, *Gauss sums, Kloosterman sums, and monodromy groups*, Annals of Mathematics Studies **116**, Princeton University Press, 1988. MR Zbl

[Katz 1989] N. M. Katz, "Perversity and exponential sums", pp. 209–259 in *Algebraic number theory*, Adv. Stud. Pure Math. **17**, Academic Press, Boston, 1989. MR Zbl

[Katz 1990] N. M. Katz, *Exponential sums and differential equations*, Annals of Mathematics Studies **124**, Princeton University Press, 1990. MR Zbl

[Katz 1996] N. M. Katz, *Rigid local systems*, Annals of Mathematics Studies **139**, Princeton University Press, 1996. MR Zbl

[Katz 2004] N. M. Katz, "Notes on G_2, determinants, and equidistribution", *Finite Fields Appl.* **10**:2 (2004), 221–269. MR Zbl

[Katz 2005] N. M. Katz, *Moments, monodromy, and perversity: a Diophantine perspective*, Annals of Mathematics Studies **159**, Princeton University Press, 2005. MR Zbl

[Katz 2018] N. Katz, "Rigid local systems on $\mathbb{A}^1$ with finite monodromy", preprint, 2018, available at http://www.math.princeton.edu/~nmk/gpconj114.pdf. With an appendix by P. H. Tiep. To appear in *Mathematika*.

[Katz and Sarnak 1999] N. M. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, American Mathematical Society Colloquium Publications **45**, American Mathematical Society, Providence, RI, 1999. MR Zbl

[Lübeck 2007] F. Lübeck, "Character degrees and their multiplicities for some groups of Lie type of rank < 9", web site, 2007, available at http://www.math.rwth-aachen.de/~Frank.Luebeck/chev/DegMult/index.html.

[Panyushev 2004] D. I. Panyushev, "Weight multiplicity free representations, $\mathfrak{g}$-endomorphism algebras, and Dynkin polynomials", *J. London Math. Soc.* (2) **69**:2 (2004), 273–290. MR Zbl

[Rasala 1977] R. Rasala, "On the minimal degrees of characters of S_n", *J. Algebra* **45**:1 (1977), 132–181. MR Zbl

[Raynaud 1994] M. Raynaud, "Revêtements de la droite affine en caractéristique $p > 0$ et conjecture d'Abhyankar", *Invent. Math.* **116**:1-3 (1994), 425–462. MR Zbl

[Serre 2003] J.-P. Serre, "On a theorem of Jordan", *Bull. Amer. Math. Soc.* (*N.S.*) **40**:4 (2003), 429–440. MR Zbl

[Tiep 2003] P. H. Tiep, "Low dimensional representations of finite quasisimple groups", pp. 277–294 in *Groups, combinatorics & geometry* (Durham, NC, 2001), edited by A. A. Ivanov et al., World Sci. Publ., River Edge, NJ, 2003. MR Zbl

[Tiep and Zalesskii 1996] P. H. Tiep and A. E. Zalesskii, "Minimal characters of the finite classical groups", *Comm. Algebra* **24**:6 (1996), 2093–2167. MR Zbl

Received 5 Oct 2017. Revised 3 Apr 2018.

ROBERT M. GURALNICK:
guralnic@usc.edu
Department of Mathematics, University of Southern California, Los Angeles, CA 90089-2532, United States

NICHOLAS M. KATZ:
nmk@math.princeton.edu
Department of Mathematics, Princeton University, Fine Hall, Princeton, NJ 08544-1000, United States

PHAM HUU TIEP:
tiep@math.rutgers.edu
Department of Mathematics, Rutgers University, Piscataway, NJ 08854-8019, United States

msp

TUNISIAN JOURNAL OF MATHEMATICS
Vol. 1, No. 3, 2019
dx.doi.org/10.2140/tunis.2019.1.321

Local estimates for Hörmander's operators with Gevrey coefficients and application to the regularity of their Gevrey vectors

Makhlouf Derridj

Given a general Hörmander's operator $P = \sum_{j=1}^{m} X_j^2 + Y + b$ in an open set $\Omega \subset \mathbb{R}^n$, where $Y, X_1, \dots, X_m$ are smooth real vector fields in Ω, $b \in C^\infty(\Omega)$, and given also an open, relatively compact set Ω_0 with $\overline{\Omega}_0 \subset \Omega$, and $s \in \mathbb{R}$, $s \geq 1$, such that the coefficients of P are in $G^s(\Omega_0)$ and P satisfies a $\frac{1}{p}$-Sobolev estimate in Ω_0, our aim is to establish local estimates reflecting local domination of ordinary derivatives by powers of P, in Ω_0. As an application, we give a direct proof of the $G^{2ps}(\Omega_0)$-regularity of any $G^s(\Omega_0)$-vector of P.

1. Introduction

The study of the regularity of analytic vectors of partial differential operators goes back to the work of T. Kotake and N. S. Narasimhan [1962] who proved the (local) analyticity of (local) analytic vectors of elliptic operators with analytic coefficients (see also [Nelson 1959] for another related context). This property, called the "iteration property" or even the "Kotake–Narasimhan property" was further studied in the following decades in more general situations (such as systems, or nonelliptic operators) and also in the Gevrey categories G^s, $s \geq 1$ ($s = 1$ corresponds to the analytic case). This was in particular the case for the class of differential operators of principal type with analytic coefficients and also, after the famous article by L. Hörmander on hypoelliptic operators of second order, the systems of real-analytic real vector fields satisfying the so-called Hörmander condition, and

MSC2010: 35B65, 35G99, 35J70, 35K65.
Keywords: Gevrey vectors, Degenerate elliptic-parabolic differential operators.

also for the Hörmander's operators themselves. A result of G. Métivier [1978] shows that the "iteration property" is not true for nonelliptic operators in the Gevrey category G^s, $s > 1$, and another one by M.S. Baouendi and Métivier [1982] gave the "iteration property" for hypoelliptic operators of principal type in the analytic setting. Concerning analytic real vector fields, satisfying Hörmander's condition, that property was shown by M. Damlakhi and B. Helffer [1980], followed by a more precise version by Helffer and C. Mattera [1980].

More recently, a series of papers studying the case of involutive systems of analytic complex vector fields, concerning analytic or more generally Gevrey vectors, have been published [Barostichi et al. 2011; Castellanos et al. 2013]. Even knowing that the "iteration property" is not true for nonelliptic operators, one can ask about the Gevrey regularity $G^{s'}$ of an s-Gevrey vector ($s \geq 1$) and even give the best s' one may obtain. Such a study is contained in the above mentioned papers. A more recent paper by N. Braun Rodrigues, G. Chinni, P. Cordaro and M. Jahnke [Braun Rodrigues et al. 2016] was partly devoted to the study of global analytic vectors for some sums of squares on a product of two tori. A little later, we studied in [Derridj $\geq$ 2019] the case of $G^k(\Omega)$-vectors of Hörmander's operators of the first kind (or degenerate elliptic) with $G^k(\Omega)$ coefficients, Ω an open set in $\mathbb{R}^n$ (see the definitions in Section 2), and $k \in \mathbb{N}^*$ (in particular analytic vectors for $k = 1$), in which we proved an optimal result (the optimality following from the work in the global case by the authors of [Braun Rodrigues et al. 2016]).

In this paper, we consider general Hörmander's operators P (of the second kind or degenerate elliptic parabolic) for which we study the existence of local estimates giving local domination of the ordinary derivatives by powers of P, when the coefficients of P are in $G^s(\Omega_0)$, $\overline{\Omega}_0 \subset \Omega$, and P satisfies a "$\frac{1}{p}$-Sobolev estimate" on Ω_0 (see Section 2). This, with our main result Theorem 4.2, is used to obtain $G^{2ps}(\Omega_0)$-regularity for $G^s(\Omega_0)$-vectors of P ($s \geq 1$), providing therefore, first a direct proof, without using the method of addition of an extra variable, and second the result, with any $s \in \mathbb{R}$, $s \geq 1$. Let us remark that, in our preceding result, as we used the above method, our result was obtained there for $s = k \in \mathbb{N}^*$.

In a forthcoming paper, we will consider operators of the first kind, for which the integer p (giving the $\frac{1}{p}$-Sobolev estimate in Ω_0) is intimately related to the vector fields $(X_1, \dots, X_m)$ and prove finer local estimates of domination by powers of P, giving, as an application, the optimal $G^{ps}(\Omega_0)$ regularity for any $G^s(\Omega_0)$-vector of P. A complete survey on results in this field until 1987 may be found in [Bolley et al. 1987] and a more very recent short one may be found in [Derridj 2017].

We recall in Section 2 some definitions and elementary facts about the operators P, Gevrey functions and Gevrey vectors. Section 3 will be devoted to preliminary lemmas and propositions as a preparation for the proof of our main theorem.

Our main theorem will be proved in Section 4 and the last section is devoted to the application of the main result to the regularity of Gevrey vectors of P.

2. Some notation and definitions

We consider a system $(Y, X_1, \dots, X_m)$ of real vector fields with smooth coefficients on an open set $\Omega \subset \mathbb{R}^n$, and

$$P = \sum_{j=1}^{m} X_j^2 + Y + b, \quad b \in C^\infty(\Omega), \tag{2-1}$$

see [Hörmander 1967; Kohn 1978; Rothschild and Stein 1976]. Let us just recall Hörmander's condition for hypoellipticity in Ω of the operator (2-1):

> The Lie algebra, $\mathrm{Lie}(Y, X_1, \dots, X_m)$, generated by the vector fields $Y, X_1, \dots, X_m$, is of maximal rank in Ω. (2-2)

Concretely, the family of brackets of all lengths of the vector fields $Y, X_1, \dots, X_m$ span at any point $x \in \Omega$ the tangent space at x.

Under the condition (2-2), Hörmander proved the following a priori subelliptic estimate which we briefly describe. The L^2-norm and Sobolev H^σ-norm are denoted by $\|\cdot\|$ and $\|\cdot\|_\sigma$.

Let us recall, below, some norms introduced by Hörmander [1967]:

$$\begin{aligned} |||v||| &= \|v\| + \textstyle\sum_{j=1}^{m} \|X_j v\|, \quad v \in \mathcal{D}(\Omega), \\ \|v\|' &= \sup\bigl\{ \|(v, w)\| : |||w||| < 1,\ w \in \mathcal{D}(\Omega) \bigr\} \le \|v\|. \end{aligned} \tag{2-3}$$

Theorem 2.1 [Hörmander 1967]. *Let $\Omega_0 \Subset \Omega$ and assume (2-2). Then there exist $\sigma > 0$ and $C > 0$ such that*

$$\|v\|_\sigma \le C(|||v||| + \|Yv\|'), \quad \text{for all } v \in \mathcal{D}(\Omega_0). \tag{2-4}$$

We call (2-4) a subelliptic estimate for P.

The constants σ and C depend on Ω_0 and $Y, X_1, \dots, X_m$. More specifically, σ depends on the length of the brackets needed in order to span the tangent space at every point of $\overline{\Omega}_0$.

Now, elementary technical manipulations give, with $C_0 > 0$,

$$\|v\|_\sigma \le C_0(\|Pv\|' + \|v\|), \quad \text{for all } v \in \mathcal{D}(\Omega_0). \tag{2-5}$$

A particular ingredient in order to get (2-5) and which we need in the sequel is the set of obvious inequalities:

$$\|X_j v\|' \le C\|v\|, \quad \forall v \in \mathcal{D}(\Omega), \text{ for some } C > 0,\ j = 1, \dots, m. \tag{2-6}$$

We want to say a word on the case $Y = 0$, or more generally the case where in (2-2) one considers the Lie algebra $\mathrm{Lie}(X_1, \dots, X_m)$ generated by $X_1, \dots, X_m$; in that case, one has a more precise estimate. We considered that case in our preceding work, obtaining an optimal result for the Gevrey regularity of k-Gevrey vectors of P (named in that case Hörmander's operator of the first kind), $k \in \mathbb{N}^*$.

Coming back to our general case (named operator of the second kind), we need to write a more precise a priori estimate than (2-5) which we need to consider in the sequel:

$$\|v\|_\sigma + \sum_{j=1}^{m} \|X_j v\| \le C_0(\|Pv\|' + \|v\|), \quad \text{for all } v \in \mathscr{D}(\Omega_0). \tag{2-7}$$

Again this is easily obtained, using (2-4).

Let us recall, in order to be complete, definitions of Gevrey functions and Gevrey vectors of a differential operator of order m (here it will be $m = 2$).

Definition 2.2. Let $s \ge 1$. The space of Gevrey functions of order s, $G^s(\Omega)$, is defined as

$$G^s(\Omega) := \Big\{u \in C^\infty(\Omega) : \forall K \Subset \Omega,\ \exists C_K > 0 \text{ s.t. } |\partial^\alpha u|_K \le C_K^{|\alpha|+1} |\alpha|!^s \\ \forall \alpha \in \mathbb{N}^n,\ |\alpha| = \textstyle\sum_{j=1}^n \alpha_j,\ \partial^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n}\Big\}. \tag{2-8}$$

Remark 2.3. It is known that for $s > 1$, one has the property of partition of unity in $G^s(\Omega)$: in particular, given two open sets Ω_1, Ω_2, with $\overline{\Omega}_1$ compact and $\overline{\Omega}_1 \subset \Omega_2$, there exists a function $\varphi \in \mathscr{D}(\Omega_2)$, $\varphi \equiv 1$ on $\overline{\Omega}_1$, $\varphi \in G^s(\Omega_2)$.

Definition 2.4. Let P be a differential operator of order m in Ω. The space $G^s(\Omega, P)$ of s-Gevrey vectors of P in Ω, $s \ge 1$, is defined as

$$G^s(\Omega, P) := \Big\{u \in L^2_{\mathrm{loc}}(\Omega) : \forall K \Subset \Omega,\ \exists C_K > 0 \text{ s.t. } \forall k \in \mathbb{N}, \\ P^k u \in L^2(K) \text{ and } \|P^k u\|_{L^2(K)} \le C_K^{k+1} (mk)!^s\Big\}. \tag{2-9}$$

As in our case, P is of order 2 and hypoelliptic, with a subelliptic estimate (2-5) or (2-7), (2-9) reduces to

$$G^s(\Omega, P) := \Big\{u \in C^\infty(\Omega) : \forall K \Subset \Omega,\ \exists C_K > 0 \\ \text{s.t. } \forall k \in \mathbb{N},\ \|P^k u\|_{L^2(K)} \le C_K^{k+1} (2k)!^s\Big\}. \tag{2-10}$$

Remark 2.5. We used in our definitions (2-9), (2-10), the commonly used L^2-norm, but in some specific situations, such as for systems of complex vectors, other norms are used [Barostichi et al. 2011; Castellanos et al. 2013].

3. Preliminary lemmas and propositions

When trying to get estimates for derivatives $\partial^\alpha u$ of a function u, knowing Pu, we are faced in particular with the study of commutators $[P, \partial^\alpha]$, $\alpha \in \mathbb{N}^n$, so, to the study of commutators $[X_j^2, \partial^\alpha]$, $[Y+b, \partial^\alpha]$. Now, one has the following equality:

$$[X_j^2, \partial^\alpha] = 2X_j[X_j, \partial^\alpha] - [X_j, [X_j, \partial^\alpha]]. \tag{3-1}$$

So we have to look closely at the commutators $[X_j, \partial^\alpha]$, $[Y+b, \partial^\alpha]$ and the double commutators $[X_j, [X_j, \partial^\alpha]]$. In order to get an estimate for the coefficients of the differential operators above, it is sufficient to consider the following basic commutators

$$\begin{aligned} &[a_\ell \partial_\ell, \partial^\alpha], \quad \ell = 1, \dots, n, \qquad a_\ell \in G^s(\Omega), \\ &[a_\ell \partial_\ell, [a_k \partial_k, \partial^\alpha]], \quad \ell, k = 1, \dots, n, \quad a_\ell, a_k \in G^s(\Omega), \end{aligned} \tag{3-2}$$

as Y and X_j are linear combinations of the basic vector fields $a_\ell \partial_\ell$, $a_\ell \in G^s(\Omega)$.

Of course, the commutator $[b, \partial^\alpha]$ is elementary. Then

$$\begin{aligned} &[b, \partial^\alpha] = -\textstyle\sum_{\beta<\alpha} \binom{\alpha}{\beta} b^{(\alpha-\beta)} \partial^\beta = \sum_{\beta<\alpha} b_{\alpha\beta} \partial^\beta, \\ &\qquad \textstyle\binom{\alpha}{\beta} = \prod_j \binom{\alpha_j}{\beta_j}, \ \beta < \alpha \ \Leftrightarrow \beta_j \le \alpha_j, \ \beta \ne \alpha, \\ &[a_\ell \partial_\ell, \partial^\alpha] = -\textstyle\sum_{\beta<\alpha} \binom{\alpha}{\beta} a_\ell^{(\alpha-\beta)} \partial^{\beta+\ell}, \\ &\qquad \beta + \ell = \big\{(\beta+\ell)_i = \beta_i, \ i \ne \ell, \ (\beta+\ell)_\ell = \beta_\ell + 1\big\}. \end{aligned} \tag{3-3}$$

Let us, now, give the expression of $[a_\ell \partial_\ell, [a_k \partial_k, \partial^\alpha]]$ or, in view of (3-3),

$$\sum_{\beta<\alpha} \binom{\alpha}{\beta} [a_k^{(\alpha-\beta)}, \partial^{\beta+k}, a_\ell \partial_\ell].$$

But for $\beta < \alpha$,

$$[a_k^{(\alpha-\beta)} \partial^{\beta+k}, a_\ell \partial_\ell] = \sum_{\gamma<\beta+k} \binom{\beta+k}{\gamma} a_k^{(\alpha-\beta)} a_\ell^{(\beta+k-\gamma)} \partial^{\gamma+\ell} - a_\ell a_k^{(\alpha-\beta+\ell)} \partial^{\beta+k}.$$

Hence we get

$$\begin{aligned} &[a_\ell \partial_\ell, [a_k \partial_k, \partial^\alpha]] \\ &\quad = \sum_{\substack{\gamma<\beta+k, \\ \beta<\alpha}} \binom{\alpha}{\beta} \binom{\beta+k}{\gamma} a_k^{(\alpha-\beta)} a_\ell^{(\beta+k-\gamma)} \partial^{\gamma+\ell} - \sum_{\beta<\alpha} \binom{\alpha}{\beta} a_\ell a_k^{(\alpha-\beta+\ell)} \partial^{\beta+k}. \end{aligned} \tag{3-4}$$

In the first sum in the second member of (3-4), we distinguish two families of γ's such that $\gamma < \beta + k$:

(i) If $\gamma_k = 0$, then $\gamma \le \beta$, so $\gamma < \alpha$ (as $\beta < \alpha$). Hence in that case

$$\begin{aligned}\textstyle\sum_{\beta<\alpha,\,\gamma<\beta+k,\,\gamma_k=0}\binom{\alpha}{\beta}\binom{\beta+k}{\gamma}a_k^{(\alpha-\beta)}a_\ell^{(\beta+k-\gamma)}\partial^{\gamma+\ell} \quad &\text{is a part of} \\ \textstyle\sum_{\beta<\alpha,\,\gamma\le\beta}\binom{\alpha}{\beta}\binom{\beta+k}{\gamma}a_k^{(\alpha-\beta)}a_\ell^{(\beta+k-\gamma)}\partial^{\gamma+\ell}.&\end{aligned} \tag{3-5}$$

(ii) If $\gamma_k \neq 0$, we set $\delta = (\delta_1, \dots, \delta_n)$ with $\delta_\rho = \gamma_\rho$ if $\rho \neq k$ and $\delta_k = \gamma_k - 1$. So $\delta < \beta < \alpha$, in particular $|\delta| \le |\alpha| - 2$ (easy to see) and $\gamma = \delta + k$. Hence

$$\begin{aligned}\textstyle\sum_{\beta<\alpha,\,\gamma<\beta+k,\gamma_k\neq0}\binom{\alpha}{\beta}\binom{\beta+k}{\gamma}a_k^{(\alpha-\beta)}a_\ell^{(\beta+k-\gamma)}\partial^{\gamma+\ell} \quad &\text{is a part of} \\ \textstyle\sum_{\beta<\alpha,\,\delta\le\beta}\binom{\alpha}{\beta}\binom{\beta+k}{\delta+k}a_k^{(\alpha-\beta)}a_\ell^{(\beta-\delta)}\partial^{\delta+\ell+k}.&\end{aligned} \tag{3-6}$$

Setting $I_{(\ell,k)} = (I_1, \dots, I_n)$ with $I_\rho = 0$ if $\rho \notin (\ell, k)$ and $I_\rho = 1$ if $\rho \in \{\ell, k\}$, the sum in the second line of (3-6) can be written as

$$\sum_{\substack{\beta<\alpha\\ \delta<\beta}}\binom{\alpha}{\beta}\binom{\beta+k}{\delta+k}a_k^{(\alpha-\beta)}\partial^{\delta+I_{(\ell,k)}} \tag{3-7}$$

So, looking at (3-4) and in view of (3-5), (3-6) and taking as 0 the other coefficients in $\sum_{\beta<\alpha,\,\gamma\le\beta}$ (in (3-5)) and $\sum_{\beta<\alpha,\,\delta<\beta}$ (in (3-6)) which are not in (3-4), we may write

$$\begin{aligned}&[a_\ell\partial_\ell, [a_k\partial_k, \partial^\alpha]] \\ &\qquad = -\sum_{\beta<\alpha} a_{\ell k\alpha\beta}\partial^{\beta+k} + \sum_{\gamma<\alpha} b_{\ell k\alpha\gamma}\partial^{\gamma+\ell} + \sum_{\substack{\delta<\alpha\\ |\delta|\le|\alpha|-2}} c_{\ell k\alpha\delta}\partial^{\delta+I_{(\ell,k)}}, \\ &\qquad\qquad \text{where}\quad a_{\ell k\alpha\beta} = \tbinom{\alpha}{\beta}a_\ell a_k^{(\alpha-\beta+\ell)}, \\ &\qquad\qquad\qquad b_{\ell k\alpha\gamma} = \textstyle\sum_{\gamma\le\beta<\alpha}\tbinom{\alpha}{\beta}\tbinom{\beta+k}{\gamma}a_k^{(\alpha-\beta)}a_\ell^{(\beta+k-\gamma)}, \\ &\qquad\qquad\qquad c_{\ell k\alpha\delta} = \textstyle\sum_{\delta<\beta<\delta}\tbinom{\alpha}{\beta}\tbinom{\beta+k}{\delta+k}a_k^{(\alpha-\beta)}a_\ell^{(\beta-\delta)}.\end{aligned} \tag{3-8}$$

Now considering a vector field X_j with smooth coefficients $a_{j\ell}$, i.e.,

$$X_j = \sum_{\ell=1}^{n} a_{j\ell}\partial_\ell = \sum_{\ell=1}^{n} Z_{j\ell},$$

we obtain from (3-3)

$$\begin{aligned}[X_j, \partial^\alpha] &= \sum_{\ell=1}^{n}[Z_{j\ell}, \partial^\alpha] = -\sum_{\substack{\beta<\alpha\\ \ell=1,\dots,n}}\binom{\alpha}{\beta}a_{j\ell}^{(\alpha-\beta)}\partial^{\beta+\ell} = \sum_{\substack{\beta<\alpha\\ \ell=1,\dots,n}} a_{j\alpha\beta\ell}\partial^{\beta+\ell}, \\ [Y, \partial^\alpha] &= \sum_{\ell=1}^{n}[b_\ell\partial_\ell, \partial^\alpha] = -\sum_{\substack{\beta<\alpha\\ \ell=1,\dots,n}}\binom{\alpha}{\beta}b_\ell^{(\alpha-\beta)}\partial^{\beta+\ell} = \sum_{\substack{\beta<\alpha\\ \ell=1,\dots,n}} b_{\alpha\beta\ell}\partial^{\beta\ell}.\end{aligned} \tag{3-9}$$

Concerning the double brackets, we obtain:

$$[X_j,[X_j\partial^\alpha]]=\textstyle\sum_{\ell,k}[Z_{j\ell},[Z_{jk},\partial^\alpha]],\quad \text{with } [Z_{j\ell},[Z_{k\ell},\partial^\alpha]] \text{ given in (3-8)},$$
$$\text{where } a_{\ell k\alpha\beta} \text{ is replaced by } a_{j\ell k\alpha\beta}, \text{ and so on.} \tag{3-10}$$

Now we assume that the coefficients of P are in $G^s(\Omega)$. So $a_{j\ell}$, b_ℓ and b are in $G^s(\Omega)$. So we have that for any compact K in Ω, there exists $C_K>0$ such that for all $\nu\in\mathbb{N}^n$,

$$|a^{(\nu)}_{j\ell}|_K+|b^{(\nu)}_\ell|_K+|b^{(\nu)}|_K\le C_K^{|\nu|+1}|\nu|!^s,\quad \ell\in\{1,\dots,n\},\ j\in\{1,\dots,m\}. \tag{3-11}$$

Writing (3-10) more concretely, we get using (3-8)

$$[X_j,[X_j,\partial^\alpha]]=\sum_{\substack{\beta<\alpha\\ k=1,\dots,n}} d_{jk\alpha\beta}\partial^{\beta+k}+\sum_{\substack{\beta<\alpha\\ k,\ell=1,\dots,n}} c_{j\ell k\alpha\beta}\partial^{\beta+\ell+k},$$
$$\text{where } d_{jk\alpha\beta}=-\textstyle\sum_{\ell=1}^n a_{j\ell k\alpha\beta}+\sum_{\ell=1}^n b_{jk\ell\alpha\beta}. \tag{3-12}$$

We want now to get estimates for the coefficients of the brackets and double brackets in (3-9) and (3-12) and the operators associated to these brackets, when the coefficients are in $G^s(\Omega)$.

Proposition 3.1. *Assume the coefficients of P are in $G^s(\Omega)$. Given any compact set K in Ω, there exists $B=B_K>0$ such that the coefficients of the operators* $[b,\partial^\alpha]$ (*given in* (3-3)), $[X_j,\partial^\alpha]$, $[Y,\partial^\alpha]$ (*given in* (3-9)), $[X_j,[X_j,\partial^\alpha]]$ (*given in* (3-12)) *satisfy the following estimates*:

$$\begin{aligned}|b_{\alpha\beta}|_K+|b_{\alpha\beta\ell}|_K+|a_{j\alpha\beta\ell}|_K+|\nabla b_{\alpha\beta}|_K&\\ +|\nabla b_{\alpha\beta\ell}|_K+|\nabla a_{j\alpha\beta\ell}|_K&\le B^{|\alpha-\beta|}\Big(\tfrac{\alpha!}{\beta!}\Big)^s,\\ |c_{j\ell\alpha\beta}|_K+|\nabla c_{j\ell k\alpha\beta}|_K&\le B^{|\alpha-\beta|}\Big(\tfrac{(\alpha+k)!}{(\beta+k)!}\Big)^s,\\ |d_{jk\alpha\beta}|_K+|\nabla d_{jk\alpha\beta}|_K&\le B^{|\alpha-\beta|}\Big((|\alpha|+1)\tfrac{\alpha!}{\beta!}\Big)^s.\end{aligned} \tag{3-13}$$

Proof. We recall first that $\beta<\alpha$ and $(\alpha+k)_\ell=\alpha_\ell$ for $\ell\ne k$ and $(\alpha+k)_k=\alpha_k+1$. The first line comes easily from the expression of the functions $b_{\alpha\beta}$, $b_{\alpha\beta\ell}$, $a_{j\alpha\beta\ell}$, and their derivatives. Note that we took $B^{|\alpha-\beta|}$ in place of $B^{|\alpha-\beta|+1}$ as $|\alpha-\beta|\ge 1$ and used (3-11). The proof for the other functions needs more work. We first use the following estimate for $a^{(\alpha-\beta)}_{jk}a^{(\beta-\delta)}_{j\ell}$, which is in the expression of $c_{j\ell k\alpha\delta}$ (see last line in (3-8)):

$$\left|a^{(\alpha-\beta)}_{jk}a^{(\beta-\delta)}_{\ell}\binom{\alpha}{\beta}\binom{\beta+k}{\delta+k}\right|\le B^{|\alpha-\beta|}(\lambda B)^{|\beta-\delta|}\left(\frac{(\alpha+k)!}{(\delta+k)!}\right)^s,\quad \lambda\ge 1. \tag{3-14}$$

For that we used

$$\frac{\alpha!}{\beta!}\frac{(\beta+k)!}{(\delta+k)!}=\frac{\alpha!(\beta_k+1)}{(\delta+k)!}\le\frac{(\alpha+k)!}{(\delta+k)!}.$$

Hence

$$|c_{j\ell k\alpha\delta}|_K \le \Big(\sum_{\delta<\beta<\alpha} B^{|\alpha-\beta|}(\lambda B)^{|\beta-\delta|}\Big)\Big(\frac{(\alpha+k)!}{(\delta+k)!}\Big)^s, \quad \lambda \ge 1.$$

As $(\lambda B)^{|\beta-\delta|} = (\lambda B)^{|\alpha-\delta|}(\lambda B)^{-|\alpha-\beta|}$, $(\delta<\beta<\alpha)$, we get:

$$|c_{j\ell k\alpha\delta}|_K \le \Big(\frac{(\alpha+k)!}{(\delta+k)!}\Big)^s (\lambda B)^{|\alpha-\delta|} \sum_{\beta<\alpha} \lambda^{-|\alpha-\beta|}.$$

Now we use the following easy lemma:

Lemma 3.2. *There exists $\epsilon_0 > 0$ (independent from α) such that if $0<\epsilon\le\epsilon_0$ then $\sum_{\beta<\alpha}\epsilon^{|\alpha-\beta|}\le 1$.*

This comes from the fact that if $\beta_j \le \alpha_j$, $j=1,\dots,n$, and $\alpha\ne\beta$ then $\sum_{j=1}^n \beta_j < \sum_{j=1}^n \alpha_j$. In fact, setting $\lambda_j = \alpha_j - \beta_j \in \mathbb{N}$, $\sum_{j=1}^n \lambda_j \ge 1$,

$$\begin{aligned}\sum \epsilon^{\sum\lambda_j} &\le \sum_{\lambda_1\ge 1}\epsilon^{\sum\lambda_j} + \cdots + \sum_{\lambda_n\ge 1}\epsilon^{\sum\lambda_j}\\ &\le \epsilon\Big(\sum_{(\lambda_2,\dots,\lambda_n)\in\mathbb{N}^{n-1}}\epsilon^{\lambda_2+\dots+\lambda_n} + \cdots + \sum_{(\lambda_1,\dots,\lambda_{n-1})\in\mathbb{N}^{n-1}}\epsilon^{\lambda_1+\dots+\lambda_{n-1}}\Big)\\ &\le \epsilon n 2^{n-1} \quad \text{if } \epsilon \le \tfrac{1}{2}.\end{aligned}$$

The lemma follows by taking $\epsilon_0 = (n2^{n-1})^{-1}$.

Coming back to our proof, we consider $\lambda_0 = n2^{n-1} = \epsilon_0^{-1}$ and replace B by $\lambda_0 B$. We get the estimate for $c_{jk\alpha\beta}$. As $d_{jk\alpha\beta} = -\sum_{\ell=1}^n (a_{j\ell k\alpha\beta} - b_{jk\ell\alpha\beta})$, we just need to bound on K the functions $a_{j\ell k\alpha\beta}$ and $b_{jk\ell\alpha\beta}$ for $\ell = 1,\dots,n$. The worst term is $b_{jk\ell\alpha\gamma}$. As we did above, we use, with $\lambda\ge 1$,

$$\Big|\binom{\alpha}{\beta}\binom{\beta+\ell}{\gamma}a_\ell^{(\alpha-\beta)}a_k^{(\beta+\ell-\gamma)}\Big| \le B^{|\alpha-\beta|+1}(\lambda B)^{|\beta-\gamma|}\Big(\frac{\alpha!}{\gamma!}(\beta_\ell+1)\Big)^s.$$

So,

$$\begin{aligned}|b_{jk\ell\alpha\gamma}| &\le \Big(\sum_{\gamma\le\beta<\alpha} B^{|\alpha-\beta|+1}(\lambda B)^{|\beta-\gamma|}\Big)\Big(\frac{(\alpha+\ell)!}{\gamma!}\Big)^s\\ &\le \Big(\frac{(\alpha+\ell)!}{\gamma!}\Big)^s B(\lambda B)^{|\alpha-\gamma|}\sum_{\beta<\alpha}\lambda^{-|\alpha-\beta|}\\ &\le B\Big(\frac{\alpha!}{\gamma!}\Big)^s(|\alpha|+1)^s(\lambda B)^{|\alpha-\gamma|}\sum_{\beta<\alpha}\lambda^{-|\alpha-\beta|}.\end{aligned}$$

Now taking $\lambda\ge\lambda_0$, B large enough, we obtain what we want (more precisely we take $\tilde{B}$ such that (as $|\alpha-\gamma|\ge 1$), $B(\lambda B)^{|\alpha-\gamma|} \le (\lambda\tilde{B})^{|\alpha-\gamma|}$, and then choose the final B as $\lambda\tilde{B}$). Concerning the derivatives of first order, we just have to apply what we did, using bounds on K, not only for the coefficients of P, but also bounds of their derivatives of first order. In order to be rigorous and complete, let us bound a derivative of $b_{j\ell k\alpha\gamma}$, which we denote by $b'_{j\ell k\alpha\gamma}$. Then we have (see (3-8))

$$b'_{j\ell k\alpha\beta} = \sum_{\gamma\le\beta<\alpha} \binom{\alpha}{\beta}\binom{\beta+k}{\gamma}\big((a_{jk}^{(\alpha-\beta)})' a_{j\ell}^{(\beta+k-\gamma)} + a_{jk}^{(\alpha-\beta)}(a_{j\ell}^{(\beta+k-\gamma)})'\big).$$

So we just have to do the same as before to get the bounds on K for the functions a'_{jk}, $a_{j\ell}$, a_{jk} and $a'_{j\ell}$. Hence taking B, greater if necessary, we obtain (3-13). ☐

As a consequence, we obtain the following:

Proposition 3.3. *Assume that the coefficients of P are in $G^s(\Omega)$. Then for every relatively compact open set Ω_0 in Ω ($\Omega_0 \Subset \Omega$), there exists $B = B(\Omega_0, P) > 0$ such that, for every τ, $0 \le \tau \le 1$, one has for $j = 1, \dots, m$, $\beta < \alpha$, $\ell, k \in \{1, \dots, n\}$,*

$$\begin{aligned}
&\|b_{\alpha\beta}v\|_\tau + \|b_{\alpha\beta\ell}v\|_\tau + \|a_{j\alpha\beta\ell}v\|_\tau \\
&\qquad\qquad \le B^{|\alpha-\beta|}\Big(\frac{\alpha!}{\beta!}\Big)^s \|v\|_\tau, \quad \beta<\alpha,\ \ell=1,\dots,n,\\
&\|c_{j\ell k\alpha\beta}v\|_\tau \le B^{|\alpha-\beta|}\Big(\frac{(\alpha+k)!}{(\beta+k)!}\Big)^s \|v\|_\tau, \\
&\|d_{jk\alpha\beta}v\|_\tau \le B^{|\alpha-\beta|}\Big(\frac{\alpha!(|\alpha|+1)}{\beta!}\Big)^s \|v\|_\tau, \quad \forall v \in \mathcal{D}(\Omega_0).
\end{aligned} \tag{3-15}$$

Proof. From inequalities (3-13), one obtains estimates (3-15) for $\tau = 0$ and $\tau = 1$. Then (3-15) follows from the cases $\tau = 0$ and $\tau = 1$ by interpolation between Sobolev spaces L^2 and H^1. ☐

Remark 3.4. Our Proposition 3.1 is a refinement of our Lemma 5.3 in [Derridj ≥ 2019]. We need this refinement in order to prove our local estimates of ordinary derivatives in terms of powers of P.

In order to begin to state what we need for our results, we make the following assumption:

$$\text{Estimate (2-4) is true with} \quad \sigma = \frac{1}{p}, \quad p \in \mathbb{N}^*. \tag{3-16}$$

Then our local estimates will use the equality $\sigma = \frac{1}{p}$. In this section, we want to give and prove the basic ones which we will use in another section in order to give a sequence of local estimates for P, assuming (2-7), (3-16) and P with coefficients in $G^s(\Omega)$, for some $s \ge 1$.

Proposition 3.5. *Assume (2-4), (3-16) and that P has smooth coefficients. Let Ω_1 be a relatively compact open subset of Ω_0. Then there exists a constant $C = C(\Omega_0, P) > 0$ such that for all $(u, \varphi) \in C^\infty(\overline{\Omega}_1) \cap \mathcal{D}(\Omega_1)$,*

$$\|\varphi u\|_\sigma \le C\Big(\|\varphi P u\|' + \sum_{|\beta|\le 2} \|\varphi^{(\beta)} u\|\Big). \tag{3-17}$$

Proof. This proposition has a simple proof. Taking $v = \varphi u$ in (2-7), we get

$$\|\varphi u\|_\sigma + \sum_{j=1}^{m} \|X_j \varphi u\| \le C_0(\|P\varphi u\|' + \|\varphi u\|),$$

taking $\Omega_0 = \Omega_1$ and C_0 as in (2-7) related to $\overline{\Omega}_0 \Subset \Omega$, we have

$$\|P\varphi u\|' \le \|\varphi P u\|' + \|[P, \varphi]u\|', \tag{3-18}$$

with

$$\begin{aligned}\|[P, \varphi]u\|' &\le 2\sum_{j=1}^{m}\big(\|X_j[X_j, \varphi]u\|' + \|X_j^2(\varphi)u\|'\big) + \|Y(\varphi)u\|' \\ &\le 2\sum_{j=1}^{m}\big(\|X_j(\varphi)u\| + \|X_j^2(\varphi)u\|\big) + \|Y(\varphi)u\|, \quad \text{from (2-3), (2-6).}\end{aligned}$$

Now as the X_j, Y are smooth in Ω and $\overline{\Omega}_0 \subset \Omega$,

$$\|X_j(\varphi)u\| \le C_1 \sum_{|\beta|\le 1} \|\varphi^{(\beta)} u\|, \quad \|X_j^2(\varphi)u\| \le C_1 \sum_{|\beta|\le 2} \|\varphi^{(\beta)} u\|. \tag{3-19}$$

So, with a suitable C, (3-17) is obtained from (3-18), (3-19). □

In order to reach estimates for ordinary derivatives, one way is to try to obtain estimates for φu in the Sobolev spaces $H^{\ell\sigma}$, $\ell = 1, \dots, p$, (3-17) corresponding to $\ell = 1$. As the basic estimate (2-6) is with H^σ (i.e., $\ell = 1$), it is natural to use the following, recalling some notation and definitions in [Derridj ≥ 2019, Section 2]: let Ω_0, be such that $\overline{\Omega}_1 \subset \Omega_0 \subset \overline{\Omega}_0 \Subset \Omega$. So we consider $\psi \in \mathcal{D}(\Omega_0)$, $\psi = 1$ on $\overline{\Omega}_1$; one way to estimate $\|v\|_{\ell\sigma}$, knowing (2-7) is to consider $\psi T_{\ell\sigma} v$, $v \in \mathcal{D}(\Omega_1)$; in fact, for all $v \in \mathcal{D}(\Omega_1)$,

$$\|v\|_{(\ell+1)\sigma} = \|\psi v\|_{(\ell+1)\sigma} = \|T_{(\ell+1)\sigma}\psi v\| = \|T_{\ell\sigma}\psi v\|_\sigma,$$

(see [Derridj ≥ 2019, 2.11-2.13]),

where $\widehat{T_\rho w}(\xi) = (1 + \|\xi\|^2)^{\rho/2}\widehat{w}(\xi)$, for all $w \in \mathcal{D}(\mathbb{R}^n)$. Hence

$$\|v\|_{(\ell+1)\sigma} \le \|[T_{\ell\sigma}, \psi]v\|_\sigma + \|\psi T_{\ell\sigma} v\|_\sigma \le C_2\|v\|_{(\ell+1)\sigma-1} + \|\psi T_{\ell\sigma} v\|_\sigma.$$

Now for $\ell = 1, \dots, p-1$, $(\ell+1)\sigma - 1 \le 0$. So

$$\|v\|_{(\ell+1)\sigma} \le \tilde{C}\|v\| + \|\psi T_{\ell\sigma} v\|_\sigma, \quad \text{for all } v \in \mathcal{D}(\Omega_1). \tag{3-20}$$

So this boils down to applying (2-7) to $\psi T_{\ell\sigma} v \in \mathcal{D}(\Omega_0)$ using the constant C_0. Then we have

$$\|v\|_{(\ell+1)\sigma} \le \tilde{C}(\|v\| + \|\psi T_{\ell\sigma} v\|_\sigma) \le \tilde{C}\bigg(\|\psi T_{\ell\sigma} v\| + \sum_j \|X_j \psi T_{\ell\sigma} v\| + \|v\|\bigg).$$

So, we get:

$$\|\varphi u\|_{(\ell+1)\sigma} \le C_0\tilde{C}\big(\|P\psi T_{\ell\sigma}\varphi u\|' + \|\psi T_{\ell\sigma}\varphi u\| + \|\varphi u\|\big),$$

yielding

$$\|\varphi u\|_{(\ell+1)\sigma} \le C_3\big(\|P\psi T_{\ell\sigma}\varphi u\|' + \|\varphi u\|_{\ell\sigma}\big), \qquad \text{for all } (u,\varphi)\in C^\infty(\overline{\Omega}_1)\times\mathfrak{D}(\Omega_1). \tag{3-21}$$

Now we provide a suitable bound for $\|P\psi T_{\ell\sigma}\varphi u\|'$ as follows:

$$\begin{aligned}
&\|P\psi T_{\ell\sigma}\varphi u\|' \\
&\quad\le \|[P,\psi T_{\ell\sigma}]\varphi u\|' + \|\psi T_{\ell\sigma}[P,\varphi]u\|' + \|\psi T_{\ell\sigma}\varphi P u\|' \\
&\quad\le \sum_{j=1}^m \big(2\|X_j[X_j,\psi T_{\ell\sigma}]\varphi u\|' + \|[X_j,[X_j,\psi T_{\ell\sigma}]]\varphi u\|\big) + \|[Y+b,\psi T_{\ell\sigma}]\varphi u\| \\
&\qquad + \sum_{j=1}^m \big(2\|\psi T_{\ell\sigma}X_j[X_j,\varphi]u\|' + \|\psi T_{\ell\sigma}[X_j[X_j,\varphi]]u\|'\big) \\
&\qquad\qquad + \|\psi T_{\ell\sigma}[Y,\varphi]u\|' + \|\psi T_{\ell\sigma}\varphi P u\|'.
\end{aligned}$$

Using again inequalities in (2-3), (2-6), we get, with some constant C which may vary from line to another,

$$\begin{aligned}
&\|P\psi T_{\ell\sigma}\varphi u\|' \\
&\quad\le C\Bigg(\|\varphi u\|_{\ell\sigma} + \sum_{j=1}^m\Big(2\|[\psi T_{\ell\sigma},X_j][X_j,\varphi]u\|' + \|X_j\psi T_{\ell\sigma}[X_j,\varphi]u\|' \\
&\qquad\qquad + \|X_j^2(\varphi)u\|_{\ell\sigma}\Big) + \|Y(\varphi)u\|_{\ell\sigma}\Bigg) + \|\psi T_{\ell\sigma}\varphi P u\|' \\
&\quad\le C\Bigg(\|\psi T_{\ell\sigma}\varphi P u\|' + \sum_{j=1}^m\big(\|X_j(\varphi)u\|_{\ell\sigma} + \|X_j^2(\varphi)u\|_{\ell\sigma} + \|\varphi u\|_{\ell\sigma} \\
&\qquad\qquad + \|Y(\varphi)u\|_{\ell\sigma}\big)\Bigg).
\end{aligned} \tag{3-22}$$

Hence, in fact, we proved the following with Ω_0 as above:

Proposition 3.6. *Under the hypotheses in Proposition 3.5, there exists a constant $C = C(\Omega_0, P) > 0$ such that*

$$\|\varphi u\|_{(\ell+1)\sigma} \le C\Bigg(\|\varphi P u\|_{\ell\sigma} + \sum_{|\beta|\le 2}\|\varphi^{(\beta)}u\|_{\ell\sigma}\Bigg), \quad \ell = 0,\dots,p-1. \tag{3-23}$$

The next step is to obtain such estimates for couples $(\partial^\alpha u,\varphi)$ for $(u,\varphi)\in C^\infty(\overline{\Omega}_1)\times\mathfrak{D}(\Omega_1)$, $\alpha\in\mathbb{N}^n$. In order to get such an estimate, let us first consider the case $\ell = 0$. We use (3-17). So

$$\|\varphi\partial^\alpha u\|_\sigma \le C\|\varphi P\partial^\alpha u\|' + \sum_{|\beta|\le 2}\|\varphi^{(\beta)}\partial^\alpha u\|.$$

Here, what is new is to use $\varphi P\partial^\alpha u = \varphi[P, \partial^\alpha]u + \varphi\partial^\alpha Pu$:

$$\|\varphi P\partial^\alpha u\|' \le \|\varphi[P, \partial^\alpha]u\|' + \|\varphi\partial^\alpha Pu\|'$$
$$\le \sum_{j=1}^{m}\big(2\|\varphi X_j[X_j, \partial^\alpha]u\|' + \|\varphi[X_j, [X_j, \partial^\alpha]]u\|'\big) + \|\varphi[Y+b, \partial^\alpha]u\|' + \|\varphi\partial^\alpha Pu\|'. \quad (3\text{-}24)$$

Writing $\varphi X_j[X_j, \partial^\alpha]u = [\varphi, X_j][X_j, \partial^\alpha]u + X_j\varphi[X_j, \partial^\alpha]u$ and using again (2-3) and (2-6), we get

$$\|\varphi P\partial^\alpha u\| \le \sum_{j=1}^{m}\big(2\|X_j(\varphi)[X_j, \partial^\alpha]u\| + \|\varphi[X_j, [X_j, \partial^\alpha]]u\|\big) + \|\varphi[Y+b, \partial^\alpha]u\| + \|\varphi\partial^\alpha Pu\|,$$

and then we have to use expressions in (3-9) and (3-12). As we have to do that in order to bound $\|\varphi\partial^\alpha u\|_{(\ell+1)\sigma}$, we will write the step after the use of (3-9), (3-12) just in this general case; replacing u by $\partial^\alpha u$ in (3-22), we get, with some constant $C > 0$, which may vary from line to line,

$$\|\varphi\partial^\alpha u\|_{(\ell+1)\sigma} \le C\bigg(\|\psi T_{\ell\sigma}\varphi P\partial^\alpha u\|' + \sum_{|\beta|\le 2}\|\varphi^{(\beta)}\partial^\alpha u\|_{\ell\sigma}\bigg). \quad (3\text{-}25)$$

Now, as we did above with (3-24), we get

$$\|\varphi\partial^\alpha u\|_{(\ell+1)\sigma} \le C\bigg(\|\psi T_{\ell\sigma}\varphi\partial^\alpha Pu\|' + \sum_{j=1}^{m}\big(\|X_j(\varphi)[X_j, \partial^\alpha]u\|_{\ell\sigma} + \|\varphi[X_j[X_j, \partial^\alpha]]u\|_{\ell\sigma}\big) + \|\varphi[Y+b, \partial^\alpha]u\|_{\ell\sigma} + \sum_{|\beta|\le 2}\|\varphi^{(\beta)}\partial^\alpha u\|_{\ell\sigma}\bigg).$$

As $\ell\sigma \le 1$, $(\ell = 0, \dots, p-1)$, we finally obtain:

Proposition 3.7. *There exists a constant $C > 0$ such that*

for all $(u, \varphi) \in C^\infty(\overline{\Omega}_1) \times \mathcal{D}(\Omega_1)$, *and all* $\alpha \in \mathbb{N}^n$,

$$\|\varphi\partial^\alpha u\|_{(\ell+1)\sigma} \le C\Bigg(\|\varphi\partial^\alpha Pu\|_{\ell\sigma} + \sum_{|\beta|\le 2}\|\varphi^{(\beta)}\partial^\alpha u\|_{\ell\sigma} + \sum_{j=1}^{m}\bigg(\sum_{|\beta|=1}\|\varphi^{(\beta)}[X_j, \partial^\alpha]u\|_{\ell\sigma} + \|\varphi[X_j, [X_j, \partial^\alpha]]u\|_{\ell\sigma}\bigg) + \|\varphi[Y+b, \partial^\alpha]u\|_{\ell\sigma}\Bigg). \quad (3\text{-}26)$$

In what follows, s is given and $s \ge 1$.

4. Local relations of domination by powers of P

We need to introduce, in this section, further notation:

For $\epsilon > 0$, $j \in \mathbb{N}$, $\gamma \in \mathbb{N}^n$, and $(u, \varphi) \in C^\infty(\overline{\Omega}_1) \times \mathcal{D}(\Omega_1)$, we set

$$N^\epsilon_{j,\gamma}(u, \varphi) = \epsilon^{|\gamma|+2j} |\gamma|!^{-s} (2j)!^{-s} \|\varphi^{(\gamma)} P^j u\|. \quad (4\text{-}1)$$

Once ϵ is fixed, we often delete ϵ and write $N^\epsilon_{j,\gamma} = N_{j,\gamma}$.

Before stating our main theorem, we give a simple useful lemma, which we will apply many times.

Lemma 4.1. *Let $(k, \beta) \in \mathbb{N} \times \mathbb{N}^n$, and $\rho \in \mathbb{N}$. Then*

$$\rho!^s N^\epsilon_{j,\gamma}(P^k u, \varphi^{(\beta)}) \le \epsilon^{-(|\beta|+2k)} (\rho + |\beta| + 2k)!^s N^\epsilon_{j+k,\gamma+\beta}(u, \varphi), \quad \textit{if } |\gamma| + 2j \le \rho. \quad (4\text{-}2)$$

Proof. From the definition in (4-1), we see that

$$N^\epsilon_{j,\gamma}(P^k u, \varphi^{(\beta)}) = \epsilon^{-(|\beta|+2k)} (|\gamma|+1)^s \cdots (|\gamma|+|\beta|)^s \big((2j+1) \cdots (2j+2k)\big)^s N_{j+k,\gamma+\beta}(u, \varphi).$$

Then, observe that, for $|\gamma| + 2j \le \rho$,

$$\rho!(|\gamma|+1) \cdots (|\gamma|+|\beta|)(2j+1) \cdots (2(j+k)) \le (\rho + |\beta| + 2k)!. \quad (4\text{-}3)$$

The proof is then finished by taking (4-3) to the power $s \ge 1$. □

Theorem 4.2. *Let $s \ge 1$ be given. Assume that the coefficients of P are in $G^s(\Omega)$ and properties (2-4) and (3-16) hold on Ω_0, $\overline{\Omega}_0 \subset \Omega$. Let Ω_1 be an open set with $\overline{\Omega}_1 \subset \Omega_0$. For every $0 < \epsilon \le 1$, there exists $M = M(\epsilon, \Omega_0, P) \ge 1$ such that for all $\alpha \in \mathbb{N}^n$, $\ell = 0, \dots, p$, $(u, \varphi) \in C^\infty(\overline{\Omega}_1) \times \mathcal{D}(\Omega_1)$,*

$$\|\varphi \partial^\alpha u\|_{\ell\sigma} \le M^{2p|\alpha|+\ell+1} (2(p|\alpha|+\ell))!^s \sum_{|\beta|+2j \le 2(p|\alpha|+\ell)} N_{j,\beta}(u, \varphi). \quad (4\text{-}4)$$

Proof. It consists of a double induction on $|\alpha| = r$ and on ℓ. More precisely, in a first step, we prove the estimates (4-2) for $\alpha = 0$. In all the proof, we will specify $(4\text{-}4)_{\alpha,\ell}$ for (4-4) when we consider the couple $(\alpha, \ell) \in \mathbb{N} \times \{0, \dots, p\}$. So we want, in this first step, to prove $(4\text{-}4)_{0,\ell}$, $\ell \in \{0, \dots, p\}$.

(A) Proof of $(4\text{-}4)_{0,\ell}$, $\ell \in \{0, \dots, p\}$: As $(4\text{-}4)_{0,0}$ is trivial, all we have to do is to make an induction on ℓ. So assume that $(4\text{-}4)_{0,i}$ is true for $i \le \ell$, $\ell \in \{0, \dots, p-1\}$. Then we want to prove $(4\text{-}4)_{0,\ell+1}$. For that we use (3-23) in Proposition 3.6 for $(u, \varphi) \in C^\infty(\overline{\Omega}_1) \times \mathcal{D}(\Omega_1)$. So in order to bound $\|\varphi u\|_{(\ell+1)\sigma}$, we just have to suitably bound $\|\varphi P u\|_{\ell\sigma}$ and $\sum_{|\beta| \le 2} \|\varphi^{(\beta)} u\|$. Hence this reduces to applying $(4\text{-}4)_{0,\ell}$

respectively to the couples (Pu,φ) and $(u,\varphi^{(\beta)})$ in $C^\infty(\bar\Omega_1)\times\mathfrak{D}(\Omega_1)$. So

$$\|\varphi Pu\|_{\ell\sigma}\le M^{\ell+1}(2\ell)!^s\sum_{|\beta|+2j\le 2\ell}N_{j,\beta}(Pu,\varphi),\quad (u,\varphi)\in C^\infty(\bar\Omega_1)\times\mathfrak{D}(\Omega_1). \tag{4-5}$$

Now we apply Lemma 4.1 with $\rho=2\ell$ in (4-2). Then

$$\begin{aligned}\|\varphi Pu\|_{\ell\sigma}&\le\epsilon^{-2}(2(\ell+1))!^sM^{\ell+1}\sum_{|\beta|+2j\le 2(\ell+1)}N_{j,\beta}(u,\varphi)\\&\le\epsilon^{-2}M^{-1}M^{(\ell+1)+1}(2(\ell+1))!^s\sum_{|\beta|+2j\le 2(\ell+1)}N_{j,\beta}(u,\varphi).\end{aligned}$$

We do exactly the same for $\|\varphi^{(\beta)}u\|_{\ell\sigma}$, as $|\beta|\le 2$:

$$\|\varphi^{(\beta)}u\|_{\ell\sigma}\le\epsilon^{-2}M^{-1}M^{(\ell+1)+1}(2(\ell+1))!^s\sum_{|\gamma|+2j\le 2(\ell+1)}N_{j,\gamma}(u,\varphi). \tag{4-6}$$

So from (3-23), (4-5) and (4-6), we get

$$\begin{aligned}&\|\varphi u\|_{(\ell+1)\sigma}\\&\quad\le C(2+n+n^2)\epsilon^{-2}M^{-1}M^{(\ell+1)+1}(2(\ell+1))!^s\sum_{|\beta|+2j\le 2(\ell+1)}N_{j,\beta}(u,\varphi).\end{aligned} \tag{4-7}$$

So we see that under the condition

$$C(2+n+n^2)\epsilon^{-2}M^{-1}\le 1\quad\big(\text{equivalently, } M\ge C(2+n+n^2)\epsilon^{-2}\big), \tag{4-8}$$

Equation $(4\text{-}4)_{0,\ell+1}$ is satisfied for all $(u,\varphi)\in C^\infty(\bar\Omega_1)\times\mathfrak{D}(\Omega_1)$.

(B) Proof of $(4\text{-}4)_{\alpha,\ell}$ for all $(\alpha,\ell)\in\mathbb{N}^n\times\{0,\dots,p\}$: All we have to do, as $(4\text{-}4)_{0,\ell}$ is true for $\ell=0,\dots,p$, is to make an induction on $|\alpha|=r$. More precisely, if $(4\text{-}4)_{\alpha,\ell}$ is true for $|\alpha|\le r$, $\ell\in\{0,\dots,p\}$, then it is true for $|\alpha|=r+1$, $\ell=0,\dots,p$. We use the same kind of proof as before: given $\alpha+i$, with $|\alpha|=r$, $i\in\{1,\dots,n\}$, we want to suitably bound $\|\varphi\partial_i\partial^\alpha u\|_{\ell\sigma}$ for $\ell\in\{0,\dots,p\}$.

(1) $\ell=0$: We use $\|\varphi\partial_i\partial^\alpha u\|\le\|\partial_i(\varphi)\partial^\alpha u\|+\|\varphi\partial^\alpha u\|_1$. Hence we just have to apply $(4\text{-}4)_{\alpha,0}$ with $(u,\varphi^{(i)})$ and $(4\text{-}4)_{\alpha,p}$ with (u,φ). We will obtain, directly, or applying also Lemma 4.1,

$$\begin{aligned}&\|\varphi^{(i)}\partial^\alpha u\|\\&\quad\le\epsilon^{-1}M^{-2p}M^{2p(|\alpha|+1)+1}(2p(|\alpha|+1))!^s\sum_{|\beta|+2j\le 2p(|\alpha|+1)}N_{j,\beta}(u,\varphi),\end{aligned} \tag{4-9}$$

$$\|\varphi\partial^\alpha u\|_{p\sigma}\le M^{-p}M^{2p(|\alpha|+1)+1}(2p(|\alpha|+1))!^s\sum_{|\beta|+2j\le 2p(|\alpha|+1)}N_{j,\beta}(u,\varphi). \tag{4-10}$$

So, summing (4-9) and (4-10), we get that if

$$M^{-p}(1+\epsilon^{-1}M^{-p}) \le 1, \tag{4-11}$$

then $(4\text{-}4)_{\alpha+i,0}$ is satisfied, for $i = 1, \dots, n$.

(2) Proof of $(4\text{-}4)_{\alpha,\ell}$, for $|\alpha| = r+1$, $\ell > 0$, i.e., $\ell \in \{1, \dots, p\}$: So assuming that $(4\text{-}4)_{\alpha,\ell}$ is true for $|\alpha| = r$, $\ell = 0, \dots, p$, and $(4\text{-}4)_{\alpha,\rho}$ is true for $|\alpha| = r+1$, $\rho \in \{1, \dots, \ell\}$, then, if $\ell < p$, we want to prove that $(4\text{-}4)_{\alpha,\ell+1}$ is true, $|\alpha| = r+1$. Now we have to use (3-26) in Proposition 3.7. So in order to suitably bound $\|\varphi\partial^\alpha u\|_{(\ell+1)\sigma}$, we are led to bound $\|\varphi\partial^\alpha Pu\|_{\ell\sigma}$, $\|\varphi^{(\beta)}\partial^\alpha u\|_{\ell\sigma}$, $|\beta| \le 2$, but also much more terms like simple brackets of X_j's and Y with ∂^α and double brackets of X_j's with ∂^α.

The proof will follow the lines of our proof for $\alpha = 0$, but here with more terms, and some are more difficult to handle than others, namely those coming from the brackets of the X_j's with ∂^α (simple and double brackets). The term $\|\varphi[Y+b, \partial^\alpha]u\|_{\ell\sigma}$ is similar to the terms $\|\varphi^{(\beta)}[X_j, \partial^\alpha]u\|$, $\beta = 0$.

(a) A bound on $\|\varphi\partial^\alpha Pu\|_{\ell\sigma}$: We apply $(4\text{-}4)_{\alpha,\ell}$ for the couple (Pu, φ):

$$\|\varphi\partial^\alpha Pu\|_{\ell\sigma} \le M^{2p|\alpha|+\ell+1}(2(p|\alpha|+\ell))!^s \sum_{|\beta|+2j\le 2(p|\alpha|+\ell)} N_{j,\beta}(Pu, \varphi).$$

Applying Lemma 4.1, we get (here $\rho = 2(p|\alpha|+\ell)$)

$$\begin{aligned}
&\|\varphi\partial^\alpha Pu\|_{\ell\sigma} \\
&\quad \le \epsilon^{-2}M^{2p|\alpha|+\ell+1}(2(p|\alpha|+\ell+1))!^s \sum_{|\beta|+2j\le 2(p|\alpha|+\ell+1)} N_{j,\beta}(u, \varphi) \\
&\quad \le (\epsilon^{-2}M^{-1})M^{2p|\alpha|+(\ell+1)+1}(2(p|\alpha|+\ell+1))!^s \\
&\qquad\qquad \cdot \sum_{|\beta|+2j\le 2(p|\alpha|+\ell+1)} N_{j,\beta}(u, \varphi).
\end{aligned} \tag{4-12}$$

(b) A bound on $\|\varphi^{(\beta)}\partial^\alpha u\|_{\ell\sigma}$, $|\beta| \le 2$: We apply $(4\text{-}4)_{\alpha,\ell}$ to the couple $(u, \varphi^{(\beta)})$:

$$\|\varphi^{(\beta)}\partial^\alpha u\|_{\ell\sigma} M^{2p|\alpha|+\ell+1}(2(p|\alpha|+\ell))!^s \sum_{|\beta|+2j\le 2(p|\alpha|+\ell)} N_{j,\beta}(u, \varphi^{(\beta)}).$$

Applying again Lemma 4.1 (as $|\beta| \le 2$), we get

$$\begin{aligned}
&\|\varphi^{(\beta)}\partial^\alpha u\|_{\ell\sigma} \\
&\quad \le \epsilon^{-2}M^{2p|\alpha|+\ell+1}(2(p|\alpha|+\ell+1))!^s \sum_{|\beta|+2j\le 2(p|\alpha|+\ell+1)} N_{j,\beta}(u, \varphi) \\
&\quad \le \epsilon^{-2}M^{-1}M^{2p|\alpha|+(\ell+1)+1}(2(p|\alpha|+\ell+1))!^s \\
&\qquad\qquad \cdot \sum_{|\beta|+2j\le 2(p|\alpha|+\ell+1)} N_{j,\beta}(u, \varphi).
\end{aligned} \tag{4-13}$$

(c) A bound on $\|\varphi^{(\beta)}[X_j, \partial^\alpha]u\|_{\ell\sigma}$, $|\beta| \le 1$, $j = 1, \dots, m$, $\|\varphi[Y+b, \partial^\alpha]u\|_{\ell\sigma}$: (It is easy to see that the term $\|\varphi[Y+b, \partial^\alpha]u\|_{\ell\sigma}$ is much simpler than the others, since it corresponds to $\beta = 0$.)

From expressions in (3-9) where we delete $j \in \{1, \dots, m\}$:

$$\|\varphi^{(\beta)}[X, \partial^\alpha]u\|_{\ell\sigma} \le \sum_{\substack{\gamma<\alpha\\ i=1,\dots,n}} \|a_{\alpha\gamma i}\varphi^{(\beta)}\partial^{\gamma+i}u\|_{\ell\sigma}. \tag{4-14}$$

Then applying estimates in (3-15) (recall that j is deleted):

$$\|\varphi^{(\beta)}[X, \partial^\alpha]u\|_{\ell\sigma} \le \sum_{\substack{\gamma<\alpha\\ i=1,\dots,n}} B^{|\alpha-\gamma|}\Big(\frac{\alpha!}{(\gamma+i)!}\Big)^s \|\varphi^{(\beta)}\partial^{\gamma+i}u\|_{\ell\sigma}. \tag{4-15}$$

As $|\gamma+i| \le |\alpha| = r+1$, we can apply (4-4)$_{\gamma+i,\ell}$ to $(u, \varphi^{(\beta)})$. So we get

$$\begin{aligned}\|\varphi^{(\beta)}[X, \partial^\alpha]u\|_{\ell\sigma} &\\ \le n \sum_{\substack{\gamma<\alpha\\ i=1,\dots,n}} \Bigg(B^{|\alpha-\gamma|} M^{2p(|\gamma|+1)+\ell+1}(2(p(|\gamma|+1)+\ell))!^s\Big(\frac{\alpha!}{\gamma!}\Big)^s &\\ \cdot \sum_{|\delta|+2j\le 2(p(|\gamma|+1)+\ell)} N_{j,\delta}(u, \varphi^{(\beta)})\Bigg).\end{aligned} \tag{4-16}$$

Using (4-2) in Lemma 4.1, with $\rho = 2(p(|\gamma|+1)+\ell)$, we find

$$(2(p(|\gamma|+1)+\ell))!^s N_{j,\delta}(u, \varphi^{(\beta)}) \le (2(p(|\gamma|+1)+\ell)+1)!^s N_{j,\delta+\beta}(u, \varphi), \tag{4-17}$$

and hence

$$\begin{aligned}(2(p(|\gamma|+1)+\ell))!^s\Big(\frac{\alpha!}{\gamma!}\Big)^s N_{j,\delta}(u, \varphi^{(\beta)}) &\\ \le (2(p(|\alpha|+\ell+1)!^s N_{j,\delta+\beta}(u, \varphi),\end{aligned} \tag{4-18}$$

for all (j, δ) such that $|\delta|+2j \le 2(p(|\gamma|+1)+\ell)+1$. So, coming back to the second member in (4-16),

$$\begin{aligned}\|\varphi^{(\beta)}[X, \partial^\alpha]u\|_{\ell\sigma} &\\ \le n \sum_{\gamma<\alpha}\Bigg(B^{|\alpha-\gamma|} M^{2p(|\gamma|+1)+\ell+1}(2(p|\alpha|+\ell+1))!^s &\\ \cdot \epsilon^{-1} \sum_{|\delta|+2j\le 2(p|\alpha|+\ell+1)} N_{j,\delta}(u, \varphi)\Bigg),\end{aligned} \tag{4-19}$$

or

$$\|\varphi^{(\beta)}[X,\partial^\alpha]u\|_{\ell\sigma} \le n(2(p|\alpha|+\ell+1))!^s \sum_{|\delta|+2j\le 2(p|\alpha|+\ell+1)} N_{j,\delta}(u,\varphi) \cdot \epsilon^{-1}\sum_{\gamma<\alpha} B^{|\alpha-\gamma|}M^{2p(|\gamma|+1)+\ell+1}. \quad (4\text{-}20)$$

The last sum is bounded as follows:

$$\sum_{\gamma<\alpha} B^{|\alpha-\beta|}M^{2p(|\gamma|+1)+\ell+1} = BM^{2p|\alpha|+\ell+1}\sum_{\gamma<\alpha}\Big(\frac{B}{M^{2p}}\Big)^{|\alpha-\gamma|-1}. \quad (4\text{-}21)$$

Now we have the following lemma:

Lemma 4.3. *There exists $\theta_0 > 0$, independent of α, such that*

$$\sum_{\gamma<\alpha}\lambda^{|\alpha-\gamma|-1} \le n+1, \quad \text{if } 0\le\lambda<\theta_0. \quad (4\text{-}22)$$

The proof of Lemma 4.3 is similar to that of Lemma 3.2 after noticing that $\sum_{\gamma<\alpha,\, |\alpha-\gamma|=1}\lambda^{|\alpha-\gamma|-1} = n$.

Remark 4.4. Lemma 4.3 is not true when one works with the sum $\sum_{|\gamma|<|\alpha|}\lambda^{|\alpha-\gamma|-1}$ as the sum $\sum_{|\gamma|<|\alpha|,\, |\alpha-\gamma|=1} 1$ is not bounded by a constant independent of α.

Applying (4-22) we see that under the condition

$$M^{2p}\ge\theta_0^{-1}B, \quad (4\text{-}23)$$

we obtain from (4-20), (4-21) and (4-22)

$$\|\varphi^{(\beta)}[X,\partial^\alpha]u\|_{\ell\sigma} \le n(n+1)\epsilon^{-1}BM^{-1}M^{2p|\alpha|+(\ell+1)+1}(2p|\alpha|+\ell+1)!^s \cdot \sum_{|\delta|+2j\le 2(p|\alpha|+\ell+1)} N_{j,\delta}(u,\varphi). \quad (4\text{-}24)$$

(d) A bound on $\|\varphi[X_j,[X_j,\partial^\alpha]]u\|_{\ell\sigma}$, $j=1,\dots,m$: As we did above we delete the index j and write $\|\varphi[X,[X,\partial^\alpha]]u\|$. Of course, we also delete j in the coefficients of the bracket. Looking at (3-12), we have two kinds of terms to study:

(i) A bound on $\sum_{\beta<\alpha,\, k=1,\dots,n}\|\varphi d_{k\alpha\beta}\partial^{\beta+k}u\|_{\ell\sigma} = E_1$: Using Equation (3-15) in Proposition 3.3, as $\ell\sigma\le 1$, we have

$$E_1 \le \sum_{\substack{\beta<\gamma\\ k=1,\dots,n}} B^{|\alpha-\beta|}\Big(\frac{\alpha!}{\beta!}(|\alpha|+1)\Big)^s M^{2\beta(|\beta|+1)+\ell+1}(2(p(|\beta|+1)+\ell))!^s \cdot \sum_{|\gamma|+2j\le 2(p(|\beta|+1)+\ell)} N_{j,\gamma}(u,\varphi). \quad (4\text{-}25)$$

We want to remark here that there is a factor $(|\alpha|+1)^s$ in (4-25), but it is compensated by the fact that one has φ, not $\varphi^{(\beta)}$, $|\beta|=1$. Precisely, we have

$$E_1 \le n \sum_{\beta<\alpha} B^{|\alpha-\beta|}\Big(\frac{(|\alpha|+1)!}{|\beta|!}\Big)^s (2p(|\beta|+1)+\ell)!^s M^{2p(|\beta|+1)+\ell+1} \cdot \sum_{|\gamma|+2j\le 2(p|\alpha|+\ell)} N_{j,\gamma}(u,\varphi), \quad \text{as } |\beta|+1\le|\alpha|. \tag{4-26}$$

Now we have the following inequality:

$$\Big(\frac{(|\alpha|+1)!}{|\beta|!}\Big)^s \big(2(p(|\beta|+1)+\ell)\big)!^s \le (2(p|\alpha|+\ell+1))!^s, \quad |\beta|+1\le|\alpha|. \tag{4-27}$$

So, from (4-26) and (4-27), we get

$$E_1 \le n(2(p|\alpha|+\ell+1))!^s \sum_{\beta<\alpha} B^{|\alpha-\beta|} M^{2p(|\beta|+1)+\ell+1} \cdot \sum_{|\gamma|+2j\le 2(p|\alpha|+\ell)} N_{j,\gamma}(u,\varphi). \tag{4-28}$$

The sum in the second member in (4-28) is the same than the sum in (4-21), replacing γ by β. So from (4-22) in Lemma 4.3, we get, under condition (4-23),

$$E_1 \le n(n+1)BM^{-1}\, M^{2p|\alpha|+(\ell+1)+1}(2p|\alpha|+\ell+1)!^s \cdot \sum_{|\gamma|+2j\le 2(p|\alpha|+\ell)} N_{j,\gamma}(u,\varphi). \tag{4-29}$$

(ii) A bound on $\sum_{\beta<\alpha,\, i,k=1,\dots,n} \|\varphi c_{ik\alpha\beta}\partial^{\beta+i+k}u\|_{\ell\sigma} = E_2$: We write the proof, for completeness, noticing however that $\partial^{\beta+i+k}$ is of the form $\partial^{\beta+I}$ with $|I|=2$. Using estimates in (3-15), we get

$$E_2 \le \sum_{\substack{\beta<\alpha\\ i,k=1,\dots,n}} B^{|\alpha-\beta|}\Big(\frac{(\alpha+k)!}{(\beta+k)!}\Big)^s \|\varphi\partial^{\beta+i+k}u\|_{\ell\sigma}. \tag{4-30}$$

Hence,

$$E_2 \le \sum_{\substack{\beta<\alpha,\ |\beta|\le|\alpha|-2\\ i,k=1,\dots,n}} B^{|\alpha-\beta|} M^{2p(|\beta|+2)+\ell+1}\Big(\frac{(\alpha+k)!}{(\beta+k)!}\Big)^s (2(p(|\beta|+2)+\ell))!^s \cdot \sum_{|\gamma|+2j\le 2(p(|\beta|+2)+\ell)} N_{j,\gamma}(u,\varphi). \tag{4-31}$$

Now we have the following:

$$\Big(\frac{(\alpha+k)!}{(\beta+k)!}\Big)^s (2(p(|\beta|+2)+\ell))!^s \le (2(p|\alpha|+\ell+1))!^s. \tag{4-32}$$

(4-32) is a consequence of $(\alpha+k)!/(\beta+k)! \le (|\alpha|+1)!/(|\beta|+1)!$. Hence

$$E_2 \le n^2(2(p|\alpha|+\ell+1))!^s \Bigg(\sum_{\substack{\beta<\alpha \\ |\beta|\le|\alpha|-2}} B^{|\alpha-\beta|} M^{2p(|\beta|+2)+\ell+1} \Bigg) \cdot \sum_{|\gamma|+2j\le 2(p|\alpha|+\ell+1)} N_{j,\gamma}(u,\varphi), \quad (4\text{-}33)$$

as $\mathrm{Card}\big\{(i,k)\in\{1,\dots,n\}\big\} = n^2$. Now, we bound

$$F = \textstyle\sum_{\beta<\alpha,\, |\beta|\le|\alpha|-2} B^{|\alpha-\beta|} M^{2p(|\beta|+2)+\ell+1}$$

with

$$F \le \sum_{\substack{\beta<\alpha \\ |\beta|\le|\alpha|-2}} \Big(\frac{B}{M^{2p}}\Big)^{|\alpha-\beta|-2} M^{2p|\alpha|+(\ell+1)+1} B^2 M^{-1}. \quad (4\text{-}34)$$

Now using Lemma 4.3 and the fact that $\sum_{\beta<\alpha,\, |\beta|=|\alpha|-2} 1 \le n^2$, we obtain

$$E_2 \le (n^2+1)^2 B^2 M^{-1}\, M^{2p|\alpha|+(\ell+1)+1} (2(p|\alpha|+\ell+1))!^s \cdot \sum_{|\delta|+2j\le 2(p|\alpha|+\ell+1)} N_{j,\delta}(u,\varphi). \quad (4\text{-}35)$$

We recall that we said in (c) that we do not consider the term

$$\|\varphi[Y+b,\partial^\alpha]\|_{\ell\sigma} \le \|\varphi[Y,\partial^\alpha]u\|_{\ell\sigma} + \|\varphi[b,\partial^\alpha]u\|_{\ell\sigma}$$

as $\|\varphi[Y,\partial^\alpha]u\|_{\ell\sigma}$ is like $\|\varphi[X_j,\partial^\alpha]u\|_{\ell\sigma}$ and $\|\varphi[b,\partial^\alpha]u\|_{\ell\sigma}$ much smaller. So in order to bound the term $\|\varphi[X_j,[X_j,\partial^\alpha]]u\|_{\ell\sigma}$, we have to collect (4-29) and (4-35), which are true under condition (4-23). We find

$$\|\varphi[X_j,[X_j,\partial^\alpha]]u\|_{\ell\sigma} \le 2(n^2+1)^2 B^2 M^{-1} M^{2p|\alpha|+(\ell+1)+1} (2(p|\alpha|+\ell+1))!^s \cdot \sum_{|\gamma|+2j\le 2(p|\alpha|+\ell)} N_{j,\gamma}(u,\varphi) \quad (4\text{-}36)$$

In order to give a bound for $\|\varphi\partial^\alpha u\|_{(\ell+1)\sigma}$, we have from (3-26) in Proposition 3.7 to take C times the bound in (a) plus $C(n^2+n+1)$ times the bound in (b) plus $Cm(n+2)$ times the bound in (c) plus Cm times the bound in (d), under, of course, the conditions on M indicated in the proofs of (a), (b), (c) and (d).

Of course, we also have to take care of the conditions needed on M for the validity of these bounds. These may be summarized as follows:

(a) (4-12) is satisfied, just under the induction hypothesis,

(b) (4-13) is satisfied, under the induction hypothesis,

(c) (4-24) is satisfied, under the induction hypothesis and (4-23),

(d) (4-36) is satisfied, under the induction hypothesis and (4-23).

So adding all these estimates with the suitable factors yields

$$\begin{aligned}\|\varphi\partial^\alpha u\|_{(\ell+1)\sigma} &\\ \le C M^{-1}(2\epsilon^{-2}+n(n+1)\epsilon^{-1}B+2(n^2+1)^2B^2)\, M^{2p|\alpha|+(\ell+1)+1}(2(p|\alpha|+\ell+1))!^s &\\ \cdot \sum_{|\gamma|+2j\le 2(p|\alpha|+\ell)} N_{j,\gamma}(u,\varphi) \quad (4\text{-}37)\end{aligned}$$

under the condition (4-23). From (4-37), we deduce that if M satisfies

$$M \ge \sup\bigl\{(\theta_0^{-1}B)^{1/2^p},\ 1,\ C\bigl(2\epsilon^{-2}+n(n+1)\epsilon^{-1}B+2(n^2+1)^2B^2\bigr)\bigr\}, \quad (4\text{-}38)$$

then $(4\text{-}4)_{\alpha,\ell+1}$ is true. So $(4\text{-}4)_{\alpha,\ell}$ is true, $|\alpha| = r+1$, for all ℓ.

Now, let us finish the proof of the theorem. Since we proved $(4\text{-}4)_{0,\ell}$, $\ell = 1, \dots, p$, and the induction

$$\bigl\{(4\text{-}4)_{\alpha,\ell},\ |\alpha| \le r,\ \ell\in\{1,\dots,p\}\bigr\} \Rightarrow \bigl\{(4\text{-}4)_{\alpha,\ell},\ |\alpha| = r+1,\ \ell\in\{1,\dots,p\}\bigr\},$$

under respectively condition ((4-8) and (4-11)) and condition (4-38), we have $M = M(\epsilon,\Omega_0,P) > 0$ so that the theorem is completely proved, when the conditions (4-8), (4-11) and (4-38) are satisfied. As $\epsilon \le 1$, we see that $M^p \ge 2\epsilon^{-1}$ implies that (4-11) holds. So, everything boils down to only the following condition:

$$\begin{aligned}M \ge \sup\Bigl\{(\theta_0^{-1}B)^{1/2^p},\ C(2+n+n^2)\epsilon^{-2},&\\ C\bigl(2\epsilon^{-2}+n(n+1)\epsilon^{-1}B+2(n^2+1)^2B^2\bigr)\Bigr\}&\\ = M(\epsilon,\theta_0,B). \quad (4\text{-}39)\end{aligned}$$

B depends on P and Ω_0 and θ_0 depends on n. Hence $M(\epsilon,\theta_0,B)$ can be written, as n is fixed, $M(\epsilon,\Omega_0,P)$. The proof of Theorem 4.2 is now complete. □

5. Gevrey regularity for Gevrey vectors

We want to give in this section an application of Theorem 4.2. In fact, we shall just use the estimates (4-4) for $\ell = 0$, which we rewrite here, for $(u,\varphi)\in C^\infty(\bar\Omega_1)\times\mathcal{D}(\Omega_1)$:

$$\|\varphi\partial^\alpha u\| \le M_\epsilon^{2p|\alpha|+1}(2p|\alpha|)!^s \sum_{|\beta|+2j\le 2p|\alpha|} N^\epsilon_{j,\beta}(u,\varphi). \quad (5\text{-}1)$$

Moreover, we want to state a theorem for operators of order m, satisfying estimates similar to (5-1), but with 2 replaced by m and with coefficients in $G^s(\Omega)$. For that purpose, we clarify the notation as m replaces 2.

Firstly, Ω and Ω_1 are as in Section 4, $\bar\Omega_1 \Subset \Omega$ and $s \ge 1$. Then define

$$N^\epsilon_{j,\beta} = \epsilon^{|\beta|+mj}|\beta|!^{-s}(mj)!^{-s}\|\varphi^{(\beta)}P^j u\|, \quad (u,\varphi)\in C^\infty(\bar\Omega_1)\times\mathcal{D}(\Omega_1). \quad (5\text{-}2)$$

Now assume that P satisfies the following estimates:

$$\forall\epsilon,\ 0<\epsilon\le 1, \exists M_\epsilon\ge 1 \quad \text{such that} \quad \forall\alpha\in\mathbb{N}^n,\ \forall(u,\varphi)\in C^\infty(\overline{\Omega}_1)\times\mathscr{D}(\Omega_1)$$
$$\|\varphi\partial^\alpha u\|\le M_\epsilon^{r|\alpha|+1}(r|\alpha|)!^s\sum_{|\beta|+mj\le r|\alpha|}N_{j,\beta}(u,\varphi)\quad \text{for some } r\in\mathbb{N}^*. \tag{5-3}$$

Now, we provide a proposition on Gevrey regularity of Gevrey vectors of P satisfying (5-3).

Proposition 5.1. *Let P be a linear partial differential operator with $G^s(\Omega)$ coefficients, of order m, satisfying (5-3) in Ω_1 with $\overline{\Omega}_1\subset\Omega$. Then any $G^s(\Omega_1)$-vector of P which is $C^\infty(\Omega)$ is in $G^{rs}(\Omega_1)$, $s\ge 1$.*

Proof. We have to distinguish between the cases $s>1$ and $s=1$.

(1) Case $s>1$: Let $u\in C^\infty(\Omega)\cap G^s(\Omega_1,P)$. In order to prove that $u\in G^{rs}(\Omega_1)$, we have to show that given any open set Ω_2 with $\overline{\Omega}_2\subset\Omega_1$, we have:

$$\exists C_{\Omega_2}>0 \text{ s.t. } \forall\alpha\in\mathbb{N}^n,\quad \|\partial^\alpha u\|_{L^2(\Omega_2)}\le C_{\Omega_2}^{|\alpha|+1}(r|\alpha|)!^s=C^{|\alpha|+1}(r|\alpha|)!^s. \tag{5-4}$$

First, we consider $\varphi\in G^s(\Omega_1)\cap\mathscr{D}(\Omega_1)$, $\varphi=1$ on Ω_2. Then

$$\|\partial^\alpha u\|_{L^2(\Omega_2)}\le\|\varphi\partial^\alpha u\|. \tag{5-5}$$

As $u\in C^\infty(\overline{\Omega}_1)$ and $\varphi\in\mathscr{D}(\Omega_1)$, we apply (5-3). So we get

$$\|\partial^\alpha u\|_{L^2(\Omega_2)}\le(CM_\epsilon)^{r|\alpha|+1}(r|\alpha|)!^s\sum_{|\beta|+mj\le r|\alpha|}N_{j,\beta}^\epsilon(u,\varphi). \tag{5-6}$$

Now as $\varphi\in G^s(\Omega_1)\cap\mathscr{D}(\Omega_1)$, there exists $A=A_\varphi$ such that

$$\sup|\varphi^{(\beta)}|\le A^{|\beta|+1}|\beta|!^s. \tag{5-7}$$

Also, from (5-2), we have

$$N_{j,\beta}^\epsilon=\epsilon^{|\beta|+mj}|\beta|!^{-s}(mj)!^{-s}\|\varphi^{(\beta)}P^ju\|. \tag{5-8}$$

Now as $u\in G^s(\Omega_1,P)$, taking $K=\operatorname{Supp}(\varphi)\subset\Omega_1$

$$\exists B=B_K>0,\quad \text{such that}\quad \|P^ju\|_{L^2(K)}\le B^{mj+1}(mj)!^s,\quad \forall j\in\mathbb{N}. \tag{5-9}$$

From (5-7), (5-8) and (5-9), we obtain

$$N_{j,\beta}^\epsilon(u,\varphi)\le\epsilon^{|\beta|+mj}A^{|\beta|+1}\,B^{mj+1}\le\epsilon^{|\beta|+mj}D^{|\beta|+mj+1} \tag{5-10}$$

for some constant D (for example $A+B$). So, with (5-6),

$$\|\partial^\alpha u\|_{L^2(\Omega_2)}\le D(CM_\epsilon)^{r|\alpha|+1}(r|\alpha|)!^s\sum_{|\beta|+mj\le r|\alpha|}(\epsilon D)^{|\beta|+mj}. \tag{5-11}$$

Now let us choose ϵ such that

$$\epsilon D = \tfrac{1}{2}. \tag{5-12}$$

Then with $M = M_{1/(2D)}$, we have

$$\|\partial^\alpha u\|_{L^2(\Omega_2)} \le D(CM)^{r|\alpha|+1}(r|\alpha|)!^s \sum_{|\beta|+mj\le r|\alpha|} \left(\tfrac{1}{2}\right)^{|\beta|+mj}. \tag{5-13}$$

We have to estimate the sum in the (5-13).

Lemma 5.2. *There exists a constant* $C_0 > 0$ *such that*

$$\sum_{|\beta|+mj\le r|\alpha|} \left(\tfrac{1}{2}\right)^{|\beta|+mj} \le C_0, \quad \forall \alpha \in \mathbb{N}^n. \tag{5-14}$$

Proof of Lemma 5.2. This simple lemma has an elementary proof. For completeness, let us give it. For $\alpha = 0$, it is trivial. So assume $|\alpha| \ge 1$. It suffices to see that

$$\sum_{|\beta|+mj\le r|\alpha|} \left(\tfrac{1}{2}\right)^{|\beta|+mj} = \sum_{k=0}^{r|\alpha|}\Bigg(\sum_{|\beta|+mj=k} 1\Bigg)\left(\tfrac{1}{2}\right)^k \le \sum_{k=0}^{r|\alpha|}(k+1)^n\left(\tfrac{1}{2}\right)^k$$
$$\le \sum_{k=0}^{\infty}(k+1)^n\left(\tfrac{1}{2}\right)^k = C_0 < +\infty. \qquad \square$$

Coming back to the proof of Proposition 5.1, we have, with some constant $C_1 > 0$, using (5-13) and (5-14), the following estimate:

$$\|\partial^\alpha u\|_{L^2(\Omega_2)} \le C_1(CM)^{r|\alpha|}(r|\alpha|)!^s \tag{5-15}$$

which shows that u is in $G^{rs}(\Omega_2)$, hence in $G^{rs}(\Omega_1)$ as Ω_2 is any relatively compact set in Ω_1.

(2) Case $s = 1$: In this case, as we have no $\varphi \in \mathcal{D}(\Omega_1)$ which is in $G^1(\Omega_1)$, we proceed by using a sequence of functions of L. Ehrenpreis associated to the couple (Ω_0, Ω_1) with $\overline{\Omega}_1 \subset \Omega_0$ and $\overline{\Omega}_0 \subset \Omega$. We state below a proposition due to Ehrenpreis, providing the precise details regarding the sequence.

Proposition 5.3 [Ehrenpreis 1960]. *Let* Ω_0, Ω_1 *be as above. Then there exists a constant* $\tilde{C} > 0$ *such that*:

$$\forall N \in \mathbb{N},\ \exists \varphi_N \in \mathcal{D}(\Omega_0),\ \varphi_N|_{\Omega_1} = 1,$$
$$\text{such that} \quad |\varphi_N^{(\beta)}| \le \tilde{C}^{|\beta|+1} N^\beta, \quad \text{for } |\beta| \le N. \tag{5-16}$$

In our proof below, in order to bound $\|\partial^\alpha u\|_{L^2(\Omega_2)}$, we use, in place of φ used in case (1), the function $\varphi_{r|\alpha|}$ given by taking $N = r|\alpha|$ in (5-16). So, as $\overline{\Omega}_2 \subset \Omega_1$, we have

$$\|\partial^\alpha u\|_{L^2(\Omega_2)} \le \|\varphi_{r|\alpha|}\partial^\alpha u\| \le M_\epsilon^{r|\alpha|+1}(r|\alpha|)! \sum_{|\beta|+mj\le r|\alpha|} N^\epsilon_{j,\beta}(u, \varphi_{r|\alpha|}). \tag{5-17}$$

Now, taking A given by:

$$A = \sup(B, \tilde{C}), \tag{5-18}$$

where B is given by (5-9) with $K = \overline{\Omega}_0$, we get

$$E_\alpha = \sum_{|\beta|+mj\le r|\alpha|} N^{\epsilon}_{j,\beta}(u, \varphi_{r|\alpha|}) \le A \sum_{|\beta|+mj\le r|\alpha|} (\epsilon A)^{|\beta|+mj} |\beta|!^{-1} (r|\alpha|)^{|\beta|}, \tag{5-19}$$

since $|\beta| \le r|\alpha|$ in the sum. Looking at the second member in (5-19), we may bound by

$$E_\alpha \le \sum_{mj\le r|\alpha|} (\epsilon A)^{mj} \sum_{|\beta|\le r|\alpha|} (\epsilon A)^{|\beta|} |\beta|!^{-1} (r|\alpha|)^{|\beta|}. \tag{5-20}$$

Hence

$$E_\alpha \le \sum_{mj\le r|\alpha|} (\epsilon A)^{mj} \sum_{k=0}^{r|\alpha|} \Big(\sum_{|\beta|=k} 1 \Big) (\epsilon A r|\alpha|)^k k!^{-1}. \tag{5-21}$$

If we choose ϵ_0 so that $\epsilon_0 A = \frac{1}{2}$, we get

$$E_\alpha \le 2 \sum_{k=0}^{r|\alpha|} (k+1)^n \Big(\frac{r|\alpha|}{2}\Big)^k k!^{-1} \le 2(r|\alpha|+1)^n \sum_k \Big(\frac{r|\alpha|}{2}\Big)^k k!^{-1} \tag{5-22}$$

So finally, with some constant $C_1 > 0$, $C_1 = C_1(n)$,

$$E_\alpha \le 2(r|\alpha|+1)^n \exp\Big(\frac{r|\alpha|}{2}\Big) \le C_1 \exp(r|\alpha|). \tag{5-23}$$

Then, coming back to (5-17), we obtain

$$\|\partial^\alpha u\|_{L^2(\Omega_2)} \le M_{\epsilon_0}^{r|\alpha|+1} (r|\alpha|)!\, C_1 \exp(r|\alpha|) = C_1 M_{\epsilon_0} (e M_{\epsilon_0})^{r|\alpha|} (r|\alpha|)!. \tag{5-24}$$

As we took any Ω_2 with $\overline{\Omega}_2 \subset \Omega_1$, $\overline{\Omega}_1 \subset \Omega_0$, we obtain that $u \in G^r(\Omega_1)$. The proof of Proposition 5.1 is complete. □

As a corollary of Theorem 4.2 and Proposition 5.1, we get:

Theorem 5.4. *Let P be a Hörmander's operator on an open set Ω in $\mathbb{R}^n$ and $s \in \mathbb{R}$, $s \ge 1$. Assume that P satisfies the estimate (2-4) in some open subset Ω_0 with $\overline{\Omega}_0 \subset \Omega$ with $\sigma = 1/p$, $p \in \mathbb{N}^*$ and that its coefficients are in $G^s(\Omega_0)$. Then $G^s(\Omega_0, P) \subset G^{2ps}(\Omega_0)$.*

We conclude this article with some final remarks.

(1) In the case $s = 1$, there is another proof, using the method of addition of an extra variable (see, for example, [Bolley et al. 1987] or [Lions and Magenes 1970]), by considering the operator $\partial_t^2 + P$ in $\mathbb{R} \times \Omega \subset \mathbb{R}^{n+1}$, which is also a Hörmander's operator in $\mathbb{R} \times \Omega$, with analytic coefficients (case $s = 1$), to which one can use the theorem of Gevrey hypoellipticity G^s for $s \ge 2p$, [Derridj and Zuily 1973].

(2) We know nothing on optimality of our result. In our preceding paper [Derridj ≥ 2019], in the case of operators of the first kind, the result was optimal: $G^k(\Omega_0, P) \subset G^{pk}(\Omega_0),\ k \in \mathbb{N}$.

(3) In a forthcoming paper, we will study the question of local relations of domination by powers of P, in the case where P is of the first kind, which will be finer, giving therefore the optimal result $G^s(\Omega_0, P) \subset G^{ps}(\Omega_0)$, p being the type of $\overline{\Omega}_0$.

Acknowledgements

I am greatly indebted to Nordine Mir for encouraging me to write the present paper when he knew about the result and for his help.

References

[Baouendi and Métivier 1982] M. S. Baouendi and G. Métivier, "Analytic vectors of hypoelliptic operators of principal type", *Amer. J. Math.* **104**:2 (1982), 287–319. MR Zbl

[Barostichi et al. 2011] R. F. Barostichi, P. D. Cordaro, and G. Petronilho, "Analytic vectors in locally integrable structures", pp. 1–14 in *Geometric analysis of several complex variables and related topics* (Marrakesh, 2010), edited by Y. Barkatou et al., Contemp. Math. **550**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl

[Bolley et al. 1987] P. Bolley, J. Camus, and L. Rodino, "Hypoellipticité analytique-Gevrey et itérés d'opérateurs", *Rend. Sem. Mat. Univ. Politec. Torino* **45**:3 (1987), 1–61. MR Zbl

[Braun Rodrigues et al. 2016] N. Braun Rodrigues, G. Chinni, P. D. Cordaro, and M. R. Jahnke, "Lower order perturbation and global analytic vectors for a class of globally analytic hypoelliptic operators", *Proc. Amer. Math. Soc.* **144**:12 (2016), 5159–5170. MR Zbl

[Castellanos et al. 2013] J. E. Castellanos, P. D. Cordaro, and G. Petronilho, "Gevrey vectors in involutive tube structures and Gevrey regularity for the solutions of certain classes of semilinear systems", *J. Anal. Math.* **119** (2013), 333–364. MR Zbl

[Damlakhi and Helffer 1980] M. Damlakhi and B. Helffer, "Analyticité et itères d'un système de champs non elliptique", *Ann. Sci. École Norm. Sup.* (4) **13**:4 (1980), 397–403. MR Zbl

[Derridj 2017] M. Derridj, "On Gevrey vectors of some partial differential operators", *Complex Var. Elliptic Equ.* **62**:10 (2017), 1474–1491. MR Zbl

[Derridj ≥ 2019] M. Derridj, "On Gevrey vectors of L. Hörmander's operators", preprint. To appear in *Trans. Amer. Math. Soc.*

[Derridj and Zuily 1973] M. Derridj and C. Zuily, "Sur la régularité Gevrey des opérateurs de Hörmander", *J. Math. Pures Appl.* (9) **52** (1973), 309–336. MR Zbl

[Ehrenpreis 1960] L. Ehrenpreis, "Solution of some problems of division, IV: Invertible and elliptic operators", *Amer. J. Math.* **82** (1960), 522–588. MR Zbl

[Helffer and Mattera 1980] B. Helffer and C. Mattera, "Analyticité et itérés réduits d'un système de champs de vecteurs", *Comm. Partial Differential Equations* **5**:10 (1980), 1065–1072. MR Zbl

[Hörmander 1967] L. Hörmander, "Hypoelliptic second order differential equations", *Acta Math.* **119** (1967), 147–171. MR Zbl

[Kohn 1978] J. J. Kohn, "Lectures on degenerate elliptic problems", pp. 89–151 in *Pseudodifferential operator with applications* (Bressanone, Italy, 1977), Liguori, Naples, 1978. MR Zbl

[Kotake and Narasimhan 1962] T. Kotake and M. S. Narasimhan, "Regularity theorems for fractional powers of a linear elliptic operator", *Bull. Soc. Math. France* **90** (1962), 449–471. MR Zbl

[Lions and Magenes 1970] J.-L. Lions and E. Magenes, *Problèmes aux limites non homogènes et applications*, vol. 3, Travaux et Recherches Mathématiques **20**, Dunod, Paris, 1970. MR Zbl

[Métivier 1978] G. Métivier, "Propriété des itérés et ellipticité", *Comm. Partial Differential Equations* **3**:9 (1978), 827–876. MR Zbl

[Nelson 1959] E. Nelson, "Analytic vectors", *Ann. of Math.* (2) **70** (1959), 572–615. MR Zbl

[Rothschild and Stein 1976] L. P. Rothschild and E. M. Stein, "Hypoelliptic differential operators and nilpotent groups", *Acta Math.* **137**:3-4 (1976), 247–320. MR Zbl

Received 24 Nov 2017. Revised 11 Dec 2017.

MAKHLOUF DERRIDJ:
maklouf.derridj@numericable.fr
78350 Les Loges en Josas, France

TUNISIAN JOURNAL OF MATHEMATICS
Vol. 1, No. 3, 2019
dx.doi.org/10.2140/tunis.2019.1.347

Generic colourful tori and inverse spectral transform for Hankel operators

Patrick Gérard and Sandrine Grellier

This paper explores the regularity properties of an inverse spectral transform for Hilbert–Schmidt Hankel operators on the unit disc. This spectral transform plays the role of action-angle variables for an integrable infinite dimensional Hamiltonian system: the cubic Szegő equation. We investigate the regularity of functions on the tori supporting the dynamics of this system, in connection with some wave turbulence phenomenon, discovered in a previous work and due to relative small gaps between the actions. We revisit this phenomenon by proving that generic smooth functions and a G_δ dense set of irregular functions do coexist on the same torus. On the other hand, we establish some uniform analytic regularity for tori corresponding to rapidly decreasing actions which satisfy some specific property ruling out the phenomenon of small gaps.

1. Introduction

1.1. ***The cubic Szegő equation.*** This paper explores the properties of some inverse spectral transformation related to an integrable infinite dimensional Hamiltonian system. Introduced in [Gérard and Grellier 2010] as a model of nondispersive evolution equation, the cubic Szegő equation reads

$$i\partial_t u = \Pi(|u|^2 u), \tag{1}$$

where $u = u(t, x)$ is a function defined for $(t, x) \in \mathbb{R} \times \mathbb{T}$, $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$, such that, for every $t \in \mathbb{R}$, $u(t, \cdot)$ belongs to the Hardy space $L^2_+(\mathbb{T})$ of L^2 functions v on $\mathbb{T}$ with only nonnegative Fourier modes,

$$\text{for all } n < 0, \quad \hat{v}(n) = 0.$$

Here

$$\hat{v}(n) = \int_0^{2\pi} v(x)\mathrm{e}^{-inx}\,\frac{dx}{2\pi}, \quad n \in \mathbb{Z}$$

P. Gérard is grateful to G. Chenevier for valuable discussions about elliptic functions. He is supported by ANR ANAE 13-BS01-0010-03.

MSC2010: primary 35B65; secondary 37K15, 47B35.

Keywords: Cubic Szegő equation, action-angle variables, integrable systems, Hankel operators, spectral analysis.

denotes the Fourier coefficient of $v \in L^2(\mathbb{T})$, and Π denotes the orthogonal projector from $L^2(\mathbb{T})$ onto $L^2_+(\mathbb{T})$:

$$\Pi\left(\sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{inx}\right) = \sum_{n=0}^{\infty} c_n \mathrm{e}^{inx}.$$

It has been proved in [Gérard and Grellier 2010] that (1) is globally well posed on Sobolev spaces $H^s_+(\mathbb{T}) := H^s(\mathbb{T}) \cap L^2_+(\mathbb{T})$ for all $s \geq \frac{1}{2}$, with conservation of the $H^{\frac{1}{2}}$ norm. Recall that, for elements in $L^2_+(\mathbb{T})$, the H^s Sobolev norm reads

$$\|v\|^2_{H^s} = \sum_{n=0}^{\infty} (1+n)^{2s} |\hat{v}(n)|^2.$$

Furthermore, it turns out that (1) enjoys an unexpected Lax pair structure, discovered in [Gérard and Grellier 2010] and studied in [Gérard and Grellier 2012; 2015; 2017]. More precisely, consider, for every $u \in H^{\frac{1}{2}}_+(\mathbb{T})$, the Hankel operator $H_u : L^2_+(\mathbb{T}) \to L^2_+(\mathbb{T})$ defined as

$$H_u(h) = \Pi(u\bar{h}).$$

Notice that H_u is an antilinear realisation of the Hankel matrix $\Gamma_{\hat{u}}$, where, for every sequence $\alpha = (\alpha_n)_{n\geq 0}$ of complex numbers, Γ_α denotes the operator on $\ell^2(\mathbb{Z}_+)$ given by the infinite matrix $(\alpha_{n+p})_{n,p\geq 0}$. Indeed, if $\mathscr{F}$ denotes the Fourier transform $v \mapsto \hat{v}$ between $L^2_+(\mathbb{T})$ and $\ell^2(\mathbb{Z}_+)$, it easy to check that

$$\mathscr{F} H_u \mathscr{F}^{-1} = \Gamma_{\hat{u}} \circ \mathcal{C},$$

where $\mathcal{C}$ denotes the complex conjugation. The Lax pair identity then reads as follows, see [Gérard and Grellier 2010]. If $s > \frac{1}{2}$ and u is a H^s_+ solution of (1), then

$$\frac{dH_u}{dt} = [B_u, H_u],$$

where B_u is a linear anti-self-adjoint operator depending on u. As a consequence, there exists a one parameter family $U(t)$ of unitary operators on $L^2_+(\mathbb{T})$ such that

$$\text{for all } t \in \mathbb{R}, \quad H_{u(t)} = U(t) H_{u(0)} U(t)^*.$$

In particular, $H^2_{u(t)} = U(t) H^2_{u(0)} U(t)^*$. Notice that H^2_u is a linear positive operator on $L^2_+(\mathbb{T})$, and that

$$\mathscr{F} H^2_u \mathscr{F}^{-1} = \Gamma_{\hat{u}} \Gamma^*_{\hat{u}},$$

thus H^2_u is a trace class operator as soon as $u \in H^{\frac{1}{2}}_+(\mathbb{T})$, with

$$\mathrm{Tr}(H^2_u) = \sum_{n=0}^{\infty} (1+n) |\hat{u}(n)|^2 = \|u\|^2_{H^{\frac{1}{2}}}.$$

Consequently, apart from 0, the spectrum of H_u^2 is made of eigenvalues, which are conservation laws of (1).

In fact, a second Lax pair for (1) holds [Gérard and Grellier 2012], which concerns the operator $K_u := S^*H_u = H_uS = H_{S^*u}$, where S denotes the shift operator on $L^2_+(\mathbb{T})$, namely multiplication by e^{ix}. Operator K_u is also a Hankel operator,

$$\mathscr{F} K_u \mathscr{F}^{-1} = \tilde{\Gamma}_{\hat{u}} \circ \mathcal{C},$$

where $\tilde{\Gamma}_\alpha$ denotes the shifted Hankel matrix $(\alpha_{n+p+1})_{n,p\geq 0}$. Again, it is possible to prove that

$$\text{for all } t \in \mathbb{R}, \quad K_{u(t)} = V(t)K_{u(0)}V(t)^*,$$

for some one parameter family $V(t)$ of unitary operators on $L^2_+(\mathbb{T})$, and consequently that the eigenvalues of K_u^2 are conservation laws of (1). Denote by $(\rho_j^2)_{j\geq 1}$ the positive eigenvalues of H_u^2 and by $(\sigma_k^2)_{k\geq 1}$ the positive eigenvalues of K_u^2, so that the ρ_j are the singular values of $\Gamma_{\hat{u}}$ and the σ_k are the singular values of $\tilde{\Gamma}_{\hat{u}}$. In view of the identity

$$K_u^2 = H_u^2 - (\,\cdot\,|u)_{L^2}u$$

and of the min-max theorem, the following interlacing property holds:

$$\rho_1 \geq \sigma_1 \geq \rho_2 \geq \sigma_2 \geq \cdots .$$

1.2. ***The spectral transform.*** If u belongs to a dense G_δ subset $H^{\frac{1}{2}}_{+,\mathrm{gen}}(\mathbb{T})$ of $H^{\frac{1}{2}}_+(\mathbb{T})$, one can establish (see [Gérard and Grellier 2012]) that

$$\rho_1 > \sigma_1 > \rho_2 > \sigma_2 > \cdots$$

We set

$$s_{2j-1} = \rho_j, \quad s_{2k} = \sigma_k, \qquad j,k \geq 1.$$

The s_r are called the singular values of the pair (H_u, K_u). Of course

$$\sum_{r=1}^{\infty} s_r^2 = \mathrm{Tr}(H_u^2) + \mathrm{Tr}(K_u^2) = \sum_{n=0}^{\infty}(1+2n)|\hat{u}(n)|^2 < \infty.$$

Conversely, given a square summable *strictly* decreasing sequence $(s_r)_{r\geq 1}$ of positive numbers, the set of $u \in H^{\frac{1}{2}}_+$ such that the s_r are the singular values of the pair (H_u, K_u), in the above sense, is an infinite dimensional torus $\mathcal{T}((s_r)_{r\geq 1})$ [Gérard and Grellier 2012] of $H^{\frac{1}{2}}_+(\mathbb{T})$. This torus is parametrised by the following explicit representation [Gérard and Grellier 2017], where we classically identify functions of $L^2_+(\mathbb{T})$ with holomorphic functions $u = u(z)$ on the unit disc such that

$$\sup_{r<1} \int_0^{2\pi} |u(r\mathrm{e}^{ix})|^2\, dx < \infty.$$

The current element of the infinite dimensional torus $\mathcal{T}((s_r)_{r\geq 1})$ is then given by

$$u(z) = \lim_{N\to\infty}\left\langle \mathscr{C}_N(z)^{-1}(\mathbb{1}_N), \mathbb{1}_N\right\rangle, \qquad |z|<1, \quad \mathbb{1}_N := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{C}^N, \tag{2}$$

where

$$\mathscr{C}_N(z) := \left(\frac{s_{2j-1}\mathrm{e}^{i\psi_{2j-1}} - zs_{2k}\mathrm{e}^{i\psi_{2k}}}{s_{2j-1}^2 - s_{2k}^2}\right)_{1\leq j,k\leq N} \tag{3}$$

and $(\psi_r)_{r\geq 1} \in \mathbb{T}^\infty$ is an arbitrary sequence of angles. Furthermore, the evolution of the new variables $(s_r, \psi_r)_{r\geq 1}$ through the dynamics of (1) is given by

$$\frac{ds_r}{dt} = 0, \quad \frac{d\psi_r}{dt} = s_r^2, \qquad r = 1, 2, \ldots.$$

A natural question is then the description of the regularity of u in these new variables. A first type of answer to this question is provided by results due to Peller and Semmes, see, e.g., [Peller 2003], which characterise the Schatten classes

$$\sum_{r\geq 1} s_r^p < \infty, \quad 0 < p < \infty,$$

in terms of the Besov spaces

$$\sum_{j=0}^{\infty} 2^j \int_{\mathbb{T}} |\Delta_j u|^p \, dx < \infty,$$

where $(\Delta_j u)_{j\geq 0}$ denotes the dyadic blocks of u. In particular, if u is smooth, then $(s_r)_{r\geq 1}$ satisfies

$$\sum_{r=1}^{\infty} s_r^p < \infty, \quad \text{for all } p < \infty.$$

However, the latter condition is far from being sufficient to control high regularity of u. In fact, Sobolev regularity H^s for $s > \frac{1}{2}$ cannot be easily described by the variables $(s_r, \psi_r)_{r\geq 1}$, as shown by the following result.

Theorem 1 [Gérard and Grellier 2017]. *There exists a dense G_δ subset of initial data in*

$$C_+^\infty(\mathbb{T}) := \bigcap_s H_+^s(\mathbb{T})$$

such that the corresponding solutions of (1) satisfies, for every $s > \frac{1}{2}$,

$$\textit{for all } M \geq 1, \qquad \limsup_{t\to\infty} \frac{\|u(t)\|_{H^s}}{|t|^M} = +\infty, \quad \liminf_{t\to\infty} \|u(t)\|_{H^s} < \infty.$$

In other words, in the $(s_r, \psi_r)_{r\geq 1}$ representation, the size of the high Sobolev norms may strongly depend on the angles $(\psi_r)_{r\geq 1}$. The goal of this paper is to investigate this phenomenon in more detail.

1.3. *Overview of the results.* Our first result claims that generic smooth functions u are located on a torus $\mathcal{T}((s_r)_{r\geq 1})$ containing also very singular functions.

Theorem 2. *There exists a dense G_δ subset $\mathcal{G}$ of $C^\infty_+(\mathbb{T})$ such that every element u of $\mathcal{G}$ belongs to $H^{\frac{1}{2}}_{+,\mathrm{gen}}(\mathbb{T})$, and the infinite dimensional torus $\mathcal{T}((s_r)_{r\geq 1})$ passing through u has a dense G_δ subset—for the $H^{\frac{1}{2}}$ topology—which is disjoint of H^s for every $s > \frac{1}{2}$.*

Theorem 2 states that, on the tori $\mathcal{T}((s_r)_{r\geq 1})$ passing through generic smooth functions, the regularity changes dramatically from C^∞ to the outside of H^s for every $s > \frac{1}{2}$. Of course, this result can be seen as a natural extension of Theorem 1 recalled above, of which we use the weaker form that tori $\mathcal{T}((s_r)_{r\geq 1})$ passing through generic smooth functions are unbounded in H^s for every $s > \frac{1}{2}$. However, in order to find singular functions on these tori, we combine it with a structure property of these tori, which we think has its own interest.

Lemma 3. *Let $s > \frac{1}{2}$ and let $(s_r)_{r\geq 1}$ be a square summable decreasing sequence of positive numbers such that the numbers s_r^2, $r \geq 1$ are linearly independent on $\mathbb{Q}$. Then we have the following alternatives*:

- *either $\mathcal{T}((s_r)_{r\geq 1})$ is a bounded subset of H^s,*
- *or $\mathcal{T}((s_r)_{r\geq 1}) \setminus H^s$ is a dense G_δ subset of $\mathcal{T}((s_r)_{r\geq 1})$ for the $H^{\frac{1}{2}}$ topology.*

The point of Theorem 2 is that, even for fast decaying singular values (s_r), the regularity of u may be spoiled by the relative smallness of the gaps $s_r - s_{r+1}$ with respect to s_r. In fact, if

$$u_N(z) = \langle \mathscr{C}_N(z)^{-1}(\mathbb{1}_N), \mathbb{1}_N \rangle, \quad \psi_r = 0,\ r = 1, 2, \ldots,$$

with the notation introduced above, then, using the positivity property of the Hankel matrices $\Gamma_{\hat{u}_N}$ and $\tilde{\Gamma}_{\hat{u}_N}$ equivalent to $\psi_r = 0$ for all r (see [Gérard and Grellier 2014; Gérard and Pushnitski 2015]), we prove in the Appendix that

$$\|u_N\|_{C^1(\mathbb{T})} \geq \sum_{j=1}^{N} \frac{s_{2j-1}s_{2j}}{s_{2j-1} - s_{2j}}. \tag{4}$$

It is then easy to find fast decaying sequences (s_r) such that the above right hand side tends to infinity as N goes to infinity, which implies that (u_N) is unbounded in $C^1(\mathbb{T})$. However, at this stage we do not know how to conclude that u is not in $C^1(\mathbb{T})$.

Our two other results state some uniform analytic regularity for tori $\mathcal{T}((s_r)_{r\geq 1})$ where the sequence $(s_r)_{r\geq 1}$ satisfies some specific property ruling out the phenomenon of small gaps.

Theorem 4. *For every $\rho > 0$, there exists $\delta_0 > 0$ such that, for any $\delta \in (0, \delta_0)$, if*

$$\text{for all } r \geq 1, \quad s_{r+1} \leq \delta s_r,$$

all functions $u \in \mathcal{T}((s_r)_{r\geq 1})$ are holomorphic and uniformly bounded in the disc $|z| < 1+\rho$. Consequently, for any initial datum corresponding to some of these functions, the solution of the cubic Szegő equation (1) is analytic in the disc of radius $1+\rho$ for all time, and is uniformly bounded in this disc. In particular, the trajectory is bounded in $C^\infty(\mathbb{T})$.

Theorem 4 applies in particular to geometric sequences $s_r = \mathrm{e}^{-rh}$ for $h > 0$ large enough. Our last result explores in more detail the case of geometric sequences $s_r = \mathrm{e}^{-rh}$, where $h > 0$ is arbitrary. In this case, we still obtain some uniform analytic regularity, but with a constraint on the angles ψ_r.

Theorem 5. *Let $h > 0$ and $\theta \in \mathbb{R}$. Assume (s_r) is given by $s_r = \mathrm{e}^{-rh}$ and (ψ_r) by $\psi_r = r\theta h$. Then there exists $\rho > 0$ such that the corresponding elements of $\mathcal{T}((s_r)_{r\geq 1})$ are holomorphic and uniformly bounded in the disc $|z| < 1+\rho$.*

We do not know whether or not geometric tori are embedded into the space of analytic functions on $\mathbb{T}$. What we are able to prove is that, for transcendental γ, we have the following alternatives:

- either there exists $\rho > 0$ such that every element of $\mathcal{T}((\gamma^r)_{r\geq 1})$ is holomorphic on the disc $|z| < 1+\rho$, with a uniform bound,
- or the nonanalytic elements of $\mathcal{T}((\gamma^r)_{r\geq 1})$ form a dense G_δ subset of $\mathcal{T}((\gamma^r)_{r\geq 1})$ for the $H^{\frac{1}{2}}$ topology.

This is a special case of an extension of Lemma 3 to analytic regularity (see Lemma 8).

1.4. ***Open problems.*** In view of the above theorems, the most natural open question is certainly to decide whether Theorem 4 can be generalised to any parameter $\delta < 1$. In particular, as we questioned above, if $0 < \gamma < 1$, is it true that the infinite dimensional torus $\mathcal{T}((\gamma^r)_{r\geq 1})$ is included in the space of analytic functions on $\mathbb{T}$?

Another question connected to Theorem 2 relies on estimate (4). Assuming that

$$\sum_{j=1}^{\infty} \frac{s_{2j-1}s_{2j}}{s_{2j-1} - s_{2j}} = \infty,$$

can one infer that the function $u \in \mathcal{T}((s_r)_{r\geq 1})$ characterised by $\psi_r = 0,\ r = 1, 2, \ldots,$ is not C^1 on $\mathbb{T}$? In view of Lemma 3, this would imply, if moreover the s_r^2 are linearly independent on $\mathbb{Q}$, that most of the points on this torus would be singular — say, not in H^2. Then it would be interesting to draw the consequences of this property for long term behaviour of solutions of the cubic Szegő equation on this torus.

1.5. ***Organisation of the paper.*** The proof of Theorem 2 is provided in Section 2 after reducing to Lemma 3 and Theorem 1. The proof of Lemma 3 combines a Baire category argument and some elementary ergodic argument for the cubic Szegő flow. Section 3 is devoted to the proof of Theorem 4, which is based on brute force estimates on matrices $\mathscr{C}_N(z)$. In Section 4, we prove Theorem 5 by a different approach relying on the theory of Toeplitz operators and a theorem by Baxter which reduces our analysis to proving that the restriction to $\mathbb{T}$ of a meromorphic function given by an explicit series, has no zero and has index 0, which can be realised using some elementary complex analysis and the Poisson summation formula. Finally, the estimate (4) is derived in the Appendix from an explicit calculation using Cauchy matrices, in the spirit of [Gérard and Grellier 2017; Gérard and Pushnitski 2018].

2. The melting pot property

In this section, we prove Theorem 2. First we reduce the proof to Lemma 3 by the following classical argument.

Lemma 6. *The set of $u \in C^\infty_+(\mathbb{T}) \cap H^{\frac{1}{2}}_{+,\mathrm{gen}}(\mathbb{T})$ such that the squares $s_r(u)^2$, $r \geq 1$ of the singular values $s_r(u)$ are linearly independent on $\mathbb{Q}$, is a dense G_δ subset of $C^\infty_+(\mathbb{T})$.*

Proof. From the proof of [Gérard and Grellier 2012, Lemma 7], we already know that $C^\infty_+(\mathbb{T}) \cap H^{\frac{1}{2}}_{+,\mathrm{gen}}(\mathbb{T})$ is a dense G_δ subset of $C^\infty_+(\mathbb{T})$. In fact, we can slightly modify the proof as follows. For every N, consider the open subset $\mathcal{O}_N$ made of functions $u \in C^\infty_+(\mathbb{T})$ such that the first singular values of H_u and K_u satisfy

$$\rho_1(u) > \sigma_1(u) > \rho_2(u) > \sigma_2(u) > \cdots \rho_N(u) > \sigma_N(u),$$

and such that any nontrivial linear combination of

$$\rho_1(u)^2, \sigma_1(u)^2, \rho_2(u)^2, \sigma_2(u)^2, \ldots, \rho_N(u)^2, \sigma_N(u)^2$$

with integer coefficients in $[-N, N]$, is not zero. Approximating elements of $C^\infty_+(\mathbb{T})$ by rational functions, and using the inverse spectral theorem of [Gérard and Grellier 2012] for rational functions, we easily obtain that $\mathcal{O}_N$ is dense. The conclusion follows from Baire's theorem. □

Intersecting the dense G_δ subset of $C^\infty_+(\mathbb{T})$ provided by this lemma with the one provided by Theorem 1 — or its weaker form, saying that the corresponding Szegő trajectories are unbounded in every $H^s,\ s > \frac{1}{2}$ — we observe that Theorem 2 is a consequence of Lemma 3, which we restate for the convenience of the reader.

Lemma 7. *Let $(s_r)_{r\geq 1}$ be a square-summable decreasing sequence of positive numbers such that the numbers $s_r^2,\ r \geq 1$ are linearly independent on $\mathbb{Q}$ and let $s > \frac{1}{2}$. Then we have the following alternatives*:

- *either $\mathcal{T}((s_r)_{r\geq 1})$ is a bounded subset of H^s*,
- *or $\mathcal{T}((s_r)_{r\geq 1}) \setminus H^s$ is a dense G_δ subset of $\mathcal{T}((s_r)_{r\geq 1})$ for the $H^{\frac{1}{2}}$ topology.*

Proof. Recall [Gérard and Grellier 2012; 2017] that, for the $H^{\frac{1}{2}}$ topology, $\mathcal{T}((s_r)_{r\geq 1})$ is homeomorphic to the infinite dimensional torus $\mathbb{T}^\infty$, endowed with the product topology, through the parametrisation given by (2) and (3). In particular, it is a compact metrisable space. For every $s > \frac{1}{2}$, the function

$$\|v\|_{H^s} = \left(\sum_{n=0}^{\infty} (1+n)^{2s} |\hat{v}(n)|^2\right)^{\frac{1}{2}}$$

is lower semicontinuous on $\mathcal{T}((s_r)_{r\geq 1})$. For every positive integer ℓ, consider

$$F_\ell = \{v \in \mathcal{T}((s_r)_{r\geq 1}) : \|v\|_{H^s} \leq \ell\}.$$

F_ℓ is a closed subset of $\mathcal{T}((s_r)_{r\geq 1})$, and the complement of the union of the F_ℓ is precisely $\mathcal{T}((s_r)_{r\geq 1}) \setminus H^s$. Hence, by the Baire theorem, either this set is a dense G_δ subset of $\mathcal{T}((s_r)_{r\geq 1})$, or there exists $\ell \geq 1$ such that F_ℓ has a nonempty interior. Assume that some F_ℓ has a nonempty interior, and let us show that $\mathcal{T}((s_r)_{r\geq 1})$ is a bounded subset of H^s. Let $(\psi_r^0)_{r\geq 1} \in \mathbb{T}^\infty$ such that the corresponding point v^0 in $\mathcal{T}((s_r)_{r\geq 1})$ lies in the interior of F_ℓ. In view of the product topology on $\mathbb{T}^\infty$, there exists some integer $N \geq 1$ and some $\varepsilon > 0$ such that all the elements of $\mathcal{T}((s_r)_{r\geq 1})$ corresponding to

$$\psi_r \in]\psi_r^0 - \varepsilon, \psi_r^0 + \varepsilon[, \quad r = 1, \dots, N,$$

form an open set U contained in F_ℓ. At this stage we appeal to the number theoretic assumption on the s_r^2, which we use classically under the form that the trajectory

$$\{(\psi_r^0 + t s_r^2)_{r=1,\dots,N} : t \in \mathbb{R}\}$$

is dense into the torus $\mathbb{T}^N$. Since, as recalled in the introduction, this trajectory is precisely the projection of the trajectory of the cubic Szegő flow Φ_t on the first N components, we infer that every element of $\mathcal{T}((s_r)_{r\geq 1})$ is contained in some open set $\Phi_t(U)$. Since the cubic Szegő equation is well-posed on H^s [Gérard and Grellier 2010], we infer that $\mathcal{T}((s_r)_{r\geq 1})$ is covered by the union of the interiors of the F_m for $m \geq 1$. By compactness, it is covered by a finite union, which precisely means that $\mathcal{T}((s_r)_{r\geq 1})$ is bounded in H^s. □

As stated in the introduction for geometric sequences, the following analogous result holds in the analytic setting.

Lemma 8. *Let $(s_r)_{r\geq 1}$ be a square summable decreasing sequence of positive numbers such that the numbers s_r^2, $r \geq 1$ are linearly independent on $\mathbb{Q}$. Then we have the following alternatives*:

- *either there exists $\rho > 0$ such that every element of $\mathcal{T}((s_r)_{r\geq 1})$ is holomorphic on the disc $|z| < 1 + \rho$, with a uniform bound*,
- *or the nonanalytic elements of $\mathcal{T}((s_r)_{r\geq 1})$ form a dense G_δ subset of $\mathcal{T}((s_r)_{r\geq 1})$ for the $H^{\frac{1}{2}}$ topology.*

Proof. The proof is an adaptation of the preceding one (Lemma 3) to the analytic setting. As, from [Gérard et al. 2015], the cubic Szegő equation propagates analyticity, the result follows from the Baire theorem applied to the closed sets

$$F_\ell := \left\{ v \in \mathcal{T}((s_r)_{r\geq 1}) \; : \; \sum_{n=0}^{\infty} \mathrm{e}^{\frac{n}{\ell}} |\hat{v}(n)| \leq \ell \right\}$$

for $\ell \geq 1$. □

3. Example of bounded analytic tori

In this section, we prove Theorem 4.

Let $u \in \mathcal{T}((s_r)_{r\geq 1})$. Recall that

$$u = \lim_{N\to\infty} u_N, \qquad \text{where } u_N(z) := \langle \mathscr{C}_N(z)^{-1} \mathbb{1}_N, \mathbb{1}_N \rangle,$$

$$\mathscr{C}_N(z) := \left(\frac{s_{2j-1}\mathrm{e}^{i\psi_{2j-1}} - z s_{2k}\mathrm{e}^{i\psi_{2k}}}{s_{2j-1}^2 - s_{2k}^2} \right)_{1\leq j,k\leq N},$$

and

$$\mathbb{1}_N = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{C}^N.$$

Our assumption is

$$s_{r+1} = \varepsilon_r s_r, \qquad r \geq 1,$$

where the sequence $(\varepsilon_r)_{r\geq 1}$ satisfies

$$0 < \varepsilon_r \leq \delta \quad \text{for some } \delta < 1.$$

Our aim is to prove that, for δ sufficiently small, the functions u_N are holomorphic and uniformly bounded in some disc of radius $1 + \rho$, where $\rho > 0$, independently of N. Our strategy is to use that $\mathscr{C}_N(0)$ is related to a Cauchy matrix, and hence, that an explicit formula for its inverse is known. We write

$$\mathscr{C}_N(z) = \mathscr{C}_N(0) - z\dot{\mathscr{C}}_N = \mathscr{C}_N(0)(I - z\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N),$$

where

$$\dot{\mathscr{C}}_N := \left(\frac{s_{2k}e^{i\psi_{2k}}}{s_{2j-1}^2 - s_{2k}^2}\right)_{1\le j,k\le N},$$

and we establish the following lemma.

Lemma 9. *For any* $0 < \delta < 1$, *there exists some constant* $C_\delta > 0$ *such that, for any* $N \ge 1$,

$$\sum_{j,k} |(\mathscr{C}_N(0)^{-1})_{j,k}| \le C_\delta s_1. \tag{5}$$

There exists a universal constant $A > 0$ *such that, for* $\delta \in \left(0, \frac{1}{2}\right)$ *and for any* $N \ge 1$,

$$\|\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N\|_{\ell^1\to\ell^1} \le A\,\delta. \tag{6}$$

Let us assume Lemma 9 proved. Take $\rho > 0$, and choose δ_0 such that $A\delta_0(1+\rho) \le \frac{1}{2}$. Hence, for any $\delta \in (0, \delta_0)$, from estimate (6),

$$(I - z\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N)$$

is invertible for any z with $|z| < 1+\rho$ and its inverse $R_N(z)$ is analytic and has uniformly bounded norm for any z with $|z| < 1+\rho$. Indeed, for any $N \ge 1$, and any z with $|z| < 1+\rho$, by the Neumann series identity,

$$\|R_N(z)\|_{\ell^1\to\ell^1} \le \sum_{k=0}^{\infty} |z|^k \|\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N\|_{\ell^1\to\ell^1}^k \le \sum_{k=0}^{\infty} 2^{-k} \le 2. \tag{7}$$

Writing

$$\mathscr{C}_N(z)^{-1} = (I - z(\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N))^{-1}\mathscr{C}_N(0)^{-1} = R_N(z)(\mathscr{C}_N(0))^{-1}$$

we get

$$u_N(z) = \left\langle R_N(z)\mathscr{C}_N(0)^{-1}(\mathbb{1}_N), \mathbb{1}_N\right\rangle.$$

Using (5) and (7), we conclude that the series defining u_N converges uniformly for $|z| < 1+\rho$. Hence u_N is analytic and uniformly bounded in the disc of radius $1+\rho$. We infer that u is as well analytic in the disc of radius $1+\rho$ and bounded on this disc.

This completes the proof of Theorem 4, modulo Lemma 9.

3.1. ***Proof of Lemma 9.*** Notice that

$$\mathscr{C}_N(0) = \operatorname{diag}(s_{2j-1}\mathrm{e}^{i\psi_{2j-1}})\mathscr{T}, \qquad \mathscr{T} := \left(\frac{1}{s_{2j-1}^2 - s_{2k}^2}\right)_{1\le j,k\le N}.$$

Since $\mathscr{T}$ is a Cauchy matrix, its inverse is explicitly known, so the inverse of $\mathscr{C}_N(0)$ is given by

$$\mathscr{C}_N(0)^{-1} = \left(\frac{(-1)^{j+k+N}\alpha_j^{(N)}\beta_k^{(N)}}{s_{2j-1}^2 - s_{2k}^2}\frac{1}{s_{2j-1}\mathrm{e}^{i\psi_{2j-1}}}\right)_{1\le k,j\le N},$$

where

$$\alpha_j^{(N)} := \frac{\prod_\ell (s_{2j-1}^2 - s_{2\ell}^2)}{\prod_{\ell<j}(s_{2\ell-1}^2 - s_{2j-1}^2)\prod_{\ell>j}(s_{2j-1}^2 - s_{2\ell-1}^2)},$$

$$\beta_k^{(N)} := \frac{\prod_\ell (s_{2\ell-1}^2 - s_{2k}^2)}{\prod_{\ell<k}(s_{2\ell}^2 - s_{2k}^2)\prod_{\ell>k}(s_{2k}^2 - s_{2\ell}^2)}.$$

In particular,

$$\begin{aligned}|\alpha_j^{(N)}| &= \prod_{\ell<j}\frac{s_{2\ell}^2}{s_{2\ell-1}^2}\prod_{\ell<j}\left(\frac{1-\prod_{r=2\ell}^{2j-2}\varepsilon_r^2}{1-\prod_{r=2\ell-1}^{2j-2}\varepsilon_r^2}\right)\prod_{\ell>j}\frac{\left(1-\prod_{r=2j-1}^{2\ell-1}\varepsilon_r^2\right)}{\left(1-\prod_{r=2j-1}^{2\ell-2}\varepsilon_r^2\right)}s_{2j-1}^2(1-\varepsilon_{2j-1}^2)\\ &\le \prod_{\ell<j}\varepsilon_{2\ell-1}^2\,\frac{s_{2j-1}^2}{\prod_{m=1}^\infty(1-\delta^{4m})}.\end{aligned}$$

Indeed, in the first line above, the factors in the second product are bounded by 1, while, in the third product, the ℓ-factor is bounded by $\frac{1}{1-\delta^{4(\ell-j)}}$. Similarly, we have

$$\begin{aligned}|\beta_k^{(N)}| &= \prod_{\ell<k}\frac{s_{2\ell-1}^2}{s_{2\ell}^2}\prod_{\ell>k}\left(\frac{1-\prod_{r=2k}^{2\ell-2}\varepsilon_r^2}{1-\prod_{r=2k}^{2\ell-1}\varepsilon_r^2}\right)\prod_{\ell<k}\frac{\left(1-\prod_{r=2\ell-1}^{2k-1}\varepsilon_r^2\right)}{\left(1-\prod_{r=2\ell}^{2k-1}\varepsilon_r^2\right)}s_{2k-1}^2(1-\varepsilon_{2k-1}^2)\\ &\le \prod_{\ell<k}\frac{1}{\varepsilon_{2\ell-1}^2}\,\frac{s_{2k-1}^2}{\prod_{m=1}^\infty(1-\delta^{4m})}.\end{aligned}$$

Setting

$$B_\delta = \frac{1}{\prod_{m=1}^\infty(1-\delta^{4m})^2},$$

we obtain

$$|(\mathscr{C}_N(0)^{-1})_{kj}| \le B_\delta \frac{s_{2j-1}s_{2k-1}^2}{|s_{2j-1}^2 - s_{2k}^2|} \prod_{\ell<j} \varepsilon_{2\ell-1}^2 \prod_{\ell<k} \frac{1}{\varepsilon_{2\ell-1}^2}$$

$$\le B_\delta \begin{cases} \frac{1}{1-\delta^{4(k-j)+2}} \frac{s_{2k-1}^2}{s_{2j-1}} \prod_{j\le\ell<k} \frac{1}{\varepsilon_{2\ell-1}^2} & \text{if } j<k, \\ \frac{1}{1-\delta^2} s_{2j-1} & \text{if } j=k, \\ \frac{1}{1-\delta^{4(j-k-1)+2}} s_{2j-1} \frac{s_{2k-1}^2}{s_{2k}^2} \prod_{k\le\ell<j} \varepsilon_{2\ell-1}^2 & \text{if } j>k. \end{cases}$$

To summarise,

$$|(\mathscr{C}_N(0)^{-1})_{kj}| \le \frac{B_\delta}{1-\delta^2} s_{2j-1} \begin{cases} \delta^{2(k-j)} & \text{if } j<k, \\ 1 & \text{if } j=k, k+1, \\ \delta^{2(j-k-1)} & \text{if } j>k+1. \end{cases} \tag{8}$$

In particular, it gives

$$\sum_{k,j} |(\mathscr{C}_N(0)^{-1})_{jk}| \le \frac{2B_\delta}{(1-\delta^2)^2} \sum_{j\le N} s_{2j-1} \le \frac{2B_\delta s_1}{(1-\delta^2)^3}.$$

This proves estimate (5).

For the second estimate, one has to consider

$$\|\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N\|_{\ell^1\to\ell^1} \le \sup_\ell \sum_k |(\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N)_{k\ell}|.$$

Recall that

$$\dot{\mathscr{C}}_N := \left(\frac{s_{2\ell}\mathrm{e}^{i\psi_{2\ell}}}{s_{2j-1}^2 - s_{2\ell}^2}\right)_{1\le j,\ell\le N}.$$

In particular,

$$|(\dot{\mathscr{C}}_N)_{j\ell}| \le \frac{1}{1-\delta^2} \begin{cases} \frac{s_{2\ell}}{s_{2j-1}^2} & \text{if } j\le\ell, \\ \frac{1}{s_{2\ell}} & \text{if } j\ge\ell+1. \end{cases}$$

As $(\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N)_{k\ell} = \sum_j (\mathscr{C}_N(0)^{-1})_{kj} (\dot{\mathscr{C}}_N)_{j\ell}$, we get from the preceding estimate (8) on $|(\mathscr{C}_N(0)^{-1})_{kj}|$ that:

- If $k>\ell$,

$$|(\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N)_{k\ell}| \le \frac{B_\delta}{(1-\delta^2)^2} \Bigg(\sum_{j\le\ell} \frac{s_{2\ell}}{s_{2j-1}} \prod_{j\le r\le k-1} \varepsilon_{2r}^2 + \sum_{\ell+1\le j\le k} \frac{s_{2j-1}}{s_{2\ell}} \prod_{j\le r\le k-1} \varepsilon_{2r}^2 + \sum_{j\ge k+1} \frac{s_{2j-1}}{s_{2\ell}} \prod_{k+1\le r\le j-1} \varepsilon_{2r-1}^2 \Bigg)$$

$$\leq \frac{B_\delta}{(1-\delta^2)^2}\Bigg(\sum_{j\leq \ell}\delta^{2(\ell-j)+1}\delta^{2(k-j)} + \sum_{\ell+1\leq j\leq k}\delta^{2(j-\ell-1)+1}\delta^{2(k-j)} + \sum_{j\geq k+1}\delta^{2(j-\ell-1)+1}\delta^{2(j-k-1)}\Bigg)$$

$$\leq \delta\frac{B_\delta}{(1-\delta^2)^2}\Bigg(2\delta^{2(k-\ell)}\sum_{s\geq 0}\delta^{4s} + \sum_{\ell+1\leq j\leq k}\delta^{2(j-\ell-1)}\delta^{2(k-j)}\Bigg),$$

then since

$$\sum_{k:\,k\geq \ell+1}\sum_{\ell+1\leq j\leq k}\delta^{2(j-\ell-1)}\delta^{2(k-j)} = \sum_{j\geq \ell+1}\sum_{k\geq j}\delta^{2(j-\ell-1)}\delta^{2(k-j)} = \frac{1}{(1-\delta^2)^2},$$

one gets

$$\sum_{k:\,k>\ell}|(\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N)_{k\ell}| \leq \frac{\delta B_\delta}{(1-\delta^2)^4}\left(\frac{1+3\delta^2}{1+\delta^2}\right).$$

- If $k<\ell$,

$$|(\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N)_{k\ell}|$$

$$\leq \frac{B_\delta}{(1-\delta^2)^2}\Bigg(\sum_{j\leq k}\frac{s_{2\ell}}{s_{2j-1}}\prod_{j\leq r\leq k-1}\varepsilon_{2r}^2 + \sum_{k+1\leq j\leq \ell}\frac{s_{2\ell}}{s_{2j-1}}\prod_{k+1\leq r\leq j-1}\varepsilon_{2r-1}^2 + \sum_{j\geq \ell+1}\frac{s_{2j-1}}{s_{2\ell}}\prod_{k+1\leq r\leq j-1}\varepsilon_{2r-1}^2\Bigg)$$

$$\leq \frac{B_\delta}{(1-\delta^2)^2}\Bigg(\sum_{j\leq k}\delta^{2(\ell-j)+1}\delta^{2(k-j)} + \sum_{k+1\leq j\leq \ell}\delta^{2(\ell-j)+1}\delta^{2(j-k-1)} + \sum_{j\geq \ell+1}\delta^{2(j-\ell-1)+1}\delta^{2(j-k-1)}\Bigg)$$

$$\leq \frac{\delta B_\delta}{(1-\delta^2)^2}\Bigg(2\delta^{2(\ell-k)}\sum_{s\geq 0}\delta^{4s} + \sum_{k+1\leq j\leq \ell}\delta^{2(\ell-j)}\delta^{2(j-k-1)}\Bigg)$$

and, as before,

$$\sum_{k:\,k<\ell}|(\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N)_{k\ell}| \leq \frac{\delta\, B_\delta}{(1-\delta^2)^4}\left(\frac{1+3\delta^2}{1+\delta^2}\right).$$

- For $k=\ell$,

$$|(\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N)_{kk}|$$

$$\leq \frac{B_\delta}{(1-\delta^2)^2}\Bigg(\sum_{j\leq \ell}\frac{s_{2\ell}}{s_{2j-1}}\prod_{j\leq r\leq k-1}\varepsilon_{2r}^2 + \sum_{j\geq k+1}\frac{s_{2j-1}}{s_{2\ell}}\prod_{k+1\leq r\leq j-1}\varepsilon_{2r-1}^2\Bigg)$$

$$\le \frac{B_\delta}{(1-\delta^2)^2}\Big(\sum_{j\le \ell}\delta^{2(\ell-j)+1}\delta^{2(k-j)}+\sum_{j\ge k+1}\delta^{2(j-\ell-1)+1}\delta^{2(j-k-1)}\Big)$$
$$\le \frac{B_\delta}{(1-\delta^2)^2}\Big(\sum_{j\le k}\delta^{4(k-j)+1}+\sum_{j\ge k+1}\delta^{4(j-k-1)+1}\Big)$$
$$=2\frac{\delta B_\delta}{(1-\delta^2)^2(1-\delta^4)}.$$

Eventually, if $\delta\le\frac{1}{2}$, say, we obtain, with a universal constant A,

$$\|(\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N)\|_{\ell^1\to\ell^1}\le \sup_\ell\sum_k |(\mathscr{C}_N(0)^{-1}\dot{\mathscr{C}}_N)_{k\ell}|\le A\delta.$$

This completes the proof of Lemma 9.

4. The totally geometric spectral data

In this section, we consider the totally geometric case and prove Theorem 5. For some fixed $h>0$ and $\theta\in\mathbb{R}$, we consider the symbol u with spectral data (s_r,ψ_r) with $s_r=\mathrm{e}^{-rh}$ and $\psi_r=r\theta h$. In particular, $s_{r+1}=s_r\mathrm{e}^{-h}$ so that, for h sufficiently large, it becomes a particular case of subgeometric spectral data treated in Theorem 4. However, the result here does not require any smallness on e^{-h}.

Our strategy here is to use Toeplitz operators and a stability result from [Baxter 1963].

4.1. *Background on Toeplitz operators.* Let us first introduce some basic notation. For a continuous function Φ on $\mathbb{T}$, we denote by $T(\Phi)$ the Toeplitz operator of symbol Φ defined on $L^2_+(\mathbb{T})$ by

$$T(\Phi)(f)=\Pi(\Phi f)$$

or equivalently, the operator defined on $\ell^2(\mathbb{N})$ by

$$(T(\Phi)((a_k)))_j:=\sum_{k=0}^{\infty}\hat{\Phi}(j-k)a_k,\quad j\in\mathbb{N}.$$

For any integer N, we denote by $T_N(\Phi)$ the truncated operator defined by

$$T_N(\Phi):=\Pi_N T(\Phi)\Pi_N.$$

Here

$$\Pi_N:\ell^2(\mathbb{N})\to\ell^2(\mathbb{N})$$

is the orthogonal projector:

$$(x_0,x_1,x_2,\dots)\mapsto(x_0,x_1,x_2,\dots,x_{N-1}).$$

The operator T_N corresponds to the $N \times N$ truncated Toeplitz matrix

$$(\hat{\Phi}(j-k))_{0\le j,k\le N-1}.$$

Recall that a sequence of $N \times N$ matrices $(A_N)_{N\ge 1}$ is said to be stable if there is an N_0 such that the matrices A_N are invertible for all $N \ge N_0$ and

$$\sup_{N\ge N_0} \|A_N^{-1}\|_{\ell^2\to\ell^2} < \infty.$$

Theorem 10 [Baxter 1963; Böttcher and Grudsky 2000]. *The sequence* $(T_N(\Phi))_{N\ge 1}$ *is stable if and only if* $T(\Phi)$ *is invertible.*

Let us emphasise that the operators are considered as operators acting on $\ell^2(\mathbb{N})$ or $L^2_+(\mathbb{T})$ so that the stability is evaluated in the $\ell^2(\mathbb{N})$ norm. The characterisation of the invertibility of Toeplitz operators is well known. We recall it for the convenience of the reader.

Theorem 11. *Let* Φ *be a continuous function on the unit circle. If* Φ *has index* 0 *and does not vanish on the circle*, *then* T_Φ *is invertible on* $L^2_+(\mathbb{T})$. *Under these hypotheses*, $\Phi = \mathrm{e}^{\varphi} = \Phi_+\overline{\Phi_-}$ *with*

$$\Phi_+ = \mathrm{e}^{\Pi(\varphi)} \quad \textit{and} \quad \Phi_- = \mathrm{e}^{\overline{(I-\Pi)(\varphi)}}$$

and the inverse of T_Φ *is given by* $T_{\Phi_+^{-1}}T_{\overline{\Phi_-}^{-1}}$.

As an immediate consequence, one gets the following characterisation of the stability of truncated Toeplitz operators.

Corollary 12. *Let* Φ *be a continuous function on the unit circle. The sequence of truncated Toeplitz operators* $(T_N(\Phi))$ *is stable if and only if* Φ *has no zero on the unit circle and has index* 0.

We are going to use this argument to prove Theorem 5.

4.2. ***Totally geometric spectral data and Toeplitz operators.*** We claim that in the case of totally geometric spectral data, the explicit formula giving u_N involves the inverse of a truncated Toeplitz operator. From direct computation, one has

$$\mathscr{C}_N(z) = \left(\frac{\omega^{2j-1} - z\omega^{2k}}{|\omega|^{4j-2} - |\omega|^{4k}}\right)_{1\le j,k\le N} = \left(\frac{1}{\bar{\omega}^{2j-1}}\frac{1 - z\omega^{2(k-j)+1}}{1-|\omega|^{4(k-j)+2}}\right)_{1\le j,k\le N},$$

where $\omega = \mathrm{e}^{-h(1-i\theta)}$. In that case, if $T_N(z)$ and $T_{N,r}(z)$ denote the matrices

$$T_N(z) = \left(\frac{1 - z\omega^{2(k-j)+1}}{1-|\omega|^{4(k-j)+2}}\right)_{1\le j,k\le N}$$

and

$$T_{N,r}(z) = \left(r^{k-j}\frac{1 - z\omega^{2(k-j)+1}}{1-|\omega|^{4(k-j)+2}}\right)_{1\le j,k\le N},$$

we get from our explicit formula, for any $r>0$,

$$u_N(z)=\langle T_N(z)^{-1}(\bar\omega^{2j-1}),\mathbb{1}\rangle=\big\langle T_{N,r}(z)^{-1}(r^{-j}\bar\omega^{2j-1})_{1\le j\le N}),\ (r^k)_{1\le k\le N}\big\rangle.$$

We consider for $|\zeta|=r,\ |\omega|^2<r<1,\ z\in\mathbb{C}$, the symbol

$$\Phi(z,\zeta):=\sum_{\ell\in\mathbb{Z}}\frac{1-z\omega^{2\ell+1}}{1-|\omega|^{4\ell+2}}\zeta^\ell.$$

The transpose of the matrix

$$\left(r^{k-j}\frac{1-z\omega^{2(k-j)+1}}{1-|\omega|^{4(k-j)+2}}\right)_{j,k\ge1}$$

corresponds to the matrix of the Toeplitz operator of symbol

$$\Phi(z,r\cdot):\zeta\mapsto\Phi(z,r\zeta).$$

We are going to prove the following result.

Proposition 13. *There exist $|\omega|^2<r<1$ and $\rho>0$ such that the function $\zeta\mapsto\Phi(z,r\zeta)$ has no zero and has index* 0 *on the unit circle, for every z such that $|z|<1+\rho$.*

Assuming this result proved, we obtain by Corollary 12 that, uniformly in z, $|z|<1+\rho$, $\|T_{N,r}(z)^{-1}\|_{\ell^2\to\ell^2}$ is bounded (or more precisely the norm of its transpose is bounded). As $|\omega|^2<r<1$, we obtain that the sequence $(u_N(z))_N$ with

$$u_N(z)=\big\langle T_{N,r}(z)^{-1}(r^{-j}\bar\omega^{2j-1})_{1\le j\le N}),\ (r^k)_{1\le k\le N}\big\rangle$$

is uniformly bounded and converges to $u(z)$ for any z, $|z|<1+\rho$. We conclude as in the previous section. This ends the proof of Theorem 5.

It remains to prove Proposition 13, which is the objective of the next subsections. As a preliminary, observe that, for $|\omega|^2<|\zeta|<1,\ \gamma=|\omega|^2$,

$$\Phi(z,\zeta)=F_\gamma(\zeta)-z\omega F_\gamma(\zeta\omega^2),$$

where

$$F_\gamma(\zeta)=\Phi(0,\zeta)=\sum_{j\in\mathbb{Z}}\frac{\zeta^j}{1-\gamma^{2j+1}},\quad \gamma=|\omega|^2. \tag{9}$$

We collect some basic properties of function F_γ in the following lemma.

Lemma 14. *The function F_γ has a meromorphic extension in $\mathbb{C}\setminus\{0\}$ given by*

$$F_\gamma(\zeta)=\sum_{\ell\in\mathbb{Z}}\frac{\gamma^\ell}{1-\zeta\gamma^{2\ell}}. \tag{10}$$

Its only poles in $\mathbb{C}\setminus\{0\}$ *are the* $\gamma^{2\ell},\ \ell\in\mathbb{Z}$ *and* $F_\gamma(\gamma^{2\ell+1})=0,\ \ell\in\mathbb{Z}$. *Furthermore*

$$F_\gamma\left(\frac{1}{\zeta}\right)=-\zeta F_\gamma(\zeta),\qquad F_\gamma\left(\frac{\zeta}{\gamma^2}\right)=\gamma F_\gamma(\zeta). \tag{11}$$

Proof. Let us give another expression of F_γ. By assumption,

$$\gamma<|\zeta|<1,$$

hence, $|\zeta|>\gamma^2$ and

$$\begin{aligned}
F_\gamma(\zeta)&=\sum_{j=0}^{\infty}\frac{\zeta^j}{1-\gamma^{2j+1}}+\sum_{j=0}^{\infty}\frac{\zeta^{-j-1}}{1-\gamma^{-2j-1}}\\
&=\sum_{j=0}^{\infty}\zeta^j\sum_{\ell=0}^{\infty}\gamma^{(2j+1)\ell}-\sum_{j=0}^{\infty}\frac{\zeta^{-j-1}\gamma^{2j+1}}{1-\gamma^{2j+1}}\\
&=\sum_{\ell=0}^{\infty}\gamma^\ell\sum_{j=0}^{\infty}(\zeta\gamma^{2\ell})^j-\gamma\zeta^{-1}\sum_{\ell=0}^{\infty}\gamma^\ell\sum_{j=0}^{\infty}(\zeta^{-1}\gamma^{2\ell+2})^j\\
&=\sum_{\ell=0}^{\infty}\frac{\gamma^\ell}{1-\zeta\gamma^{2\ell}}-\sum_{\ell=0}^{\infty}\frac{\gamma^{\ell+1}}{\zeta-\gamma^{2\ell+2}},
\end{aligned}$$

and we obtain (10). The other properties are elementary consequences of this equality. □

Remark 15. Set $\gamma=\mathrm{e}^{-\pi\tau},\ \tau>0$. From the second identity (11), we observe that the meromorphic function

$$G_\tau(w)=\mathrm{e}^{2i\pi w}(F_\gamma(\mathrm{e}^{2i\pi w}))^2$$

satisfies

$$\text{for all } \lambda\in\mathbb{Z}+i\tau\mathbb{Z},\qquad G_\tau(w+\lambda)=G_\tau(w),$$

which means that G_τ is an elliptic function relative to the lattice $\mathbb{Z}+i\tau\mathbb{Z}$. Since G_τ has only double poles at the lattice points, with singularity

$$\frac{1}{(\zeta-1)^2}\sim-\frac{1}{4\pi^2w^2}$$

at $w=0$, and since it cancels at points $i\frac{\tau}{2}+\mathbb{Z}+i\tau\mathbb{Z}$, we infer that

$$G_\tau(w)=-\frac{1}{4\pi^2}\Big(\wp_\tau(w)-\wp_\tau\Big(i\frac{\tau}{2}\Big)\Big),$$

where

$$\wp_\tau(w)=\frac{1}{w^2}+\sum_{\lambda\neq 0}\Big(\frac{1}{(w-\lambda)^2}-\frac{1}{\lambda^2}\Big)$$

denotes the Weierstrass $\mathfrak{P}$ function relative to the lattice $\mathbb{Z}+i\tau\mathbb{Z}$. See, e.g., [Saks and Zygmund 1952].

4.3. *Ruling out the zeroes on the unit circle.* In this section, we prove the following lemma.

Lemma 16. *There exists $\rho>0$ such that $\Phi(z,\zeta)$ does not vanish in a neighbourhood of the circle $|\zeta|=1$ for any z such that $|z|\le 1+\rho$.*

Lemma 16 is a consequence of the following result.

Lemma 17. *For every $\gamma\in(0,1)$,*

$$\gamma^{\frac{1}{2}}\max_{|\zeta|=\gamma}|F_\gamma(\zeta)|<\min_{|\zeta|=1}|F_\gamma(\zeta)|.$$

Proof. First of all we rewrite both sides of the above inequality. If $\zeta=\mathrm{e}^{i\theta}$,

$$\begin{aligned}F_\gamma(\zeta)&=\sum_{k=0}^{\infty}\frac{\gamma^k}{1-\gamma^{2k}\mathrm{e}^{i\theta}}+\sum_{\ell=1}^{\infty}\frac{\gamma^{-\ell}}{1-\gamma^{-2\ell}\mathrm{e}^{i\theta}}\\&=\sum_{k=0}^{\infty}\frac{\gamma^k}{1-\gamma^{2k}\mathrm{e}^{i\theta}}+\sum_{\ell=1}^{\infty}\frac{\gamma^{\ell}}{\gamma^{2\ell}-\mathrm{e}^{i\theta}}\\&=\frac{1}{1-\mathrm{e}^{i\theta}}+\sum_{\ell=1}^{\infty}\frac{\gamma^\ell(1+\gamma^{2\ell})(1-\mathrm{e}^{-i\theta})}{1+\gamma^{4\ell}-2\gamma^{2\ell}\cos\theta}\\&=(1-\mathrm{e}^{-i\theta})\left(\frac{1}{2(1-\cos\theta)}+\sum_{\ell=1}^{\infty}\frac{\gamma^\ell(1+\gamma^{2\ell})}{1+\gamma^{4\ell}-2\gamma^{2\ell}\cos\theta}\right),\end{aligned}$$

hence

$$\begin{aligned}|F_\gamma(\zeta)|&=\frac{1}{2|\sin(\theta/2)|}+2|\sin(\theta/2)|\sum_{\ell=1}^{\infty}\frac{\gamma^\ell(1+\gamma^{2\ell})}{1+\gamma^{4\ell}-2\gamma^{2\ell}\cos\theta}\\&=|\sin(\theta/2)|\sum_{\ell\in\mathbb{Z}}\frac{\gamma^\ell(1+\gamma^{2\ell})}{1+\gamma^{4\ell}-2\gamma^{2\ell}\cos\theta}.\end{aligned}$$

Similarly, if $\zeta=\gamma\mathrm{e}^{i\varphi}$, we have

$$\begin{aligned}F_\gamma(\zeta)&=\sum_{k=0}^{\infty}\frac{\gamma^k}{1-\gamma^{2k+1}\mathrm{e}^{i\varphi}}+\sum_{\ell=0}^{\infty}\frac{\gamma^{-\ell-1}}{1-\gamma^{-2\ell-1}\mathrm{e}^{i\varphi}}\\&=\sum_{k=0}^{\infty}\frac{\gamma^k}{1-\gamma^{2k+1}\mathrm{e}^{i\varphi}}+\sum_{\ell=0}^{\infty}\frac{\gamma^{\ell}}{\gamma^{2\ell+1}-\mathrm{e}^{i\varphi}}\\&=\sum_{\ell=0}^{\infty}\frac{\gamma^\ell(1+\gamma^{2\ell+1})(1-\mathrm{e}^{-i\varphi})}{1+\gamma^{4\ell+2}-2\gamma^{2\ell+1}\cos\varphi},\end{aligned}$$

so that

$$|F_\gamma(\zeta)| = 2|\sin(\varphi/2)| \sum_{\ell=0}^{\infty} \frac{\gamma^\ell(1+\gamma^{2\ell+1})}{1+\gamma^{4\ell+2}-2\gamma^{2\ell+1}\cos\varphi}$$
$$= |\sin(\varphi/2)| \sum_{\ell\in\mathbb{Z}} \frac{\gamma^\ell(1+\gamma^{2\ell+1})}{1+\gamma^{4\ell+2}-2\gamma^{2\ell+1}\cos\varphi}.$$

Consequently,

$$\min_{|\zeta|=1}|F_\gamma(\zeta)| - \gamma^{\frac{1}{2}} \max_{|\zeta|=\gamma}|F_\gamma(\zeta)| = \min_{\theta\in\mathbb{T}}|\sin(\theta/2)| \sum_{\ell\in\mathbb{Z}} \frac{\gamma^\ell(1+\gamma^{2\ell})}{1+\gamma^{4\ell}-2\gamma^{2\ell}\cos\theta}$$
$$- \max_{\varphi\in\mathbb{T}}|\sin(\varphi/2)| \sum_{\ell\in\mathbb{Z}} \frac{\gamma^{\ell+1/2}(1+\gamma^{2\ell+1})}{1+\gamma^{4\ell+2}-2\gamma^{2\ell+1}\cos\varphi}$$

Set, for $x\in\mathbb{R}$, $\theta\in\mathbb{T}\setminus\{0\}$,

$$f_{\gamma,\theta}(x) = |\sin(\theta/2)| \frac{\gamma^x(1+\gamma^{2x})}{1+\gamma^{4x}-2\gamma^{2x}\cos\theta}.$$

Then we are reduced to proving that

$$\inf_{\theta\in\mathbb{T}\setminus\{0\}} \sum_{k\in\mathbb{Z}} f_{\gamma,\theta}(k) - \sup_{\varphi\in\mathbb{T}\setminus\{0\}} \sum_{k\in\mathbb{Z}} f_{\gamma,\varphi}\big(k+\tfrac{1}{2}\big) > 0.$$

Applying the Poisson summation formula, we have

$$\sum_{k\in\mathbb{Z}} f_{\gamma,\theta}(k) = \sum_{n\in\mathbb{Z}} \hat{f}_{\gamma,\theta}(2\pi n), \qquad \sum_{k\in\mathbb{Z}} f_{\gamma,\varphi}\big(k+\tfrac{1}{2}\big) = \sum_{n\in\mathbb{Z}} (-1)^n \hat{f}_{\gamma,\varphi}(2\pi n),$$

where

$$\hat{f}_{\gamma,\theta}(\xi) = |\sin(\theta/2)| \int_{\mathbb{R}} \frac{\gamma^x(1+\gamma^{2x})}{1+\gamma^{4x}-2\gamma^{2x}\cos\theta} \mathrm{e}^{-ix\xi}\,dx$$
$$= \frac{|\sin(\theta/2)|}{|\log\gamma|} \int_0^\infty \frac{(1+t^2)t^{-i\xi/\log\gamma}}{1+t^4-2t^2\cos\theta}\,dt$$
$$= \frac{|\sin(\theta/2)|}{2|\log\gamma|} \int_0^\infty \frac{(1+y)y^{-i\xi/2\log\gamma-\frac{1}{2}}}{1+y^2-2y\cos\theta}\,dy,$$

where we have set $t=\gamma^x$, $y=t^2$. We calculate the above integral by introducing the holomorphic function

$$g(z) = \frac{|\sin(\theta/2)|}{2|\log\gamma|} \frac{(1+z)z^{-i\xi/2\log\gamma-\frac{1}{2}}}{1+z^2-2z\cos\theta},$$

on the domain $\mathbb{C}\setminus\mathbb{R}_+$, where the argument of z belongs to $(0,2\pi)$. Integrating on the contour of Figure 1 and making $R\to\infty$, $\varepsilon\to 0$, we obtain, by the residue

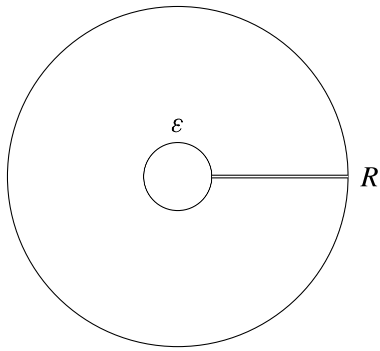

Figure 1. Contour for the proof of Lemma 17.

theorem, assuming $\theta \in (0, 2\pi)$ with no loss of generality,

$$\begin{aligned}\hat{f}_{\gamma,\theta}(\xi)(1+\mathrm{e}^{\pi\xi/\log\gamma}) &= 2i\pi\big[\mathrm{Res}(g(z), z=\mathrm{e}^{i\theta}) + \mathrm{Res}(g(z), z=\mathrm{e}^{-i\theta})\big] \\ &= \frac{i\pi\,\sin(\theta/2)}{|\log\gamma|}\left(\frac{2\cos(\theta/2)}{2i\,\sin\theta}\mathrm{e}^{\theta\xi/2\log\gamma} + \frac{2\cos(\theta/2)}{2i\,\sin\theta}\mathrm{e}^{(2\pi-\theta)\xi/2\log\gamma}\right) \\ &= \frac{\pi}{2|\log\gamma|}\big(\mathrm{e}^{\theta\xi/2\log\gamma} + \mathrm{e}^{(2\pi-\theta)\xi/2\log\gamma}\big).\end{aligned}$$

We infer

$$\hat{f}_{\gamma,\theta}(\xi) = \frac{\pi}{2|\log\gamma|}\frac{\cosh\big((\pi-\theta)\xi/(2\log\gamma)\big)}{\cosh\big(\pi\xi/(2\log\gamma)\big)}, \quad \theta \in (0, 2\pi).$$

Finally, for $\theta, \varphi \in (0, 2\pi)$,

$$\begin{aligned}\sum_{k\in\mathbb{Z}} f_{\gamma,\theta}(k) - \sum_{k\in\mathbb{Z}} f_{\gamma,\varphi}\big(k+\tfrac{1}{2}\big) &= \frac{\pi}{2|\log\gamma|}\Bigg(\sum_{n\in\mathbb{Z}}\frac{\cosh\big((\pi-\theta)\pi n/(\log\gamma)\big)}{\cosh\big(\pi^2 n/(\log\gamma)\big)} \\ &\qquad\qquad - \sum_{n\in\mathbb{Z}}(-1)^n\frac{\cosh\big((\pi-\varphi)\pi n/(\log\gamma)\big)}{\cosh\big(\pi^2 n/(\log\gamma)\big)}\Bigg) \\ &= \frac{\pi}{|\log\gamma|}\Bigg(\sum_{n=1}^{\infty}\frac{\cosh\big((\pi-\theta)\pi n/(\log\gamma)\big)}{\cosh\big(\pi^2 n/(\log\gamma)\big)} \\ &\qquad\qquad + \sum_{n=1}^{\infty}(-1)^{n+1}\frac{\cosh\big((\pi-\varphi)\pi n/(\log\gamma)\big)}{\cosh\big(\pi^2 n/(\log\gamma)\big)}\Bigg) \\ &\geq \frac{\pi}{|\log\gamma|}\sum_{n=1}^{\infty}\frac{1}{\cosh\big(\pi^2 n/(\log\gamma)\big)},\end{aligned}$$

since the second series is an alternating series of the form

$$\sum_{n=1}^{\infty}(-1)^{n+1}a_n,$$

with a_n decaying to 0 as $n \to \infty$. Therefore

$$\min_{|\zeta|=1}|F_\gamma(\zeta)| - \gamma^{\frac{1}{2}}\max_{|\zeta|=\gamma}|F_\gamma(\zeta)| \geq \frac{\pi}{|\log\gamma|}\sum_{n=1}^{\infty}\frac{1}{\cosh\big(\pi^2 n/(\log\gamma)\big)} > 0. \qquad \square$$

Lemma 17 implies that Φ has no zeroes for $|\zeta| = 1$ and $|z| \leq 1$. By continuity, it has no zeroes in a neighbourhood of this set. Hence Lemma 16 is proved.

4.4. *Studying the index.* Let us first recall the definition of the index. For $0 < R < \infty$, we denote by $\mathcal{C}_R$ the circle

$$\{z \in \mathbb{C} : |z| = R\}.$$

Let f be a holomorphic function near $\mathcal{C}_R$, with no zero on $\mathcal{C}_R$. The index on $\mathcal{C}_R$ around 0 of f is given by

$$\mathrm{Ind}_{f(\mathcal{C}_R)}(0) := \frac{1}{2i\pi}\int_{\mathcal{C}_R}\frac{f'(\zeta)}{f(\zeta)}\,d\zeta.$$

In this section, we prove the following lemma.

Lemma 18. *For any $r < 1$ sufficiently close to 1, the function*

$$\zeta \mapsto F_\gamma(r\zeta)$$

has index zero on the unit circle.

Notice that $\Phi(0, r\zeta) = F_\gamma(r\zeta)$. As the index is valued in $\mathbb{Z}$ and the map $z \mapsto \Phi(z, \zeta)$ is smooth, Lemma 18 implies that the index of $\zeta \mapsto \Phi(z, r\zeta)$ is zero for any z with $|z| \leq 1 + \rho$ as long as r is sufficiently close to 1.

Corollary 19. *For any $r < 1$ sufficiently close to 1, the function*

$$\zeta \mapsto \Phi(z, r\zeta)$$

has index zero for any z.

This corollary will complete the proof of Proposition 13.

Proof of Lemma 18. We could use Remark 15 in order to reduce to properties of the Weierstrass $\mathfrak{P}$ function. However, for the convenience of the reader, we prefer to give a self-contained proof. Let us assume that R is chosen so that $R \neq \gamma^{2\ell}$, $\ell \in \mathbb{Z}$ and $F_\gamma \neq 0$ on $\mathcal{C}_R$. We consider the index of F_γ on $\mathcal{C}_R$ around 0:

$$I(R) := \mathrm{Ind}_{F_\gamma(\mathcal{C}_R)}(0) := \frac{1}{2i\pi}\int_{\mathcal{C}_R} \frac{F'_\gamma(\zeta)}{F_\gamma(\zeta)}\, d\zeta.$$

The statement of Lemma 18 is equivalent to

$$I(1^-) := \lim_{R\to 1^-} I(R) = 0.$$

By definition, I is valued in $\mathbb{Z}$ and is continuous on the intervals corresponding to the circles avoiding the zeroes and the poles of F_γ. From properties (11), one has

$$I(R) + I\Big(\frac{1}{R}\Big) = -1, \quad I(R\gamma^2) = I(R). \tag{12}$$

In particular,

$$I(R) + I\Big(\frac{\gamma^2}{R}\Big) = -1 \tag{13}$$

and

$$I(1^+) = I((\gamma^2)^+), \tag{14}$$

where $I(r^\pm) = \lim_{t\to r^\pm} I(t)$. We are going to compute $I((\gamma^2)^+)$ in another way, using the zeroes and the poles of F_γ.

Let us first collect some basic relations. Let n be the number of zeroes in the annulus

$$\{z \in \mathbb{C} : \gamma < |z| < 1\}.$$

Since there are no poles inside this annulus, one has

$$n = I(1^-) - I(\gamma^+). \tag{15}$$

From Equation (13) with $R = 1^-$ and $R = \gamma^+$,

$$I(1^-) + I((\gamma^2)^+) = -1 \quad \text{and} \quad I(\gamma^+) + I(\gamma^-) = -1.$$

Subtracting these equalities gives $I(1^-) - I(\gamma^+) = I(\gamma^-) - I((\gamma^2)^+)$, hence

$$n = I(\gamma^-) - I((\gamma^2)^+). \tag{16}$$

Denote by m the number of zeroes on $\mathcal{C}_\gamma$. As γ is a zero of F_γ, $m \geq 1$, and

$$m = I(\gamma^+) - I(\gamma^-) \tag{17}$$

since there is no pole on $\mathcal{C}_\gamma$. Denote by N the number of zeroes on $\mathcal{C}_1$. Then

$$I(1^+) - I(1^-) = N - 1, \tag{18}$$

since 1 is the only pole on $\mathcal{C}_1$.

Now, we compute $I((\gamma^2)^+)$:

$$\begin{aligned} I((\gamma^2)^+) &= I(\gamma^-) - n && \text{from (16)} \\ &= I(\gamma^+) - m - n && \text{from (17)} \\ &= I(1^-) - n - m - n && \text{from (15)} \\ &= I(1^+) - (N-1) - m - 2n && \text{from (18).} \end{aligned}$$

Recalling (14), we conclude that $N + 2n + m = 1$, so $n = 0$ and $N + m = 1$. Since $m \geq 1$, this implies $N = 0$ and $m = 1$. From (18) and the equality $I(1^+) + I(1^-) = -1$ (Equation (13) with $R = 1^+$), one concludes $I(1^+) = -1$ and $I(1^-) = 0$ as required. □

Appendix: A formula for the C^1 norm

Let $u \in L^2_+(\mathbb{T})$ be a rational function corresponding to the finite list of singular values $\rho_1 > \sigma_1 > \cdots > \rho_N > \sigma_N$ and angles $\psi_r = 0$ for $r = 1, \ldots, 2N$. Then we checked in [Gérard and Grellier 2012; 2014] that this cancellation of the angles precisely corresponds to the positivity of the operators $\Gamma_{\hat{u}}$ and $\tilde{\Gamma}_{\hat{u}}$ on $\ell^2(\mathbb{Z}_+)$. The representation formula (2), (3) then reduces to

$$u(z) = \langle \mathscr{C}_N(z)^{-1}(\mathbb{1}_N), \mathbb{1}_N \rangle,$$

with

$$\mathscr{C}_N(z) := \left(\frac{s_{2j-1} - s_{2k} z}{s_{2j-1}^2 - s_{2k}^2} \right)_{1 \leq j,k \leq N}, \tag{19}$$

Furthermore, the positivity of the Hankel matrices $\Gamma_{\hat{u}}$ and $\tilde{\Gamma}_{\hat{u}}$ implies the positivity of the Fourier coefficients of u, since, denoting by $(e_n)_{n \geq 0}$ the canonical basis of $\ell^2(\mathbb{Z}_+)$,

$$\langle \Gamma_{\hat{u}} e_n, e_n \rangle = \hat{u}(2n), \quad \langle \tilde{\Gamma}_{\hat{u}} e_n, e_n \rangle = \hat{u}(2n+1).$$

Therefore the C^1 norm of u on $\mathbb{T}$ is given by

$$S(u) := \sum_{n=1}^{\infty} n \hat{u}(n).$$

The lemma below explicitly computes $S(u)$.

Lemma 20. $$S(u) = \sum_{k=1}^{N} \sigma_k \left(\prod_{j=1}^{N} \frac{\rho_j + \sigma_k}{\rho_j - \sigma_k} \right) \left(\prod_{\ell \neq k} \frac{\sigma_k + \sigma_\ell}{\sigma_\ell - \sigma_k} \right),$$

where every term in the above sum is positive.

Proof. We have

$$S(u) = u'(1) = \left\langle \dot{\mathscr{C}} \mathscr{C}(1)^{-1}(\mathbb{1}), {}^t\mathscr{C}(1)^{-1}(\mathbb{1}) \right\rangle,$$

with

$$\mathscr{C}(1) := \left(\frac{1}{\rho_j + \sigma_k}\right)_{1\le j,k\le N}, \quad \dot{\mathscr{C}} := \left(\frac{\sigma_k}{\rho_j^2 - \sigma_k^2}\right)_{1\le j,k\le N}.$$

Notice that $\mathscr{C}(1)$ is a Cauchy matrix, so that the expression of $\mathscr{C}(1)^{-1}(\mathbb{1})$ is explicit. We have

$$\mathscr{C}(1)^{-1}(\mathbb{1}) = \left(\frac{\prod_{j=1}^N(\rho_j + \sigma_k)}{\prod_{\ell\ne k}(\sigma_k - \sigma_\ell)}\right)_{1\le k\le N}. \tag{20}$$

Let us give a simple proof of this formula, inspired from calculations in [Gérard and Pushnitski 2018]. Denote by x_k, $k = 1, \dots, N$, the components of $\mathscr{C}(1)^{-1}(\mathbb{1})$. We have

$$\sum_{k=1}^N \frac{x_k}{\rho_j + \sigma_k} = 1, \quad j = 1, \dots, N.$$

Consider the polynomial functions

$$Q(\rho) := \prod_{k=1}^N(\rho + \sigma_k), \quad P(\rho) := Q(\rho)\sum_{k=1}^N \frac{x_k}{\rho + \sigma_k}.$$

Then Q has degree N, P has degree at most $N-1$ and

$$P(\rho_j) = Q(\rho_j), \quad j = 1, \dots N.$$

Since $Q-P$ is a unitary polynomial of degree N which cancels at ρ_j, $j = 1, \dots, N$, we have

$$Q(\rho) - P(\rho) = \prod_{j=1}^N(\rho - \rho_j).$$

Consequently,

$$P(-\sigma_k) = -\prod_{j=1}^N(-\sigma_k - \rho_j) = (-1)^{N-1}\prod_{j=1}^N(\sigma_k + \rho_j).$$

Since

$$x_k = \frac{P(-\sigma_k)}{Q'(-\sigma_k)},$$

this yields (20). Similarly, we have

$$^t\mathscr{C}(1)^{-1}(\mathbb{1}) = \left(\frac{\prod_{\ell=1}^N(\rho_j + \sigma_\ell)}{\prod_{i\ne j}(\rho_j - \rho_i)}\right)_{1\le j\le N}. \tag{21}$$

Coming back to the proof of Lemma 20, we have, in view of (20) and (21),

$$S(u) = \sum_{j,k=1}^{N} \mu_{jk}^{(N)}, \quad \mu_{jk}^{(N)} := \sigma_k \frac{\rho_j + \sigma_k}{\rho_j - \sigma_k} \left(\prod_{i \neq j} \frac{\rho_i + \sigma_k}{\rho_j - \rho_i} \right) \left(\prod_{\ell \neq k} \frac{\rho_j + \sigma_\ell}{\sigma_k - \sigma_\ell} \right).$$

Multiplying and dividing $\mu_{jk}^{(N)}$ by $\prod_{i \neq j} (\rho_i - \sigma_k)$, we have, for every k,

$$\sum_{j=1}^{N} \mu_{jk}^{(N)} = \frac{\sigma_k R(\sigma_k)}{\prod_{\ell \neq k} (\sigma_k - \sigma_\ell)} \prod_{i=1}^{N} \frac{\rho_i + \sigma_k}{\rho_i - \sigma_k},$$

with

$$R(\sigma) = \sum_{j=1}^{N} \prod_{i \neq j} \frac{\rho_i - \sigma}{\rho_j - \rho_i} \prod_{\ell \neq k} (\rho_j + \sigma_\ell).$$

Notice that, for every $j = 1, \ldots, N$,

$$R(\rho_j) = (-1)^{N-1} \prod_{\ell \neq k} (\rho_j + \sigma_\ell).$$

Since R has degree $N - 1$, we infer

$$R(\sigma) = (-1)^{N-1} \prod_{\ell \neq k} (\sigma + \sigma_\ell),$$

so that

$$\sum_{j=1}^{N} \mu_{jk}^{(N)} = \frac{\sigma_k (-1)^{N-1} \prod_{\ell \neq k} (\sigma_k + \sigma_\ell)}{\prod_{\ell \neq k} (\sigma_k - \sigma_\ell)} \prod_{i=1}^{N} \frac{\rho_i + \sigma_k}{\rho_i - \sigma_k},$$

which is the claimed formula. The positivity of each term is an easy consequence of the inequalities $\rho_1 > \sigma_1 > \rho_2 > \sigma_2 > \cdots$. □

As a consequence of Lemma 20, we retain the following inequality, obtained after discarding most of the factors bigger than 1 in each of the products.

Corollary 21. $$\|u\|_{C^1} \geq \sum_{k=1}^{N} \frac{\sigma_k (\rho_k + \sigma_k)}{\rho_k - \sigma_k}.$$

Notice that this implies inequality (4). Unfortunately, at this stage we do not have arguments allowing us to extend this inequality to nonrational functions u, which would imply that $u \notin C^1$ if

$$\sum_{k=1}^{\infty} \frac{\rho_k \sigma_k}{\rho_k - \sigma_k} = \infty.$$

References

[Baxter 1963] G. Baxter, "A norm inequality for a "finite-section" Wiener–Hopf equation", *Illinois J. Math.* **7** (1963), 97–103. MR Zbl

[Böttcher and Grudsky 2000] A. Böttcher and S. M. Grudsky, *Toeplitz matrices, asymptotic linear algebra, and functional analysis*, Birkhäuser, Basel, 2000. MR Zbl

[Gérard and Grellier 2010] P. Gérard and S. Grellier, "The cubic Szegő equation", *Ann. Sci. Éc. Norm. Supér.* (4) **43**:5 (2010), 761–810. MR Zbl

[Gérard and Grellier 2012] P. Gérard and S. Grellier, "Invariant tori for the cubic Szegő equation", *Invent. Math.* **187**:3 (2012), 707–754. MR Zbl

[Gérard and Grellier 2014] P. Gérard and S. Grellier, "Inverse spectral problems for compact Hankel operators", *J. Inst. Math. Jussieu* **13**:2 (2014), 273–301. MR Zbl

[Gérard and Grellier 2015] P. Gérard and S. Grellier, "An explicit formula for the cubic Szegő equation", *Trans. Amer. Math. Soc.* **367**:4 (2015), 2979–2995. MR Zbl

[Gérard and Grellier 2017] P. Gérard and S. Grellier, *The cubic Szegő equation and Hankel operators*, Astérisque **389**, Société Mathématique de France, Paris, 2017. MR Zbl

[Gérard and Pushnitski 2015] P. Gérard and A. Pushnitski, "An inverse problem for self-adjoint positive Hankel operators", *Int. Math. Res. Not.* **2015**:13 (2015), 4505–4535. MR Zbl

[Gérard and Pushnitski 2018] P. Gérard and A. Pushnitski, "Inverse spectral theory for a class of non-compact Hankel operators", preprint, 2018. To appear in *Mathematika*. arXiv

[Gérard et al. 2015] P. Gérard, Y. Guo, and E. S. Titi, "On the radius of analyticity of solutions to the cubic Szegő equation", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **32**:1 (2015), 97–108. MR Zbl

[Peller 2003] V. V. Peller, *Hankel operators and their applications*, Springer, 2003. MR Zbl

[Saks and Zygmund 1952] S. Saks and A. Zygmund, *Analytic functions*, Monografie Matematyczne **XXVIII**, Polskie Towarzystwo Matematyczne, Warsaw, 1952. MR Zbl

Received 5 Dec 2017.

PATRICK GÉRARD:

Patrick.Gerard@math.u-psud.fr
Laboratoire de Mathematiques D'Orsay, Université Paris Sud XI, CNRS, Orsay, France

SANDRINE GRELLIER:

Sandrine.Grellier@univ-orleans.fr
Universite d'Orleans, Institut Denis Poisson, Bâtiment de Mathematiques, Orleans, France

TUNISIAN JOURNAL OF MATHEMATICS
Vol. 1, No. 3, 2019
https://doi.org/10.2140/tunis.2019.1.373

Ramification groups of coverings and valuations

Takeshi Saito

Dedicated to the memory of Michel Raynaud

We give a purely scheme theoretic construction of the filtration by ramification groups of the Galois group of a covering. The valuation need not be discrete but the normalizations are required to be locally of complete intersection.

A correction was submitted on 12 Jan 2022 and posted online on 17 Jan 2022.

For a Galois extension of a complete discrete valuation field with not necessarily perfect residue field, the filtration by ramification groups on the Galois group is defined in a joint article [Abbes and Saito 2002] with Ahmed Abbes. Although the definition there is based on rigid geometry, it was later observed that the use of rigid geometry can be avoided and the conventional language of schemes suffices [Saito 2009]. In this article, we reformulate the construction in [Abbes and Saito 2002] in the language of schemes. As a byproduct, we give a generalization for ramified finite Galois coverings of normal and universally Japanese noetherian schemes and valuations not necessarily discrete.

All the ideas are present in the 2002 article, possibly in different formulation. As in that article, the main ingredients in the definition of ramification groups are the following: First, we interpret a subgroup as a quotient of the fiber functor with a cocartesian property, Proposition 1.4.2. Thus, the definition of ramification groups is a consequence of a construction of quotients of the fiber functor, indexed by elements of the rational value group of valuation.

The required quotients of the fiber functor are constructed as the sets of connected components of geometric fibers of dilatations [Abbes and Saito 2011; Saito 2009] defined by an immersion of the covering to a smooth scheme over the base scheme. Here a crucial ingredient is the reduced fiber theorem of Bosch,

MSC2010: primary 14B25; secondary 11S15.
Keywords: ramification groups, valuation rings.

Lütkebohmert and Raynaud [Bosch et al. 1995] recalled in Theorem 1.2.5. This specializes to the finiteness theorem of Grauert and Remmert in the classical case where the base is a discrete valuation ring. A variant of the filtration is defined using the underlying sets of geometric fibers of quasifinite schemes without using the sets of connected components.

To prove the basic properties of ramification groups stated in Theorem 3.3.1 including the rationality of breaks, semicontinuity etc., a key ingredient is a generalization due to Temkin [2011] of the semistable reduction theorem of curves recalled in Theorem 1.3.5.

Let X be a normal noetherian scheme and let $U \subset X$ be a dense open subscheme. The Zariski–Riemann space $\widetilde{X}$ is defined as the inverse limit of proper schemes X' over X such that $U' = U \times_X X' \to U$ is an isomorphism. Points of $\widetilde{X}$ on the boundary $\widetilde{X} - U$ correspond bijectively to the inverse limits of the images of the closed points by the liftings of the morphisms $T = \operatorname{Spec} A \to X$ for valuation rings $A \subsetneqq K = k(t)$ for points $t \in U$ such that $T \times_X U$ consists of the single point t.

Let $W \to U$ be a finite étale connected Galois covering of the Galois group G. We will construct in Theorem 3.3.1 filtrations (G_T^{γ}) and $(G_T^{\gamma+})$ on G by ramification groups for a morphism $T \to X$ as above indexed by the positive part

$$(0, \infty)_{\Gamma_{\mathbb{Q}}} \subset \Gamma_{\mathbb{Q}} = \Gamma \otimes \mathbb{Q}$$

for the value group $\Gamma = K^{\times}/A^{\times}$. To complete the definition, we need to assume that for every intermediate covering $V \to U$, the normalization Y of X in V is locally of complete intersection over X to assure the cocartesian property in Proposition 1.4.2. The required cocartesian property Proposition 3.1.2 is then a consequence of a lifting property in commutative algebra recalled in Proposition 1.1.5.

The definition depends on X, not only on $W \to U$. In other words, for a normal noetherian scheme X' over X as above, the filtrations (G_T^{γ}) and $(G_T^{\gamma+})$ defined for X and those for X' may be different. This arises from the fact that the formation of the normalization Y need not commute with base change $X' \to X$. To obtain a definition depending only on $W \to U$, one would need to take the inverse limit with respect to X'. This requires that the normalizations over T to be locally of complete intersection.

By Proposition 1.4.2, the definition of the filtrations (G_T^{γ}) and $(G_T^{\gamma+})$ are reduced to the construction of surjections $F_T^{\infty} \to F_T^{\gamma}$ and $F_T^{\infty} \to F_T^{\gamma+}$ for a fiber functor F_T^{∞}. To define them, for each intermediate covering $V \to U$, we take an embedding $Y \to Q$ of the normalization to a smooth scheme over X. Further taking a ramified covering and a blow-up X', we find an effective Cartier divisor $R' \subset X'$ and a lifting $T' \to X'$ of $T \to X$ such that the valuation $v'(R')$ of R' is γ for each $\gamma \in \Gamma_{\mathbb{Q}}$. Then, we define a dilatation $Q'^{(R')}$ over X' to be the normalization of an open subscheme $Q'^{[R']}$ of the blow-up of the base change $Q' = Q \times_X X'$ at the closed subscheme $Y \times_X R' \subset Q \times_X X'$. To obtain a construction independent of the choice of X', we apply the reduced fiber theorem of Bosch–Lütkebohmert–Raynaud for $Q'^{(R')} \to X'$ to be flat and to have reduced geometric fibers.

Now the desired functor $F_T^{\gamma}(Y/X)$ is defined as the set of connected components of the geometric fiber of $Q'^{(R')} \to X'$ at the image of the closed point by $T' \to X'$. We recover the construction in [Abbes and Saito 2002] in the classical case where $X = T$ is the spectrum of a complete discrete valuation ring as we show in Lemma 3.3.2 using Example 2.1.1(1) and Remark 1.1.2. Its variant $F_T^{\gamma+}(Y/X)$ is defined more simply as the geometric fiber of the inverse image $Y' \times_{Q'^{[R']}} Q'^{(R')}$ with respect to the morphism $Y' = Y \times_X X' \to Q'^{[R']}$ lifting the original immersion $Y \to Q$. The fact that the construction is independent of the choice of immersion $Y \to Q$ is based on a homotopy invariance of dilatations proved in Proposition 2.1.5.

To study the behavior of the functors F_T^{γ} and $F_T^{\gamma+}$ thus defined for the variable γ, we use a semistable curve C over X defined by $st = f$ for a nonzero divisor f on X defining an effective Cartier divisor $D \subset X$ such that $D \cap U = \varnothing$ as a parameter space for γ. Let $\widetilde{D} \subset C$ denote the effective Cartier divisor defined by t. Then, for $\gamma \in [0, v(D)]_{\Gamma_{\mathbb{Q}}}$, there is a lifting $T' \to C$ of $T \to X$ such that $v'(\widetilde{D}) = \gamma$. Using this together with a local description (Proposition 1.3.3) of Cartier divisors on a semistable curve over a normal noetherian scheme and a combination of the reduced fiber theorem and the semistable reduction theorem over a general base scheme, we derive basic properties of F_T^{γ} and $F_T^{\gamma+}$ in Proposition 3.1.8 to prove Theorems 3.2.6 and 3.3.1.

Convention. In this article, we assume that for a noetherian scheme X, the normalization of the reduced part of a scheme of finite type over X remains to be of finite type over X. This property is satisfied if X is of finite type over a field, $\mathbb{Z}$, or a complete discrete valuation ring, for example.

1. Preliminaries

1.1. *Connected components.*

Definition 1.1.1 [EGA IV$_2$ 1965, définition (6.8.1)]. Let $f : X \to S$ be a flat morphism locally of finite presentation of schemes. We say that f is *reduced* if for every geometric point s of S, the geometric fiber X_s is *reduced*.

In [SGA 1 1971, exposé X, définition 1.1], reduced morphism is called separable morphism. A morphism f of finite presentation is étale if and only if f is quasifinite, flat and reduced.

We study the sets of connected components of geometric fibers of a flat and reduced morphism of finite type. Let S be a scheme and let s and t be geometric points of S. Let $S_{(s)}$ denote the strict localization. A specialization $s \leftarrow t$ of geometric points means a morphism $S_{(s)} \leftarrow t$ over S.

Assume that S is noetherian. Let $X \to S$ be a flat and reduced morphism of finite type and let $s \leftarrow t$ be a specialization of geometric points of S. We define the cospecialization mapping

$$\pi_0(X_s) \to \pi_0(X_t) \tag{1-1}$$

as follows. By replacing S by the closure of the image of t, we may assume that S is integral and that t is above the generic point η of S. By replacing S further by a quasifinite scheme over S such that the function field is a finite extension of $\kappa(\eta)$ in $\kappa(t)$, we may assume that the canonical mapping $\pi_0(X_t) \to \pi_0(X_\eta)$ is a bijection. Let $U \subset S$ be a dense open subset such that the canonical mapping $\pi_0(X_\eta) \to \pi_0(X_U)$ is a bijection. Then, by [EGA IV$_4$ 1967, corollaire (18.9.11)], the canonical mapping $\pi_0(X_U) \to \pi_0(X)$ is also a bijection. Thus, we define the cospecialization mapping (1-1) to be the composition

$$\pi_0(X_s) \to \pi_0(X) \xleftarrow{\simeq} \pi_0(X_\eta) \xleftarrow{\simeq} \pi_0(X_t).$$

We say that the sets of connected components of geometric fibers of $X \to S$ are locally constant if for every specialization $s \leftarrow t$ of geometric points of S, the cospecialization mapping $\pi_0(X_s) \to \pi_0(X_t)$ is a bijection. By [EGA IV$_3$ 1966, théorème (9.7.7)] and by noetherian induction, there exists a finite stratification $S = \coprod_i S_i$ by locally closed subschemes such that the sets of connected components of geometric fibers of the base change $X \times_S S_i \to S_i$ are locally constant for every i. We call this fact that the sets of connected components of geometric fibers of $X \to S$ are constructible.

Remark 1.1.2. Let $S = \operatorname{Spec} \mathcal{O}_K$ for a discrete valuation ring $\mathcal{O}_K$ and let $X = \operatorname{Spec} A$ be an affine scheme of finite type over S. Let $\bar{s} \to S$ be a geometric closed point. Let $\mathfrak{X} = \operatorname{Spf} \hat{A}$ be the formal completion along the closed fiber and let $\mathfrak{X}_{\overline{K}} = \operatorname{Sp} \hat{A} \otimes_{\mathcal{O}_K} \overline{K}$ be the associated affinoid variety over an algebraic closure $\overline{K}$ of the fraction field K of $\mathcal{O}_K$. If X is flat and reduced over S, the cospecialization mapping $\pi_0(X_{\bar{s}}) \to \pi_0(\mathfrak{X}_{\overline{K}})$ is a bijection.

Let $Y \to S$ be another flat and reduced morphism of finite type and let $f : X \to Y$ be a morphism over S. The cospecialization mappings (1-1) form a commutative diagram

$$\begin{array}{ccc} \pi_0(X_s) & \longrightarrow & \pi_0(X_t) \\ \downarrow & & \downarrow \\ \pi_0(Y_s) & \longrightarrow & \pi_0(Y_t) \end{array} \tag{1-2}$$

Lemma 1.1.3. *Let $f : X \to Y$ be a morphism of schemes of finite type over a noetherian scheme S. Assume that X is étale over S and that Y is flat and reduced over S. Let A denote the subset of X consisting of the images of geometric points x of X satisfying the following condition*:

> *Let s be the geometric point of S defined as the image of x and let $C \subset Y_s$ be the connected component of the fiber containing the image of x. Then, $f_s^{-1}(C) \subset X_s$ consists of a single point x.*

Then A is closed.

Proof. By the constructibility of connected components of geometric fibers of Y, the subset $A \subset X$ is constructible. For a specialization $s \leftarrow t$ of geometric points of S, the upper horizontal arrow in the commutative diagram

$$\begin{array}{ccc} X_s & \longrightarrow & X_t \\ \downarrow & & \downarrow \\ \pi_0(Y_s) & \longrightarrow & \pi_0(Y_t) \end{array}$$

is an injection since $X \to S$ is étale. Hence A is closed under specialization and is closed. □

We have specialization mappings going the other way for proper morphisms. Let X be a proper scheme over S. Let $s \leftarrow t$ be a specialization of geometric points of S. Then, the inclusion $X_s \to X \times_S S_{(s)}$ induces a bijection $\pi_0(X_s) \to \pi_0(X \times_S S_{(s)})$ by [SGA 4½ 1977, IV proposition (2.1)]. Its composition with the mapping $\pi_0(X_t) \to \pi_0(X \times_S S_{(s)})$ induced by the morphism $X_t \to X \times_S S_{(s)}$ defines the specialization mapping

$$\pi_0(X_s) \leftarrow \pi_0(X_t). \tag{1-3}$$

For a morphism $X \to Y$ of proper schemes over S, the specialization mappings make a commutative diagram

$$\begin{array}{ccc} \pi_0(X_s) & \longleftarrow & \pi_0(X_t) \\ \downarrow & & \downarrow \\ \pi_0(Y_s) & \longleftarrow & \pi_0(Y_t). \end{array} \tag{1-4}$$

Lemma 1.1.4. *Let $f : X \to Y$ be a finite unramified morphism of schemes. Let B denote the subset of X consisting of the images of geometric points x of X satisfying the following condition*:

> *For the geometric point y of Y defined as the image of x, the fiber $X \times_Y y$ consists of a single point x.*

Then, B is open.

Proof. The complement $X - B$ equals the image of the complement $X \times_Y X - X$ of the diagonal by a projection. Since $X \to Y$ is unramified, the complement $X \times_Y X - X \subset X \times_Y X$ is closed. Since the projection $X \times_Y X \to X$ is finite, the image $X - B$ is closed. □

Proposition 1.1.5. *Let*

$$\begin{array}{ccc} Z' & \longrightarrow & X' \\ \downarrow & \square & \downarrow f \\ Z & \longrightarrow & X \end{array} \tag{1-5}$$

be a cartesian diagram of noetherian schemes. Assume that X is normal, the horizontal arrows are closed immersion, the right vertical arrow is quasifinite and the left vertical arrow is finite. Assume further that there exists a dense open subscheme $U \subset X$ such that $U' = U \times_X X' \to U$ is faithfully flat and that $U' \subset X'$ is also dense.

(1) *Let $C \subset Z$ be an irreducible closed subset and let $C' \subset f^{-1}(C)$ be an irreducible component. Then, $C' \to C$ is surjective.*

(2) *Let $C \subset Z$ be a connected closed subset and let $C' \subset f^{-1}(C)$ be a connected component. Then, $C' \to C$ is surjective.*

Proof. (1) By replacing U by a dense open subscheme if necessary, we may assume that $U' \to U$ is finite. By Zariski's main theorem, there exists a scheme $\bar{X}'$ finite over X containing X' as an open subscheme. By replacing $\bar{X}'$ by the closure of U', we may assume that U' is dense in $\bar{X}'$. Since U' is closed in $\bar{X}' \times_X U$, we have $\bar{X}' \times_X U = U'$. Since $Z' = (\bar{X}' \times_X Z) \cap X'$ is closed and open in $\bar{X}' \times_X Z$, by replacing X' by $\bar{X}'$, we may assume that f is finite.

Since f is a closed mapping, it suffices to show that the generic point z of C is the image of the generic point z' of C'. Let x' be a point of C'. Replacing X by an affine neighborhood of $x = f(x') \in C$, we may assume $X = \operatorname{Spec} A$ and $X' = \operatorname{Spec} B$ are affine. Then, the assumption implies that $A \to B$ is an injection and B is finite over A. Since x is a point of the closure $C = \overline{\{z\}}$, the assertion follows from [Bourbaki 1985, Chapter V, Section 2.4, Theorem 3].

(2) Let $C_1 \subset C$ be an irreducible component such that $C_1 \cap f(C')$ is not empty. Then, there exists an irreducible component C'_1 of $f^{-1}(C_1) \subset f^{-1}(C)$ such that $C'_1 \cap C'$ is not empty. By (1), we have $C_1 = f(C'_1)$. Since C' is a connected component of $f^{-1}(C)$ and $C'_1 \cap C' \neq \varnothing$, we have $C'_1 \subset C'$ and hence $C_1 = f(C'_1) \subset f(C')$. Thus, the complement $C - f(C')$ is the union of irreducible components of C not meeting $f(C')$ and is closed. Since $f(C') \subset C$ is also closed and is nonempty, we have $C = f(C')$. □

Corollary 1.1.6. *Let*

$$\begin{array}{ccccccc} Z' & \longrightarrow & X' & \longleftarrow & Y'_1 & \xleftarrow{g'} & Y' \\ \downarrow & \square & \downarrow f & \square & \downarrow f_1 & & \downarrow f' \\ Z & \longrightarrow & X & \longleftarrow & Y_1 & \xleftarrow{g} & Y \end{array} \tag{1-6}$$

be a commutative diagram of noetherian schemes such that the left square is cartesian and satisfies the conditions in Proposition 1.1.5. Assume that $Y_1 \subset X$ is a closed subscheme, that the middle square is cartesian and that the four arrows in the right square are finite. Assume that there exists a dense open subscheme $V_1 \subset Y_1$ such that $V = V_1 \times_{Y_1} Y \subset Y$ is also dense and that $g|_V : V \to V_1$ and $g'|_{V'} : V' = V \times_Y Y' \to V'_1 = V_1 \times_{Y_1} Y'_1$ are isomorphisms.

(1) *For any irreducible (resp. connected) component C of Y, we have $f^{-1}(g(C)) = g'(f'^{-1}(C))$. Consequently, we have $f^{-1}(g(Y)) = g'(Y')$.*

(2) *Suppose that the mapping $Z \times_X Y \to \pi_0(Y)$ is a bijection. Then, the diagram*

$$\begin{array}{ccc} Z' \cap Y'_1 & \longleftarrow & Z' \times_{X'} Y' \\ \downarrow & & \downarrow \\ Z \cap Y_1 & \longleftarrow & Z \times_X Y \end{array} \tag{1-7}$$

of underlying sets induces a surjection $Z' \times_{X'} Y' \to (Z' \cap Y'_1) \times_{Z \cap Y_1} (Z \times_X Y)$ of sets. If $Z \times_X Y \to Z \cap Y_1$ is surjective, then $Z' \times_{X'} Y' \to Z' \cap Y'_1$ is also surjective. Further, if $Y' \to Y$ is surjective, then $Z' \cap Y'_1 \to Z \cap Y_1$ is also surjective and the diagram (1-7) is a cocartesian diagram of underlying sets.

(3) *The diagram*

$$\begin{array}{ccc} \pi_0(Z') & \longleftarrow & Z' \cap Y'_1 \\ \downarrow & & \downarrow \\ \pi_0(Z) & \longleftarrow & Z \cap Y_1 \end{array} \tag{1-8}$$

of sets induces a surjection $Z' \cap Y'_1 \to \pi_0(Z') \times_{\pi_0(Z)} (Z \cap Y_1)$ of sets. If $Z \cap Y_1 \to \pi_0(Z)$ is surjective, then $Z' \cap Y'_1 \to \pi_0(Z')$ is also surjective. Further if $Z' \cap Y'_1 \to Z \cap Y_1$ is surjective, the diagram (1-8) is a cocartesian diagram of sets.

Proof. (1) Let $C \subset Y$ be an irreducible component. The inclusion $f^{-1}(g(C)) \supset g'(f'^{-1}(C))$ is clear. We show the other inclusion. Since V is dense in Y, the intersection $C \cap V$ and hence its image $g(C) \cap V_1$ are not empty. Let C' be an irreducible component of $f^{-1}(g(C)) \subset Y'_1$. Since $Y'_1 \to Y_1$ is finite and $g(C) \subset Y_1$ is an irreducible closed subset, we have $g(C) = f(C')$ by Proposition 1.1.5(1). Since $f(C' \cap V'_1) = f(C') \cap V_1 = g(C) \cap V_1$ is not empty, $C' \cap V'_1 = g'(g'^{-1}(C' \cap V'_1))$ is also nonempty and hence is dense in C'. Since $g'^{-1}(C' \cap V_1') = g'^{-1}(C') \cap V' \subset f'^{-1}(C)$ and since $g' : Y' \to X'$ is proper, we have $C' \subset g'(f'^{-1}(C))$.

Since a connected component of Y and Y itself are unions of irreducible components of Y, the remaining assertions follow from the assertion for irreducible components.

(2) Let $z' \in Z' \cap Y_1'$ and $y \in Z \times_X Y$ be points satisfying $f(z') = g(y)$ in $Z \cap Y_1$. Let $C \subset Y$ be the unique connected component containing y. Since $z' \in f^{-1}(g(C)) = g'(f'^{-1}(C))$ by (1), there exists a point $y' \in Z' \times_{X'} f'^{-1}(C) \subset Z' \times_{X'} Y'$ such that $z' = g'(y')$. Since $f'(y') \in Z \times_X Y$ is a unique point contained in $C \in \pi_0(Y)$, we have $y = f'(y')$. Thus, $(z', y) \in (Z' \cap Y_1') \times_{Z \cap Y_1} (Z \times_X Y)$ is the image of $y' \in Z' \times_{X'} Y'$.

If $Z \times_X Y \to Z \cap Y_1$ is surjective, then $(Z' \cap Y_1') \times_{Z \cap Y_1} (Z \times_X Y) \to Z' \cap Y_1'$ is surjective and hence the first assertion implies the surjectivity of $Z' \times_{X'} Y' \to Z' \cap Y_1'$.

If both $Z \times_X Y \to Z \cap Y_1$ and $Y' \to Y$ are surjective, then $Z' \times_{X'} Y' = Z \times_X Y' \to Z \times_X Y$ is also surjective and hence by the commutative diagram (1-7), the mapping $Z' \cap Y_1' \to Z \cap Y_1$ is a surjection. This implies that the diagram (1-7) with $Z' \times_{X'} Y'$ replaced by $(Z' \cap Y_1') \times_{Z \cap Y_1} (Z \times_X Y)$ is a cocartesian diagram of underlying sets. Hence the surjectivity of $Z' \times_{X'} Y' \to Z' \cap Y_1' \times_{Z \cap Y_1} (Z \times_X Y)$ implies that the diagram (1-7) is a cocartesian diagram of underlying sets.

(3) Let $C' \subset Z'$ be a connected component and let $z \in Z \cap Y_1$ be a point such that the connected component $C \subset Z$ satisfying $f(C') \subset C$ contains z. Since $f(C') = C$ by Proposition 1.1.5(2), the intersection $C' \cap f^{-1}(z) \subset Z' \cap Y_1'$ is not empty. Hence $(C', z) \in \pi_0(Z') \times_{\pi_0(Z)} (Z \cap Y_1)$ is in the image of $C' \cap f^{-1}(z) \subset Z' \cap Y_1'$.

The remaining assertions are proved similarly as in (2). □

1.2. ***Flat and reduced morphisms.*** Let $k \geqq 0$ be an integer. Recall that a noetherian scheme X satisfies the condition (R_k) if for every point $x \in X$ of $\dim \mathcal{O}_{X,x} \leqq k$, the local ring $\mathcal{O}_{X,x}$ is regular [EGA IV$_2$ 1965, définition (5.8.2)]. Recall also that a noetherian scheme X satisfies the condition (S_k) if for every point $x \in X$, we have $\operatorname{prof} \mathcal{O}_{X,x} \geqq \inf(k, \dim \mathcal{O}_{X,x})$ [EGA IV$_2$ 1965, définition (5.7.2)].

Proposition 1.2.1. *Let $f : X \to S$ be a flat morphism of finite type of noetherian schemes and let $k \geqq 0$ be an integer. We define a function $k : S \to \mathbb{N}$ by $k(s) = \max(k - \dim \mathcal{O}_{S,s}, 0)$.*

(1) *If S satisfies the condition (R_k) and if the fiber $X_s = X \times_S s$ satisfies $(\mathrm{R}_{k(s)})$ for every $s \in S$, then X satisfies the condition (R_k).*

(2) *If X satisfies the condition (R_k) and if $f : X \to S$ is faithfully flat, then S satisfies the condition (R_k).*

Proof. (1) Assume $\dim \mathcal{O}_{X,x} \leqq k$ and set $s = f(x)$. Then, we have $\dim \mathcal{O}_{S,s} \leqq \dim \mathcal{O}_{X,x} \leqq k$ and $\dim \mathcal{O}_{X_s,x} = \dim \mathcal{O}_{X,x} - \dim \mathcal{O}_{S,s} \leqq k(s)$ by [EGA IV$_2$ 1965, proposition (6.1.1)]. Hence $\mathcal{O}_{S,s}$ and $\mathcal{O}_{X_s,x}$ are regular by the assumption. Thus $\mathcal{O}_{X,x}$ is regular by [EGA IV$_1$ 1964, chapitre $0_{\rm IV}$ proposition (17.3.3)(ii)].

(2) This follows from [EGA IV$_2$ 1965, proposition (6.5.3)(i)]. □

Proposition 1.2.2. *Let $f : X \to S$ be a flat morphism of finite type of noetherian schemes and let $k \geqq 0$ be an integer. Let the function $k : S \to \mathbb{N}$ be as in Proposition 1.2.1.*

(1) *If S satisfies the condition* (S_k) *and if the fiber X_s satisfies* $(\mathrm{S}_{k(s)})$ *for every $s \in S$, then X satisfies the condition* (S_k).

(2) *If X satisfies the condition* (S_k) *and if $f : X \to S$ is faithfully flat, then S satisfies the condition* (S_k).

(3) *If X satisfies the condition* (S_k) *and if S is of Cohen–Macaulay, then the fiber X_s satisfies* $(\mathrm{S}_{k(s)})$ *for every $s \in S$.*

Proof. (1) Let $x \in X$ and $s = f(x)$. Then, we have $\operatorname{prof} \mathcal{O}_{S,s} \geqq \inf(k, \dim \mathcal{O}_{S,s})$ and $\operatorname{prof} \mathcal{O}_{X_s,x} \geqq \inf(k(s), \dim \mathcal{O}_{X_s,x})$ by the assumption. By $\dim \mathcal{O}_{X_s,x} = \dim \mathcal{O}_{X,x} - \dim \mathcal{O}_{S,s}$ [EGA IV$_2$ 1965, proposition (6.1.1)], we have

$$\inf(k, \dim \mathcal{O}_{S,s}) + \inf(k(s), \dim \mathcal{O}_{X_s,x}) = \inf(k, \dim \mathcal{O}_{X,x}).$$

Hence the claim follows from $\operatorname{prof} \mathcal{O}_{X,x} = \operatorname{prof} \mathcal{O}_{S,s} + \operatorname{prof} \mathcal{O}_{X_s,x}$ [EGA IV$_2$ 1965, proposition (6.3.1)].

(2) This follows from [EGA IV$_2$ 1965, proposition (6.4.1)(i)].

(3) Let $x \in X$ and $s = f(x)$. Then by the assumption, we have

$$\operatorname{prof} \mathcal{O}_{X,x} \geqq \inf(k, \dim \mathcal{O}_{X,x}) \quad \text{and} \quad \operatorname{prof} \mathcal{O}_{S,s} = \dim \mathcal{O}_{S,s}.$$

By $\operatorname{prof} \mathcal{O}_{X_s,x} = \operatorname{prof} \mathcal{O}_{X,x} - \operatorname{prof} \mathcal{O}_{S,s} \geqq 0$ [EGA IV$_2$ 1965, proposition (6.3.1)] and $\dim \mathcal{O}_{X_s,x} = \dim \mathcal{O}_{X,x} - \dim \mathcal{O}_{S,s}$ [EGA IV$_2$ 1965, proposition (6.1.1)] we have $\operatorname{prof} \mathcal{O}_{X_s,x} \geqq \inf(k - \dim \mathcal{O}_{S,s}, \dim \mathcal{O}_{X_s,x}) \geqq k(s)$ and the assertion follows. □

Corollary 1.2.3. *Let $f : X \to S$ be a flat morphism of finite type of noetherian schemes and let $U \subset X$ be the largest open subset smooth over S.*

(1) *Assume that the fiber X_s is reduced for every $s \in S$. Assume further that S is normal and that for the generic point s of each irreducible component, X_s is normal. Then X is normal.*

(2) *For $s \in S$ and a geometric point $\bar{s}$ above s, we consider the following conditions:*

 (i) *The geometric fiber $X_{\bar{s}}$ is reduced.*
 (ii) *U_s is dense in X_s.*

Then, we have (i) $\Rightarrow$ (ii). *Conversely, if X is normal and S is regular of dimension* $\leqq 1$, *then we have* (ii) $\Rightarrow$ (i).

Proof. (1) By Serre's criterion [EGA IV$_2$ 1965, théorème (5.8.6)], S satisfies (R_2) and (S_1). By [EGA IV$_2$ 1965, proposition (5.8.5)], every fiber X_s satisfies (R_1) and (S_0). Further if s is the generic point of an irreducible component, the fiber X_s satisfies (R_2) and (S_1). Since the function $k(s)$ for $k = 2$ satisfies $k(s) \leqq 1$ unless s is the generic point an irreducible component and $k(s) = 2$ for such point, the scheme X satisfies the conditions (R_2) and (S_1) by Propositions 1.2.1(1) and 1.2.2(1). Thus the assertion follows by [EGA IV$_2$ 1965, théorème (5.8.6)].

(2) (i) $\Rightarrow$ (ii): Since $X_{\bar{s}}$ is reduced, there exists a dense open subset $V \subset X_{\bar{s}}$ smooth over $\bar{s}$. Since f is flat, the image of V in X_s is a subset of U_s.

(ii) $\Rightarrow$ (i): Since X satisfies (S_2) and S is Cohen–Macaulay of dimension $\leqq 1$, the fiber X_s satisfies (S_1) by Proposition 1.2.2(3). Hence the geometric fiber $X_{\bar{s}}$ also satisfies (S_1) by [EGA IV$_2$ 1965, proposition (6.7.7)]. By (ii), $X_{\bar{s}}$ satisfies (R_0). Hence the assertion follow from [EGA IV$_2$ 1965, proposition (5.8.5)]. □

Lemma 1.2.4. *Let S be a noetherian scheme and let $f : Y \to X$ be a quasifinite morphism of schemes of finite type over S. Assume that X is smooth over S and that Y is flat and reduced over S. Assume that there exist dense open subschemes $U \subset S$ and $U \times_S X \subset W \subset X$ such that $Y \times_X W \to W$ is étale and that for every point $s \in S$, the inverse image $f_s^{-1}(W_s) \subset Y_s = Y \times_S s$ of $W_s = W \times_S s \subset X_s = X \times_S s$ by $f_s : Y_s \to X_s$ is dense. Then, $Y \to X$ is étale.*

Proof. If S is regular, the assumption that $Y \times_X W \to W$ is étale and Corollary 1.2.3(1) implies that the quasifinite morphism $Y \to X$ of normal noetherian schemes is étale in codimension $\leqq 1$. Since X is regular, the assertion follows from the purity theorem of Zariski–Nagata.

Since X and Y are flat over S, it suffices to show that for every point $s \in S$, the morphism $Y_s = Y \times_S s \to X_s$ is étale. Let $S' \to S$ be the normalization of the blow-up at the closure of $s \in S$. Then, there exists a point $s' \in S'$ above $s \in S$ such that the local ring $\mathcal{O}_{S',s'}$ is a discrete valuation ring. Since the assumption is preserved by the base change $\operatorname{Spec} \mathcal{O}_{S',s'} \to S$, the morphism $Y_{s'} = Y \times_S s' \to X_{s'} = X \times_S s'$ is étale. Hence $Y_s \to X_s$ is also étale as required. □

The following statement is a combination of the reduced fiber theorem and the flattening theorem.

Theorem 1.2.5 [Bosch et al. 1995, Theorem 2.1′; Raynaud and Gruson 1971, théorème (5.2.2)]. *Let S be a noetherian scheme and let $U \subset S$ be a schematically dense open subscheme. Let X be a scheme of finite type over S such that $X_U = X \times_S U$ is schematically dense in X and that $X_U \to U$ is flat and reduced.*

Then there exists a commutative diagram

$$\begin{array}{ccc} X & \longleftarrow & X' \\ \downarrow & & \downarrow \\ S & \longleftarrow & S' \end{array} \tag{1-9}$$

of schemes satisfying the following conditions:

(i) *The morphism $S' \to S$ is the composition of a blow-up $S^* \to S$ with center supported in $S - U$ and a faithfully flat morphism $S' \to S^*$ of finite type such that $U' = S' \times_S U \to U$ is étale.*

(ii) *The morphism $X' \to S'$ is* ***flat*** *and* ***reduced****. The induced morphism $X' \to X \times_S S'$ is finite and its restriction $X' \times_{S'} U' \to X \times_S U'$ is an isomorphism.*

If $X_U \to U$ is smooth and if S' is normal, then X' is the normalization of $X \times_S S'$ by Corollary 1.2.3(1). If $X_U \to U$ is étale, the first condition in (ii) implies that $X' \to S'$ is étale.

For the morphism $S' \to S$ satisfying the condition (i) in Theorem 1.2.5, we have the following variant of the valuative criterion.

Lemma 1.2.6. *Let S be a scheme and let U be a dense open subscheme. Let $S_1 \to S$ be a proper morphism such that $U_1 = U \times_S S_1 \to U$ is an isomorphism and let $S' \to S_1$ be a quasifinite faithfully flat morphism. Let $t \in U$, let $A \subset K = k(t)$ be a valuation ring and let $T = \operatorname{Spec} A \to S$ be a morphism extending $t \to U$. Then, there exist $t' \in U' = U \times_S S'$ above t, a valuation ring $A' \subset K' = k(t')$ such that $A = A' \cap K$ and a commutative diagram*

$$\begin{array}{ccc} T' & \longrightarrow & S' \\ \downarrow & & \downarrow \\ T & \longrightarrow & S \end{array} \tag{1-10}$$

for $T' = \operatorname{Spec} A'$. Further, if $t = T \times_S U$, then we have $t' = T' \times_{S'} U'$.

Proof. Since $S_1 \to S$ is proper and $U_1 \to U$ is an isomorphism, the morphism $T \to S$ is uniquely lifted to $T \to S_1$ by the valuative criterion of properness. Let $x_1 \in T \times_{S_1} S'$ be a closed point and let $t' \in t \times_{S_1} S'$ be a point above t such that x_1 is contained in the closure $T_1 = \overline{\{t'\}} \subset T \times_{S_1} S'$ with the reduced scheme structure. Let $A' \subset k(t')$ be a valuation ring dominating the local ring $\mathcal{O}_{T_1,x_1}$. Then, we have the commutative diagram (1-10) for $T' = \operatorname{Spec} A'$.

Since t' is the unique point of $t \times_T T'$, the equality $t = T \times_S U$ implies $t' = T' \times_{S'} U'$. □

1.3. ***Semistable curves.*** Let S be a scheme. Recall that a flat separated scheme X of finite presentation over S is a semistable curve, if every geometric fiber is purely of dimension 1 and has at most nodes as singularities.

Example 1.3.1. Let S be a scheme and let $D \subset S$ be an effective Cartier divisor. Let $C' \to \mathbb{A}^1_S$ be the blow-up at $D \subset S \subset \mathbb{A}^1_S$ regarded as a closed subscheme by the 0-section. Then, the complement $C_D \subset C'$ of the proper transform of the 0-section is a semistable curve over S and is smooth over the complement $U = S - D$. The exceptional divisor $\widetilde{D} \subset C_D$ is an effective Cartier divisor satisfying $0 \leqq \widetilde{D} \leqq D \times_S C_D$. The difference $D \times_S C_D - \widetilde{D}$ equals the proper transform of $\mathbb{A}^1_D$.

If $S = \operatorname{Spec} A$ is affine, $\mathbb{A}^1_S = \operatorname{Spec} A[t]$ and if D is defined by a nonzero divisor $f \in A$, we have $C_D = \operatorname{Spec} A[s,t]/(st - f)$ and $\widetilde{D} \subset C_D$ is defined by t.

Lemma 1.3.2. *Let S be a scheme and let $U \subset S$ be a schematically dense open subscheme. Let C be a separated flat scheme of finite presentation over S such that the base change $C_U = C \times_S U$ is a smooth curve over U. Then, the following conditions are equivalent*:

(1) *C is a semistable curve over S.*

(2) *Étale locally on C and on S, there exist an effective Cartier divisor $D \subset S$ such that $D \cap U$ is empty and an étale morphism $C \to C_D$ over S to the semistable curve C_D defined in Example 1.3.1.*

Proof. This is a special case of [SGA 7_{II} 1973, corollaire 1.3.2]. □

Let S be a normal noetherian scheme and let $j : U = S - D \to S$ be the open immersion of the complement of an effective Cartier divisor D. Let $i : D \to S$ be the closed immersion and let $\pi_D : \bar{D} \to D$ denote the normalization. Then, the valuations at the generic points of irreducible components of D define an exact sequence $0 \to \mathbb{G}_{m,S} \to j_*\mathbb{G}_{m,U} \to i_*\pi_{D*}\mathbb{Z}_{\bar{D}}$ of étale sheaves on S.

Let $f : C = C_D \to S$ be the semistable curve over S defined in Example 1.3.1. Let $\tilde{j} : U_C = C \times_S U \to C$ denote the open immersion and let $\tilde{i} : D_C = C \times_S D \to C$ denote the closed immersion. Let $A \subset C$ be the exceptional divisor and let $B = D_C - A \subset C$ be the effective Cartier divisor defined as the proper transform of $\mathbb{A}^1_D$. Let $a : A \to C$ and $b : B \to C$ and $e : E = A \cap B \to C$ denote the closed immersions. Then, the Cartier divisors $A, B, D_C \subset C$ defines a commutative diagram

$$\begin{array}{ccc} f^*i_*\mathbb{Z} & \longrightarrow & a_*\mathbb{Z} \oplus b_*\mathbb{Z} \\ \downarrow & & \downarrow \\ f^*(j_*\mathbb{G}_{m,U}/\mathbb{G}_{m,S}) & \longrightarrow & \tilde{j}_*\mathbb{G}_{m,U_C}/\mathbb{G}_{m,C} \end{array} \tag{1-11}$$

of étale sheaves on C.

Proposition 1.3.3. *Let S be a normal noetherian scheme and let $D \subset S$ be an effective Cartier divisor. Let $f : C = C_D \to S$ be the semistable curve defined in Example 1.3.1. Then, the diagram (1-11) induces an exact sequence*

$$0 \to f^*i_*\mathbb{Z} \to f^*(j_*\mathbb{G}_{m,U}/\mathbb{G}_{m,S}) \oplus (a_*\mathbb{Z} \oplus b_*\mathbb{Z}) \to \tilde{j}_*\mathbb{G}_{m,U_C}/\mathbb{G}_{m,C} \to 0 \quad (1\text{-}12)$$

of étale sheaves on D_C.

Proof. Let z be a geometric point of C; we will show the exactness of the stalks of (1-12) at z. Replacing S by the strict localization at the image x of z, we may assume that S is strict local and that x is the closed point. For $t \in S = S_{(x)}$, the Milnor fiber $C_{(z)} \times_S t$ at t of the strict localization $C_{(z)}$ at z is geometrically connected by [EGA IV$_4$ 1967, théorème (18.9.7)]. Further, if $z \in E$ and if $t \in D$, the fiber at t of $C_{(z)} - E_{(z)}$ has 2 geometrically connected components.

First, we consider the case where C is smooth over S at z. Then, since the Milnor fiber $C_{(z),t}$ is connected, the canonical morphism $f^*i_*\mathbb{Z}_{\bar{D}} \to i_{C*}\mathbb{Z}_{\bar{D}_C}$ is an isomorphism. Hence, the stalk of the lower horizontal arrow (1-11) at z is an injection. Further, this is a surjection by flat descent.

We assume that $C \to S$ is not smooth at z. Let $\widetilde{D}$ be a Cartier divisor of $C_{(z)}$ supported on $D_{C_{(z)}} = C_{(z)} \times_S D$. Then similarly as above, there exists a Cartier divisor D_1 on S supported on D such that $D_0 = \widetilde{D} - f^*D_1$ is supported on the inverse image of A. Define a $\mathbb{Z}$-valued function n on $y \in E_{(z)} = D$ as the intersection number of D_0 with the fiber $B \times_S y$. We show that the function n is constant. By adding some multiple of A to $\widetilde{D}$ if necessary, we may assume that D_0 is an effective Cartier divisor of C supported on A. Since B is flat over D, the pull-back $D_0 \times_C B$ is an effective Cartier divisor of B finite flat over D by [EGA IV$_1$ 1964, 0_{IV} proposition (15.1.16) c)⇒b)]. Hence the function n is constant. Thus we have $\widetilde{D} = f^*D_1 + n \cdot A$ and the exactness of the stalks of (1-12) at z follows. □

Corollary 1.3.4. *Let S be a normal noetherian scheme and let $C \to S$ be a semistable curve. Let $x \in S$ be a point and let $z \in C \times_S x$ be a singular point of the fiber. Assume that z is contained in the intersection of two irreducible components C_1 and C_2 of $C \times_S x$. Let $s_1 : S \to C$ and $s_2 : S \to C$ be sections meeting with the smooth parts of C_1 and C_2 respectively.*

Let $U \subset S$ be a dense open subscheme such that $C_U = C \times_S U$ is smooth over U and let $\widetilde{D} \subset C$ be an effective Cartier divisor such that $\widetilde{D} \cap C_U$ is empty. Define effective Cartier divisors $D_1 = s_1^*\widetilde{D}$ and $D_2 = s_2^*\widetilde{D}$ of S as the pull-back of $\widetilde{D}$.

Then, on a neighborhood of x, we have either $D_1 \leqq D_2$ or $D_2 \leqq D_1$. Suppose we have $D_1 \leqq D_2$ on a neighborhood of x. Then, we have $D_1 \times_S C \leqq \widetilde{D} \leqq D_2 \times_S C$ on a neighborhood of z.

Proof. In the notation of the proof of Proposition 1.3.3, we have $\widetilde{D} = f^*D_1 + nA$ for an integer n on an étale neighborhood of z. Hence the assertion follows. □

We recall a combination the flattening theorem and a strong version of the semistable reduction theorem for curves over a general base scheme.

Theorem 1.3.5 [Raynaud and Gruson 1971, théorème (5.2.2); Temkin 2011, Theorem 2.3.3]. *Let S be a noetherian scheme and let $U \subset S$ be a schematically dense open subscheme. Let $C \to S$ be a separated morphism of finite type such that $C \times_S U \to U$ is a smooth relative curve and that $C \times_S U \subset C$ is schematically dense. Then, there exists a commutative diagram*

$$\begin{array}{ccc} C & \longleftarrow & C' \\ \big\downarrow & & \big\downarrow \\ S & \longleftarrow & S' \end{array}$$

of schemes satisfying the following conditions:

(i) *The morphism $S' \to S$ is the composition of a proper modification $S_1 \to S$ such that $U_1 = U \times_S S_1 \to U$ is an isomorphism and a faithfully flat morphism $S' \to S_1$ such that $U' = U \times_S S' \to U_1$ is étale and $U' \subset S'$ is schematically dense.*

(ii) *The morphism $C' \to S'$ is a* semistable *curve and the morphism $C' \to C \times_S S'$ is a proper modification such that $C' \times_{S'} U' \to C \times_S U'$ is an isomorphism.*

Corollary 1.3.6. *Let S be a noetherian scheme and let $U \subset S$ be a schematically dense open subscheme. Let $C \to S$ be a separated morphism of finite type such that $C_U = C \times_S U \to U$ is a smooth relative curve and that $C_U \subset C$ is schematically dense. Let $X \to C$ be a separated morphism of finite type such that $X_U = X \times_S U \subset X$ is schematically dense and that $X_U \to C_U$ is flat and reduced. Then, there exists a commutative diagram*

$$\begin{array}{ccc} X & \longleftarrow & X' \\ \big\downarrow & & \big\downarrow \\ C & \longleftarrow & C' \\ \big\downarrow & & \big\downarrow \\ S & \longleftarrow & S' \end{array}$$

of schemes satisfying the following conditions:

(i) *The morphism $S' \to S$ is the composition of a proper modification $S_1 \to S$ such that $U_1 = U \times_S S_1 \to U$ is an isomorphism and a faithfully flat morphism $S' \to S_1$ such that $U' = U \times_S S' \to U_1$ is étale and $U' \subset S'$ is schematically dense.*

(ii) *The morphism $C' \to S'$ is a semistable curve and the morphism $C' \to C \times_S S'$ is the composition of a proper modification $C'_0 \to C \times_S S'$ such that $C'_0 \times_{S'} U' \to$*

$C \times_S U'$ is an isomorphism, a faithfully flat morphism $C'_1 \to C'_0$ such that $C'_1 \times_{S'} U' \to C'_0 \times_{S'} U'$ is étale and of a proper modification $C' \to C'_1$ such that $C' \times_{S'} U' \to C'_1 \times_{S'} U'$ is an isomorphism.

(iii) *The morphism $X' \to C'$ is flat and reduced, the morphism $X' \to X \times_C C'$ is finite and $X' \times_{S'} U' \to X \times_C C' \times_{S'} U'$ is an isomorphism.*

Proof. By the reduced fiber theorem (Theorem 1.2.5) applied to $X \to C$, there exists a commutative diagram

$$\begin{array}{ccc} X & \longleftarrow & X_1 \\ \downarrow & & \downarrow \\ C & \longleftarrow & C_1 \end{array}$$

satisfying the conditions (i) and (ii) of Theorem 1.2.5. Since $C_1 \times_S U \to C \times_S U$ is étale and $C_1 \times_S U \subset C_1$ is schematically dense, by the combination of the stable reduction theorem and the flattening theorem (Theorem 1.3.5), there exists a commutative diagram

$$\begin{array}{ccc} C_1 & \longleftarrow & C' \\ \downarrow & & \downarrow \\ S & \longleftarrow & S' \end{array}$$

satisfying the conditions (i) and (ii) of Theorem 1.3.5.

We show that $X' = X_1 \times_{C_1} C' \to C' \to S'$ satisfy the required conditions. By the construction, $S' \to S$ satisfies the condition (i) and $C' \to S'$ is a semistable curve. Since $C_1 \to C$ is obtained by applying Theorem 1.2.5 and $C' \to S'$ is obtained by applying Theorem 1.3.5, the composition $C' \to C'_1 = C_1 \times_S S' \to C \times_S S'$ satisfies the condition in (ii). Finally, the base change $X' \to C'$ of a flat and reduced morphism $X_1 \to C_1$ is flat and reduced. Since $X' \to C'$ is obtained by applying Theorem 1.2.5, the morphism $X' \to X \times_C C'$ satisfies the condition (iii). □

1.4. ***Subgroups and fiber functor.*** For a finite group G, let (Finite G-sets) denote the category of finite sets with left G-actions.

Definition 1.4.1. We say that a category C is a *finite Galois category* if there exist a finite group G and an equivalence of categories $F : C \to$ (Finite G-sets). If $F : C \to$ (Finite G-sets) is an equivalence of categories, we say that G is the *Galois group* of the finite Galois category C and call the functor F itself or the composition $C \to$ (Finite-sets) with the forgetful functor also denoted by F a *fiber functor* of C.

We say that a morphism $F \to F'$ of functors $F, F' : C \to$ (Finite-sets) is a surjection if $F(X) \to F'(X)$ is a surjection for every object X of C. For a subgroup $H \subset G$ and for a fiber functor $F : C \to$ (Finite G-sets), let F_H denote the functor $C \to$ (Finite-sets) defined by $F_H(X) = H\backslash F(X)$. The canonical morphism $F \to F_H$ is a surjection.

Surjections $F \to F_H$ are characterized as follows.

Proposition 1.4.2 (cf. [Abbes and Saito 2002, Proposition 2.1]). *Let C be a finite Galois category of the Galois group G and let $F : C \to$ (Finite-sets) be a fiber functor. Let $F' : C \to$ (Finite-sets) be another functor and let $F \to F'$ be a surjection of functors. Then, the following conditions are equivalent*:

(1) *For every surjection $X \to Y$ in C, the diagram*

$$\begin{array}{ccc} F(X) & \longrightarrow & F'(X) \\ \downarrow & & \downarrow \\ F(Y) & \longrightarrow & F'(Y) \end{array} \tag{1-13}$$

is a cocartesian diagram of finite sets. For every pair of objects X and Y of C, the morphism $F'(X) \amalg F'(Y) \to F'(X \amalg Y)$ is a bijection.

(2) *There exists a subgroup $H \subset G$ such that $F \to F'$ induces an isomorphism $F_H \to F'$.*

Proof. (1) $\Rightarrow$ (2): We may assume $C =$ (Finite G-sets) and F is the forgetful functor. For $X = G$, the mapping $F(G) = G \to F'(G)$ is a surjection of finite sets. Define an equivalence relation $\sim$ on G by requiring that $G/{\sim} \to F'(G)$ be a bijection and set $H = \{x \in G \mid x \sim e\}$. Then, since the group G acts on the object G of C by the right action, the relation $x \sim y$ is equivalent to $xy^{-1} \in H$. Since $\sim$ is an equivalence relation, the transitivity implies that H is stable under the multiplication, the reflexivity implies $e \in H$ and the symmetry implies that H is stable under the inverse. Hence H is a subgroup and the surjection $F(G) = G \to F'(G)$ induces a bijection $H\backslash G \to F'(G)$.

Let X be an object of $C =$ (Finite G-sets) and regard $G \times X$ as a G-set by the left action on G. Then, since the functor F' preserves the disjoint union, we have a canonical isomorphism $F'(G \times X) \to F'(G) \times X \to (H\backslash G) \times X$. Further, the cocartesian diagram (1-13) for the surjection $G \times X \to X$ in C defined by the action of G is given by

$$\begin{array}{ccc} G \times X & \longrightarrow & (H\backslash G) \times X \\ \downarrow & & \downarrow \\ X & \longrightarrow & F'(X) \end{array} \tag{1-14}$$

Thus we obtain a bijection $H\backslash X \to F'(X)$.

(2) $\Rightarrow$ (1): This is clear. □

Corollary 1.4.3. *Let the notation be as in Proposition 1.4.2 and let G' be a quotient group. Let $C' \subset C$ be the full subcategory consisting of objects X such that $F(X)$ are G'-sets. Then the subgroup $H' \subset G'$ defined by the surjection $F|_{C'} \to F'|_{C'}$ of the restrictions of the functors equals the image of $H \subset G$ in G'.*

Proof. If a G-set X is a G'-set, the quotient $H\backslash X$ is $H'\backslash X$. □

Corollary 1.4.4. *Let C be a finite Galois category of Galois group G and let $F : C \to$ (Finite G-sets) be a fiber functor. Let $G' \to G$ be a morphism of groups and let F also denote the functor $C \to$ (Finite G'-sets) defined as the composition defined by $G' \to G$. Let $F' : C \to$ (Finite G'-sets) be another functor and let $F \to F'$ be a surjection of functors such that the composition with the forgetful functor satisfies the condition* (1) *in Proposition 1.4.2.*

Let $H \subset G$ be the subgroup satisfying the condition (2) *in Proposition 1.4.2 and let $G'_1 \subset G$ be the image of $G' \to G$. Then, the functor F' induces a functor $C \to$* (Finite G'_1-sets) *and $G'_1 \subset G$ is a subgroup of the normalizer $N_G(H)$ of H.*

Proof. For an object X of C, $F(X)$ regarded as a G'-set is a G'_1-set. Since $F(X) \to F'(X)$ is a surjection of G'-sets, $F'(X)$ is also a G'_1-set. Since the left action of $G'_1 \subset G$ on the G-set $F(G) = G$ induces an action on $F'(G) = H\backslash G$, the subgroup H is normalized by G'_1. □

2. Dilatations

2.1. *Functoriality of dilatations.* Let X be a noetherian scheme and we consider morphisms

$$D \to X \leftarrow Q \leftarrow Y \tag{2-1}$$

of separated schemes of finite type over X satisfying the following condition:

(i) $D \subset X$, $D_Y = D \times_X Y \subset Y$ and $D_Q = D \times_X Q \subset Q$ are effective Cartier divisors and $Y \to Q$ is a closed immersion.

In later subsections, we will further assume the following condition:

(ii) X is normal and Q is smooth over X.

We give examples of constructions of Q for a given Y over X.

Example 2.1.1. Assume that X and Y are separated schemes of finite type over a noetherian scheme S.

(1) Assume $S = \operatorname{Spec} A$ and $Y = \operatorname{Spec} B$ are affine. Then, taking a surjection $A[T_1, \ldots, T_n] \to B$, we obtain a closed immersion $Y \to Q = \mathbb{A}^n_S \times_S X$.

(2) Assume that Y is smooth over S. Then, $Q = Y \times_S X \to X$ is smooth and the canonical morphism $Y \to Q = Y \times_S X$ is a closed immersion.

(3) Assume that $\pi : Y \to X$ is finite flat and define a vector bundle Q over X by the symmetric $\mathcal{O}_X$-algebra $S^\bullet \pi_* \mathcal{O}_Y$. Then the canonical surjection $S^\bullet \pi_* \mathcal{O}_Y \to \pi_* \mathcal{O}_Y$ defines a closed immersion $Y \to Q$.

For morphisms (2-1) satisfying the condition (i) above, we construct a commutative diagram

$$\begin{array}{ccccc} \bar{Y} & \longrightarrow & Q^{(D)} & & \\ \downarrow & & \downarrow & \searrow & \\ Y & \longrightarrow & Q^{[D]} & \longrightarrow & Q \end{array} \tag{2-2}$$

of schemes over X as follows. Let $\mathcal{I}_D \subset \mathcal{O}_X$ and $\mathcal{I}_Y \subset \mathcal{O}_Q$ be the ideal sheaves defining the closed subschemes $D \subset X$ and $Y \subset Q$. Let $Q' \to Q$ be the blow-up at $D_Y = D \times_X Y \subset Q$ and define the dilatation $Q^{[D]}$ at $Y \to Q$ and D to be the largest open subset of Q' where $\mathcal{I}_D \mathcal{O}_{Q'} \supset \mathcal{I}_Y \mathcal{O}_{Q'}$. Since D_Y is a divisor of Y, by the functoriality of blow-up, the immersion $Y \to Q$ is uniquely lifted to a closed immersion $Y \to Q^{[D]}$. Let $\bar{Y}$ and $Q^{(D)}$ be the normalizations of Y and $Q^{[D]}$ and let $\bar{Y} \to Q^{(D)}$ be the morphism induced by the morphism $Y \to Q^{[D]}$. If there is a risk of confusion, we also write $Q^{[D]}$ and $Q^{(D)}$ as $Q^{[D.Y]}$ and $Q^{(D.Y)}$ in order to make Y explicit.

Locally, if $Q = \operatorname{Spec} A$ and $Y = \operatorname{Spec} A/I$ are affine and if $D \subset X$ is defined by a nonzero divisor f, we have

$$Q^{[D]} = \operatorname{Spec} A[I/f] \tag{2-3}$$

for the subring $A[I/f] \subset A[1/f]$ and the immersion $Y \to Q^{[D]}$ is defined by the isomorphism $A[I/f]/(I/f)A[I/f] \to A/I$.

Example 2.1.2. Let X be a noetherian scheme and let $D \subset X$ be an effective Cartier divisor.

(1) Let Q be a smooth separated scheme over X and let $s : X \to Q$ be a section. Let $Y = s(X) \subset Q$ be the closed subscheme. Then, $Q^{[D]}$ is smooth over X. If X is normal, the canonical morphism $Q^{(D)} \to Q^{[D]}$ is an isomorphism.

(2) Assume that X is normal. Let Q be a smooth curve over X and let $s_1, \dots, s_n : X \to Q$ be sections. Define a closed subscheme $Y \subset Q$ as the sum $\sum_{i=1}^n s_i(X)$ of the sections regarded as effective Cartier divisors of Q. Assume that $D \subset s_n^*(s_i(X))$ for $i = 1, \dots, n-1$. Then $Q^{(nD)} \to X$ is smooth and $Y \times_{Q^{[nD]}} Q^{(nD)} \subset Q^{(nD)}$ is the sum $\sum_{i=1}^n \tilde{s}_i(X)$ of the sections $\tilde{s}_i : X \to Q^{(nD)}$ lifting $s_i : X \to Q$.

In fact, we may assume that $X = \operatorname{Spec} A$ is affine and, locally on Q, take an étale morphism $Q \to \mathbb{A}^1_X$. Then, we may assume that $Q = \mathbb{A}^1_X = \operatorname{Spec} A[T]$ and Y is defined by $P = \prod_{i=1}^n (T - a_i)$ for $a_i \in A$. We may further assume that D is defined by a nonzero divisor $a \in A$ dividing $a_1, \dots, a_n$. Then, we have

$Q^{[nD]} = \operatorname{Spec} A[T][P/a^n]$ and $T' = T/a$ satisfies $\prod_{i=1}^n (T' - a_i/a) = P/a^n$ in $A[T][1/a]$. Hence we have $Q^{(nD)} = \operatorname{Spec} A[T']$ and this equals $Q^{[D.s_n(X)]}$ and is smooth over X. The section $Y \to Q^{[nD]}$ is defined by $P/a^n = 0$ and hence $Y \times_{Q^{[nD]}} Q^{(nD)} \subset Q^{(nD)}$ is defined by $A[T']/\prod_{i=1}^n (T' - a_i/a)$.

We study the base change $Q^{[D]} \times_X D$.

Lemma 2.1.3. (1) *The canonical morphism $Q^{[D]} \to Q$ induces*

$$Q^{[D]} \times_X D = Q^{[D]} \times_Q D_Y \to D_Y. \tag{2-4}$$

(2) *If $Y \to Q$ is a regular immersion and if $T_Y Q$ and $T_D X$ denote the normal bundles, we have a canonical isomorphism*

$$T_Y Q(-D_Y) \times_Y D_Y = (T_Y Q \times_Y D_Y) \otimes (T_D X \times_D D_Y)^{\otimes -1} \to Q^{[D]} \times_X D. \tag{2-5}$$

The isomorphism (2-5) depends only on the restriction $D_Y \to Q$ and not on $Y \to Q$ itself.

(3) *Assume that Q is smooth over X and $X = Y \to Q$ is a section. Let $T(Q/X)$ denote the relative tangent bundle defined by the symmetric $\mathcal{O}_Q$-algebra $S^\bullet_{\mathcal{O}_Q} \Omega^1_{Q/X}$. Then, we have a canonical isomorphism*

$$T(Q/X)(-D) \times_Q D = (T(Q/X) \times_Q D) \otimes T_D X^{\otimes -1} \to Q^{[D]} \times_X D. \tag{2-6}$$

The isomorphism (2-6) depends only on the restriction $D \to Q$ and not on the section $X \to Q$ itself.

Proof. (1) Since $\mathcal{I}_D \mathcal{O}_{Q^{[D]}} \supset \mathcal{I}_Y \mathcal{O}_{Q^{[D]}}$ on $Q^{[D]}$ by the definition of $Q^{[D]}$, we have $Q^{[D]} \times_X D = Q^{[D]} \times_Q D_Y$. Hence, we obtain a morphism $Q^{[D]} \times_X D \to D_Y$.

(2) Assume that $Y \to Q$ is a regular immersion. Then, $D_Y \to Q$ is also a regular immersion and the normal bundle $T_{D_Y} Q$ fits in an exact sequence

$$0 \to T_{D_Y} D_Q \to T_{D_Y} Q \to T_D X \times_D D_Y \to 0$$

depending only on $D \to X$ and $D_Y \to Q$ and not on $Y \to Q$. Let $Q' \to Q$ be the blow-up at $D_Y \subset Q$. Then, the exceptional divisor $Q' \times_Q D_Y$ is canonically identified with the projective space bundle $\mathbb{P}(T_{D_Y} Q)$ over D_Y. Its open subset $Q^{[D]} \times_Q D_Y$ is identified as in (2-5) since $T_{D_Y} D_Q = T_Y Q \times_Y D_Y$.

(3) Since the normal bundle $T_X Q$ is canonically identified with the restriction $T(Q/X) \times_Q X$ of the relative tangent bundle, the assertion follows from (2). □

We give a sufficient condition for the morphism $\overline{Y} \to Q^{(D)}$ to be an immersion.

Lemma 2.1.4. *Assume that X and $Y - D_Y$ are normal and let $\pi : \overline{Y} \to Y$ be the normalization. Assume that $\overline{Y} \to X$ is étale and that $\pi_* \mathcal{O}_{\overline{Y}} / \mathcal{O}_Y$ is an $\mathcal{O}_{D_Y}$-module. Then, the finite morphism $\overline{Y} \to Q^{(2D)}$ is a closed immersion.*

Proof. Since the assertion is étale local on Q and X, we may assume that $Y \to X$ is finite and that the étale covering $\overline{Y} \to X$ is split. We may further assume that X, Y and Q are affine and that D is defined by a nonzero divisor f on X. Let $Y = \operatorname{Spec} A$, $\overline{Y} = \operatorname{Spec} \bar{A}$, $Q = \operatorname{Spec} B$, $Q^{[2D]} = \operatorname{Spec} B^{[2D]}$, $Q^{(2D)} = \operatorname{Spec} B^{(2D)}$ for $A = B/I$, $B^{[2D]} = B[I/f^2] \subset B[1/f]$ and the normalization $B^{(2D)}$ of $B^{[2D]}$. Since $\overline{Y} \to X$ is a split étale covering, it suffices to show that for every idempotent $e \in \bar{A}$, there exists a lifting $\tilde{e} \in B^{(2D)}$.

Since $\bar{A}/A$ is annihilated by f, the product $fe = g$ is an element of A. Let $\tilde{g} \in B$ be a lifting of g. Since $e^2 = e$, the element $h = \tilde{g}^2 - f\tilde{g} \in B$ is contained in I and hence $h/f^2 \in B[1/f]$ is an element of $B^{[2D]}$. Thus $\tilde{e} = \tilde{g}/f \in B[1/f]$ is a root of the polynomial $T^2 - T - h/f^2 \in B^{[2D]}[T]$ and is an element of $B^{(2D)}$. Since $\tilde{e}$ is a lifting of e, the assertion follows. □

We study the functoriality of the construction. We consider a commutative diagram

$$\begin{array}{ccccccccc} D \times_X X' & \overset{\subset}{\hookrightarrow} & D' & \overset{\subset}{\hookrightarrow} & X' & \longleftarrow & Q' & \longleftarrow & Y' \\ \downarrow & & & & \downarrow & & \downarrow & & \downarrow \\ D & & \overset{\subset}{\hookrightarrow} & & X & \longleftarrow & Q & \longleftarrow & Y \end{array} \tag{2-7}$$

of schemes such that the both lines satisfy the condition (i) on the diagram (2-1). Then, by the functoriality of dilatations and normalizations, we obtain a commutative diagram

$$\begin{array}{ccccccc} Y' & \longrightarrow & Q'^{[D']} & \longleftarrow & Q'^{(D')} & \longleftarrow & \overline{Y}' \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ Y & \longrightarrow & Q^{[D]} & \longleftarrow & Q^{(D)} & \longleftarrow & \overline{Y}. \end{array} \tag{2-8}$$

The diagram (2-8) induces a morphism

$$Q'^{(D')} \times_{Q'^{[D']}} Y' \to Q^{(D)} \times_{Q^{[D]}} Y. \tag{2-9}$$

Let $\bar{x}$ be a geometric point of D and let $\bar{x}'$ be a geometric point of $D \times_X X'$ above $\bar{x}$. Then the diagram (2-8) also induces a mapping

$$\pi_0(Q'^{(D')}_{\bar{x}'}) \to \pi_0(Q^{(D)}_{\bar{x}}) \tag{2-10}$$

of the sets of connected components of the geometric fibers.

First we study the dependence on Q.

Proposition 2.1.5. *Suppose $X = X'$, $Y = Y'$ and $D = D'$ and let $\bar{x}$ be a geometric point of D.*

(1) *Assume that $Q' \to Q$ is smooth and let $T = T(Q'/Q)$ denote the relative tangent bundle of Q' over Q. Then $Q'^{[D]} \to Q^{[D]}$ is also smooth and there exists a cartesian diagram*

$$\begin{array}{ccc} T(-D) \times_{Q'} D_Y & \longleftarrow & Q'^{[D]} \times_X D \\ \downarrow & \square & \downarrow \\ D_Y & \xleftarrow{(2\text{-}4)} & Q^{[D]} \times_X D \end{array} \tag{2-11}$$

(2) *Assume that Q and Q' are smooth over X. Then, the square*

$$\begin{array}{ccc} Q'^{[D]} & \longleftarrow & Q'^{(D)} \\ \downarrow & & \downarrow \\ Q^{[D]} & \longleftarrow & Q^{(D)} \end{array} \tag{2-12}$$

is cartesian. The induced morphism $Q'^{(D)} \times_{Q'^{[D]}} Y \to Q^{(D)} \times_{Q^{[D]}} Y$ (2-9) is an isomorphism over Y and the induced mapping $\pi_0(Q'^{(D)}_{\bar{x}}) \to \pi_0(Q^{(D)}_{\bar{x}})$ (2-10) is a bijection.

Proof. (1) First, we show the case where $Q' \to Q$ admits a section $Q \to Q'$ extending $Y \to Q'$. The section $Q \to Q'$ defines a section $Q^{[D]} \to Q' \times_Q Q^{[D]}$. Define $(Q' \times_Q Q^{[D]})^{[D_{Q^{[D]}} \cdot Q^{[D]}]}$ to be the dilatation of $Q' \times_Q Q^{[D]}$ for the section $Q^{[D]} \to Q' \times_Q Q^{[D]}$ and a divisor $D_{Q^{[D]}} = D \times_X Q^{[D]}$ over $Q^{[D]}$. We show that the canonical morphism $Q'^{[D]} \to Q' \times_Q Q^{[D]}$ induces an isomorphism

$$Q'^{[D]} \to (Q' \times_Q Q^{[D]})^{[D_{Q^{[D]}} \cdot Q^{[D]}]}. \tag{2-13}$$

Since the question is étale local on Q', we may assume that $Q' = \mathbb{A}^n_Q$ and the section $Q \to Q'$ is the 0-section. Further, we may assume that $Q = \operatorname{Spec} A$ and $Y = \operatorname{Spec} A/I$ are affine and that $D \subset X$ is defined by a nonzero divisor f on X. We set $A' = A[T_1, \dots, T_n]$ and $Q' = \operatorname{Spec} A'$. The 0-section $Q \to Q'$ is defined by the ideal $J = (T_1, \dots, T_n) \subset A'$. We have $Q^{[D]} = \operatorname{Spec} A[I/f]$ and $Q'^{[D]} = \operatorname{Spec} A'[I'/f]$ for $I' = IA' + J$. Since $A'[I'/f] = A[I/f][T_1/f, \dots, T_n/f]$ as a subring of $A'[1/f]$, we obtain an isomorphism (2-13).

By the isomorphism (2-13) and Example 2.1.2(1), the morphism $Q'^{[D]} \to Q^{[D]}$ is smooth. Further, by Lemma 2.1.3(3), we obtain a cartesian diagram (2-11), depending only on $D \to X$, $D_Y \to Q$ and $D_Y \to Q'$ but not on the choice of section $Q \to Q'$ extending $Y \to Q'$.

We prove the general case. Since $Q' \to Q$ has a section on $Y \subset Q$, locally on Q, there exist a closed subscheme $Q_1 \subset Q'$ étale over Q such that $Y \to Q'$ is induced by $Y \to Q_1$. For the smoothness of $Q'^{[D]} \to Q^{[D]}$, since the assertion is étale local, we may assume that $Q_1 = Q$ is a section. Hence the smoothness

$Q'^{[D]} \to Q^{[D]}$ follows. Further since the cartesian diagram (2-11) defined étale locally is independent of the choice of section, we obtain (2-11) for Q' by patching.

(2) First, we show the case where $Q' \to Q$ is smooth. Then by (1), $Q'^{[D]} \to Q^{[D]}$ is also smooth and the fibered product $Q^{(D)} \times_{Q^{[D]}} Q'^{[D]}$ is normal. Hence the square (2-12) is cartesian and the morphism (2-9) is an isomorphism. By the cartesian squares (2-12) and (2-11), $Q'^{(D)}_{\bar{x}}$ is a vector bundle over $Q^{(D)}_{\bar{x}}$. Hence (2-10) is a bijection.

We show the general case. A morphism $f : Q' \to Q$ is decomposed as the composition of the projection $\mathrm{pr}_2 : Q' \times_X Q \to Q$ and a section of the projection $\mathrm{pr}_1 : Q' \times_X Q \to Q'$. Hence, the cartesian squares (2-12) and the bijections (2-10) for the projections imply those for f respectively. The cartesian square (2-12) for f implies an isomorphism (2-9) for f. □

Corollary 2.1.6. *Assume that Q and Q' are smooth over X. Then, the morphism $Q'^{(D')} \times_{Q'^{[D']}} Y' \to Q^{(D)} \times_{Q^{[D]}} Y$ (2-9) is independent of $Q' \to Q$. Let $\bar{x}$ be a geometric point of D and let $\bar{x}'$ be a geometric point of D' above $\bar{x}$. Then the mapping $\pi_0(Q'^{(D')}_{\bar{x}'}) \to \pi_0(Q^{(D)}_{\bar{x}})$ (2-10) is independent of morphism $Q' \to Q$.*

Proof. Decompose a morphism $Q' \to Q$ as $Q' \to Q' \times_X Q \to Q$. Then the isomorphism (2-9) and the bijection (2-10) for $Q' \to Q' \times_X Q$ are the inverses of those for the projection $Q' \times_X Q \to Q'$. Hence the assertion follows. □

By the canonical isomorphism (2-9), the finite scheme $Y \times_{Q^{[D]}} Q^{(D)}$ over Y is independent of Q. We write it as $Y^{(D)}$.

Lemma 2.1.7. *Suppose that the squares*

$$\begin{array}{ccc} D' & \longrightarrow & X' \\ \downarrow & \square & \downarrow \\ D & \longrightarrow & X \end{array} \qquad \begin{array}{ccc} Q' & \longrightarrow & Y' \\ \downarrow & \square & \downarrow \\ Q & \longrightarrow & Y \end{array}$$

are cartesian.

(1) *The morphism $Q'^{[D']} \to Q^{[D]} \times_Q Q'$ is a closed immersion and $Q'^{(D')} \to Q^{(D)} \times_Q Q'$ is finite. Consequently, the morphism $Q'^{(D')} \times_{Q'} Y' \to Q^{(D)} \times_Q Y$ is finite if $Y' \to Y$ is finite. Further, if Q and Q' are normal, then $Q'^{(D')}$ equals the normalization of $Q^{(D)} \times_Q Q'$ in $Q' - D' \times_{X'} Q'$.*

(2) *If $Q' \to Q$ is flat, the square*

$$\begin{array}{ccc} Q'^{[D']} & \longrightarrow & Q' \\ \downarrow & \square & \downarrow \\ Q^{[D]} & \longrightarrow & Q \end{array}$$

is cartesian.

Proof. Since the assertion is local on a neighborhood of $Y' \subset Q'$, we may assume that $Q = \operatorname{Spec} A$, $Y = \operatorname{Spec} A/I$, $Q' = \operatorname{Spec} A'$ and $Y' = \operatorname{Spec} A'/IA'$ are affine and that D is defined by a nonzero divisor f on X. Then, we have $Q^{[D]} = \operatorname{Spec} A[I/f]$ and $Q'^{[D']} = \operatorname{Spec} A'[IA'/f]$.

(1) Since $A[I/f] \otimes_A A' \to A'[IA'/f]$ is a surjection, the morphism $Q'^{[D']} \to Q^{[D]} \times_Q Q'$ is a closed immersion. The remaining assertions follow from this immediately.

(2) If $A \to A'$ is flat, the injection $A[I/f] \to A[1/f]$ induces an injection

$$A' \otimes_A A[I/f] \to A' \otimes_A A[1/f] = A'[1/f].$$

Hence the surjection $A' \otimes_A A[I/f] \to A'[IA'/f]$ is an isomorphism. □

The construction of $Q^{(D)}$ commutes with base change if $Q^{(D)} \to X$ is flat and reduced.

Lemma 2.1.8. *Suppose that the diagram (2-7) is cartesian and $D' = D \times_X X'$. Assume that one of the following conditions is satisfied*:

(i) *X' is normal, $Q \to X$ is smooth and $Q^{(D)} \to X$ is flat and reduced.*

(ii) *$X' \to X$ is smooth.*

Then the square

$$\begin{array}{ccc} Q^{(D)} & \longleftarrow & Q'^{(D')} \\ \downarrow & & \downarrow \\ X & \longleftarrow & X' \end{array} \tag{2-14}$$

is cartesian.

Proof. By Lemma 2.1.7(1), $Q'^{(D')}$ is the normalization of $Q^{(D)} \times_X X'$. If the condition (i) is satisfied, then $Q^{(D)} \times_X X'$ is normal by Corollary 1.2.3(1). If $X' \to X$ is smooth, then $Q^{(D)} \times_X X'$ is smooth over $Q^{(D)}$ and is normal. Hence the square (2-14) is cartesian in both cases. □

We study the dependence on D and show that the canonical morphism contracts the closed fiber.

Lemma 2.1.9. *Suppose $X = X'$, $Y = Y'$ and $Q = Q'$, and that $D_1 = D' - D$ is an effective Cartier divisor of X. Then, the morphism $Q^{[D']} \to Q^{[D]}$ (resp. $Q^{(D')} \to Q^{(D)}$) induces a morphism $Q^{[D']} \times_Q D_{1,Y} \to D_{1,Y} \subset Y \subset Q^{[D]}$ (resp. $Q^{(D')} \times_Q D_{1,Y} \to Q^{(D)} \times_{Q^{[D]}} D_{1,Y} \subset Q^{(D)}$).*

Proof. We consider the immersion $Y \to Q^{[D]}$ lifting $Y \to Q$. Then, the morphism $Q^{[D']} \to Q^{[D]}$ induces an isomorphism $Q^{[D']} \to (Q^{[D]})^{[D_1]}$ to the dilatation $(Q^{[D]})^{[D_1]}$ of $Q^{[D]}$ for $Y \to Q^{[D]}$ and $D_1 \subset X$. Hence the morphism (2-4) defines a morphism $Q^{[D']} \times_Q D_{1,Y} \to D_{1,Y}$. The assertion for $Q^{(D')}$ follows from this. □

2.2. *Dilatations and complete intersection.* We give a condition for the right square in (2-7) to be cartesian.

Lemma 2.2.1. *Let S be a noetherian scheme and let $Q \to P$ be a quasifinite morphism of smooth schemes of finite type over S. If $Q \to P$ is flat on dense open subschemes, then $Q \to P$ is flat and locally of complete intersection of relative virtual dimension* 0.

Proof. Let $U \subset P$ and $V \subset Q$ be dense open subschemes such that $V \to U$ is flat. Then the relative dimension of $V \to S$ is the same as that of $U \to S$. Hence, we may assume that the relative dimensions of $P \to S$ and $Q \to S$ are the same integer n.

The morphism $Q \to P$ is the composition of the graph $Q \to Q \times_S P$ and the projection $Q \times_S P \to P$. For every point $x \in P$, the fiber $Q \times_P x \to Q \times_S x$ is a regular immersion of codimension n. Hence by [EGA IV$_3$ 1966, proposition (15.1.16) c)$\Rightarrow$b)] applied to the immersion $Q \to Q \times_S P$ over P, the immersion $Q \to Q \times_S P$ is also a regular immersion of codimension n and $Q \to P$ is flat. □

Lemma 2.2.2. *Let S be a noetherian scheme and let $Y \to X$ be a morphism of schemes of finite type over S.*

(1) *Suppose that there exists a cartesian diagram*

$$\begin{array}{ccc} Q & \longleftarrow & Y \\ \downarrow & \square & \downarrow \\ P & \longleftarrow & X \end{array} \tag{2-15}$$

of schemes of finite type over S satisfying the following conditions:

> *P and Q are smooth over S and $Q \to P$ is quasifinite and is flat on dense open subschemes. The horizontal arrows are closed immersions.*

Then $Y \to X$ is quasifinite, flat and locally of complete intersection of relative virtual dimension 0.

(2) *Conversely, suppose that $Y \to X$ is finite (resp. quasifinite) and locally of complete intersection of relative virtual dimension* 0. *Then $Y \to X$ is flat and, locally on X (resp. locally on X and on Y), there exists a cartesian diagram* (2-15) *satisfying the following conditions*:

> *P and Q are smooth of the same relative dimension over S and $Q \to P$ is quasifinite and flat. The horizontal arrows are closed immersions.*

Proof. (1) By Lemma 2.2.1, the quasifinite morphism $Q \to P$ is flat and locally of complete intersection. Hence $Y \to X$ is also quasifinite, flat and locally of complete intersection of relative virtual dimension 0.

(2) Since the assertion is local, we may assume that S, X and Y are affine. Take a closed immersion

$$Q_1 = \mathbb{A}^m_X \leftarrow Y.$$

Since the immersion $Y \to Q_1$ is a regular immersion of codimension m and since $Y \to X$ is finite (resp. quasifinite), after shrinking X (resp. Q_1 and Y), we may assume that the ideal defining $Y \subset Q_1$ is generated by m sections $f_1, \dots, f_m$ of $\mathcal{O}_{Q_1}$. Also take a closed immersion $P_1 = \mathbb{A}^n_S \leftarrow X$ and an open subscheme $Q \subset \mathbb{A}^m_{P_1}$ to obtain a cartesian diagram

$$\begin{array}{ccccc} Q & \longleftarrow & Q_1 & \longleftarrow & Y \\ \downarrow & \square & \downarrow & & \\ P_1 & \longleftarrow & X & & \end{array} \tag{2-16}$$

Taking sections $\tilde{f}_1, \dots, \tilde{f}_m$ of $\mathcal{O}_Q$ lifting $f_1, \dots, f_m$ after shrinking Q if necessary, define a morphism $Q \to P = \mathbb{A}^m_{P_1}$. Then, we obtain a cartesian diagram

$$\begin{array}{ccccc} Q & \longleftarrow & Q_1 & \longleftarrow & Y \\ \downarrow & \square & \downarrow & \square & \downarrow \\ P & \longleftarrow & \mathbb{A}^m_X & \longleftarrow & X \end{array} \tag{2-17}$$

where the lower right horizontal arrow $\mathbb{A}^m_X \to X$ is the 0-section.

The schemes $P = \mathbb{A}^{n+m}_S$ and $Q \subset \mathbb{A}^{n+m}_S$ are smooth over S. Since $Y \to X$ is quasifinite, after replacing Q by a neighborhood of Y if necessary, the morphism $Q \to P$ is quasifinite. Since Q and P are smooth of the same relative dimension over S, the morphism $Q \to P$ is flat on dense open subschemes. By Lemma 2.2.1, the quasifinite morphism $Q \to P$ is flat and hence $Y \to X$ is also flat. $\square$

We give examples of construction of the diagram (2-15).

Example 2.2.3. Assume that X and Y are schemes of finite type over a noetherian scheme S.

(1) Assume $X = \operatorname{Spec} A$ and $Y = \operatorname{Spec} B$ are affine. Let

$$A[T_1, \dots, T_n]/(f_1, \dots, f_n) \to B$$

be an isomorphism and define a morphism

$$Q = \mathbb{A}^n_X = \operatorname{Spec} A[T_1, \dots, T_n] \to P = \mathbb{A}^n_X$$

by $f_1, \dots, f_n$. Then, we obtain a cartesian diagram (2-15) by defining the section $X \to P = \mathbb{A}^n_X$ to be the 0-section.

(2) Assume that X and Y are smooth over a noetherian scheme S. Then, we obtain a cartesian diagram

$$\begin{array}{ccc} Y & \longrightarrow & Q = Y \times_S X \\ \big\downarrow & \square & \big\downarrow \\ X & \longrightarrow & P = X \times_S X \end{array}$$

Consider a cartesian diagram (2-15) satisfying the conditions of Lemma 2.2.2(1) and let $D \subset X$ be an effective Cartier divisor. Assume that $Q^{(D)} \to P^{(D)}$ is étale on a neighborhood of $Q^{(D)} \times_X D$. Let $\bar{x}$ be a geometric point of D and let $0_{\bar{x}}$ denote the geometric point above the origin of the vector space $P^{(D)}_{\bar{x}}$ over $\bar{x}$. Then, since $Q^{(D)}_{\bar{x}} \to P^{(D)}_{\bar{x}}$ is finite étale, we have an action of the fundamental group $\pi_1(P^{(D)}_{\bar{x}}, 0_{\bar{x}})$ on

$$Y^{(D)}_{\bar{x}} = Q^{(D)}_{\bar{x}} \times_{P^{(D)}_{\bar{x}}} 0_{\bar{x}}.$$

The action on $Y^{(D)}_{\bar{x}}$ is compatible with the canonical mapping $Y^{(D)}_{\bar{x}} \to \pi_0(Q^{(D)}_{\bar{x}})$ with respect to the trivial action on $\pi_0(Q^{(D)}_{\bar{x}})$ and is transitive on the inverse image of each element of $\pi_0(Q^{(D)}_{\bar{x}})$.

Since $Q^{[D]} \to P^{[D]} \times_P Q$ is an isomorphism by Lemma 2.1.7(2), for a geometric point $\bar{y}$ of $Y_{\bar{x}}$ and for the geometric point $0_{\bar{y}}$ of $Q^{[D]}_{\bar{y}}$ above $P^{(D)}_{\bar{x}}$, we have canonical isomorphisms $Q^{[D]}_{\bar{y}} = Q^{[D]} \times_Q \bar{y} \to P^{[D]}_{\bar{x}} = P^{(D)}_{\bar{x}}$ and $\pi_1(Q^{[D]}_{\bar{y}}, 0_{\bar{y}}) \to \pi_1(P^{(D)}_{\bar{x}}, 0_{\bar{x}})$. The action of $\pi_1(P^{(D)}_{\bar{x}}, 0_{\bar{x}})$ on $Y^{(D)}_{\bar{x}}$ is compatible with the action of $\pi_1(Q^{[D]}_{\bar{y}}, 0_{\bar{y}})$ on $Y^{(D)}_{\bar{x}} \times_{Y_{\bar{x}}} \bar{y}$. For a morphism $Q' \to Q$, the canonical morphism $\pi_1(Q'^{[D]}_{\bar{y}}, 0_{\bar{y}}) \to \pi_1(Q^{[D]}_{\bar{y}}, 0_{\bar{y}})$ is compatible with the actions on $Y^{(D)}_{\bar{x}} \times_{Y_{\bar{x}}} \bar{y}$.

We study the relation between the étaleness of $Q^{(D)} \to P^{(D)}$ and the annihilator of $\mathcal{O}_{Y^{(D)}} \otimes_{\mathcal{O}_Y} \Omega^1_{Y/X}$.

Lemma 2.2.4. *Let*

$$\begin{array}{ccc} Q & \longleftarrow & Y \\ \big\downarrow & \square & \big\downarrow \\ P & \longleftarrow & X \end{array}$$

be a cartesian diagram of separated schemes of finite type over X. Assume that P and Q are smooth over X and that the vertical arrows are quasifinite and flat.

Assume that there exists an effective Cartier divisor $D_1 \subset D = D_1 + D_0$ of X such that $\mathcal{O}_{Y^{(D)}} \otimes_{\mathcal{O}_Y} \Omega^1_{Y/X}$ is annihilated by $\mathcal{I}_{D_1} \subset \mathcal{O}_X$ and that we have an equality $D_0 = D$ of underlying sets. Then, there exists an open neighborhood $W \subset Q^{[D]}$ of $Q^{[D]} \times_X D$ such that $Q^{[D]} \to P^{[D]}$ is étale on $W - (Q^{[D]} \times_X D)$.

Proof. It suffices to show that each irreducible component $Z \subset Q^{[D]}$ of the inverse image of the support of $\Omega^1_{Q/P}$ is either a subset of $Q^{[D]} \times_X D$ or does not meet $Q^{[D]} \times_X D$, since $Q^{[D]} \to Q$ is an isomorphism on the complement of the inverse images of D. Assume that Z is not a subset of $Q^{[D]} \times_X D$ but does meet $Q^{[D]} \times_X D$

and regard Z as an integral closed subscheme of $Q^{[D]}$. Then, $D \times_X Z \subset Z$ is a nonempty effective Cartier divisor.

Since the assertion is étale local on Q and X, we may assume that $Y \to X$ is faithfully flat and finite. Let $T_0 \subset Z \times_X Y^{(D)}$ be the closure of the complement $Z \times_X Y^{(D)} - D \times_X (Z \times_X Y^{(D)})$ and let T be its normalization. Then, since $Y \to X$ is finite surjective, $T \to Z$ is also finite surjective. Hence $D_T = D \times_X T \subset T$ is a nonempty effective Cartier divisor.

By the assumption that $\mathcal{O}_{Y^{(D)}} \otimes_{\mathcal{O}_Y} \Omega^1_{Y/X}$ is annihilated by $\mathcal{I}_{D_1} \subset \mathcal{O}_X$, the $\mathcal{O}_T$-module $\mathcal{O}_T \otimes_{\mathcal{O}_Y} \Omega^1_{Y/X}$ is annihilated by $\mathcal{I}_{D_1} \cdot \mathcal{O}_T$. Since D_T is a scheme over $Q^{[D]} \times_X D$, we have an isomorphism $\mathcal{O}_{D_T} \otimes_{\mathcal{O}_Q} \Omega^1_{Q/P} \to \mathcal{O}_{D_T} \otimes_{\mathcal{O}_Y} \Omega^1_{Y/X}$ by Lemma 2.1.3(1). Thus $\mathcal{O}_{D_T} \otimes_{\mathcal{O}_Q} \Omega^1_{Q/P}$ is also annihilated by $\mathcal{I}_{D_1} \cdot \mathcal{O}_{D_T}$. Since $D = D_1 + D_0$, this means an inclusion $\mathcal{I}_{D_1} \cdot \mathcal{O}_T \otimes_{\mathcal{O}_Q} \Omega^1_{Q/P} \subset \mathcal{I}_{D_0} \cdot \mathcal{I}_{D_1} \cdot \mathcal{O}_T \otimes_{\mathcal{O}_Q} \Omega^1_{Q/P}$. By Nakayama's lemma, we have $\mathcal{I}_{D_1} \cdot \mathcal{O}_T \otimes_{\mathcal{O}_Q} \Omega^1_{Q/P} = 0$ on a neighborhood of $D_0 \times_X T$.

Since Z is a subset of the inverse image of support of $\Omega^1_{Q/P}$, the annihilator ideal of $\mathcal{O}_T \otimes_{\mathcal{O}_Q} \Omega^1_{Q/P}$ is 0. This contradicts to that $D_0 \times_X T = D_T$ is nonempty. $\square$

Lemma 2.2.5. *Assume X is normal and let*

$$\begin{array}{ccc} Q & \longleftarrow & Y \\ \downarrow & \square & \downarrow \\ P & \longleftarrow & X \end{array}$$

be a cartesian diagram of separated schemes of finite type over X. Assume that P and Q are smooth over X and that the vertical arrows are quasifinite and flat.

Let Y_0 be a closed subscheme of Y étale over X satisfying an equality $D_{Y_0} = D_Y$ of underlying sets and let $\mathcal{J}_0 \subset \mathcal{O}_{D_Y}$ be the nilpotent ideal defining $D_{Y_0} \subset D_Y$. Let $n \geqq 1$ be an integer satisfying $\mathcal{J}_0^n = 0$ and let $D_0 \subset D$ be an effective Cartier divisor on X satisfying $nD_0 \leqq D$.

Assume that $Y^{(D)} = Y \times_{Q^{[D]}} Q^{(D)}$ is étale over X. Then $\mathcal{O}_{Y^{(D)}} \otimes_{\mathcal{O}_Y} \Omega^1_{Y/X}$ is annihilated by the ideal $\mathcal{I}_{D-D_0} \subset \mathcal{O}_X$ defining $D - D_0 \subset X$.

Proof. Let $\mathcal{I} \subset \mathcal{I}_0 \subset \mathcal{O}_Q$ and $\mathcal{I}_D \subset \mathcal{I}_{D_0} \subset \mathcal{O}_X$ be the ideals defining the closed subschemes $Y_0 \subset Y \subset Q$ and $D_0 \subset D \subset X$. Let $Y_0^{(n)} \subset Q$ denote the closed scheme defined by the ideal $\mathcal{I}_0^n \subset \mathcal{O}_Q$. Let $Q^{[D_0 \cdot Y_0]} \to Q$ denote the dilatation for $Y_0 \to Q$ and D_0. We also define a dilatation $Q^{[nD_0 \cdot Y_0^{(n)}]} \to Q$ for $Y_0^{(n)} \to Q$ and nD_0.

Since Y_0 is étale over X, the scheme $Q^{[D_0 \cdot Y_0]}$ is smooth over X by Example 2.1.2(1) and equals its normalization $Q^{(D_0 \cdot Y_0)}$. The canonical morphism $Q^{[D_0 \cdot Y_0]} \to Q^{[nD_0 \cdot Y_0^{(n)}]}$ is finite and induces an isomorphism $Q^{(D_0 \cdot Y_0)} \to Q^{(nD_0 \cdot Y_0^{(n)})}$ on the normalizations.

By the assumptions $\mathcal{J}_0^n = 0$ and $nD_0 \leqq D$, we have $\mathcal{I}_0^n \subset \mathcal{I} + \mathcal{I}_D \subset \mathcal{I} + \mathcal{I}_{nD_0}$. Hence we have a morphism $Q^{[nD_0]} \to Q^{[nD_0 \cdot Y_0^{(n)}]}$. Further, by $nD_0 \leqq D$, we obtain a morphism $Q^{(D)} \to Q^{(D_0 \cdot Y_0)}$ of normalizations.

The dilatation $P^{[D]}$ of P for the section $X \to P$ and D is smooth over X by Example 2.1.2(1) and hence is equal to the normalization $P^{(D)}$. Since $Y^{(D)} \to X$ is étale and since each square of the diagram

$$\begin{array}{ccccc}
Y^{(D)} & \longrightarrow & Q^{(D)} & & \\
\downarrow & & \downarrow & & \\
Y & \longrightarrow & Q^{[D]} & \longrightarrow & Q \\
\downarrow & & \downarrow & & \downarrow \\
X & \longrightarrow & P^{[D]} & \longrightarrow & P
\end{array}$$

is cartesian by Lemma 2.1.7(2), the quasifinite morphism $Q^{(D)} \to P^{(D)}$ of normal schemes is étale on a neighborhood $W \subset Q^{(D)}$ of $Y^{(D)}$ by [EGA IV$_4$ 1967, théorème (18.10.16)].

The commutative diagram

$$\begin{array}{ccccc}
Q^{(D)} & \longrightarrow & Q^{(D_0 \cdot Y_0)} & \longrightarrow & Q \\
\downarrow & & & & \downarrow \\
P^{(D)} & & \longrightarrow & & P
\end{array}$$

of schemes defines a commutative diagram

$$\begin{array}{ccccc}
\mathcal{O}_W \otimes \Omega^1_{Q^{(D)}/X} & \longleftarrow & \mathcal{O}_W \otimes \Omega^1_{Q^{(D_0 \cdot Y_0)}/X} & \longleftarrow & \mathcal{O}_W \otimes \Omega^1_{Q/X} \\
\uparrow & & & & \uparrow \\
\mathcal{O}_W \otimes \Omega^1_{P^{(D)}/X} & & \longleftarrow & & \mathcal{O}_W \otimes \Omega^1_{P/X}
\end{array}$$

of locally free $\mathcal{O}_W$-modules. Since $Q^{(D)} \to P^{(D)}$ is étale on W, the left vertical arrow is an isomorphism.

Since $X \to X$ and $Y_0 \to X$ are étale, the lower horizontal arrow (resp. the upper right horizontal arrow) induces an isomorphism $\mathcal{O}_W \otimes \Omega^1_{P/X} \to \mathcal{I}_D \cdot \mathcal{O}_W \otimes \Omega^1_{P^{(D)}/X}$ (resp. $\mathcal{O}_W \otimes \Omega^1_{Q/X} \to \mathcal{I}_{D_0} \cdot \mathcal{O}_W \otimes \Omega^1_{Q^{(D_0 \cdot Y_0)}/X}$). Hence $\mathcal{I}_{D-D_0} \cdot \mathcal{O}_W \otimes \Omega^1_{Q/X} = \mathcal{I}_D \cdot \mathcal{O}_W \otimes \Omega^1_{Q^{(D_0 \cdot Y_0)}/X}$ is contained in the image of $\mathcal{O}_W \otimes \Omega^1_{P/X}$. Or equivalently, $\mathcal{O}_W \otimes_{\mathcal{O}_Q} \Omega^1_{Q/P}$ is annihilated by $\mathcal{I}_{D-D_0}$. Hence its pull-back $\mathcal{O}_{Y^{(D)}} \otimes_{\mathcal{O}_Y} \Omega^1_{Y/X}$ is also annihilated by $\mathcal{I}_{D-D_0}$. □

3. Ramification

3.1. ***Ramification of quasifinite schemes.*** Let X be a normal noetherian scheme and let D be an effective Cartier divisor of X. Let Y be a quasifinite scheme over X such that $D_Y = D \times_X Y \subset Y$ is a Cartier divisor.

Locally on X, there exists a smooth scheme Q over X and a closed immersion $Y \to Q$ over X. Then, by Proposition 2.1.5 and Corollary 2.1.6, the scheme $Y^{(D)}$ over Y defined as $Y \times_{Q^{[D]}} Q^{(D)}$ is canonically independent of Q. Hence a finite scheme $Y^{(D)}$ over Y is defined by patching. Similarly, for a geometric point $\bar{x}$ above $x \in D$, the set $\pi_0(Q^{(D)}_{\bar{x}})$ of connected components of the geometric fiber is canonically independent of Q.

Definition 3.1.1. Let X be a normal noetherian scheme and let D be an effective Cartier divisor of X. Let Y be a quasifinite scheme over X such that $D_Y = D \times_X Y \subset Y$ is a Cartier divisor and let $\overline{Y}$ be the normalization of Y. Let $\bar{x}$ be a geometric point above a point $x \in D$.

By taking a closed immersion $Y \to Q$ to a smooth scheme Q over X defined on a neighborhood of x, we define finite sets $F^D_{\bar{x}}(Y/X)$ and $F^{D+}_{\bar{x}}(Y/X)$ by

$$F^D_{\bar{x}}(Y/X) = \pi_0(Q^{(D)}_{\bar{x}}), \qquad F^{D+}_{\bar{x}}(Y/X) = Y^{(D)}_{\bar{x}} \tag{3-1}$$

equipped with canonical mappings

$$\begin{array}{ccc} \overline{Y}_{\bar{x}} & \xrightarrow{\varphi^{D+}_{\bar{x}}} & F^{D+}_{\bar{x}}(Y/X) \\ {\scriptstyle \varphi^D_{\bar{x}}}\downarrow & \swarrow & \downarrow \\ F^D_{\bar{x}}(Y/X) & \longrightarrow & Y_{\bar{x}} \end{array} \tag{3-2}$$

induced by the morphisms in (2-2):

$$\begin{array}{ccc} \overline{Y} & \longrightarrow & Y^{(D)} \\ \downarrow & \swarrow & \downarrow \\ Q^{(D)} & \longrightarrow & Q \end{array}$$

We consider a commutative diagram

$$\begin{array}{ccccccccc} Y' & \longrightarrow & X' & \overset{\supset}{\hookleftarrow} & D' & \overset{\supset}{\hookleftarrow} & D \times_X X' & \longleftarrow & \bar{x}' \\ \downarrow & & \downarrow & & & & \downarrow & & \downarrow \\ Y & \longrightarrow & X & & \overset{\supset}{\hookleftarrow} & & D & \longleftarrow & \bar{x} \end{array} \tag{3-3}$$

of noetherian schemes. We assume that X' is normal, $D' \subset X'$ is an effective Cartier divisor, Y' is quasifinite over X' and that $D'_{Y'} \subset Y'$ is an effective Cartier divisor.

Then, the commutative diagram (2-8) induces a commutative diagram

$$\begin{array}{ccccccc} \overline{Y}'_{\bar{x}'} & \xrightarrow{\varphi^{D'+}_{\bar{x}'}} & F^{D'+}_{\bar{x}'}(Y'/X') & \longrightarrow & F^{D'}_{\bar{x}'}(Y'/X') & \longrightarrow & Y'_{\bar{x}'} \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \overline{Y}_{\bar{x}} & \xrightarrow{\varphi^{D+}_{\bar{x}}} & F^{D+}_{\bar{x}}(Y/X) & \longrightarrow & F^{D}_{\bar{x}}(Y/X) & \longrightarrow & Y_{\bar{x}}. \end{array} \tag{3-4}$$

For effective Cartier divisors D and D' of a scheme X defined by the ideal sheaves $\mathcal{I}_D, \mathcal{I}_{D'} \subset \mathcal{O}_X$ and for $x \in D$, we write $D < D'$ at x if we have a strict inclusion $\mathcal{I}_{D,x} \supsetneqq \mathcal{I}_{D',x}$. If $X = X'$, $Y = Y'$, $\bar{x} = \bar{x}'$ and if $D < D'$ at the image x of $\bar{x}$ as Cartier divisors, further we have an arrow $F^{D'}_{\bar{x}}(Y/X) \to F^{D+}_{\bar{x}}(Y/X)$ making the two triangles obtained by dividing the middle square commutative by Lemma 2.1.9.

Proposition 3.1.2. *Assume that $Y \to X$ is quasifinite, flat and locally of complete intersection and that the normalization $\overline{Y}$ of Y is étale over X.*

(1) *The arrows in diagram (3-2)*

$$\begin{array}{ccc} \overline{Y}_{\bar{x}} & \xrightarrow{\varphi^{D+}_{\bar{x}}} & F^{D+}_{\bar{x}}(Y/X) \\ {\scriptstyle \varphi^{D}_{\bar{x}}}\downarrow & \swarrow & \downarrow \\ F^{D}_{\bar{x}}(Y/X) & \longrightarrow & Y_{\bar{x}} \end{array}$$

are surjections.

(2) *Let $Y' \to Y$ be a surjective morphism locally of complete intersection of quasifinite and flat schemes over X. Assume that the normalization $\overline{Y}'$ of Y' is étale over X. Then, the diagram*

$$\begin{array}{ccccccc} \overline{Y}'_{\bar{x}} & \xrightarrow{\varphi^{D+}_{\bar{x}}} & F^{D+}_{\bar{x}}(Y'/X) & \longrightarrow & F^{D}_{\bar{x}}(Y'/X) & \longrightarrow & Y'_{\bar{x}} \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \overline{Y}_{\bar{x}} & \xrightarrow{\varphi^{D+}_{\bar{x}}} & F^{D+}_{\bar{x}}(Y/X) & \longrightarrow & F^{D}_{\bar{x}}(Y/X) & \longrightarrow & Y_{\bar{x}} \end{array} \tag{3-5}$$

is a cocartesian diagram of surjections.

Proof. By replacing X by the strict localization $X_{(\bar{x})}$, we may assume that $\bar{x} \to X$ is a closed immersion and that $Y \to X$ is finite.

(1) By Lemma 2.2.2(2), we may assume that there exist smooth schemes P and Q over X and a cartesian diagram

$$\begin{array}{ccc} Y & \longrightarrow & Q \\ \downarrow & \square & \downarrow \\ X & \longrightarrow & P \end{array}$$

of schemes over X such that the horizontal arrows are closed immersions and that the vertical arrows are quasifinite and flat. We verify that the diagram

$$\begin{array}{ccccccc} Q^{(D)}_{\bar{x}} & \longrightarrow & Q^{(D)} & \longleftarrow & Y^{(D)} & \longleftarrow & \overline{Y} \\ \downarrow & \square & \downarrow & & \downarrow & & \downarrow \\ P^{(D)}_{\bar{x}} & \longrightarrow & P^{(D)} & \longleftarrow & X & = & X \end{array} \tag{3-6}$$

satisfies the assumptions in Corollary 1.1.6. Since $P^{[D]} \to X$ is smooth, we have $P^{(D)} = P^{[D]}$. By Lemma 2.1.7.2, the diagram

$$\begin{array}{ccccc} Q & \longleftarrow & Q^{[D]} & \longleftarrow & Y \\ \downarrow & \square & \downarrow & \square & \downarrow \\ P & \longleftarrow & P^{[D]} & \longleftarrow & X \end{array}$$

is cartesian. Hence the middle square in (3-6) is also cartesian.

The diagram (3-6) satisfies the finiteness assumption in Corollary 1.1.6, by Lemma 2.1.7(1). Since $X = X_{(\bar{x})}$ is strictly local, the assumption that the canonical mapping $\bar{x} \to \pi_0(\overline{X})$ is a bijection is satisfied. Since $P^{(D)}_{\bar{x}}$ is a vector space over $\bar{x}$ and is connected, the mapping $\bar{x} \to P^{(D)}_{\bar{x}} \cap X \to \pi_0(P^{(D)}_{\bar{x}})$ are bijections of sets consisting of single elements. We may assume that the finite étale morphism $\overline{Y} \to X$ is surjective since if otherwise the assertion is trivial. Hence by Corollary 1.1.6(2) (resp. (3)), the mapping $\overline{Y}_{\bar{x}} \to Y^{(D)}_{\bar{x}} = F^{D+}_{\bar{x}}(Y/X)$ $\big($resp. $F^{D+}_{\bar{x}}(Y/X) = Y^{(D)}_{\bar{x}} \to \pi_0(Q^{(D)}_{\bar{x}}) = F^{D}_{\bar{x}}(Y/X)\big)$ is surjective.

Similarly, applying Corollary 1.1.6(2) to the diagram

$$\begin{array}{ccccccc} Y_{\bar{x}} & \longrightarrow & Y & \longleftarrow & Y & \longleftarrow & \overline{Y} \\ \downarrow & \square & \downarrow & \square & \downarrow & & \downarrow \\ \bar{x} & \longrightarrow & X & \longleftarrow & X & = & X \end{array}$$

we see that $\overline{Y}_{\bar{x}} \to Y_{\bar{x}}$ is a surjection.

(2) By Lemma 2.2.2(2), we may assume that there exists smooth schemes Q and Q' over X and a cartesian diagram

$$\begin{array}{ccc} Y' & \longrightarrow & Q' \\ \downarrow & & \downarrow \\ Y & \longrightarrow & Q \end{array}$$

of schemes over X such that the horizontal arrows are closed immersions and that the vertical arrows are quasifinite and flat.

We verify that the diagram

$$\begin{array}{ccccccc} Q'^{(D)}_{\bar{x}} & \longrightarrow & Q'^{(D)} & \longleftarrow & Y'^{(D)} & \longleftarrow & \overline{Y}' \\ \downarrow & \square & \downarrow & & \downarrow & & \downarrow \\ Q^{(D)}_{\bar{x}} & \longrightarrow & Q^{(D)} & \longleftarrow & Y^{(D)} & \longleftarrow & \overline{Y} \end{array}$$

satisfies the assumptions in Corollary 1.1.6. The middle square is cartesian by Lemma 2.1.7(2). The finiteness assumption in Corollary 1.1.6 is satisfied by Lemma 2.1.7(1). Since the finite étale covering $\overline{Y} \to X$ is split and X is connected, the assumption that the canonical mapping $\overline{Y}_{\bar{x}} \to \pi_0(\overline{Y})$ is a bijection is satisfied. By (1), $\overline{Y}_{\bar{x}} \to Y^{(D)}_{\bar{x}} \to \pi_0(Q^{(D)}_{\bar{x}})$ are surjective. We may assume that Y and Y' are finite over X. Since $Y' \to Y$ is surjective, the morphism $\overline{Y}' \to \overline{Y}$ of finite étale schemes over X is also surjective. Hence by Corollary 1.1.6(2) (resp. (3)), the right square (resp. the middle square) of (3-5) is a cocartesian diagram of surjections.

Similarly, applying Corollary 1.1.6(2) to the diagram

$$\begin{array}{ccccccc} Y'_{\bar{x}} & \longrightarrow & Y' & \longleftarrow & Y' & \longleftarrow & \overline{Y}' \\ \downarrow & \square & \downarrow & \square & \downarrow & & \downarrow \\ Y_{\bar{x}} & \longrightarrow & Y & \longleftarrow & Y & \longleftarrow & \overline{Y} \end{array}$$

we see that the big rectangle in (3-5) is a cocartesian diagram of surjections. □

Corollary 3.1.3. *Assume that $Y \to X$ is locally of complete intersection and that the normalization $\overline{Y}$ is étale over X. Let P and Q be smooth schemes over X and let*

$$\begin{array}{ccc} Y & \longrightarrow & Q \\ \downarrow & \square & \downarrow \\ X & \longrightarrow & P \end{array}$$

be a cartesian diagram of schemes over X such that the horizontal arrows are closed immersions and that the vertical arrows are quasifinite and flat. Then, the mapping $\overline{Y}_{\bar{x}} \to F^{D+}_{\bar{x}}(Y/X)$ is an injection on the inverse image of $y \in Y$ if and only if $Q^{(D)} \to P^{(D)}$ is étale on the inverse image of y by $Y^{(D)} \to Y$.

Proof. Since the assertion is étale local, we may assume that $Y \to X$ and $Q \to P$ are finite and that y is the unique point of the inverse image of x. Then, by Proposition 3.1.2(1), $\overline{Y}_{\bar{x}} \to Y^{(D)}_{\bar{x}} = F^{D+}_{\bar{x}}(Y/X) \subset Q^{(D)}_{\bar{x}}$ is a bijection of finite sets. Hence $Q^{(D)} \to P^{(D)}$ is étale at x by [EGA IV$_4$ 1967, théorème (18.10.16)]. □

Definition 3.1.4. Let X be a normal noetherian scheme and let $U \subset X$ be a dense open subscheme. Let Y be a quasifinite scheme over X such that $V = U \times_X Y \to U$ is étale. Let D be an effective Cartier divisor of X such that $U \cap D$ is empty and that $D_Y = D \times_X Y$ is an effective Cartier divisor.

(1) For $x \in D$, we consider the following condition on X, Y and D:

(RF) There exist an open neighborhood W of $x \in X$, a smooth scheme Q over W and a closed immersion $Y \times_X W \to Q$ such that the normalization $\overline{Y}$ of Y is étale over W and that the normalization $Q^{(D)}$ of the dilatation $Q^{[D]}$ is flat and reduced over W.

If the condition (RF) is satisfied at every $x \in D$, we say that Y over X satisfies the condition (RF) for D.

(2) Let $x \in D$ and assume that Y over X satisfies the condition (RF) for D at x. Let y be a point of $\overline{Y} \times_X x \subset \overline{Y} \times_X D$. We say that the ramification of $Y \to X$ is bounded by D (resp. by $D+$) at y, if the mapping $\varphi_{\bar{x}}^{D} : \overline{Y}_{\bar{x}} \to F_{\bar{x}}^{D}(Y/X)$ $\big($resp. $\varphi_{\bar{x}}^{D+} : \overline{Y}_{\bar{x}} \to F_{\bar{x}}^{D+}(Y/X)\big)$ is an injection on the inverse image of y.

We say that the ramification of $Y \to X$ is bounded by D (resp. by $D+$) at x, if the mapping $\varphi_{\bar{x}}^{D} : \overline{Y}_{\bar{x}} \to F_{\bar{x}}^{D}(Y/X)$ (resp. $\varphi_{\bar{x}}^{D+} : \overline{Y}_{\bar{x}} \to F_{\bar{x}}^{D+}(Y/X)$) is an injection.

If ramification is bounded by D, it is bounded by $D+$. We show that the condition (RF) is independent of the choice of Q.

Lemma 3.1.5. *Let X be a normal noetherian scheme and let $U \subset X$ be a dense open subscheme. Let Y be a quasifinite scheme over X such that $V = U \times_X Y \to U$ is étale. Let D be an effective Cartier divisor of X such that $U \cap D$ is empty and that $D_Y \subset Y$ is an effective Cartier divisor. Let $x \in D$.*

(1) *Assume that Y over X satisfies (RF) for D at x. Let $W \subset X$ be an open neighborhood of x, let Q be a smooth scheme over W and let $Y \times_X W \to Q$ be a closed immersion. Then, there exists an open neighborhood $W' \subset W$ of x, such that $(Q \times_W W')^{(D \times_X W')} \to W'$ is flat and reduced.*

(2) *Let $X' \to X$ be a morphism of normal noetherian scheme such that $U' = U \times_X X'$ is a dense open subscheme and that $D'_{Y'} = D_Y \times_X X' \subset Y' = Y \times_X X'$ is an effective Cartier divisor. Let x' be a point of $D' = D \times_X X'$ above x. We consider the following conditions*:

(i) *Y over X satisfies (RF) for D at x.*

(ii) *Y' over X' satisfies (RF) for D' at x'.*

We have (i) $\Rightarrow$ (ii). *Conversely, if $X' \to X$ is smooth at x', we have* (ii) $\Rightarrow$ (i).

Proof. (1) Set $D_W = D \times_X W$. After shrinking W if necessary, we may assume that there exist a smooth scheme Q_0 over W and a closed immersion $Y \times_X W \to Q_0$ such that $Q_0^{(D_W)} \to W$ is flat and reduced. Since $Q^{(D_W)} \leftarrow (Q \times_W Q_0)^{(D_W)} \to Q_0^{(D_W)}$ are smooth by Proposition 2.1.5, the assertion follows.

(2) (i)⇒(ii): This follows from Lemma 2.1.8.

(ii)⇒(i): After shrinking X' if necessary, we may assume that $X' \to X$ is smooth. Let W be an open neighborhood of x, let $Y \times_X W \to Q$ be a closed immersion to a smooth scheme Q over W and let $W' = W \times_X X'$. Then the morphism

$$(Q \times_W W')^{(D' \times_{X'} W')} \to Q^{(D \times_X W)} \times_W W'$$

is an isomorphism by Lemma 2.1.8. Hence the assertion follows. □

Lemma 3.1.6. *Let X be a normal noetherian scheme and let $U \subset X$ be a dense open subscheme. Let Y be a quasifinite scheme over X such that $V = U \times_X Y \to U$ is étale. Let D and D', with $D \subset D'$, be effective Cartier divisors of X such that $U \cap D'$ is empty and that $D'_Y \subset Y$ is an effective Cartier divisor. Let $x \in D$ and assume that Y over X satisfies (RF) for D and D' at x.*

Let $y \in Y$ be a point above x. If the ramification of Y over X is bounded by $D+$ at y and if $D < D'$ at x, then the ramification of Y over X is bounded by D' at y.

Proof. It follows from Lemma 2.1.9. □

Lemma 3.1.7. *Let X be a normal noetherian scheme and let $U \subset X$ be a dense open subscheme. Let Y be a quasifinite scheme over X such that $V = U \times_X Y \to U$ is étale. Let D be an effective Cartier divisor of X such that $U \cap D$ is empty and that $D_Y \subset Y$ is an effective Cartier divisor. Assume that Y over X satisfies the condition (RF) for D.*

Let $S \subset D_Y$ (resp. $S^+ \subset D_Y$) denote the subset consisting of points $y \in D_Y$ where the ramification of $Y \to X$ is bounded by D (resp. by $D+$).

(1) *We have $S \subset S^+$.*

(2) *The subset $S \subset D_Y$ is closed and the subset $S^+ \subset D_Y$ is open.*

Proof. (1) It follows from the commutative diagram (3-2).

(2) By Lemma 1.1.3 applied to $\overline{Y} \to Q^{(D)}$, we see that S is closed. Similarly, by Lemma 1.1.4 applied to $\overline{Y} \to Y^{(D)}$ we see that S^+ is open. □

Proposition 3.1.8. *Let X be a normal noetherian scheme and let $U \subset X$ be a dense open subscheme. Let Y be a quasifinite scheme over X such that $V = U \times_X Y \to U$ is étale. Let D be an effective Cartier divisor of X such that $U \cap D$ is empty and that $D_Y \subset Y$ is an effective Cartier divisor.*

Let C be a semistable curve over X such that $C_U = C \times_X U \to U$ is smooth. Let $x \in X$ be a point of D and let $z \in C$ be a singular point of the fiber C_x. Assume that there exist two irreducible components C_1 and C_2 of the fiber C_x meeting at z and let ζ_1 and ζ_2 be their generic points. Let $D_1 \subset D_2$ be effective Cartier divisors *on X and let $\widetilde{D} \subset C$ be an effective Cartier divisor such that $D_1 < D_2$ at x and that $\widetilde{D} = D_i \times_X C = D_{i,C}$ on a neighborhood of ζ_i for $i = 1, 2$.*

Assume that $Y_C = Y \times_X C$ over C satisfies the condition (RF) for $\widetilde{D}$ at z.

(1) *Y over X satisfies the condition (RF) for D_1 and D_2 at x.*

(2) *We have a commutative diagram*

$$\begin{array}{ccccc}
F_{\bar{x}}^{D_2+}(Y/X) & \longrightarrow & F_{\bar{z}}^{\widetilde{D}+}(Y_C/C) & \longrightarrow & F_{\bar{x}}^{D_1+}(Y/X) \\
\downarrow & \nearrow & \downarrow & \nearrow & \downarrow \\
F_{\bar{x}}^{D_2}(Y/X) & \longrightarrow & F_{\bar{z}}^{\widetilde{D}}(Y_C/C) & \longrightarrow & F_{\bar{x}}^{D_1}(Y/X)
\end{array} \tag{3-7}$$

(3) *The lower left horizontal arrow $F_{\bar{x}}^{D_2}(Y/X) \to F_{\bar{z}}^{\widetilde{D}}(Y_C/C)$ in (3-7) is an injection. The upper right horizontal arrow $F_{\bar{z}}^{\widetilde{D}+}(Y_C/C) \to F_{\bar{x}}^{D_1+}(Y/X)$ in (3-7) is an injection on the image of $\overline{Y}_{\bar{x}}$.*

Proof. (1) Since ζ_1 and ζ_2 are contained in any open neighborhood of z, the scheme Y_C over C satisfies (RF) for $\widetilde{D}$ at ζ_1 and ζ_2. Since $C \to X$ is smooth at ζ_1 and ζ_2, the scheme Y over X satisfies (RF) for D_1 and D_2 at x by Lemma 3.1.5(2).

(2) Let $D_{1,C}$ and $D_{2,C}$ be the pull-backs of D_1 and D_2 to C. Then, we have $D_{1,C} < \widetilde{D} < D_{2,C}$ at z. Hence by (3-4) with the slant arrow added, we obtain a commutative diagram

$$\begin{array}{ccccc}
F_{\bar{z}}^{D_{2,C}+}(Y_C/C) & \longrightarrow & F_{\bar{z}}^{\widetilde{D}+}(Y_C/C) & \longrightarrow & F_{\bar{z}}^{D_{1,C}+}(Y_C/C) \\
\downarrow & \nearrow & \downarrow & \nearrow & \downarrow \\
F_{\bar{z}}^{D_{2,C}}(Y_C/C) & \longrightarrow & F_{\bar{z}}^{\widetilde{D}}(Y_C/C) & \longrightarrow & F_{\bar{z}}^{D_{1,C}}(Y_C/C)
\end{array} \tag{3-8}$$

Since Y over X satisfies (RF) for D_1 and D_2 at x by (1), the pull-back defines canonical isomorphisms from the left and right columns of (3-7) to those of (3-8) by Lemma 2.1.8. Thus we obtain (3-7).

(3) By functoriality of cospecialization mappings, we obtain a commutative diagram

$$\begin{array}{ccccc}
F_{\bar{\zeta}_2}^{D_{2,C}}(Y_C/C) & \xleftarrow{\text{cosp.}} & F_{\bar{z}}^{D_{2,C}}(Y_C/C) & \longleftarrow & F_{\bar{x}}^{D_2}(Y/X) \\
\downarrow & & \downarrow & \swarrow & \\
F_{\bar{\zeta}_2}^{\widetilde{D}}(Y_C/C) & \xleftarrow{\text{cosp.}} & F_{\bar{z}}^{\widetilde{D}}(Y_C/C). & &
\end{array} \tag{3-9}$$

By Lemma 2.1.8 and by $\widetilde{D} = D_{2,C}$ at ζ_2, the composition $F_{\bar{x}}^{D_2}(Y/X) \to F_{\bar{\zeta}_2}^{\widetilde{D}}(Y_C/C)$ is a bijection. Hence $F_{\bar{x}}^{D_2}(Y/X) \to F_{\bar{z}}^{\widetilde{D}}(Y_C/C)$ is injective.

Since the second assertion is étale local on X, we may assume that $Y \to X$ is finite. By functoriality of specialization mappings, we obtain a commutative diagram

$$
\begin{array}{ccccccc}
 & & F^{\widetilde{D}+}_{\bar{z}}(Y_C/C) & \xleftarrow{\text{sp.}} & F^{\widetilde{D}+}_{\bar{\zeta}_1}(Y_C/C) & \longleftarrow & \overline{Y}_{\bar{x}} \\
 & \swarrow & \downarrow & & \downarrow & \swarrow & \\
F^{D_1+}_{\bar{x}}(Y/X) & \longrightarrow & F^{D_{1,C}+}_{\bar{z}}(Y_C/C) & \xleftarrow{\text{sp.}} & F^{D_{1,C}+}_{\bar{\zeta}_1}(Y_C/C). & &
\end{array}
$$

The vertical arrow $F^{\widetilde{D}+}_{\bar{z}}(Y_C/C) \to F^{D_{1,C}+}_{\bar{z}}(Y_C/C)$ is an injection on the image of $\overline{Y}_{\bar{x}}$, since the composition $F^{\widetilde{D}+}_{\bar{\zeta}_1}(Y_C/C) \to F^{D_{1,C}+}_{\bar{z}}(Y_C/C)$ is a bijection. Hence the assertion follows. □

3.2. ***Ramification and valuations.*** In the rest of the article, A denotes a valuation ring and K denotes its fraction field. Let $v : K^\times \to \Gamma = K^\times/A^\times$ denote the valuation.

Definition 3.2.1. Let X be a normal separated noetherian scheme, let $U \subset X$ be a dense open subscheme and let A be a valuation ring. We say that a morphism $T = \operatorname{Spec} A \to X$ is *U-external* if $T \times_X U$ consists of a single point t.

For a morphism $T = \operatorname{Spec} A \to X$ and an effective Cartier divisor $D \subset X$, let $v(D) \in \Gamma$ denote the valuation $v(f)$ of a nonzero divisor f defining $D \subset X$ on a neighborhood of the image of T.

Let $\widetilde{X} = \varprojlim X'$ be the inverse limit of proper schemes $X' \to X$ such that $U' = U \times_X X' \to U$ is an isomorphism. Then, points of $\widetilde{X} - U$ correspond bijectively to the inverse limits of the images of the closed points by the liftings of U-external morphisms $T \to X$ defined by valuation rings of the residue fields of points of U by [Fujiwara and Kato 2018, Theorem E.2.11].

Lemma 3.2.2. *Let X be a normal noetherian scheme, let $U \subset X$ be a dense open subscheme, let $t \in U$ be a point, let $A \subsetneqq K = k(t)$ be a valuation ring and let $T = \operatorname{Spec} A \to X$ be a U-external morphism.*

(1) *Let $g \in \Gamma(U', \mathcal{O}^\times_{U'})$ be an invertible function defined on an open neighborhood $U' \subset U$ of $t \in U$ such that $v(g) = \gamma \geqq 0$. Then, there exists a normal scheme X' of finite type over X such that $U \times_X X' = U'$, g is extended to a nonzero divisor on X' defining an effective Cartier divisor $R' \subset X'$, and $U' = X' - D'$ is the complement of an effective Cartier divisor $D' \subset X'$ and a U'-external morphism $T \to X'$ lifting $T \to X$ and $v(R') = \gamma$.*

(2) *Let K' be a finite separable extension of $K = k(t)$ and let $A' \subsetneqq K'$ be a valuation ring such that $A' \cap K = A$. Set $T' = \operatorname{Spec} A'$ and let $\gamma > 0$ be a*

positive element of the value group Γ' of A'. Then, there exist a commutative diagram

$$\begin{array}{ccccc} U' & \longrightarrow & X' & \longleftarrow & T' \\ \downarrow & \square & \downarrow & & \downarrow \\ U & \longrightarrow & X & \longleftarrow & T \end{array}$$

of schemes, a point $t' \in U'$ above t, an isomorphism $K' \to k(t')$ over K and an effective Cartier divisor R' of X' satisfying the following conditions (i)–(iv):

(i) *X' is a normal scheme of finite type over X.*
(ii) *The left square is cartesian and U' is a dense open subscheme of X' étale over U.*
(iii) *$T' \to X'$ is a U'-external morphism extending $t' \to U'$.*
(iv) *$R' \cap U' = \varnothing$ and $v'(R') = \gamma$.*

(3) *Let*

$$\begin{array}{ccccccc} U' & \longrightarrow & X' & \longleftarrow & T' & \longleftarrow & \bar{x}' \\ \swarrow & & \swarrow & & \swarrow & & \swarrow \\ U_1 & \longrightarrow & X_1 & \longleftarrow & T_1 & \longleftarrow & \bar{x}_1 \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ U & \longrightarrow & X & \longleftarrow & T & \longleftarrow & \bar{x} \end{array}$$

be a commutative diagram, let $t_1 \in U_1$ and $t' \in U'$ be points above $t \in U$ and let $R_1 \subset X_1$ and $R' \subset X'$ be effective Cartier divisors satisfying the following conditions (a)–(d):

(a) *X_1 and X' are normal noetherian schemes and $X_1 \to X$ is of finite type.*
(b) *The left square and the left parallelogram are cartesian and $U_1 \to U$ is étale. The open subschemes $U_1 \subset X_1$ and $U' \subset X'$ are dense.*
(c) *$T_1 = \mathrm{Spec}\, A_1$ and $T' = \mathrm{Spec}\, A'$ for valuation rings $A_1 \subsetneqq K_1 = k(t_1)$ and $A' \subsetneqq K' = k(t')$ satisfying $A_1 \cap K = A' \cap K = A$. The morphism $T_1 \to X_1$ is U_1-external and $T' \to X'$ is U'-external.*
(d) *$R_1 \cap U_1$ and $R' \cap U'$ are empty and we have $v_1(R_1) \leqq v'(R')$ in $\Gamma'_{\mathbb{Q}}$.*

Then, there exist a commutative diagram

$$\begin{array}{ccccccc} U'_1 & \longrightarrow & X'_1 & \longleftarrow & T'_1 & \longleftarrow & \bar{x}'_1 \\ \| & & \downarrow & & \downarrow & & \downarrow \\ U' \times_U U_1 & \longrightarrow & X' \times_X X_1 & \longleftarrow & T' \times_T T_1 & \longleftarrow & \bar{x}' \times_{\bar{x}} \bar{x}_1 \end{array}$$

and $t'_1 \in U'_1$ above t satisfying the following conditions (i)–(iv):

(i) *X'_1 is a normal scheme of finite type over X'.*

(ii) *U'_1 is $U' \times_{X'} X'_1$ and is a dense open subscheme of X'_1.*

(iii) *$T'_1 = \mathrm{Spec}\, A'_1$ for a valuation ring $A'_1 \subset K'_1 = k(t'_1)$.*

(iv) *For the pull-backs $R'_1 = R_1 \times_{X_1} X'_1$ and $R'_2 = R' \times_{X'} X'_1$, we have $R'_1 \leqq R'_2$.*

Proof. (1) Let Z and Z' be closed subschemes such that $U = X - Z$ and $U' = X - Z'$. By replacing X by the normalization of the blow-up at Z and at Z' and by the valuative criterion of properness, we may assume that $U = X - D$ and $U' = X - D'$ are the complements of effective Cartier divisors $D, D' \subset X$.

Let $x \in X$ be the image of the closed point of T and let $W = \mathrm{Spec}\, B \subset X$ be an open neighborhood of x such that $D \cap W$, $D' \cap W$ are principal divisors defined by $f, f' \in B$. Then we have $U' \cap W = W - D' \cap W = \mathrm{Spec}\, B[1/f']$. Set $g = h/f'^n \in B[1/f']$. The function g and hence also $h \in B$ are also invertible on $U' \cap W$. Set $\alpha = v(f), \alpha' = v(f') \in \Gamma$. Since $T \to X$ is U-external, we have $\Gamma^+[1/\alpha] = \Gamma$. Hence after replacing f by its power, we may assume that $\alpha' \leqq \alpha$.

Let $W' \to W$ be the normalization of the blow-up at the ideals (f'^n, h) and (f, f'). Since f, f' and h are invertible on $U' \cap W$, the morphism $W' \to W$ induces an isomorphism $U' \times_{X'} W' \to U' \cap W$. Since $W' \to W$ is proper, the morphism $T \to W$ is uniquely lifted to $T \to W'$. Since the generic point $t \in T$ is the unique point of $U \times_X T \supset (U' \times_{X'} W') \times_{W'} T$, the morphism $T \to W'$ is U'-external.

Let $x' \in W'$ be the image of the closed point of T. Since the ideals (f'^n, h) and (f, f') of $\mathcal{O}_{W,x'}$ are principal ideals and since $v(h) \geqq v(f'^n)$ and $v(f') \geqq v(f)$, there exists an open neighborhood X' of $x' \in W$ such that $U' \subset X'$, where we have inclusions $(f'^n) \supset (h)$ and $(f) \supset (f')$. Then, $g = h/f'^n$ defines a Cartier divisor R' on X' satisfying $R' \cap U' = \varnothing$ and $v(R') = \gamma$. We also have an inclusion $U \times_X X' = X' - D \times_X X' \subset X' - D' \times_X X' = U' \times_X X' = U'$. Since the other inclusion is obvious, we have $U' = U \times_X X'$.

(2) We may take an étale scheme $U_1 \to U$ such that $t' = \mathrm{Spec}\, K' = t \times_U U_1$ and a finite scheme $X_1 \to X$ containing U_1 as a dense open scheme. After shrinking U_1 if necessary, we may take an invertible function $g \in \Gamma(U_1, \mathcal{O}_{U_1}^\times)$ such that $\gamma = v'(g)$. Since T' is a localization of the normalization of $T \times_X X_1$, the morphism $t' \to U_1 \subset X_1$ is uniquely extended to $T' \to X_1$.

Then, by (1) applied to the open subschemes $U_1 \subset U \times_X X_1 \subset X_1$, to the morphism $T' \to X_1$ and to the invertible function $g \in \Gamma(U_1, \mathcal{O}_{U_1}^\times)$, the assertion follows.

(3) Let $T_{(\bar{x})}$, $T_{1,(\bar{x}_1)}$ and $T'_{(\bar{x}')}$ denote the strict localizations. We take a point $\tilde{t}'_1 \in T'_{(\bar{x}')} \times_{T_{(\bar{x})}} T_{1,(\bar{x}_1)}$ above the generic point of $T'_{(\bar{x}')}$. Then the normalization $\tilde{T}'_1$ of $T'_{(\bar{x}')}$ in $\tilde{t}'_1$ is $\tilde{T}'_1 = \mathrm{Spec}\, A_1'^{sh}$ for a strictly local valuation ring $A_1'^{sh}$. Let $t'_1 \in t' \times_t t_1 \subset T' \times_T T_1$ be the image of $\tilde{t}'_1$ and set $K'_1 = k(t'_1)$ and $A'_1 = A_1'^{sh} \cap K'_1$. Let $T'_1 = \mathrm{Spec}\, A'_1$ and $\bar{x}'_1$ be the geometric point of T'_1 defined by a geometric closed point of $\tilde{T}'_1$.

Let X'_0 be the normalization of $X' \times_X X_1$ in $U'_1 = U' \times_U U_1$. Define effective Cartier divisors of X'_0 by $R'_{0,1} = R_1 \times_{X_1} X'_0$ and $R'_{0,2} = R' \times_{X'} X'_0$. Let $\bar{X}'_1 \to X'_0$ be the normalization of the blow-up at $R'_0 \cap R'_1 = R'_0 \times_{X'_0} R'_1$ and define effective Cartier divisors of $\bar{X}'_1$ by $\bar{R}'_1 = R_1 \times_{X_1} \bar{X}'_1$ and $\bar{R}'_2 = R' \times_{X'} \bar{X}'_1$. Since $\bar{X}'_1 \to X' \times_X X_1$ is proper, the morphism $t'_1 \to t' \times_t t_1 \subset U' \times_U U_1$ is uniquely lifted to $T'_1 \to \bar{X}'_1$ by the valuative criterion of properness.

Let $x'_1 \in \bar{X}'_1$ be the image of the closed point of T'_1. The intersection $\bar{R}'_1 \cap \bar{R}'_2 \subset \bar{X}'_1$ is the exceptional divisor and hence is an effective Cartier divisor. Since $v'_1(\bar{R}'_1) \leqq v'_1(\bar{R}'_2)$, on an open neighborhood $X'_1 \subset \bar{X}'_1$ of x'_1, we have $\bar{R}'_1 \cap \bar{R}'_2 = \bar{R}'_1 \leqq \bar{R}'_2$ by Nakayama's lemma. □

Let X be a normal noetherian scheme and let $U \subset X$ be a dense open subscheme. Let $t \in U$ and let $T = \operatorname{Spec} A \to X$ be a U-external morphism defined by a valuation ring $A \subsetneqq K = k(t)$ of the residue field at a point $t \in U$. Let $\bar{x}$ and $\bar{t}$ be geometric points of T supported on the closed point and on the generic point respectively. Recall that $T_{(\bar{x})}$ denotes the strict localization and that a specialization $\bar{x} \leftarrow \bar{t}$ is a morphism $T_{(\bar{x})} \leftarrow \bar{t}$ of schemes.

Let A' be a valuation ring and let $T' = \operatorname{Spec} A' \to T$ be a faithfully flat morphism. We identify Γ as a subgroup of the value group Γ' of A' by the canonical injection $\Gamma \to \Gamma'$. Let $\bar{x}'$ and $\bar{t}'$ be geometric points of T' above $\bar{x}$ and $\bar{t}$ respectively. We say that a specialization $\bar{x}' \leftarrow \bar{t}'$ is a lifting of $\bar{x} \leftarrow \bar{t}$ if the diagram

$$\begin{array}{ccccccc} \bar{x}' & \longrightarrow & T' & \longleftarrow & T'_{(\bar{x}')} & \longleftarrow & \bar{t}' \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \bar{x} & \longrightarrow & T & \longleftarrow & T_{(\bar{x})} & \longleftarrow & \bar{t} \end{array}$$

is commutative.

We consider a commutative diagram

$$\begin{array}{ccc} X' & \longleftarrow & T' \\ \downarrow & & \downarrow \\ X & \longleftarrow & T \end{array} \tag{3-10}$$

of schemes equipped with an effective Cartier divisor $R' \subset X'$ and a lifting $\bar{x}' \leftarrow \bar{t}'$ to T' of the specialization $\bar{x} \leftarrow \bar{t}$ satisfying the following conditions (i)–(iii):

(i) X' is a normal noetherian scheme of finite type over X such that $U' = U \times_X X' \subset X'$ is a dense open subscheme étale over U.

(ii) $T' = \operatorname{Spec} A' \to X'$ is a U'-external morphism defined by a valuation ring $A' \subsetneqq K' = k(t')$ of the residue field at a point $t' \in U'$ above t such that $A' \cap K = A$.

(iii) $R' \cap U' = \varnothing$ and $v'(R') = \gamma$ in the value group Γ' of A'.

For elements $\alpha \leqq \beta$ of a totally ordered group Γ, let $(\alpha, \beta)_\Gamma \subset \Gamma$ denote the subset $\{\gamma \in \Gamma \mid \alpha < \gamma < \beta\}$. Similarly, we define $(\alpha, \beta]_\Gamma$, $(\alpha, \infty)_\Gamma \subset \Gamma$ etc.

Definition 3.2.3. Let X be a normal noetherian scheme and let $U \subset X$ be a dense open subscheme. Let $t \in U$, let $A \subsetneqq k(t)$ be a valuation ring of the residue field at t and let $T = \operatorname{Spec} A \to X$ be a U-external morphism. Let $\gamma \in (0, \infty)_{\Gamma_{\mathbb{Q}}}$ for $\Gamma_{\mathbb{Q}} = \Gamma \otimes \mathbb{Q}$. Let Y be a quasifinite flat scheme over X such that $V = Y \times_X U \to U$ is étale.

We define a commutative diagram

$$\begin{array}{ccc} F_T^{\infty}(Y/X) & \xrightarrow{\varphi_T^{\gamma+}} & F_T^{\gamma+}(Y/X) \\ {\scriptstyle \varphi_T^{\gamma}}\downarrow & \swarrow & \downarrow \\ F_T^{\gamma}(Y/X) & \longrightarrow & F_T^{0+}(Y/X) \end{array} \tag{3-11}$$

as the inverse limit of

$$\begin{array}{ccc} \overline{Y}'_{\bar{x}'} & \xrightarrow{\varphi_{\bar{x}'}^{R'+}} & F_{\bar{x}'}^{R'+}(Y'/X') \\ {\scriptstyle \varphi_{\bar{x}'}^{R'}}\downarrow & \swarrow & \downarrow \\ F_{\bar{x}'}^{R'}(Y'/X') & \longrightarrow & Y_{\bar{x}} \end{array} \tag{3-12}$$

for commutative diagrams (3-10) satisfying the conditions (i)–(iii) and for $Y' = Y \times_X X'$.

We say that the ramification of Y over X at T is bounded by γ (resp. by $\gamma+$) if $F_T^{\infty}(Y/X) \to F_T^{\gamma}(Y/X)$ (resp. $F_T^{\infty}(Y/X) \to F_T^{\gamma+}(Y/X)$) is an injection.

By Lemma 3.2.2, the limit is a filtered limit.

Lemma 3.2.4. (1) *There exist a commutative diagram* (3-10) *satisfying the conditions* (i)–(iii), *an effective Cartier divisor $R' \subset X'$ satisfying $R' \cap U' = \varnothing$ and $x' \in R'$ such that Y' over X' satisfies* (RF) *for R' at the image $x' \in R'$ of the closed point of T'.*

(2) *For $x' \in R' \subset X'$ satisfying the condition in* (1), *the canonical morphism from* (3-11) *to* (3-12) *is an isomorphism. The diagram* (3-11) *is a diagram of finite sets.*

Proof. (1) By Lemma 3.2.2(1), after replacing X by a normal scheme of finite type over X if necessary, we may assume that there exist an effective Cartier divisor $R \subset X$ such that $v(R) = \gamma$ and a closed immersion $Y \to Q$ over X to a smooth scheme Q over X. Applying Theorem 1.2.5 and the remark following it to $Y \to X$ and to $Q^{(R)} \to X$ and taking the normalizations, we obtain a morphism $X' \to X$ of finite type of normal noetherian schemes satisfying the following properties: The morphism $X' \to X$ is the composition of a blow-up $X^* \to X$ with center supported in $X - U$ and a faithfully flat morphism $X' \to X^*$ of finite type such that $U' = X' \times_X U \to U$ is étale. The normalization of $Y \times_X X'$ is étale over X'. The morphism $Q'^{(R')} \to X'$ is flat and reduced. Hence Y' over X' satisfies the condition (RF) for R'. The morphism $T \to X$ is lifted to $T' \to X'$ by Lemma 1.2.6.

(2) By (1) and Lemma 3.2.2, among commutative diagrams (3-10) those such that the base change $Y' = Y \times_X X'$ over X' satisfies the condition (RF) for R' at x' are cofinal. Hence the assertion follows from Lemma 2.1.8. □

We study functoriality of the construction of $F_T^{\gamma}(Y/X)$ and $F_T^{\gamma+}(Y/X)$. We consider a commutative diagram

$$\begin{array}{ccccccccc} Y' & \longrightarrow & X' & \longleftarrow & T' & \longleftarrow & \bar{x}' & \longleftarrow & \bar{t}' \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ Y & \longrightarrow & X & \longleftarrow & T & \longleftarrow & \bar{x} & \longleftarrow & \bar{t} \end{array} \tag{3-13}$$

together with dense open subschemes $U \subset X$ and $U' \subset U \times_X X' \subset X'$ and $\gamma \in (0, \infty)_{\Gamma_{\mathbb{Q}}}$ and $\gamma' \in (0, \infty)_{\Gamma'_{\mathbb{Q}}}$ satisfying the following properties:

(i) $X' \to X$ is a morphism of normal noetherian schemes.

(ii) $T = \operatorname{Spec} A \to X$ and $T' = \operatorname{Spec} A' \to X'$ are U-external and U'-external morphisms for valuation rings $A \subsetneqq K = k(t)$ and $A' \subsetneqq K' = k(t')$ of the residue fields at $t \in U$ and $t' \in U'$. The morphism $T' \to T$ is faithfully flat.

(iii) $Y \to X$ and $Y' \to X'$ are quasifinite and flat morphisms such that $Y \times_X U \to U$ and $Y' \times_{X'} U' \to U'$ are étale.

(iv) $\gamma \leqq \gamma'$.

(v) $\bar{x}' \leftarrow \bar{t}'$ is a lifting of $\bar{x} \leftarrow \bar{t}$.

Lemma 3.2.5. *We keep the notation above.*

(1) *We have a commutative diagram*

$$\begin{array}{ccccccc} F_{T'}^{\infty}(Y'/X') & \longrightarrow & F_{T'}^{\gamma'+}(Y'/X') & \longrightarrow & F_{T'}^{\gamma'}(Y'/X') & \longrightarrow & F_{T'}^{0+}(Y'/X') \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ F_T^{\infty}(Y/X) & \longrightarrow & F_T^{\gamma+}(Y/X) & \longrightarrow & F_T^{\gamma}(Y/X) & \longrightarrow & F_T^{0+}(Y/X) \end{array} \tag{3-14}$$

of finite sets. Further if $\gamma < \gamma'$, we have an arrow

$$F_T^{\gamma'}(Y'/X') \to F_T^{\gamma+}(Y/X)$$

making the two triangles obtained by dividing the middle square commutative.

(2) *If the left square in (3-13) is cartesian and if $\gamma = \gamma'$, the vertical arrows in (3-14) are bijections.*

Proof. By Lemma 3.2.4(1), we may assume that there exists an effective Cartier divisor $R \subset X$ such that $R \cap U = \varnothing$ and $v(R) = \gamma$ and that Y over X satisfies the condition (RF) for R. Further by Lemma 3.2.4(1) and Lemma 3.2.2(3), we may assume that there exists an effective Cartier divisor $R' \subset X'$ such that $R' \cap U' = \varnothing$, $v'(R') = \gamma$ and $R' \geqq R \times_X X'$ and that Y' over X' satisfies the condition (RF) for R'. Then, by Lemma 3.2.4(2), we may identify $F_T^{\gamma}(Y/X) = F_{\bar{x}}^{R}(Y/X)$, $F_T^{\gamma+}(Y/X) = F_{\bar{x}}^{R+}(Y/X)$ and $F_{T'}^{\gamma'}(Y'/X') = F_{\bar{x}'}^{R'}(Y'/X')$, $F_{T'}^{\gamma'+}(Y'/X') = F_{\bar{x}'}^{R'+}(Y'/X')$.

(1) The assertion now follows from the functoriality of dilatation (3-4).

(2) In the notation above, we may further assume that $R' = R \times_X X'$. Hence the assertion follows from Lemma 2.1.8. □

Let T^h be the henselization at the closed point $x \in T$ and let $t^h \in T^h$ denote the generic point. Then, the absolute Galois group $D_T = \mathrm{Gal}(\bar{t}/t^h)$ acts on the specialization $\bar{x} \leftarrow \bar{t}$ of geometric points of T. Hence the commutative diagram (3-11) admits a canonical action of D_T.

Theorem 3.2.6. *Let the notation be as in Definition 3.2.3. Then, there exist an element $\beta_0 \in (0, \infty)_{\Gamma_{\mathbb{Q}}}$ and finite pairs $(\alpha_i, \beta_i)_{i \in I}$ of elements of $[0, \beta_0]_{\Gamma_{\mathbb{Q}}}$ satisfying the following properties* (i)–(iii):

(i) $[0, \beta_0]_{\Gamma_{\mathbb{Q}}} = \bigcup_{i \in I} [\alpha_i, \beta_i]_{\Gamma_{\mathbb{Q}}}$.

(ii) *For $\gamma > \beta_0$ (resp. $\gamma \geqq \beta_0$), $F_T^{\gamma}(Y/X) \leftarrow F_T^{\infty}(Y/X)$ (resp. $F_T^{\gamma+}(Y/X) \leftarrow F_T^{\infty}(Y/X)$) is an injection.*

(iii) *Let $i \in I$ and $\gamma \in (\alpha_i, \beta_i)_{\Gamma_{\mathbb{Q}}}$. Then, $F_T^{\gamma}(Y/X) \leftarrow F_T^{\beta_i}(Y/X)$ is an injection and $F_T^{\alpha_i+}(Y/X) \leftarrow F_T^{\gamma+}(Y/X)$ is an injection on the image of $F_T^{\infty}(Y/X)$.*

Proof. Since we may take base change, we may assume that $Y \to X$ is finite and that the normalization $\overline{Y} \to X$ is finite étale. Hence by Lemma 2.1.4, we may assume that there exists an effective Cartier divisor $R \subset X$ such that $R \cap U = \varnothing$ and $\overline{Y} \to Y^{(R)}$ is a closed immersion.

Set $\beta_0 = v(R) \in \Gamma$. Then, by Lemma 3.2.4, after replacing X if necessary, we may assume that Y over X satisfies the condition (RF) for R. Since $\overline{Y} \to Y^{(R)}$ is a closed immersion and $F_T^{\infty}(Y/X) \to \overline{Y}_{\bar{x}}$ is a bijection, $\overline{Y}_{\bar{x}} = F_T^{\infty}(Y/X) \to Y_{\bar{x}}^{(R)} = F_T^{\beta_0+}(Y/X)$ is an injection. For $\gamma > \beta_0$, the composition

$$F_T^{\beta_0+}(Y/X) \leftarrow F_T^{\gamma}(Y/X) \leftarrow F_T^{\gamma+}(Y/X) \leftarrow F_T^{\infty}(Y/X)$$

is an injection. Hence the condition (ii) is satisfied.

Let Q be a smooth scheme over X and let $Y \to Q$ be a closed immersion. As in Example 1.3.1, we define a semistable curve $C_R \to X$ by the effective Cartier divisor $R \subset X$. Define an effective Cartier divisor $\widetilde{R} \subset C_R$ to be the exceptional divisor. Applying Corollary 1.3.6 to $(Q \times_X C_R)^{[\widetilde{R}]} \to C_R \to X$ and taking the normalizations, we obtain a commutative diagram

$$\begin{array}{ccc} C_R & \longleftarrow & C' \\ \big\downarrow & & \big\downarrow \\ X & \longleftarrow & X' \end{array}$$

where $Y_{C'} = Y \times_X C'$ over C' satisfies the condition (RF) for $R' = \widetilde{R} \times_{C_R} C'$ and $C' \to X'$ is a semistable curve.

By Lemma 1.2.6, there exist a finite extension K' of K and a valuation ring A' such that $A = A' \cap K$ and that $T \to X$ is lifted to $T' = \operatorname{Spec} A' \to X'$. Let $x' \in X'$ denote the image of the closed point of T'. Further, for $\gamma \in [0, \beta_0]_{\Gamma_{\mathbb{Q}}}$, after replacing K' by a finite extension if necessary, we may assume that γ is an element of $[0, \beta_0]_{\Gamma'}$.

Let I_1 be the set of irreducible components of the fiber $C' \times_{X'} x'$. For $i \in I_1$, let $C_i \subset C' \times_{X'} x'$ denote the corresponding connected component. Let I_2 denote the set of singular points of the fiber $C' \times_{X'} x'$. For $i \in I_2$, let $z_i \subset C' \times_{X'} x'$ denote the corresponding singular point. Set $I = I_1 \amalg I_2$.

Since the assertion is étale local on X', we may assume that for each $i \in I_1$, there exists a section $s_i : X' \to C'$. For $i \in I_1$, set $\alpha_i = \beta_i = v'(s_i^* R') \in \Gamma'^{+}$. Since $\alpha_i = v'(\widetilde{R})$ for the composition $T' \to X' \to C' \to C_R$, we have $\alpha_i \in [0, \beta_0]_{\Gamma_{\mathbb{Q}}}$. For $i \in I_2$, if z_i is contained in two irreducible components C_{i_1} and C_{i_2} such that $\alpha_{i_1} \leqq \alpha_{i_2} \in \Gamma'^{+}$, we define $\alpha_i = \alpha_{i_1} \leqq \beta_i = \alpha_{i_2} \in [0, \beta_0]_{\Gamma_{\mathbb{Q}}}$. If z_i is contained in a unique irreducible component C_{i_1}, we define $\alpha_i = \beta_i = \alpha_{i_1} \in [0, \beta_0]_{\Gamma_{\mathbb{Q}}}$.

We show that the condition (i) is satisfied. Since $\alpha_i, \beta_i \in [0, \beta_0]_{\Gamma_{\mathbb{Q}}}$ for $i \in I$, we have the inclusion

$$[0, \beta_0]_{\Gamma_{\mathbb{Q}}} \supset \textstyle\bigcup_{i \in I} [\alpha_i, \beta_i]_{\Gamma_{\mathbb{Q}}}.$$

Let γ be an element of $[0, \beta_0]_{\Gamma_{\mathbb{Q}}}$. Then, we may assume $\gamma \in [0, \beta_0]_{\Gamma'}$. Then, since $T' \to X$ has a lifting to $T' \to C_R$ such that $v'(\widetilde{R}) = \gamma$ and since $C' \to C_R \times_X X'$ is proper and birational, there exists a unique lifting $T' \to C'$ of $T' \to C_R$ by the valuative criterion. If the image of the closed point by $T' \to C'$ is contained in the smooth part $C_i \cap C_x'^{\mathrm{sm}}$ of an irreducible component $C_i \subset C'_x$ for $i \in I_1$, then we have $\gamma = \alpha_i$. If the image of the closed point by $T \to C'$ is the singular point $z_i \in C'_x$ for $i \in I_2$, then we have $\gamma \in [\alpha_i, \beta_i]_{\Gamma_{\mathbb{Q}}}$ by Corollary 1.3.4. Thus, the condition (i) is also satisfied.

We show that the condition (iii) is satisfied. For $i \in I_1$ or $i \in I_2$ such that $\alpha_i = \beta_i$, there is nothing to prove. Assume that $i \in I_2$ and z_i is contained in two irreducible components C_{i_1} and C_{i_2} such that $\alpha_i = \alpha_{i_1} < \beta_i = \alpha_{i_2} \in \Gamma'^{+}$ and let $\gamma \in (\alpha_i, \beta_i)_{\Gamma_{\mathbb{Q}}}$. Then, we may assume $\gamma \in (\alpha_i, \beta_i)_{\Gamma'}$. By Corollary 1.3.4, after replacing T' by an extension if necessary, we may take a morphism $T' \to C'$ such that the image of the closed point $x' \in T'$ is z_i and $v'(R') = \gamma$. Since $F_T^{\gamma}(Y/X) = F_{\bar{z}_i}^{R'}(Y_{C'}/C')$ and $F_T^{\gamma+}(Y/X) = F_{\bar{z}_i}^{R'+}(Y_{C'}/C')$ by Lemmas 3.2.5(2) and 3.2.4(2), the assertion follows from Proposition 3.1.8. □

We study some variants.

Let X be a normal noetherian scheme, let U be a dense open subscheme and let $V \to U$ be a finite étale morphism. We consider a cartesian diagram

$$\begin{array}{ccc} Y' & \longleftarrow & V \\ \downarrow & \square & \downarrow \\ X' & \longleftarrow & U \end{array} \tag{3-15}$$

of schemes of finite type over X satisfying the following conditions: the horizontal arrows are dense open immersions, X' is normal, $X' \to X$ is a proper birational morphism inducing the identity on U and Y' is finite flat over X'.

Let $A \subset K = k(t)$ be a valuation ring of the residue field at a point $t \in U$ and let $T = \operatorname{Spec} A \to X$ be a U-external morphism. Let $x \in T$ denote the closed point and let $\bar{x}$ be a geometric point above x. For $\gamma \in \Gamma_{\mathbb{Q},>0}$, we define

$$\begin{array}{ccc} F_T^{\infty}(V/U) & \longrightarrow & F_T^{\gamma+}(V/U) \\ \downarrow & \swarrow & \downarrow \\ F_T^{\gamma}(V/U) & \longrightarrow & F_T^{0+}(V/U) \end{array} \tag{3-16}$$

to be the inverse limit of

$$\begin{array}{ccc} F_T^{\infty}(Y'/X') & \longrightarrow & F_T^{\gamma+}(Y'/X') \\ \downarrow & \swarrow & \downarrow \\ F_T^{\gamma}(Y'/X') & \longrightarrow & F_T^{0+}(Y'/X'). \end{array} \tag{3-17}$$

Let T_V denote the normalization of T in $V \times_X T$. For X' in (3-15), let $X'_T \subset X'$ denote the reduced closed subscheme supported on the closure of $t \in U \subset X'$ and let $x' \in X'_T$ denote the image of the unique morphism $T \to X'$ lifting $T \to X$. Then, since $A = \varinjlim_{X' \to X} \mathcal{O}_{X'_T, x}$, we have $F_T^{0+}(V/U) = T_V \times_T \bar{x}$.

Lemma 3.2.7. *Suppose that the normalization T_V of T in $V \times_X T$ is finite and flat over T. Then, there exists a finite and flat $Y' \to X'$ such that $T_V = Y' \times_{X'} T$. For such $Y' \to X'$, the diagram (3-16) is isomorphic to (3-17).*

Proof. Since $A = \varinjlim_{X' \to X} \mathcal{O}_{X'_T, x}$ in the notation above, the existence of finite flat $Y' \to X'$ such that $T_V = Y' \times_{X'} T$ follows. By the flattening theorem [Raynaud and Gruson 1971, théorème (5.2.2)], such $Y' \to X'$ are cofinal among commutative diagrams (3-15). Hence the assertion follows from Lemma 3.2.5(2). □

For a normal noetherian scheme X, a formal $\mathbb{Q}$-linear combination $R = \sum_i r_i D_i$ with positive coefficients $r_i \geqq 0$ of irreducible closed subsets D_i of codimension 1 is called an effective $\mathbb{Q}$-Cartier divisor if a nonzero multiple is an effective Cartier divisor. The union $\bigcup_i D_i$ for $r_i > 0$ is called the support of R. For an open subset $U \subset X$, if U does not meet the support of R, we write $R \cap U = \varnothing$ by abuse of notation. For a U-external morphism $T = \operatorname{Spec} A \to X$, the valuation $v(R)$ is defined as an element of $[0, \infty)_{\Gamma_{\mathbb{Q}}}$.

Definition 3.2.8. Let X be a normal noetherian scheme and let $U \subset X$ be a dense open subscheme. Let Y be a quasifinite flat scheme over X such that $V = Y \times_X U \to U$ is finite étale. Let R be an effective $\mathbb{Q}$-Cartier divisor of X such that $U \cap R$ is empty and let $x \in X$ be a point contained in the support of R.

We say that the ramification of Y over X is bounded by R (resp. by $R+$) at x, if for every U-external morphism $T \to X$, the ramification of $Y \to X$ is bounded by $v(R)$ (resp. by $v(R)+$) in the sense of Definition 3.2.3.

Lemma 3.2.9. *Let the notation be as in Definition 3.2.8. Then, the following conditions* (1), (1′) *and* (2) *are equivalent*:

(1) *The ramification of $Y \to X$ is bounded by R (resp. by $R+$) in the sense of Definition 3.2.8.*

(1′) *The condition in Definition 3.2.8 with T restricted to be a discrete valuation ring is satisfied.*

(2) *For every morphism $f : X' \to X$ of finite type, of normal noetherian schemes such that $U' = U \times_X X' \to U$ is étale, that $R' = f^*R$ is an effective Cartier divisor and that $Y' = Y \times_X X' \to X'$ satisfies the condition (RF) in Definition 3.1.4 for R', the ramification of $Y' \to X'$ is bounded by R' (resp. by $R'+$) at every point of R' in the sense of Definition 3.1.4.*

Proof. (1′)⇒(2): Let $X' \to X$ be as in (2) and let $x' \in R'$ be a point. Let $X'_1 \to X'$ be the normalization of the blow-up at the closure of x'. Then, the local ring $A' = \mathcal{O}_{X'_1, x'_1}$ at the generic point x'_1 of an irreducible component of the inverse image of x' is a discrete valuation ring. The morphism $T' = \operatorname{Spec} A' \to X'_1 \to X'$ is U'-external and the image of the closed point is x'.

For $\gamma' = v'(R')$, by Lemma 3.2.4(2), the commutative diagram (3-12) is canonically identified with

$$\begin{array}{ccc}
F^{\infty}_{T'}(Y'/X') & \xrightarrow{\varphi^{\gamma'+}_{T'}} & F^{\gamma'+}_{T'}(Y'/X') \\
{\scriptstyle \varphi^{\gamma'}_{T'}}\downarrow & \swarrow & \downarrow \\
F^{\gamma'}_{T'}(Y'/X') & \longrightarrow & Y'_{\bar{x}'}
\end{array}$$

Further, this commutative diagram is canonically identified with (3-11) for $\gamma = v(R)$ by Lemma 3.2.5(2). Hence the assertion follows.

(2)$\Rightarrow$(1): Let $T \to X$ be a U-external morphism and let $\gamma = v(R)$. Then by Lemma 3.2.4(2), the commutative diagram (3-11) is canonically identified with (3-12). Hence the assertion follows.

(1) $\Rightarrow$ (1′): This implication is obvious. □

Proposition 3.2.10. *Let the notation be as in Definition 3.2.8 and assume that the ramification of Y over X is bounded by $R+$. Assume that Y is locally of complete intersection over X and let*

$$\begin{array}{ccc} Q & \longleftarrow & Y \\ \downarrow & \square & \downarrow \\ P & \longleftarrow & X \end{array}$$

be a cartesian diagram of schemes over X such that P and Q are smooth over X, the vertical arrows are quasifinite and flat and the horizontal arrows are closed immersions.

Let X' be a normal noetherian scheme over X such that $R' = R \times_X X'$ is an effective Cartier divisor and $Y' = Y \times_X X'$ over X' satisfies the condition (RF) for R'.

Then, the morphism $Q'^{(R')} \to P'^{(R')}$ is étale on a neighborhood of $Q'^{(R')} \times_{X'} R'$.

This implies [Saito 2009, Lemma 1.13 6)$\Rightarrow$4)] since $Q'^{(R')} \times_{X'} R' \to P'^{(R')} \times_{X'} R'$ is finite by Lemma 2.1.7(1).

Proof. First, we show that we may assume that there exist a closed subscheme $Y'_0 \subset Y'$ étale over X', an integer $n \geqq 1$ and an effective Cartier divisor $D'_0 \subset R'$ satisfying the following conditions: We have an equality $R'_{Y'_0} = R'_{Y'}$ of underlying sets. Let $\mathcal{J}'_0 \subset \mathcal{O}_{R'_{Y'}}$ be the nilpotent ideal defining $R'_{Y'_0} \subset R'_{Y'}$. Then, we have $\mathcal{J}'^n_0 = 0$ and $(n+1)D'_0 = R'$.

Under the condition (RF), the formation of $Q'^{(R')} \to P'^{(R')}$ commutes with base change by Lemma 2.1.8 and Example 2.1.2(1). Since $Q'^{(R')}$ and $P'^{(R')}$ are flat over X', the étaleness of $Q'^{(R')} \to P'^{(R')}$ is checked fiberwise. Hence, we may take base change. Let $x' \in R'$ be a point and let $X'' \to X'$ be the normalization of the blow-up at the closure of x'. Then, there exists a point $x'' \in X''$ above x' such that the local ring $\mathcal{O}_{X'',x''}$ is a discrete valuation ring. Hence, by replacing X' by $\operatorname{Spec} \mathcal{O}_{X'',x''}$, we may assume that X' is the spectrum of a discrete valuation ring.

Then, we may assume that $Y' \subset Q'$ is a union of sections $X' \to Q'$. There exists a disjoint union $Y'_0 \subset Y'$ of sections such that we have an equality $R'_{Y'_0} = R'_{Y'}$ of underlying sets. Let $n \geqq 1$ be an integer satisfying $\mathcal{J}'^n_0 = 0$ in the notation above. After replacing X' by a ramified covering if necessary, there exists an effective Cartier divisor D'_0 of X' satisfying $(n+1)D'_0 = R'$.

The finite morphism $Y'^{(R')} \to X'$ is étale by Corollary 3.1.3. Hence by the existence of Y'_0, D'_0 and n and by Lemma 2.2.5, the $\mathcal{O}_{Y'^{(R')}}$-module $\mathcal{O}_{Y'^{(R')}} \otimes_{\mathcal{O}_Y} \Omega^1_{Y/X}$ is annihilated by $\mathcal{I}_{nD'_0}$. Hence by Lemma 2.2.4, there exists an open neighborhood $W_1 \subset Q'^{(R')}$ of $Q'^{(R')} \times_{X'} R'$ such that $Q'^{(R')} \to P'^{(R')}$ is étale on $W_1 - (Q'^{(R')} \times_{X'} R')$.

The morphism $Q'^{(R')} \to P'^{(R')}$ is étale also on a neighborhood W_2 of $Y'^{(R')} \subset Q'^{(R')}$. Since the vector bundle $P'^{(R')} \times_{X'} R' \to R'$ has irreducible fibers, $W_2 \subset Q'^{(R')}$ is dense in the fiber of every point of R' by Proposition 1.1.5(1). Hence the assertion follows from Lemma 1.2.4. □

3.3. *Ramification groups.*

Theorem 3.3.1. *Let X be a connected normal noetherian scheme and let $U \subset X$ be a dense open subscheme. Let G be a finite group, $W \to U$ be a connected G-torsor and let C be the category of finite étale schemes over U trivialized by W. Assume that for every morphism $V_1 \to V_2$ of C, the morphism $Y_1 \to Y_2$ of normalizations of X in V_1 and in V_2 is locally of complete intersection.*

Let $t \in U$ and let $T = \operatorname{Spec} A \to X$ be a U-external morphism for a valuation ring $A \subsetneqq K = k(t)$. Let $\bar{x}$ (resp. $\bar{t}$) be a geometric point above the closed point x (resp. the generic point t) of T and let $\bar{x} \leftarrow \bar{t}$ be a specialization. Fix a lifting of $\bar{x}$ to the normalization T_W of T in $W \times_X T$ and let $I_{\bar{x}} \subset G$ be the inertia group at the image of the lifting of $\bar{x}$ to the normalization Y_W of X in W by $T_W \to Y_W$.

For an object V of C, let Y denote the normalization of X in V and consider the fiber functor sending V to $F^{\infty}_T(Y/X)$.

(1) *There exist decreasing filtrations $G^{\gamma}_T \supset G^{\gamma+}_T$ of G indexed by $\gamma \in (0,\infty)_{\Gamma_{\mathbb{Q}}}$ such that, for every object V of C, the canonical surjections $F^{\infty}_T(Y/X) \to F^{\gamma+}_T(Y/X) \to F^{\gamma}_T(Y/X)$ induce bijections*

$$G^{\gamma+}_T \backslash F^{\infty}_T(Y/X) \to F^{\gamma+}_T(Y/X), \quad G^{\gamma}_T \backslash F^{\infty}_T(Y/X) \to F^{\gamma}_T(Y/X). \tag{3-18}$$

For $I_{\bar{x}} = G^{0+}_T$, the mapping

$$G^{0+}_T \backslash F^{\infty}_T(Y/X) \to F^{0+}_T(Y/X) \tag{3-19}$$

is a bijection.

(2) *There exists a finite increasing sequence $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_n$ of elements of $[0,\infty)_{\Gamma_{\mathbb{Q}}}$ such that we have*

$$\begin{aligned} G^{\alpha_{i-1}+} = G^{\gamma} = G^{\gamma+} = G^{\alpha_i} \quad & \text{for } \gamma \in (\alpha_{i-1},\alpha_i)_{\Gamma_{\mathbb{Q}}},\ 1 \leqq i \leqq n, \\ G^{\alpha_n+} = G^{\gamma} = G^{\gamma+} = 1 \quad & \text{for } \gamma \in (\alpha_n,\infty)_{\Gamma_{\mathbb{Q}}}. \end{aligned} \tag{3-20}$$

(3) *Let $D_T \subset G$ be the decomposition group of T in $W \times_X T$. Then, D_T normalizes G^{γ} and $G^{\gamma+}$.*

Proof. (1) Let $V' \to V$ be a morphism in the category C and let $Y' \to Y$ be the morphism of normalizations of X. By Proposition 3.1.2 and Lemma 3.2.4(2), the diagram

$$\begin{array}{ccccccc} F_T^{\infty}(Y'/X) & \longrightarrow & F_T^{\gamma+}(Y'/X) & \longrightarrow & F_T^{\gamma}(Y'/X) & \longrightarrow & F_T^{0+}(Y'/X) \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ F_T^{\infty}(Y/X) & \longrightarrow & F_T^{\gamma+}(Y/X) & \longrightarrow & F_T^{\gamma}(Y/X) & \longrightarrow & F_T^{0+}(Y/X) \end{array} \tag{3-21}$$

is a cocartesian diagram of surjections. Further, the functors F_T^{γ} and $F_T^{\gamma+}$ preserve disjoint unions. Hence by Proposition 1.4.2, we obtain filtrations $(G_T^{\gamma})_{\gamma}$ and $(G_T^{\gamma+})_{\gamma}$ indexed by $\gamma \in (0, \infty)_{\Gamma_{\mathbb{Q}}}$ characterized by the bijections (3-18). For $\gamma = 0$, the bijection (3-19) follows from $F_T^{0+}(Y/X) = Y_{\bar{x}}$.

(2) Since C has only finitely many connected objects and

$$F_T^{\infty}(Y/X) \to F_T^{\gamma+}(Y/X) \to F_T^{\gamma}(Y/X)$$

are surjections, the claim follows from Theorem 3.2.6.

(3) Since the surjections $F_T^{\infty}(Y/X) \to F_T^{\gamma+}(Y/X) \to F_T^{\gamma}(Y/X)$ are compatible with the actions of $D_T \subset G$, the subgroup $D_T \subset D_{\bar{x}}$ normalizes G^{γ} and $G^{\gamma+}$ by Corollary 1.4.4. □

By the definition of the filtrations, the ramification of Y/X at T is bounded by γ (resp. by $\gamma+$) if and only if the action of G_T^{γ} (resp. of $G_T^{\gamma+}$) on $F_T^{\infty}(Y/X)$ is trivial. By Corollary 1.4.3, the filtrations (G^{γ}) and $(G^{\gamma+})$ are compatible with quotients. We have the following functoriality. Let

$$\begin{array}{ccc} X' & \longleftarrow & T' \\ \downarrow & & \downarrow \\ X & \longleftarrow & T \end{array}$$

be a commutative diagram of schemes. Assume that $X' \to X$ is a morphism of normal connected noetherian schemes and let $U' \subset U \times_X X' \subset X'$ be a dense open subscheme. The horizontal arrows $T \to X$ and $T' \to X'$ are U-external and U'-external and the vertical arrow $T' \to T$ is faithfully flat. Let W' be a connected G'-torsor over U' for a finite group G' and let $W' \to W$ be a morphism over $U' \to U$ compatible with a morphism $G' \to G$ of finite groups. Assume that $W' \to U'$ satisfies the complete intersection property as in Theorem 3.3.1 and let $(G'^{\gamma'})$ and $(G'^{\gamma'+})$ be the filtrations of G' indexed by $\gamma' \in (0, \infty)_{\Gamma'_{\mathbb{Q}}}$. Then, for $\gamma \in (0, \infty)_{\Gamma_{\mathbb{Q}}}$, the morphism $G' \to G$ induces

$$G'^{\gamma} \to G^{\gamma}, \quad G'^{\gamma+} \to G^{\gamma+} \tag{3-22}$$

by the functoriality Lemma 3.2.5(1).

We consider a variant. Let $A \subsetneqq K$ be a valuation ring and let L be a finite Galois extension of K of Galois group G. We define a filtration of G by ramification groups under the following assumptions: For every intermediate extension $K \subset M \subset L$, the normalization A_M of A in M is a valuation ring finite flat and of complete intersection over A. There exist an irreducible normal noetherian scheme X such that K is the residue field at the generic point t and a morphism $T = \operatorname{Spec} A \to X$ extending $t \to X$.

Let $T \to X$ be as above and let M be an intermediate extension. Then by Lemma 3.2.2(1), Spec A_M is the limit of $T \times_{X'} Y'$ for normal schemes X' of finite type over X equipped with a lifting $T \to X'$ of $T \to X$ and finite flat schemes $Y' \to X'$ such that $U' = U \times_X X' \to U$ is an isomorphism for a dense open subscheme $U \subset X$ and $U' \times_{X'} Y' \to U'$ is a finite étale covering corresponding to M. Since A_M is assumed to be finite flat locally of complete intersection over A, there exists a finite flat scheme $Y'_M \to X'$ locally of complete intersection such that $U' \times_{X'} Y'_M \to U'$ is finite étale and $T \times_{X'} Y'_M = \operatorname{Spec} A_M$.

Thus, there exist a dense open subscheme $U \subset X$ and a normal scheme X' of finite type over X satisfying the following conditions: The morphism $U' = U \times_X X' \to U$ is an isomorphism. The morphism $T \to X$ is lifted to $T \to X'$. For every intermediate extension M, there exists a finite flat scheme $Y'_M \to X'$ locally of complete intersection such that $U' \times_{X'} Y'_M \to U'$ is finite étale and $T \times_{X'} Y'_M = \operatorname{Spec} A_M$.

Then applying Theorem 3.3.1, we obtain filtrations (G_T^{γ}) and $(G_T^{\gamma+})$ by normal subgroups of $G = D_T$ indexed by $(0, \infty)_{\Gamma_{\mathbb{Q}}}$.

In the rest of the article, we consider the case where $X = T = \operatorname{Spec} \mathcal{O}_K$ for a complete discrete valuation ring $\mathcal{O}_K$. For a finite Galois extension of the fraction field K of the Galois group G, the decreasing filtrations $(G^r)_{r>0}$ and $(G^{r+})_{r\geqq 0}$ by normal subgroups indexed by rational numbers are defined.

Lemma 3.3.2. *Let K be a complete discrete valuation field and let L be a finite Galois extension of the Galois group $G = \operatorname{Gal}(L/K)$. Then, the filtration by ramification groups of G defined in [Abbes and Saito 2002] is the same as that defined here.*

Proof. Let M be an intermediate extension and let $Y = \operatorname{Spec} \mathcal{O}_M \to Q = \mathbb{A}^n_{\mathcal{O}_K}$ be a closed immersion defined by taking a system of generators of $\mathcal{O}_M$ over $\mathcal{O}_K$ as in Example 2.1.1(1). Then, the affinoid varieties used in the definition in [Abbes and Saito 2002] are the generic fibers of the formal completions of dilatations of $Q^{(r)}$. Since the geometric connected components of the affinoid varieties are canonically identified with those of the closed fibers as in Remark 1.1.2, the assertion follows. □

Let L be a finite separable extension of degree n of K and let $Y = \operatorname{Spec} \mathcal{O}_L$ for the integer ring $\mathcal{O}_L$. We recall the classical case where $\mathcal{O}_L$ is generated by one element over $\mathcal{O}_K$, using the Herbrand function. Take a closed immersion

$$Y = \operatorname{Spec} \mathcal{O}_L \to Q = \mathbb{A}^1_X = \operatorname{Spec} \mathcal{O}_K[T],$$

and let $P \in \mathcal{O}_K[T]$ be the monic polynomial such that we have an isomorphism $\mathcal{O}_K[T]/(P) \to \mathcal{O}_L$.

Let K' be a finite separable extension containing the Galois closure of L and let $X' = \operatorname{Spec} \mathcal{O}_{K'}$. Let $v' : K' \to \mathbb{Q} \cup \{\infty\}$ be the valuation extending the normalized valuation of K. Let $r > 0$ be a rational number in the image of v' and let $R' \subset X'$ be the effective Cartier divisor such that $v'(R') = r$. Let $Q' \supset Y'$ be the base change of $Q \supset Y$ by $X' \to X$ and let $Q'^{(r)} = Q'^{(R')}$ denote the dilatation. We compute $Q'^{(r)}$ using the Herbrand function, whose definition we briefly recall.

Decompose P as $P = \prod_{i=1}^n (T - a_i)$ in $\mathcal{O}_{K'}[T]$ and set $b_i = a_i - a_n \in \mathcal{O}_{K'}$. Set $P(T_1 + a_n) = \prod_{i=1}^n (T_1 - b_i) = T_1^n + c_1 T_1^{n-1} + \cdots + c_{n-1} T_1$ in $\mathcal{O}_{K'}[T_1]$. Changing the numbering if necessary, we assume that the valuations $s_i = v'(b_i) \in \mathbb{Q}$ are increasing in i. Note that the increasing sequence $s_0 = 0 \leqq s_1 \leqq \cdots \leqq s_{n-1} < s_n = \infty$ is independent of the choice of a_n. The valuation $v'(c_{n-1}) = \sum_{k=1}^{n-1} s_k$ equals the valuation $v'(D_{L/K})$ of the different $D_{L/K}$. It is further equal to the length of the $\mathcal{O}_L$-module $\Omega^1_{\mathcal{O}_L/\mathcal{O}_K}$ divided by the ramification index $e_{L/K}$ by [Serre 1968, Chapter III §6 corollaire 2 à la proposition 11].

The largest piecewise linear convex continuous function

$$p : [0, n-1] \to [0, v'(D_{L/K})]$$

such that the graph is below the points $(0, 0)$ and $(k, v(c_k))$ for $k = 1, \ldots, n-1$ is defined by

$$p(x) = \sum_{i=1}^{k-1} s_i + s_k(x - k + 1) \tag{3-23}$$

on $[k-1, k]$ for $k = 1, \ldots, n-1$. The graph of p is the *Newton polygon* of the polynomial $P(T_1 + a_n)$. The *Herbrand function* $\varphi : [0, \infty) \to [0, \infty)$ is a piecewise linear concave continuous function defined by

$$\varphi(s) = \sum_{i=1}^{n-1} \min(s_i, s) + s. \tag{3-24}$$

We have

$$\varphi(s) = \sum_{i=1}^{k-1} (n - i + 1) \cdot (s_i - s_{i-1}) + (n - k + 1) \cdot (s - s_{k-1}) \tag{3-25}$$

on $[s_{k-1}, s_k)$ for $k = 1, \ldots, n$.

Example 3.3.3. Let $s \in (s_{k-1}, s_k]_{\mathbb{Q}}$, $r = \varphi(s)$ and let t be an element of a finite separable extension K' of K such that $v(t) = s$. By (3-25) and Example 2.1.2, $Q'^{(r)}$ is obtained as an iterated dilatation defined inductively by $Q'_0 = Q'$,

$$Q'_i = Q'^{((n-(i-1))\cdot(s_i - s_{i-1}))}_{i-1} \quad \text{for } 0 < i < k \qquad \text{and} \qquad Q'^{(r)} = Q'^{((n-(k-1))\cdot(s-s_{k-1}))}_{k-1}.$$

Hence $Q'^{(r)} \to X'$ is smooth. Let $C \subset Q'^{(r)} \times_{X'} x'$ be the connected component meeting the section $s'_n : X' \to Q'^{(r)}$ lifting $s_n : X \to Q$ defined by $T = a_n$. Then, $\operatorname{Spec} \mathcal{O}_{K'}[T']$ for $T' = T_1/t$ is a neighborhood of $C \subset Q'^{(r)}$. Further, on $\operatorname{Spec} \mathcal{O}_{K'}[T']$, the closed subscheme $Y'^{(r)} \subset Q'^{(r)}$ is defined by $\prod_{i=k}^{n} (T' - b_i/t)$.

Consequently, the surjections

$$\overline{Y}_{\bar{x}} = \{a_1, \dots, a_n\} \to F^r_X(Y/X)$$

and

$$\overline{Y}_{\bar{x}} \to F^{r+}_X(Y/X)$$

are given by the equivalence relations $v'(a_i - a_j) \geqq s$ and $v'(a_i - a_j) > s$, respectively. In particular, $r_{L/K} = \varphi(s_{n-1}) = v'(D_{L/K}) + s_{n-1}$ is the unique rational number r such that the ramification of Y over X is bounded by $r+$ but not by r.

We give a slightly simplified proof of the proposition below giving characterizations of unramified extensions and tamely ramified extensions.

Lemma 3.3.4 [Serre 1968, chapitre III §7 proposition 13; Abbes and Saito 2002, Proposition A.3]. *Let L be a finite separable extension of a complete discrete valuation field K. Assume that $\mathcal{O}_L$ is generated by one element over $\mathcal{O}_K$ and let $r_{L/K} = \varphi(s_{n-1}) = v'(D_{L/K}) + s_{n-1}$ be as in Example 3.3.3.*

(1) *The following conditions are equivalent*:
   (i) *L is an unramified extension of K.*
   (ii) $r_{L/K} = 0$.
   (iii) $r_{L/K} < 1$.

(2) *The following conditions are equivalent*:
   (i) *L is a tamely ramified extension of K.*
   (ii) $r_{L/K} = 0$ *or* 1.
   (iii) $r_{L/K} \leqq 1$.

Proof. By [Abbes and Saito 2002, Proposition A.3], we have $v'(D_{L/K}) \geqq 1 - 1/e_{L/K}$ and equality holds if and only if L/K is tamely ramified. We have $s_{n-1} \geqq 0$ and equality holds if and only if L is unramified. If L is ramified, we have $s_{n-1} \geqq 1/e_{L/K}$ and equality holds if and only if L is tamely ramified. The assertions follows from these observations. □

Proposition 3.3.5 [Abbes and Saito 2002, Proposition 6.8]. *Let L be a finite separable extension of a complete discrete valuation field K.*

(1) *The following conditions are equivalent*:
   (i) *L is an unramified extension of K.*
   (ii) *The ramification of L over K is bounded by* 1.

(2) *The following conditions are equivalent*:
   (i) *L is a tamely ramified extension of K.*
   (ii) *The ramification of L over K is bounded by* $1+$.

Proof. For both (1) and (2), (i)$\Rightarrow$(ii) follows from Example 3.3.3 and Lemma 3.3.4, since $\mathcal{O}_L$ is generated by one element over $\mathcal{O}_K$.

We show (ii)$\Rightarrow$(i).

(1) Let L be a finite separable extension such that the ramification over K is bounded by 1 and assume that L was ramified over K.

Let G be the Galois group of a Galois closure of L over K and let $1 \subsetneqq I \subset G = \mathrm{Gal}(L/K)$ be the inertia subgroup. By replacing K and L by the subextensions corresponding to I and to a maximal subgroup $H \subsetneqq I$, we may assume that L is a cyclic extension of prime degree since I is solvable.

Then, either the ramification index $e_{L/K}$ is 1 and the residue extension is a purely inseparable extension of degree p or L is totally ramified extension. Hence $\mathcal{O}_L$ is generated by one element and the assertion follows from Example 3.3.3 and Lemma 3.3.4.

(2) If the integer ring $\mathcal{O}_L$ is generated by one element over $\mathcal{O}_K$, the assertion follows from Example 3.3.3 and Lemma 3.3.4. We prove the general case by reducing to this case by contradiction.

Let L be a finite separable extension such that the ramification over K is bounded by $1+$ and assume that L was wildly ramified over K.

Let G be the Galois group of a Galois closure of L over K and let $1 \subsetneqq P \subset I \subset G = \mathrm{Gal}(L/K)$ be the wild inertia subgroup and the inertia subgroup. By replacing K and L by the subextensions corresponding to I and to a maximal subgroup $H \subsetneqq P$, we may assume that $[L : K] = mp$ for an integer m prime to p.

Since an algebraic closure $\widetilde{F}$ of the residue field F of K is a perfect closure of the separable closure, we may construct a henselian separable algebraic extension $\widetilde{K}$ of ramification index 1 of residue field $\widetilde{F}$ as a limit $\varinjlim K_\lambda$ of finite separable extensions of ramification index 1. Since the composition $L\widetilde{K}$ is a totally ramified extension of $\widetilde{K}$, there exists a finite separable extension $K' = K_\lambda$ of ramification index 1 such that $L' = LK_\lambda$ is a totally ramified extension of K'.

By the functoriality (3-22), the ramification of L' over K' is bounded by $1+$. Since L' is totally ramified over K', the integer ring $\mathcal{O}_{L'}$ is generated by one element over $\mathcal{O}_{K'}$. Hence, L' is tamely ramified over K' and we have $[L' : K'] = m$.

By construction, there exists a sequence $K \subset K_0 \subset K_1 \subset \cdots \subset K_n = K'$ such that K_0 is an unramified extension of K and that K_i is an extension of K_{i-1} of degree p of ramification index 1 with inseparable residue field extension for each $i = 1, \ldots, n$. Since $[LK_0 : K_0] = mp$, we have $n > 0$. By taking the smallest such n, we may assume $[LK_{n-1} : K_{n-1}] = mp$.

Further, by the functoriality (3-22), we may replace K and L by K_{n-1} and LK_{n-1}. Hence, we may assume that $[K' : K] = p$ and $K' \subset L$. Since $[K' : K] = p$, the integer ring $\mathcal{O}_{K'}$ is generated by one element over $\mathcal{O}_K$. Since $K' \subset L$, the ramification of K' over K is bounded by $1+$. Hence K' is tamely ramified over K. This contradicts the assumption that the residue field extension of K' over K is inseparable. □

Acknowledgements

The author heartily thanks Ahmed Abbes for collaboration on ramification theory and encouragement on writing this article. He also thanks an anonymous referee for careful reading and helpful comments. This work is supported by JSPS Grants-in-Aid for Scientific Research (A) 26247002.

References

[Abbes and Saito 2002] A. Abbes and T. Saito, "Ramification of local fields with imperfect residue fields", *Amer. J. Math.* **124**:5 (2002), 879–920. MR Zbl

[Abbes and Saito 2011] A. Abbes and T. Saito, "Ramification and cleanliness", *Tohoku Math. J.* (2) **63**:4 (2011), 775–853. MR Zbl

[Bosch et al. 1995] S. Bosch, W. Lütkebohmert, and M. Raynaud, "Formal and rigid geometry, IV: The reduced fibre theorem", *Invent. Math.* **119**:2 (1995), 361–398. MR Zbl

[Bourbaki 1985] N. Bourbaki, *Éléments de mathématique: Algébre commutative, chapitres 5 á 7*, Masson, Paris, 1985. MR Zbl

[EGA IV$_1$ 1964] A. Grothendieck and J. A. Dieudonné, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, I", *Inst. Hautes Études Sci. Publ. Math.* **20** (1964), 5–259. MR Zbl

[EGA IV$_2$ 1965] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, II", *Inst. Hautes Études Sci. Publ. Math.* **24** (1965), 5–231. MR Zbl

[EGA IV$_3$ 1966] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III", *Inst. Hautes Études Sci. Publ. Math.* **28** (1966), 5–255. MR Zbl

[EGA IV$_4$ 1967] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, IV", *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 5–361. MR Zbl

[Fujiwara and Kato 2018] K. Fujiwara and F. Kato, *Foundations of rigid geometry, I*, European Mathematical Society, Zürich, 2018. MR Zbl

[Raynaud and Gruson 1971] M. Raynaud and L. Gruson, "Critères de platitude et de projectivité: techniques de "platification" d'un module", *Invent. Math.* **13** (1971), 1–89. MR Zbl

[Saito 2009] T. Saito, "Wild ramification and the characteristic cycle of an l-adic sheaf", *J. Inst. Math. Jussieu* **8**:4 (2009), 769–829. MR Zbl

[Serre 1968] J.-P. Serre, *Corps locaux*, 2ième ed., Publications de l'Université de Nancago **VIII**, Hermann, Paris, 1968. MR Zbl

[SGA 1 1971] A. Grothendieck, *Revêtements étales et groupe fondamental* (Séminaire de Géométrie Algébrique du Bois Marie 1960–1961), Lecture Notes in Math. **224**, Springer, 1971. MR Zbl

[SGA $4\frac{1}{2}$ 1977] P. Deligne, "Cohomologie étale, les points de départ", pp. 4–75 in *Cohomologie étale* (Séminaire de Géométrie Algébrique du Bois Marie), edited by P. Deligne, Lecture Notes in Math. **569**, Springer, 1977. MR Zbl

[SGA 7_{II} 1973] P. Deligne, "La formule de Picard–Lefschetz", exposé XV, pp. 165–196 in *Groupes de monodromie en géométrie algébrique, II: Exposés X–XXII* (Séminaire de Géométrie Algébrique du Bois Marie 1967–1969), edited by P. Deligne and N. Katz, Lecture Notes in Math. **340**, Springer, 1973. MR Zbl

[Temkin 2011] M. Temkin, "Relative Riemann–Zariski spaces", *Israel J. Math.* **185** (2011), 1–42. MR Zbl

Received 8 Jan 2018. Revised 8 May 2018.

TAKESHI SAITO:

t-saito@ms.u-tokyo.ac.jp
Department of Mathematical Sciences, University of Tokyo, Japan

TUNISIAN JOURNAL OF MATHEMATICS
Vol. 1, No. 3, 2019
dx.doi.org/10.2140/tunis.2019.1.427

Almost sure local well-posedness for the supercritical quintic NLS

Justin T. Brereton

This paper studies the quintic nonlinear Schrödinger equation on $\mathbb{R}^d$ with randomized initial data below the critical regularity $H^{(d-1)/2}$ for $d \geq 3$. The main result is a proof of almost sure local well-posedness given a Wiener randomization of the data in H^s for $s \in \left(\frac{1}{2}(d-2), \frac{1}{2}(d-1)\right)$. The argument further develops the techniques introduced in the work of Á. Bényi, T. Oh and O. Pocovnicu on the cubic problem. The paper concludes with a condition for almost sure global well-posedness.

1. Introduction

Consider the Cauchy problem for the nonlinear Schrödinger equation. Given initial data $\phi \in H^s(\mathbb{R}^d)$, for $(t,x) \in \mathbb{R} \times \mathbb{R}^d$ the solution $u(t,x) \in \mathbb{C}$ satisfies

$$\begin{aligned} iu_t + \Delta u &= \pm |u|^{p-1} u, \\ u|_{t=0} &= \phi, \end{aligned} \tag{1}$$

where $+$ and $-$ correspond to the defocusing and focusing cases, respectively. This equation has conserved mass and energy

$$M(t) = \frac{1}{2} \int_{\mathbb{R}^d} |u(t,x)|^2 \, dx,$$

$$E(t) = \frac{1}{2} \int_{\mathbb{R}^d} |\nabla u(t,x)|^2 \, dx \pm \frac{1}{p+1} \int_{\mathbb{R}^d} |u(t,x)|^{p+1} \, dx.$$

The NLS equation is also invariant under a dilation symmetry. Given $u(t,x)$ that solves (1), the function $u_\lambda(t,x) = \lambda^{2/(p-1)} u(\lambda^2 t, \lambda x)$ is a solution with rescaled initial data for every λ. Furthermore there is a Sobolev index $s_c = \frac{d}{2} - \frac{2}{p-1}$ such that the homogeneous Sobolev norm $\|u_\lambda\|_{\dot{H}^{s_c}}$ is constant under this scaling. This index s_c is known as the scaling critical index, and when $\frac{d}{2} - \frac{2}{p-1} = s_c = 1$ the problem is known as energy critical, since the energy scales like $\dot{H}^{s_c} = \dot{H}^1$.

MSC2010: primary 35K55; secondary 35R60.
Keywords: almost sure well-posedness, supercritical, dispersive, PDEs, NLS equation.

Given initial data $\phi \in H^s(\mathbb{R}^d)$, the problem is called subcritical when $s > s_c$ and supercritical when $s < s_c$.

In addition, special pairs of exponents (q, r) satisfying the bounds $2 \le q, r \le \infty$ and $(q, r, d) \neq (2, \infty, 2)$ are called Schrödinger-admissible if

$$\frac{2}{q} + \frac{d}{r} = \frac{d}{2}. \tag{2}$$

For such a pair we have the well known Strichartz estimate

$$\|S(t)\phi\|_{L_t^q L_x^r(\mathbb{R}\times\mathbb{R}^d)} \le C\|\phi\|_{L^2(\mathbb{R}^d)}, \tag{3}$$

where $S(t)$ denotes the linear Schrödinger semigroup operator $e^{it\Delta}$ that corresponds to solving the linear Schrödinger equation for time t; see [Strichartz 1977; Yajima 1987].

It is known that the NLS equation is ill-posed in the supercritical case; for such s one can construct special initial data $\phi \in H^s(\mathbb{R}^d)$ such that for every $T > 0$, (1) has no solution on $(-T, T)$ that stays in $H^s(\mathbb{R}^d)$, as demonstrated in [Alazard and Carles 2009; Christ et al. 2003]. Though local well-posedness is not guaranteed, it is important to determine if there are solutions for most supercritical initial data ϕ. This leads one to investigate the problem of almost sure well-posedness for initial data chosen for supercritical randomized initial data. Pocovnicu, Bényi and Oh have proven almost sure local well-posedness for the energy critical $\mathbb{R}^4$ problem using $X^{s,b}$ spaces in [Bényi et al. 2015a]. They then proved a separate result for the cubic equation for all $d \ge 3$ using U^p and V^p spaces and their adaptations for the Schrödinger equation in [Bényi et al. 2015b].

In this paper we adapt the techniques of [Bényi et al. 2015a; 2015b] in order to prove local well-posedness in the quintic case for dimension $d \ge 3$. Following [Bényi et al. 2015a], we apply a Wiener randomization to the initial data $\phi \in H^s(\mathbb{R}^d)$. This randomization method takes a function $\phi \in H^s(\mathbb{R}^d)$ and for each ω in a probability space Ω produces a randomized function

$$\phi^\omega = \sum_{n\in\mathbb{Z}^d} g_n(\omega)\eta(D-n)\phi \tag{4}$$

that is in $H^s(\mathbb{R}^d)$ with probability 1 but gains regularity with probability 0. The $g_n(\omega)$ are mean zero, i.i.d. complex random variables that are required to satisfy a decay condition, the Gaussian being such a random variable. The term $\eta(D-n)$ is a Fourier multiplier whose symbol approximates the characteristic function of the unit cube centered at n in frequency space.

In Section 2 we present several previously known probabilistic bounds on the Wiener randomization ϕ^ω of $\phi \in H^s(\mathbb{R}^d)$ as well as its linear Schrödinger evolution $S(t)\phi^\omega$. One of these is a probabilistic bound on $\|\langle\nabla\rangle^s S(t)\phi^\omega\|_{L^q L^r(I\times\mathbb{R}^d)}$

for arbitrarily large values of q, r. For large enough values of q, r this is a norm that scales subcritically, which means we can approach almost sure local well-posedness as if it is a subcritical problem.

Our main result is the almost sure local well-posedness of (1) with initial data ϕ^ω chosen via the Wiener randomization of any $\phi \in H^s(\mathbb{R}^d)$:

Theorem 1.1. *Fix $d \geq 3$ and $s \in \left(\frac{1}{2}(d-2), \frac{1}{2}(d-1)\right)$. Given $\phi \in H^s(\mathbb{R}^d)$ with Wiener randomization ϕ^ω, $\omega \in \Omega$, the quintic nonlinear Schrödinger equation is almost surely locally well-posed. More specifically, there exist $c_1, c_2, \theta > 0$ such that for sufficiently small $T \ll 1$, there is a set $\Omega_T \subset \Omega$ such that*

$$P(\Omega_T) \geq 1 - c_1 e^{-c_2/T^\theta \|\phi\|_{H^s}}$$

and for each $\omega \in \Omega_T$, the initial value problem

$$iu_t + \Delta u = \pm |u|^4 u,$$
$$u(0) = \phi^\omega$$

has a unique solution in the function class $C((-T, T) \to H^s(\mathbb{R}^d))$.

We now provide a brief outline of the proof. In Section 3 we define the Littlewood–Paley projection operator, as well as the U^2 and V^2 spaces and their Schrödinger analogues developed by Koch, Tataru and others in [Hadac et al. 2009; Herr et al. 2011]. In Section 4 we present Strichartz estimates as well as a bilinear estimate for these spaces. The next step is to split the NLS solution u into its linear part $z(t) = S(t)\phi^\omega$ and nonlinear part

$$v(t) = \pm \int_0^t -iS(t-t')\big[|v+z|^4(v+z)\big](t')\,dt', \tag{5}$$

the integral term of Duhamel's formula. Our probabilistic bounds tells us that z almost surely has the same regularity as the initial data ϕ^ω. Therefore the linear part of the solution is almost surely in the supercritical space $H^s(\mathbb{R}^d)$, and it remains to prove existence of the nonlinear part $v(t)$. As mentioned earlier, $z(t)$ is bounded in subcritical norms, which means we can treat our linear solution $z(t)$ as a subcritical perturbative term in the Cauchy problem

$$\begin{aligned} iv_t + \Delta v &= \pm(v+z)|v+z|^4, \\ v(0) &= 0 \end{aligned} \tag{6}$$

that is satisfied by the nonlinear part v.

This means almost sure local well-posedness of $v(t)$ is essentially a subcritical problem. We prove local existence of the nonlinear part $v(t)$ using a fixed point argument based on doing a frequency decomposition of $v(t)$ and bounding it at each frequency.

Global well-posedness is a much harder problem. There is yet to be a proof of almost sure global well-posedness of any supercritical NLS problem. Pocovnicu, Bényi and Oh proved almost sure global well-posedness of $v \in H^1(\mathbb{R}^4)$ for the cubic problem under the assumption that there is a probabilistic bound on $\|v\|_{L^\infty H^1(\mathbb{R}\times\mathbb{R}^4)}$ in [Bényi et al. 2015b]. It seems difficult to prove such a bound.

One could probably prove a similar result for the 3-dimensional quintic problem, the energy critical dimension for the quintic problem. Instead we prove almost sure global well-posedness of v in the subcritical space $S^{1+c}(\mathbb{R}\times\mathbb{R}^3)$ assuming the norm $\|v\|_{L^{10}L^{10}([-T,T]\times\mathbb{R}^3)}$ does not blow up in finite time. This means that a probabilistic a priori estimate for $\|v\|_{L^{10}L^{10}([-T,T]\times\mathbb{R}^3)}$ implies almost sure global well-posedness as expressed in the following result:

Theorem 1.2. *Assume* $\frac{7}{8} < s < 1$ *and* $0 < c < \frac{1}{8}$. *Suppose we have a probabilistic a priori estimate for* $\|v\|_{L^{10}L^{10}([-T,T]\times\mathbb{R}^3)}$, *meaning for every* $T, R > 0$ *there is a function* $\alpha(T, R)$ *and a set* $\Omega'_{T,R}$ *such that*

- *for any* $\omega \in \Omega'_{T,R}$, *if the solution* $v(t)$ *to* (6) *exists on* $(-T, T)$ *then we have the bound*
$$\|v\|_{L^{10}L^{10}([-T,T]\times\mathbb{R}^3)} < R;$$
- $P(\Omega'_{T,R}) \geq 1-\alpha(T, R)$;
- *for all* $T > 0$, $\lim_{R\to\infty} \alpha(T, R) = 0$.

Then given $\phi \in H^s(\mathbb{R}^3)$ *with Wiener randomization* ϕ^ω, *the initial value problem*

$$iu_t + \Delta u = \pm|u|^4 u,$$
$$u(0) = \phi^\omega$$

is almost surely globally well-posed, meaning there is a set $\Omega_{T,R} \subset \Omega$ *and constants* $c_1, c_2, c_3 > 0$ *such that*

$$P(\Omega_{T,R}) \geq 1 - c_1 e^{-c_2 R^2} - c_3 \alpha(T, R)$$

and for any $\omega \in \Omega_{T,R}$ *the above equation has a unique solution in the function class* $C((-T, T) \to H^s(\mathbb{R}^d))$ *with* $v(t) \in H^{1+c}(\mathbb{R}^3)$ *for any time* $t \in (-T, T)$.

2. Randomization of initial data and probabilistic estimates

Our method of randomization is the Wiener decomposition of the frequency space that was used in [Bényi et al. 2015a] and first introduced in [Zhang and Fang 2012]. Consider a Schwartz class function $\psi \in \mathcal{S}(\mathbb{R}^d)$ that approximates the cube of unit length centered at the origin in $\mathbb{R}^d$, meaning that ψ is supported on $[-1, 1]^d$ and $\sum_{n\in\mathbb{Z}^d} \psi(\xi - n)$ is identically 1. Then for each n, define the Fourier multiplier η as

$$\eta(D-n)u(x) = \mathcal{F}^{-1}[\psi(\xi-n)\mathcal{F}u]. \tag{7}$$

Note that this satisfies $\sum_{n\in\mathbb{Z}^d}\eta(D-n)u(x)=u(x)$. This provides a decomposition of the function u into pieces whose frequencies are localized to cubes.

The idea is then to consider a function $\phi\in H^s(\mathbb{R}^d)$ and for each ω from a probability space Ω create a randomized function $\sum_{n\in\mathbb{Z}^d} g_n(\omega)\eta(D-n)\phi$ for some random variables g_n. For each $n\in\mathbb{Z}^d$, let μ_n and ν_n be probability distributions on $\mathbb{R}$, symmetric about 0, such that for some constant c we have

$$\left|\int_{\mathbb{R}^d} e^{\lambda x}\,d\mu_n(x)\right|\le e^{c\lambda^2}\quad\text{and}\quad\left|\int_{\mathbb{R}^d} e^{\lambda x}\,d\nu_n(x)\right|\le e^{c\lambda^2} \tag{8}$$

for all $n\in\mathbb{Z}^d$, $\lambda\in\mathbb{R}$. A Gaussian random variable would be an example of a random variable with these properties. Then define each g_n to be an independent, mean zero, complex random variable on Ω such that $\mathrm{Re}(g_n)$ and $\mathrm{Im}(g_n)$ have distributions μ_n and ν_n. We define the Wiener randomization ϕ^ω of $\phi\in H^s(\mathbb{R}^d)$ to be

$$\phi^\omega=\sum_{n\in\mathbb{Z}^d} g_n(\omega)\eta(D-n)\phi. \tag{9}$$

The main advantage derived from the Wiener randomization is improved $L^p(\mathbb{R}^d)$ estimates on the randomized initial data ϕ^ω off a small set, as a result of a stronger Bernstein's inequality. Despite only requiring that ϕ be in H^s, the randomized ϕ^ω is in $L^p(\mathbb{R}^d)$ with probability 1. In addition, we have a probabilistic bound on $\|\phi^\omega\|_{H^s(\mathbb{R}^d)}$, which implies that $\phi^\omega\in H^s(\mathbb{R}^d)$ almost surely.

We have the following key bounds on ϕ^ω and its linear Schrödinger evolution with proofs from [Bényi et al. 2015a]. We prove the first estimate while omitting the proofs of the second and third estimates. For all $R>0, s>0$ and $\phi\in H^s(\mathbb{R}^d)$, we have

$$P\big(\|\phi^\omega\|_{H^s(\mathbb{R}^d)}>R\big)\le c_1e^{-c_2R^2/\|\phi\|^2_{H^s(\mathbb{R}^d)}},$$
$$P\big(\|S(t)\phi^\omega\|_{L^q_tL^r_x([0,T]\times\mathbb{R}^d)}>R\big)\le c_1e^{-c_2R^2/T^{2/q}\|\phi\|^2_{L^2(\mathbb{R}^d)}},$$
$$P\big(\|\phi^\omega\|_{L^p(\mathbb{R}^d)}>R\big)\le c_1e^{-c_2R^2/\|\phi\|_{L^2(\mathbb{R}^d)}}.$$

Lemma 2.1. *Given $\phi\in H^s$ with randomization ϕ^ω, for all $R>0$ there exist positive constants c_1,c_2 such that*

$$P\big(\|\phi^\omega\|_{H^s(\mathbb{R}^d)}>R\big)\le c_1e^{-c_2R^2/\|\phi\|^2_{H^s(\mathbb{R}^d)}}. \tag{10}$$

Proof. The proof is taken from [Bényi et al. 2015a]. By Minkowski's inequality, we have for $p\ge 2$,

$$\begin{aligned}\mathbb{E}\big[\|\phi^\omega\|^p_{H^s(\mathbb{R}^d)}\big]&\le\big(\big\|\|\langle\nabla\rangle^s\phi^\omega\|_{L^p(\Omega)}\big\|_{L^2(\mathbb{R}^d)}\big)^p\\&=\left(\left\|\left\|\sum_{n\in\mathbb{Z}^d} g_n(\omega)\langle\nabla\rangle^s\eta(D-n)\phi\right\|_{L^p(\Omega)}\right\|_{L^2(\mathbb{R}^d)}\right)^p.\end{aligned} \tag{11}$$

By a well known lemma on sums of random variables, stated as Lemma 2.1 in [Bényi et al. 2015a] and proven in [Burq and Tzvetkov 2008], and the fact that Fourier multipliers commute, we have

$$\begin{aligned}\mathbb{E}\big[\|\phi^\omega\|^p_{H^s(\mathbb{R}^d)}\big] &\le C\big(\big\|\sqrt{p}\,\|g_n\langle\nabla\rangle^s\eta(D-n)\phi\|_{l^2(n\in\mathbb{Z}^d)}\big\|_{L^2(\mathbb{R}^d)}\big)^p,\\ \mathbb{E}\big[\|\phi^\omega\|^p_{H^s(\mathbb{R}^d)}\big] &\le C\big(\sqrt{p}\,\|\phi\|_{H^s(\mathbb{R}^d)}\big)^p.\end{aligned} \tag{12}$$

So by Markov's inequality,

$$\begin{aligned}R^p P\big(\|\phi^\omega\|_{H^s(\mathbb{R}^d)} > R\big) &\le C\big(\sqrt{p}\,\|\phi\|_{H^s(\mathbb{R}^d)}\big)^p,\\ P\big(\|\phi^\omega\|_{H^s(\mathbb{R}^d)} > R\big) &\le \frac{\big(C_0\sqrt{p}\,\|\phi\|_{H^s(\mathbb{R}^d)}\big)^p}{R^p}.\end{aligned} \tag{13}$$

Now let $p = \big(R/\big(C_0 e\|\phi\|_{H^s}\big)\big)^2$ with C_0 taken from above. There are two cases.

- $p < 2$: In this case we cannot use the above work because it assumes $p \ge 2$ for Minkowski's inequality. Letting $c_2 = 1/(C_0^2 e^2)$ we have $e^{-c_2 R^2/\|\phi\|_{H^s}} \ge e^{-2}$. Now choosing $c_1 \ge e^2$ we have

$$P\big(\|\phi^\omega\|_{H^s(\mathbb{R}^d)} > R\big) \le 1 \le c_1 e^{-2} \le c_1 e^{-c_2 R^2/\|\phi\|^2_{H^s}}, \tag{14}$$

  since every probabilistic outcome has probability less than 1.

- $p \ge 2$: From the definition of p above and (13), we have

$$P\big(\|\phi^\omega\|_{H^s(\mathbb{R}^d)} > R\big) \le e^{-p} \le e^{-c_2 R^2/\|\phi\|^2_{H^s}}. \tag{15}$$

In both cases the lemma is proven. □

Lemma 2.2. *Given $\phi \in H^s$ with randomization ϕ^ω, for all $R > 0$ there exist positive constants c_1, c_2 such that*

$$P\big(\|S(t)\phi^\omega\|_{L^q_t L^r_x([0,T]\times\mathbb{R}^d)} > R\big) \le c_1 e^{-c_2 R^2/T^{2/q}\|\phi\|^2_{L^2}}. \tag{16}$$

After multiplying R by a small power of T we have the following corollary.

Corollary 2.3. *For small $\theta \in \big[0, \frac{1}{q}\big)$ and $R > 0$ there exist c_1, c_2 such that*

$$\begin{aligned}P\big(\|S(t)\phi^\omega\|_{L^q_t L^r_x([0,T]\times\mathbb{R}^d)} > T^\theta R\big) &\le c_1 e^{-c_2 R^2/T^{2/q-2\theta}\|\phi\|^2_{L^2}}\\ &\le c_1 e^{-c_2 R^2/\|\phi\|^2_{L^2}}.\end{aligned} \tag{17}$$

Placing derivatives inside the norm and noting that derivatives commute with Fourier multipliers such as $S(t)$ and the map $\phi \to \phi^\omega$, we have our main bound:

Theorem 2.4. *Given small $\theta \in \left[0, \frac{1}{q}\right)$ and ϕ^ω chosen according to a Wiener randomization, for all $R > 0$ there exist c_1, c_2 such that*

$$P\left(\|\langle\nabla\rangle^s S(t)\phi^\omega\|_{L_t^q L_x^r([0,T]\times\mathbb{R}^d)} > T^\theta R\right) \le c_1 e^{-c_2 R^2/T^{2/q-2\theta}\|\phi\|_{H^s}^2} \le c_1 e^{-c_2 R^2/\|\phi\|_{H^s}^2}. \tag{18}$$

This bound will be crucial in the proof of local well-posedness. This gives us as much integrability as we want in bounding a linear solution, which means the linear solution is bounded in subcritical norms, allowing us to treat local well-posedness like a subcritical problem.

Lemma 2.5. *Given $\phi \in H^s$ with randomization ϕ^ω, for all $R > 0$ there exist positive constants c_1, c_2 such that*:

$$P\left(\|\phi^\omega\|_{L^p(\mathbb{R}^d)} > R\right) \le c_1 e^{-c_2 R^2/\|\phi\|_{L^2}^2}. \tag{19}$$

Proof. The proofs of Theorem 2.4 and Lemma 2.5 can be found in [Bényi et al. 2015a]. They utilize the same basic argument as above, with some extra steps. Each proof exploits an improvement of Bernstein's inequality that results from the Wiener randomization. Note that $g_n(\omega)\eta(D-n)\phi$ has Fourier transform supported on the unit cube centered at n. Therefore, $e^{inx}g_n(\omega)\eta(D-n)\phi$ has Fourier transform supported on the unit cube centered at the origin. Bernstein's inequality implies that

$$\|e^{inx}g_n(\omega)\eta(D-n)\phi\|_{L^p} \lesssim \|e^{inx}g_n(\omega)\eta(D-n)\phi\|_{L^2} \tag{20}$$

with no loss of regularity, since multiplying by e^{inx} does not affect the L^p norm. So we obtain the bound $\|g_n(\omega)\eta(D-n)\phi\|_{L^p} \lesssim \|g_n(\omega)\eta(D-n)\phi\|_{L^2}$. This is the key ingredient in the proof that allows one to bound the higher L^p norm of ϕ^ω with high probability while only assuming that $\phi \in L^2$. □

3. Littlewood–Paley theory and function spaces

3A. ***Littlewood–Paley theory and dyadic decompositions.*** In the fixed point proof we will take the linear and nonlinear parts of our solution and dyadically decompose each into a sum of Littlewood–Paley projections. Given a smooth bump function ψ such that $\psi(\xi) = 1$ for $|\xi| \le 1$ and $\psi(\xi) = 0$ for $|\xi| \ge 2$, we have the following definition from the Littlewood–Paley theory.

Definition 3.1. Given dyadic N and a function $f \in L^2$ we define its *projection* $P_{\le N} f$ to be the Fourier multiplier such that

$$\widehat{P_{\le N} f}(\xi) = \psi\left(\frac{\xi}{N}\right)\hat{f}(\xi).$$

Of course the definition applies to a much wider range of distributions, but in this paper we need only consider functions in L^2 or H^s for some $s > 0$.

Note that $\widehat{P_{\leq N} f}$ is supported on the set $|\xi| \leq 2N$. Now we define the projection P_N that localizes to frequencies in the interval $[N/2, 2N]$.

Definition 3.2. We define $P_1 = P_{\leq 1}$ and for dyadic $N > 1$,

$$P_N f = P_{\leq N} f - P_{\leq N/2} f.$$

This defines the projection $P_N f$ with frequencies between $N/2$ and $2N$. Also, we have $\sum_N P_N f = f$, so this is indeed a decomposition.

The above info and other results on Littlewood–Paley theory can be found in the appendix of [Tao 2006].

3B. *Strichartz spaces.* In this and the following section we introduce the function spaces needed to prove well-posedness. We start with the standard Strichartz spaces: $S^s(I \times \mathbb{R}^d)$ and $N^s(I \times \mathbb{R}^d)$. Let q, r be a Schrödinger-admissible pair. Given an interval $I = [t_0, t_1]$ we define $S^s(I \times \mathbb{R}^d)$ to be the set of measurable functions bounded in the norm

$$\|u\|_{S^s(I\times\mathbb{R}^d)} = \sup_{(q,r) \text{ admissible}} \|\langle\nabla\rangle^s u\|_{L^q L^r(I\times\mathbb{R}^d)}.$$

We also define $N^{-s}(I \times \mathbb{R}^d)$ to be the dual space of $S^s(I \times \mathbb{R}^d)$, which satisfies the bound

$$\|u\|_{N^s(I\times\mathbb{R}^d)} \leq \inf_{(q,r) \text{ admissible}} \|\langle\nabla\rangle^s u\|_{L^{q'} L^{r'}(I\times\mathbb{R}^d)}.$$

The key relation between the Strichartz norms is the Strichartz estimate for solutions to the nonlinear Schrödinger equation. Suppose u is a solution to $iu_t + \Delta u = F$. Then

$$\|u\|_{S^s([t_0,t_1]\times\mathbb{R}^d)} \lesssim \|u(t_0)\|_{H^s(\mathbb{R}^d)} + \|F\|_{N^s([t_0,t_1]\times\mathbb{R}^d)}. \tag{21}$$

3C. *U^p and V^p spaces.* Our analysis requires a norm that measures how close a function is to a linear solution to the Schrödinger equation, so we use the U^p and V^p spaces of Koch and Tataru introduced in [Hadac et al. 2009; Herr et al. 2011]. We start by defining a U^p atom, and then the U^p and V^p spaces. Suppose $1 \leq p < \infty$ and $-\infty < t_0 < t_1 < \cdots < t_n \leq \infty$ is a partition of the real line. We denote the characteristic function of the k-th interval of this partition by $\chi_{[t_{k-1},t_k)}$.

Definition 3.3. A U^p *atom* is a step function into a Sobolev space $a(t) : \mathbb{R} \to H^s(\mathbb{R}^d)$ of the form

$$a = \sum_{k=1}^{n} \phi_k \chi_{[t_{k-1},t_k)}, \tag{22}$$

where $\sum_{k=1}^n \|\phi_k\|^p_{H^s(\mathbb{R}^d)} = 1$.

The definition applies to any Hilbert space H, but we will only need it for Sobolev spaces in this paper.

Definition 3.4. The space $U^p(\mathbb{R}; H^s)$ is the set of measurable functions bounded in the associated norm:

$$\|u\|_{U^p(\mathbb{R};H^s)} = \inf_{U^p \text{ atoms } a_j} \left\{ \sum_j |\lambda_j| : u = \sum_j \lambda_j a_j \right\}. \tag{23}$$

For the V^p spaces we continue to partition the real line, and take our norm to be the p-variation of the given function.

Definition 3.5. The space $V^p(\mathbb{R}; H^s)$ is the set of functions bounded under the V^p norm:

$$\|u\|_{V^p(\mathbb{R};H^s)} = \sup_{\text{partitions } t_k} \left(\sum_{k=1}^{n} \|u(t_k) - u(t_{k-1})\|^p_{H^s(\mathbb{R}^d)} \right)^{1/p}. \tag{24}$$

In addition, given an interval I, the norms $\|u\|_{U^p(I;H^s)}$, $\|u\|_{V^p(I;H^s)}$ and any of the following norms are defined as the restriction norms. For example,

$$\|u\|_{U^p(I;H^s)} = \inf_{\substack{w(t)=u(t),\ t\in I \\ w(\infty)=0=w(-\infty)}} \|w\|_{U^p(\mathbb{R};H^s)}. \tag{25}$$

Now we want to create a norm that measures how close our function is to a linear solution to the Schrödinger equation, much like in the definition of the $X^{s,b}$ spaces. If u is a linear solution then $S(-t)u$ is a function that is constant in time with $\|S(-t)u\|_{U^2(I;H^s)}$ and $\|S(-t)u\|_{V^2(I;H^s)}$ norms bounded by $\|u\|_{H^s}$. We define the $U^p_\Delta H^s$ and $V^p_\Delta H^s$ norms as

$$\|u\|_{U^p_\Delta H^s(\mathbb{R};H^s)} = \|S(-t)u\|_{U^p(\mathbb{R};H^s)},$$
$$\|u\|_{V^p_\Delta H^s(\mathbb{R};H^s)} = \|S(-t)u\|_{V^p(\mathbb{R};H^s)},$$

and the spaces $U^p_\Delta H^s$ and $V^p_\Delta H^s$ are defined as the set of measurable functions $u : \mathbb{R} \to H^s(\mathbb{R}^d)$ bounded in the $U^p_\Delta H^s$ and $V^p_\Delta H^s$ norms, respectively. These are useful spaces; however, in our proof we will rely on dyadic decomposition and will need to apply these norms at specific frequencies, so it is more useful to do computations in a slightly different norm adapted to dyadic decompositions.

Definition 3.6. We define the X^s and Y^s norms, and associated spaces, as follows:

$$\begin{aligned} \|u\|_{X^s(\mathbb{R})} &= \left(\sum_N N^{2s} \|P_N u\|^2_{U^2_\Delta L^2} \right)^{1/2}, \\ \|u\|_{Y^s(\mathbb{R})} &= \left(\sum_N N^{2s} \|P_N u\|^2_{V^2_\Delta L^2} \right)^{1/2}. \end{aligned} \tag{26}$$

Note that these norms are a little stronger than those above. They bound the closeness of the function u to a solution to the linear equation at each frequency, not just generally. Note that we immediately have the embedding $X^s \hookrightarrow Y^s$ as well as the bound $\|S(t)\phi\|_{X^s(\mathbb{R};H^s)} \le \|\phi\|_{H^s(\mathbb{R}^d)}$. This bound means that these spaces are well suited to studying the linear problem.

In addition, we define the following norm for the nonhomogeneous term that will allow us to exploit duality:

$$\|F\|_{M^s(I)} = \left\| \int_{t_0}^{t} S(t-t')F(t')\,dt' \right\|_{X^s(I)}. \tag{27}$$

This is equivalent to the dual norm of Y^s, and we have the bound

$$\|F\|_{M^s(I)} \le \sup_{\|v\|_{Y^s(I)}=1} \int_I \int_{\mathbb{R}^d} F(t,x)v(t,x)\,dx\,dt \tag{28}$$

as Lemma 3.5 in [Bényi et al. 2015b]. This is equivalent to

$$\|F\|_{M^s(I)} \le \sup_{\|v\|_{Y^0(I)}=1} \int_I \int_{\mathbb{R}^d} \langle\nabla\rangle^s F(t,x)v(x,t)\,dx\,dt. \tag{29}$$

In addition, we have a bound analogous to the Strichartz estimate (21) for the M^s norm. Suppose $u(t,x)$ is a solution to the Cauchy problem

$$\begin{aligned} iu_t + \Delta u &= F, \\ u|_{t=0} &= u(0) \end{aligned} \tag{30}$$

on the interval I. Then we have the bound

$$\|u\|_{X^s(I)} \lesssim \|u(0)\|_{H^s(\mathbb{R}^d)} + \|F\|_{M^s(I)}. \tag{31}$$

4. Strichartz estimates

Lemma 4.1. *Let q, r be a Schrödinger-admissible pair.*

(i) *Given an interval I, for any $u \in Y^0(I)$ we have*

$$\|u\|_{L_t^q L_x^r(I\times\mathbb{R}^d)} \lesssim \|u\|_{Y^0(I)}. \tag{32}$$

(ii) *Given an interval I and $p \ge \frac{2(d+2)}{d}$, for any $u \in Y^{d/2-(d+2)/p}(I)$ we have*

$$\begin{aligned} \|u\|_{L_t^p L_x^p(I\times\mathbb{R}^d)} &\lesssim \big\| |\nabla|^{d/2-(d+2)/p} u \big\|_{Y^0(I)} \\ &\lesssim \|u\|_{Y^{d/2-(d+2)/p}(I)}. \end{aligned} \tag{33}$$

Proof. The proof of the first statement is in [Bényi et al. 2015b, Lemma 3.5]. To prove the second statement note that for $\frac{1}{p} = \frac{1}{r} - \frac{k}{d}$, Sobolev embedding implies that

$$\|u\|_{L^p(\mathbb{R}^d)} \lesssim \big\||\nabla|^k u\big\|_{L^r(\mathbb{R}^d)}. \tag{34}$$

Then taking the $L_t^p(I)$ norm of both sides we have

$$\|u\|_{L^p_{t,x}(I\times\mathbb{R}^d)} \lesssim \big\||\nabla|^k u\big\|_{L_t^p L_x^r(I\times\mathbb{R}^d)}. \tag{35}$$

Then by part (i) we have, for $\frac{2}{p}+\frac{d}{r}=\frac{d}{2}$,

$$\big\||\nabla|^k u\big\|_{L_t^p L_x^r(I\times\mathbb{R}^d)} \lesssim \big\||\nabla|^k u\big\|_{Y^0(I)}. \tag{36}$$

This proves the desired inequality in $\mathbb{R}^d$ with exponents that satisfy $\frac{2}{p}+\frac{d}{r}=\frac{d}{2}$ and $\frac{1}{p}=\frac{1}{r}-\frac{k}{d}$. Substituting, we get $k=\frac{d}{2}-\frac{d+2}{p}$. □

By selecting $q=r=\frac{2(d+2)}{d}$ and $p=2(d+2)$, we obtain the following corollaries:

Corollary 4.2. *For all $u\in Y^0(I)$ we have*

$$\|u\|_{L^{2(d+2)/d}_{t,x}(I\times\mathbb{R}^d)} \lesssim \|u\|_{Y^0(I)}. \tag{37}$$

Corollary 4.3. *For all $u\in Y^{(d-1)/2}(I)$ we have*

$$\|u\|_{L^{2(d+2)}_{t,x}(I\times\mathbb{R}^d)} \lesssim \big\||\nabla|^{(d-1)/2}u\big\|_{Y^0(I)} \lesssim \|u\|_{Y^{(d-1)/2}(I)}. \tag{38}$$

Lastly, the following is a bilinear projection lemma that gives an L^2 bound on the bilinear L^2 norm of projections at different frequencies from [Bourgain 1998; Ozawa and Tsutsumi 1998]. In addition there is a version adapted to the Schrödinger equation from [Visan 2006].

Lemma 4.4. *For dyadic $N_1\le N_2$ and $\phi_1,\phi_2\in L^2$ we have*

$$\|P_{N_1}S(t)\phi_1 P_{N_2}S(t)\phi_2\|_{L^2(I\times\mathbb{R}^d)} \lesssim N_1^{(d-1)/2}N_2^{-1/2}\|P_{N_1}\phi_1\|_{L^2(\mathbb{R}^d)}\|P_{N_2}\phi_2\|_{L^2(\mathbb{R}^d)}. \tag{39}$$

Corollary 4.5. *For $N_1\le N_2$ and $u_1,u_2\in Y^0(I)$ we have*

$$\|P_{N_1}u_1P_{N_2}u_2\|_{L^2(I\times\mathbb{R}^d)} \lesssim N_1^{(d-1)/2-}N_2^{-1/2+}\|P_{N_1}u_1\|_{Y^0(I)}\|P_{N_2}u_2\|_{Y^0(I)}. \tag{40}$$

Proof. The proof is found in [Bényi et al. 2015b] as Lemma 3.5. □

This will be a key ingredient in the proof of local well-posedness because it allows us to gain half a derivative from higher frequency terms. In addition, we use the following 3-dimensional bilinear estimate that solely consists of Strichartz norms.

Theorem 4.6. *For dyadic $N_1 \le N_2$ and any small $\delta > 0$ we have*

$$\begin{aligned}\|P_{N_1}u_1 P_{N_2}u_2\|_{L^2(I\times\mathbb{R}^3)} &\lesssim N_1^{(d-1)/2-\delta} N_2^{-1/2+\delta}\big(\|P_{N_1}u_1(0)\|_{L^2(\mathbb{R}^3)} \\ &\qquad + \|(i\partial_t+\Delta)P_{N_2}u_2\|_{L^{3/2}L^{18/13}(I\times\mathbb{R}^3)}\big) \\ &\qquad \times \big(\|P_{N_2}u_2(0)\|_{L^2(\mathbb{R}^3)} \\ &\qquad + \|(i\partial_t+\Delta)P_{N_2}u_2\|_{L^{3/2}L^{18/13}(I\times\mathbb{R}^3)}\big). \end{aligned} \tag{41}$$

Proof. The proof is found in [Visan 2006] as Lemma 2.5. □

This will be a key ingredient in the proof of Theorem 1.2 in Section 6.

5. Almost sure local well-posedness

We now begin the proof of Theorem 1.1. Given some $\phi \in H^s(\mathbb{R}^d)$, let ϕ^ω be its Wiener randomization, and recall that $z(t) = S(t)\phi^\omega$ denotes the linear part of the NLS solution and $v(t)$ is the solution to (6).

Even though we do not have long term bounds on the H^s norm of v, we know that $v(0) = 0$. Exploiting our probabilistic bound on $z(t)$ in subcritical norms, we show that for $\rho \in \big(\frac{d-1}{2}, s+\frac{1}{2}\big)$ the norm $\|v\|_{X^\rho((-T,T))}$ is bounded for small enough time T.

Our method is a fixed point argument. We define

$$\Gamma v(t) = \pm \int_0^t -iS(t-t')\big[|v+z|^4(v+z)\big](t')\,dt' \tag{42}$$

and note that v is a solution if and only if $\Gamma v = v$. We now prove the following proposition, which is the bulk of our fixed point argument.

Proposition 5.1. *Assume s and ρ satisfy the bounds*

$$\frac{d}{2} > s + \frac{1}{2} > \rho > \frac{d-1}{2}. \tag{43}$$

Given $\phi \in H^s(\mathbb{R}^d)$ with randomization ϕ^ω, there exists small $\theta > 0$ such that for every $R > 0$ and sufficiently small $T \ll 1$ depending on R, we have

- $\|\Gamma v\|_{X^\rho([0,T))} \lesssim T^\theta\big(\|v\|^5_{X^\rho([0,T))} + R^5\big)$ *off a set of measure* $c_1 e^{-c_2 R^2/\|\phi\|^2_{H^s}}$,
- $\|\Gamma v_1 - \Gamma v_2\|_{X^\rho([0,T))}$
$$\lesssim T^\theta\big(R^4 + \|v_1\|^4_{X^\rho([0,T))} + \|v_2\|^4_{X^\rho([0,T))}\big)\|v_1 - v_2\|_{X^\rho([0,T))}$$
off a set of measure $c_1 e^{-c_2 R^2/\|\phi\|^2_{H^s}}$.

This stems from Theorem 2.4, which tells us that for $\theta < \frac{1}{q}$ we have

$$P\big(\|\langle\nabla\rangle^s z\|_{L^q_t L^r_x([0,T)\times\mathbb{R}^d)} \le T^\theta R\big) \ge 1 - c_1 e^{-c_2 R^2/\|\phi\|^2_{H^s}}, \tag{44}$$

allowing us to gain a factor of T.

Proof. We only prove the bound on $\|\Gamma v\|_{X^\rho([0,T))}$, as the proof of the bound on $\|\Gamma v_1 - \Gamma v_2\|_{X^\rho([0,T))}$ is similar. For dyadic $N \geq 1$ define

$$\begin{aligned}\Gamma_N(v) &= P_{\leq N}\Gamma(v)\\ &= P_{\leq N}\left(\pm\int_0^t -iS(t-t')\big[|v+z|^4(v+z)\big](t')\,dt'\right)\\ &= \pm\int_0^t -iS(t-t')P_{\leq N}\big[|v+z|^4(v+z)(t')\big]\,dt'.\end{aligned}$$

By (29) we have

$$\begin{aligned}\|\Gamma_N v\|_{X^\rho} &= \big\|P_{\leq N}\big[(v+z)|v+z|^4\big]\big\|_{M^\rho}\\ &\leq \sup_{\|v_6\|_{Y^0}\leq 1}\int_0^T\int_{\mathbb{R}^d}\langle\nabla^\rho\rangle|v+z|^4(v+z)\,\overline{P_{\leq N}v_6}\,dx\,dt.\end{aligned}$$

Now, noting that

$$\begin{aligned}\|\Gamma v\|_{X^\rho} &= \lim_{N\to\infty}\|\Gamma_N v\|_{X^\rho}\\ &= \sup_{\|v_6\|_{Y^0}\leq 1}\int_0^T\int_{\mathbb{R}^d}\langle\nabla^\rho\rangle|v+z|^4(v+z)\overline{v_6}\,dx\,dt,\end{aligned}$$

it suffices to show that for small $\theta > 0$ this integral is $\leq CT^\theta(R^5 + \|v\|_{X^\rho}^5)\|v_6\|_{Y^0}$ off a set of measure $c_1e^{-c_2R^2/\|\phi\|_{H^s}^2}$. We do this by proving the bound

$$\int_0^T\int_{\mathbb{R}^d}\langle\nabla\rangle^\rho|v+z|^4(v+z)\overline{v_6}\,dx\,dt \leq CT^\theta(R^5 + \|v\|_{X^\rho}^5)\|v_6\|_{Y^0} \tag{45}$$

via case by case analysis of terms of the form $\langle\nabla\rangle^\rho[w_1w_2w_3w_4w_5]v_6$, where each w_i is either $v_i = v$ or $z_i = z$ (or its complex conjugate), and each is dyadically decomposed into $\sum_{N_i\geq 1,\,\text{dyadic}}P_{N_i}v_i$ and $\sum_{N_j\geq 1,\,\text{dyadic}}P_{N_j}z_j$. Dyadic decomposition allows us to assume the derivatives are placed on the highest frequency term, or split them between two comparably high frequency terms. Also, we just write w_i instead of $P_{N_i}w_i$ as we sum over dyadic integers $N_i \geq 1$.

We consider four main cases based on whether each w_i is a v_i or z_i, and which two terms have the highest frequencies:

- Case 1: All five terms are v.
- Case 2: At least one term is a v and it has one of the two highest frequencies.
- Case 3: The two highest frequencies are on z terms.
- Case 4: The two highest frequencies are on a z term and the v_6 term.

These four cases are then divided into smaller subcases:

• Case 1: $v_1 v_2 v_3 v_4 v_5 v_6$.

In this case all terms are v. We do not do dyadic decompositions; instead we cut the frequency space into 5 pieces based on which frequency is largest, and assume without loss of generality that ξ_1 is. We split into two cases, based on the value of ρ, which determines which exponents we can use in Hölder's inequality.

(1a) $\rho < \frac{d}{2} - \frac{1}{4}$.

Noting that $\rho < \frac{d}{2} - \frac{1}{4}$, we apply Hölder's inequality with t exponents

$$\tau_1 = \frac{2(d+2)}{d(2d-4\rho-1)}, \quad \tau_2 = \cdots = \tau_5 = \frac{4(d+2)}{d+2-d(d-2\rho)}, \quad \tau_6 = \frac{2(d+2)}{d}$$

and x exponents

$$\sigma_1 = \frac{2(d+2)}{8\rho+4-3d}, \qquad \sigma_2 = \cdots = \sigma_5 = \frac{2(d+2)}{d-2\rho}, \qquad \sigma_6 = \frac{2(d+2)}{d}$$

and Lemma 4.1:

$$\begin{aligned} I &= \int_0^T \int_{\mathbb{R}^d} \langle \nabla \rangle^\rho v_1 v_2 v_3 v_4 v_5 v_6 \, dx \, dt \\ &\le \| \langle \nabla \rangle^\rho v_1 \|_{L^{\tau_1} L^{\sigma_1}} \|v_2\|_{L^{\tau_2} L^{\sigma_2}} \|v_3\|_{L^{\tau_3} L^{\sigma_3}} \|v_4\|_{L^{\tau_4} L^{\sigma_4}} \|v_5\|_{L^{\tau_5} L^{\sigma_5}} \|v_6\|_{L^{\tau_6}} \\ &\le \|v_1\|_{Y^\rho} T^\theta \|v_2\|_{L^{\sigma_2} L^{\sigma_2}} \|v_3\|_{L^{\sigma_3} L^{\sigma_3}} \|v_4\|_{L^{\sigma_4} L^{\sigma_4}} \|v_5\|_{L^{\sigma_5} L^{\sigma_5}} \|v_6\|_{Y^0} \\ &\le T^\theta \prod_{i=1}^{5} \|v_i\|_{Y^\rho} \|v_6\|_{Y^0} \end{aligned}$$

for some $\theta > 0$.

(1b) $\frac{d}{2} - \frac{1}{4} \le \rho < s + \frac{1}{2}$.

Noting that $\rho \ge \frac{d}{2} - \frac{1}{4}$ we apply Hölder's inequality with t exponents $\tau_1 = \infty$, $\tau_2 = \cdots = \tau_5 = 8(d+2)/(d+4)$, $\tau_6 = 2(d+2)/d$ and x exponents $\sigma_1 = 2$, $\sigma_2 = \cdots = \sigma_5 = 4(d+2)$, $\sigma_6 = 2(d+2)/d$ and Lemma 4.1:

$$\begin{aligned} I &= \int_0^T \int_{\mathbb{R}^d} \langle \nabla \rangle^\rho v_1 v_2 v_3 v_4 v_5 v_6 \, dx \, dt \\ &\le \| \langle \nabla \rangle^\rho v_1 \|_{L^{\tau_1} L^{\sigma_1}} \|v_2\|_{L^{\tau_2} L^{\sigma_2}} \|v_3\|_{L^{\tau_3} L^{\sigma_3}} \|v_4\|_{L^{\tau_4} L^{\sigma_4}} \|v_5\|_{L^{\tau_5} L^{\sigma_5}} \|v_6\|_{L^{\tau_6}} \\ &\le \|v_1\|_{Y^\rho} T^\theta \|v_2\|_{L^{\sigma_2}} \|v_3\|_{L^{\sigma_3}} \|v_4\|_{L^{\sigma_4}} \|v_5\|_{L^{\sigma_5}} \|v_6\|_{Y^0} \\ &\le T^\theta \|v_1\|_{Y^\rho} \prod_{i=2}^{5} \|v_i\|_{Y^{(d/2)-(1/4)}} \|v_6\|_{Y^0} \end{aligned}$$

for some $\theta > 0$.

• Case 2: $v_1 w_2 w_3 w_4 z_5 v_6$, $N_1 \gtrsim N_2, N_3, N_4, N_5$.

In this case there is at least one v term and the highest frequency term is a v. Therefore we can assume the derivatives fall on the v_1 term with the highest frequency.

(2a) w_2, w_3, w_4 are all z terms, $N_5 \geq N_4 \geq N_3 \geq N_2 \geq N_1^{1/2(d-1)}$.

We have assumed that v_1 has the highest frequency: $N_1 \geq N_2, N_3, N_4, N_5$. Now we apply Hölder's inequality, Lemma 4.1 and our probabilistic bound on the linear term, Theorem 2.4, and note that $s > \frac{1}{2}$. Setting $\tau = 2(d+2)/d$ and $\sigma = 2(d+2)$, we have

$$
\begin{aligned}
I &= \int_0^T \int_{\mathbb{R}^d} \langle \nabla \rangle^\rho v_1 z_2 z_3 z_4 z_5 v_6 \, dx \, dt \\
&\leq \|\langle \nabla \rangle^\rho v_1\|_{L^\tau} \|z_2\|_{L^\sigma} \|z_3\|_{L^\sigma} \|z_4\|_{L^\sigma} \|z_5\|_{L^\sigma} \|v_6\|_{L^\tau} \\
&\leq \|v_1\|_{Y^\rho} (N_2 N_3 N_4 N_5)^{-s} \prod_{i=2}^{5} \|\langle \nabla \rangle^s z_i\|_{L^\sigma} \|v_6\|_{Y^0} \\
&\leq \|v_1\|_{Y^\rho} (N_2 N_3 N_4 N_5)^{-1/2} \prod_{i=2}^{5} \|\langle \nabla \rangle^s z_i\|_{L^\sigma} \|v_6\|_{Y^0} \\
&\leq \|v_1\|_{Y^\rho} (N_2)^{-2} \prod_{i=2}^{5} \|\langle \nabla \rangle^s z_i\|_{L^\sigma} \|v_6\|_{Y^0} \\
&\leq \|v_1\|_{Y^\rho} (N_1)^{-1/(d-1)} \prod_{i=2}^{5} \|\langle \nabla \rangle^s z_i\|_{L^\sigma} \|v_6\|_{Y^0}.
\end{aligned}
$$

Noting that N_1 is the highest frequency, the sum over all frequencies is bounded by $\|v\|_{Y^\rho} T^\theta R^4 \|v_6\|_{Y^0}$ off a set of measure $c_1 e^{-c_2 R^2/\|\phi\|_{H^s}^2}$.

(2b) w_2, w_3, w_4 are all z terms, $N_2 \leq N_1^{1/2(d-1)}$.

We apply Hölder's inequality, Theorem 2.4, Lemma 4.1 and our bilinear estimate Corollary 4.5, utilizing the assumption that $N_2 \leq N_1^{1/2(d-1)}$. This time with $\tau = 2(d+2)/d$ and $\sigma = 3(d+2)$, we have

$$
\begin{aligned}
I &= \int_0^T \int_{\mathbb{R}^d} \langle \nabla \rangle^\rho v_1 z_2 z_3 z_4 z_5 v_6 \, dx \, dt \\
&\leq \|\langle \nabla \rangle^\rho v_1 z_2\|_{L^2} \|z_3\|_{L^\sigma} \|z_4\|_{L^\sigma} \|z_5\|_{L^\sigma} \|v_6\|_{L^\tau} \\
&\leq N_1^{\frac{-1}{2}+} \|v_1\|_{Y^\rho} N_2^{\frac{1}{2}(d-1)-} \|z_2\|_{Y^0} \|z_3\|_{L^\sigma} \|z_4\|_{L^\sigma} \|z_5\|_{L^\sigma} \|v_6\|_{L^\tau} \\
&\leq N_1^{\frac{-1}{2}+} \|v_1\|_{Y^\rho} N_1^{\frac{1}{4}} \|z_2\|_{Y^0} T^{0+} R^3 \|v_6\|_{Y^0} \\
&\leq N_1^{\frac{-1}{4}+} \|v_1\|_{Y^\rho} T^{0+} R^4 \|v_6\|_{Y^0},
\end{aligned}
$$

which is $\leq \|v\|_{Y^\rho} T^{0+} R^4$ off a set of small measure.

(2c) $w_2 = v_2$ is a v term, and the others can be anything.

In this case we still have $N_1 \geq N_i, i = 2, \dots, 6$. Applying Hölder's inequality, Lemma 4.1, Theorem 2.4 and Corollary 4.5 with $\tau = 2(d+2)$ and $\sigma = 4(d+2)$, we have

$$\begin{aligned} I &= \int_0^T \int_{\mathbb{R}^d} \langle\nabla\rangle^\rho v_1 w_2 w_3 w_4 z_5 v_6 \, dx \, dt \\ &\leq \|\langle\nabla\rangle^\rho v_1 v_2\|_{L^2} \|w_3\|_{L^\tau} \|w_4\|_{L^\sigma} \|z_5\|_{L^\sigma} \|v_6\|_{L^{\tau/d}} \\ &\leq N_1^{\frac{-1}{2}+} \|v_1\|_{Y^\rho} N_2^{\frac{1}{2}(d-1)-} \|v_2\|_{Y^0} \|w_3\|_{L^\tau} \|w_4\|_{L^\sigma} \|z_5\|_{L^\sigma} \|v_6\|_{Y^0} \\ &\leq N_1^{\frac{-1}{2}+} \|v_1\|_{Y^\rho} \|v_2\|_{Y^{\frac{1}{2}(d-1)}} \|w_3\|_{L^\tau} \|w_4\|_{L^\sigma} \|z_5\|_{L^\sigma} \|v_6\|_{Y^0}. \end{aligned}$$

Now if w_3 is a v term, then $\|w_3\|_{L^\tau} \lesssim \|v_3\|_{Y^{(d-1)/2}}$ as required. If w_3 is a z term, then $\|w_3\|_{L^\tau} \leq T^{0+} R$ off a set of small measure. So either way this term is bounded.

If w_4 is a z term then, again, $\|w_4\|_{L^\sigma} \leq T^{0+} R$ off a set of small measure. The only trouble is if w_4 is a v term, in which case our inequality only gives us

$$\begin{aligned} \|w_4\|_{L^\sigma} &\lesssim \big\| |\nabla|^{(d/2)-(1/4)} v_4 \big\|_{Y^0} \\ &\lesssim N_4^{1/4} \|v_4\|_{Y^{(d-1)/2}}. \end{aligned}$$

There is an extra quarter derivative; however, since N_1 is the biggest frequency we have $N_4^{1/4} \leq N_1^{1/4}$, which is absorbed by the $N_1^{-1/2+}$ term.

Therefore each term in this case is bounded by $T^{0+}(\|v\|_{X^\rho}^5 + R^5)\|v_6\|_{Y^0}$ off a set of measure $c_1 e^{-c_2 R^2/\|\phi\|_{H^s}^2}$.

• Case 3: $w_1 w_2 w_3 z_4 z_5 v_6$, $N_4 \sim N_5 \gtrsim N_1, N_2, N_3, N_6$.

In this case the two biggest frequencies are on z terms, z_4 and z_5. The first three terms are denoted w_i, $i = 1, 2, 3$ and represent either v or z. Assume without loss of generality that $N_1 \leq N_2 \leq \cdots \leq N_4 \sim N_5 \geq N_6$. Applying Hölder's inequality for exponents

$$\tau_1 = \cdots = \tau_3 = 2(d+2), \qquad \tau_4 = \tau_5 = \frac{4(d+2)}{d+1}, \qquad \tau_6 = \frac{2(d+2)}{d},$$

Lemma 4.1 and Theorem 2.4 we have

$$\begin{aligned} &\int_0^T \int_{\mathbb{R}^d} w_1 w_2 w_3 z_4 \langle\nabla\rangle^\rho z_5 v_6 \, dx \, dt \\ &\qquad \leq \|w_1\|_{L^{\tau_1}} \|w_2\|_{L^{\tau_2}} \|w_3\|_{L^{\tau_3}} \|\langle\nabla\rangle^{\rho/2} z_4\|_{L^{\tau_4}} \|\langle\nabla\rangle^{\rho/2} z_5\|_{L^{\tau_5}} \|v_6\|_{L^{\tau_6}}. \end{aligned}$$

For $\frac{1}{2}\rho < s$ this term is bounded by

$$T^{0+} R^2 (\|v\|_{Y^{(d-1)/2}}^3 + T^{0+} R^3) \|v_6\|_{Y^0}$$

off a set of measure $c_1 e^{-c_2 R^2/\|\phi\|_{H^s}^2}$. Note that $\rho < s + \frac{1}{2} < 2s$ so ρ satisfies the requirements.

• Case 4: $w_1 w_2 w_3 w_4 z_5 v_6$, $N_5 \sim N_6 \gtrsim N_1, N_2, N_3, N_4$.

This is the biggest case by far and we divide it into several subcases based on how many v terms there are.

(4a) $z_1 z_2 z_3 z_4 z_5 v_6$, $N_5 \sim N_6 \gtrsim N_1, N_2, N_3, N_4$.

Assume without loss of generality $N_1 \le N_2 \le N_3 \le N_4 \lesssim N_5 \sim N_6$. By Hölder's inequality with exponents $(2, 8, 8, 8, 8)$, Corollary 4.5 and Theorem 2.4, we have

$$\begin{aligned}
I &= \int_0^T \int_{\mathbb{R}^d} z_1 z_2 z_3 z_4 \langle\nabla\rangle^\rho z_5 v_6 \, dx \, dt \\
&\le \|z_1 v_6\|_{L^2} \|z_2\|_{L^8} \|z_3\|_{L^8} \|z_4\|_{L^8} \|\langle\nabla\rangle^\rho z_5\|_{L^8} \\
&\lesssim N_1^{\frac{1}{2}(d-1)-s+} N_5^{\frac{-1}{2}+} \|\langle\nabla\rangle^s z_1\|_{Y^0} \|v_6\|_{Y^0} \|z_2\|_{L^8} \|z_3\|_{L^8} \|z_4\|_{L^8} \|\langle\nabla\rangle^\rho z_5\|_{L^8} \\
&\lesssim N_1^{\frac{1}{2}(d-1)+} (N_1 N_2 N_3 N_4)^{-s} N_5^{\rho+\frac{-1}{2}-s+} \prod_{i=2}^{5} \|\langle\nabla\rangle^s z_i\|_{Y^0} \|\langle\nabla\rangle^s z_i\|_{L^8} \|v_6\|_{Y^0}.
\end{aligned}$$

When $s + \frac{1}{2} > \rho$ and $s > \frac{1}{8}(d-1)$ the powers of the frequencies are negative and the sum is bounded by $T^{0+} R^5 \|v_6\|_{Y^0}$. We have assumed $s + \frac{1}{2} > \rho \ge \frac{1}{2}(d-1)$ in the statement of the theorem, and note that for $d \ge 3$, $\frac{1}{2}(d-2) > \frac{1}{8}(d-1)$ and therefore we only require $s > \frac{1}{2}(d-2)$; however,

$$s > \rho - \tfrac{1}{2} \ge \tfrac{1}{2}(d-2).$$

Thus, this term is bounded by $T^{0+} R^5 \|v_6\|_{Y^0}$ off a set of measure $c_1 e^{-c_2 R^2/\|\phi\|_{H^s}^2}$.

In all following cases, we can assume there is at least one v, at least one z and that N_5 and N_6 are the highest frequencies.

(4b) $v_1 z_2 z_3 z_4 z_5 v_6$, $N_5 \sim N_6 \gtrsim N_1, N_2, N_3, N_4$.

Assume without loss of generality that $N_2 \le N_3 \le N_4 \le N_5 \sim N_6 \ge N_1$.

Noting that $N_2 \le N_3, N_4$, we set $\tau = 6(d+2)/(d+1)$ and apply Hölder's inequality, Corollary 4.5 and Theorem 2.4 to obtain

$$\begin{aligned}
I &= \int_0^T \int_{\mathbb{R}^d} v_1 z_2 z_3 z_4 \langle\nabla\rangle^\rho z_5 v_6 \, dx \, dt \\
&\le \|z_2 v_6\|_{L^2} \|v_1\|_{L^{2(d+2)}} \|z_3\|_{L^\tau} \|z_4\|_{L^\tau} \|\langle\nabla\rangle^\rho z_5\|_{L^\tau} \\
&\lesssim N_6^{\frac{-1}{2}+} N_2^{\frac{1}{2}(d-1)-} \|z_2\|_{Y^0} \|v_6\|_{Y^0} \|v_1\|_{Y^{\frac{1}{2}(d-1)}} (N_3 N_4)^{-s} N_5^{\rho-s} R^3 \\
&\lesssim N_6^{\rho-s-\frac{1}{2}+} N_2^{\frac{1}{2}(d-1)-s-} (N_3 N_4)^{-s} T^{0+} R^4 \|v_1\|_{Y^{\frac{1}{2}(d-1)}} \|v_6\|_{Y^0} \\
&\lesssim N_6^{\rho-s-\frac{1}{2}+} (N_2 N_3 N_4)^{\frac{1}{6}(d-1)-s} \|v_1\|_{Y^{\frac{1}{2}(d-1)}} T^{0+} R^4 \|v_6\|_{Y^0}.
\end{aligned}$$

When $s > \frac{1}{6}(d-1)$ the powers of the frequencies are negative, and the sum is bounded by $\|v\|_{Y^{(d-1)/2}} T^{0+} R^4 \|v_6\|_{Y^0}$. And indeed, $s > \frac{1}{6}(d-1)$, since we have $s > \frac{1}{2}(d-2)$ and $d \geq 3$.

(4c) $v_1 v_2 z_3 z_4 z_5 v_6$, $N_5 \sim N_6 \gtrsim N_1, N_2, N_3, N_4$.

Assume without loss of generality that $N_1 \geq N_2$, $N_3 \leq N_4 \leq N_5 \sim N_6$. With $\tau = 6(d+2)/(d+1)$, by Hölder's inequality, Corollary 4.5 and Theorem 2.4 we have

$$\begin{aligned}
I &= \int_0^T \int_{\mathbb{R}^d} v_1 v_2 z_3 z_4 \langle\nabla\rangle^\rho z_5 v_6 \, dx \, dt, \\
&\leq \|v_1 v_6\|_{L^2} \|v_2\|_{L^{2(d+2)}} \|z_3\|_{L^\tau} \|z_4\|_{L^\tau} \|\langle\nabla\rangle^\rho z_5\|_{L^\tau} \\
&\lesssim N_6^{\frac{-1}{2}+} N_1^{\frac{1}{2}(d-1)-} \|v_1\|_{Y^0} \|v_6\|_{Y^0} \|v_2\|_{Y^{\frac{1}{2}(d-1)}} (N_3 N_4)^{-s} (N_5)^\rho T^{0+} R^3 \\
&\lesssim N_1^{0-} (N_3 N_4)^{-s} N_5^{\rho-s-\frac{1}{2}+} \|v_1\|_{Y^{\frac{1}{2}(d-1)}} \|v_2\|_{Y^{\frac{1}{2}(d-1)}} T^{0+} R^3 \|v_6\|_{Y^0} \\
&\lesssim N_1^{0-} N_2^{0-} (N_3 N_4)^{-s} N_5^{\rho-s-\frac{1}{2}+} \|v_1\|_{Y^{\frac{1}{2}(d-1)}} \|v_2\|_{Y^{\frac{1}{2}(d-1)}} T^{0+} R^3 \|v_6\|_{Y^0}.
\end{aligned}$$

Since $s + \frac{1}{2} > \rho$, all powers of the frequencies are negative and the sum is bounded by $\|v\|^2_{Y^{(d-1)/2}} T^{0+} R^3 \|v_6\|_{Y^0}$.

(4d) $v_1 v_2 v_3 z_4 z_5 v_6$, $N_5 \sim N_6 \gtrsim N_1, N_2, N_3, N_4$.

Assume without loss of generality that $N_1 \geq N_2 \geq N_3$, $N_4 \leq N_5$. Setting $\tau = 2(d+2)$ and $\sigma = 4(d+2)/d$, by Hölder's inequality, Corollary 4.5 and Theorem 2.4 we have

$$\begin{aligned}
I &= \int_0^T \int_{\mathbb{R}^d} v_1 v_2 v_3 z_4 \langle\nabla\rangle^\rho z_5 v_6 \, dx \, dt, \\
&\leq \|v_1 v_6\|_{L^2} \|v_2\|_{L^\tau} \|v_3\|_{L^\tau} \|z_4\|_{L^\sigma} \|\langle\nabla\rangle^\rho z_5\|_{L^\sigma} \\
&\lesssim N_6^{\frac{-1}{2}+} N_1^{\frac{1}{2}(d-1)-} \|v_1\|_{Y^0} \|v_6\|_{Y^0} \|v_2\|_{Y^{\frac{1}{2}(d-1)}} \|v_3\|_{Y^{\frac{1}{2}(d-1)}} N_4^{-s} N_5^{\rho-s} T^{0+} R^2 \\
&\lesssim N_1^{0-} N_4^{-s} N_5^{\rho-s-\frac{1}{2}+} \|v_1\|_{Y^{\frac{1}{2}(d-1)}} \|v_2\|_{Y^{\frac{1}{2}(d-1)}} \|v_3\|_{Y^{\frac{1}{2}(d-1)}} T^{0+} R^2 \|v_6\|_{Y^0}.
\end{aligned}$$

Since $s + \frac{1}{2} > \rho$, all powers of the frequencies are negative and the sum is bounded by $\|v\|^3_{Y^{(d-1)/2}} T^{0+} R^2 \|v_6\|_{Y^0}$.

(4e) $v_1 v_2 v_3 v_4 z_5 v_6$, $N_5 \sim N_6 \gtrsim N_1, N_2, N_3, N_4$.

Assume without loss of generality that $N_1 \leq N_2 \leq N_3 \leq N_4 \leq N_5$. Again with $\tau = 2(d+2)$ and $\sigma = 4(d+2)/d$, by Hölder's inequality, Corollary 4.5 and Theorem 2.4 we have

$$
\begin{aligned}
I &= \int_0^T \int_{\mathbb{R}^d} v_1 v_2 v_3 v_4 \langle \nabla \rangle^{\rho} z_5 v_6 \, dx \, dt, \\
&\leq \|v_1 v_6\|_{L^2} \|v_2\|_{L^\tau} \|v_3\|_{L^\tau} \|v_4\|_{L^\sigma} \|\langle \nabla \rangle^{\rho} z_5\|_{L^\sigma} \\
&\lesssim N_6^{\frac{-1}{2}+} N_1^{\frac{1}{2}(d-1)-} \\
&\qquad \times \|v_1\|_{Y^0} \|v_6\|_{Y^0} \|v_2\|_{Y^{\frac{1}{2}(d-1)}} \|v_3\|_{Y^{\frac{1}{2}(d-1)}} \|v_4\|_{Y^{\frac{1}{4}d}} N_5^{\rho-s} T^{0+} R \\
&\lesssim N_1^{0-} N_4^{-\frac{1}{4}(d-2)} N_5^{\rho-s-\frac{1}{2}+} \\
&\qquad \times \|v_1\|_{Y^{\frac{1}{2}(d-1)}} \|v_2\|_{Y^{\frac{1}{2}(d-1)}} \|v_3\|_{Y^{\frac{1}{2}(d-1)}} \|v_4\|_{Y^{\frac{1}{2}(d-1)}} T^{0+} R \|v_6\|_{Y^0}.
\end{aligned}
$$

Since $s + \frac{1}{2} > \rho$, all powers of the frequencies are negative and the sum is bounded by $\|v\|^4_{Y^{(d-1)/2}} T^{0+} R \|v_6\|_{Y^0}$ off a set of small measure.

In each case the term is bounded by $CT^{\theta}(R^5 + \|v\|^5_{Y^{\rho}})\|v_6\|_{Y^0}$, for some $\theta > 0$. This completes the proof of the first part of the proposition. The proof of

$$
\|\Gamma v_1 - \Gamma v_2\|_{X^{\rho}} \leq CT^{\theta}(R^4 + \|v_1\|^4_{X^{\rho}} + \|v_2\|^4_{X^{\rho}})\|v_1 - v_2\|_{X^{\rho}}
$$

off a set of measure $c_1 e^{-c_2 R^2/\|\phi\|^2_{H^s}}$ is similar and is omitted. □

Using this key proposition we can close the fixed point argument in the final theorem.

Proof of Theorem 1.1. Let B_r denote the ball of radius r in $X^{\rho}([0, T))$ with $\frac{1}{2}d > s + \frac{1}{2} > \rho \geq \frac{1}{2}(d-1)$ as in the previous proposition. We claim that for small enough T and small but fixed r the map Γ is a contraction on B_r outside a set of measure $c_1 e^{-c_2 R^2/\|\phi\|^2_{H^s}}$. See Section 1.6 of [Tao 2006] for an overview of contraction based fixed point arguments.

To apply the theory for fix point arguments we require, off a small set, the contraction conditions

- $\|\Gamma v\|_{X^{\rho}([0,T))} \leq r$ for $v \in B_r$,
- $\|\Gamma v_1 - \Gamma v_2\|_{X^{\rho}([0,T))} \leq \frac{1}{2}\|v_1 - v_2\|_{X^{\rho}([0,T))}$.

By the bounds from the proposition, we have, for all R and some fixed constant C, $\|\Gamma v\|_{X^{\rho}} \leq CT^{\theta}(R^5 + r^5)$ and $\|\Gamma v_1 - \Gamma v_2\|_{X^{\rho}} \leq C\|v_1 - v_2\|_{X^{\rho}} T^{\theta}(2r^4 + R^4)$ off a set of measure $c_1 e^{-c_2 R^2/\|\phi\|_{H^s}}$.

The contraction conditions are satisfied if we select r, R, T such that

$$
\begin{aligned}
r &\leq R, \\
CT^{\theta} R^5 &\leq \frac{r}{8}.
\end{aligned} \tag{46}
$$

We can fix a value of r to satisfy the first bound. Selecting T such that $T \sim R^{-5/\theta}$

the second bound of a contraction is satisfied, and we conclude that the map Γ has a fixed point in B_r.

Therefore, for sufficiently small T, the equation $\Gamma v = v$ has a solution in B for every ϕ^ω off this set of measure $c_1 e^{-c_2 R^2/\|\phi\|_{H^s}^2}$. Setting $\alpha = -2\theta/5$, there exists a set $\Omega_T \subset \Omega$ of measure $\geq 1 - c_1 e^{-c_2/T^\alpha \|\phi\|_{H^s}^2}$ such that for $t \in [0, T)$ the Duhamel equation

$$v(t) = \pm \int_0^t -iS(t-t')\big[|v+z|^4(v+z)\big](t')\,dt' \tag{47}$$

has a unique solution in $X^\rho([0, T))$. The same argument proves the existence of a solution in $X^\rho((-T, 0])$ on a set of the same measure. Taking $u(t) = S(t)\phi + v(t)$ we have a solution on the interval $(-T, T)$ in the class

$$H^s(\mathbb{R}^d) + C((-T, T) \to H^\rho(\mathbb{R}^3)) \subset H^s(\mathbb{R}^d). \qquad \square$$

6. A condition for global well-posedness

We now present the proof of Theorem 1.2. The proof relies upon the following proposition.

Proposition 6.1. *Suppose* $0 < c < \frac{1}{8}$, $\frac{7}{8} < s < 1$ *and* $\|\phi^\omega\|_{H^s(\mathbb{R}^3)} < R$. *There exists a small positive constant* $\epsilon \ll \frac{1}{R}$ *such that for any interval* $[t_1, t_2]$ *satisfying*

$$|t_1 - t_2| \leq 1, \quad \|v\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)} < \epsilon \quad \textit{and} \quad \|\langle\nabla\rangle^s z\|_{L^q L^r([t_1,t_2]\times\mathbb{R}^3)} < \epsilon$$

for the pairs $(q, r) \in \big\{(10, 10), \big(\frac{15}{2}, \frac{15}{7}\big), \big(\frac{30}{7}, 15\big)\big\}$, *we have*

$$\|v\|_{S^{1+c}([t_1,t_2]\times\mathbb{R}^3)} \lesssim \|v(t_1)\|_{H^{1+c}(\mathbb{R}^3)} + C(\epsilon).$$

We first give the proof of Theorem 1.2 given that Proposition 6.1 is true. The rest of the paper is devoted to proving Proposition 6.1.

Proof of Theorem 1.2. Assume Proposition 6.1 and the hypothesis of Theorem 1.2, that there exists such a function α. Fix values T, R and a set $\Omega'_{T,R}$ satisfying the properties outlined in Theorem 1.2. By Theorem 2.4 and Lemma 2.1 there is a set $\Omega_{T,R} \subset \Omega'_{T,R}$ of measure at least $1 - c_1 e^{-c_2 R^2/T\|\phi\|_{H^s}} - \alpha(T, R)$ such that for any $\omega \in \Omega_{T,R}$ and $(q, r) \in \big\{(\infty, 2), (10, 10), \big(\frac{15}{2}, \frac{15}{7}\big), \big(\frac{30}{7}, 15\big)\big\}$ we have

$$\|\langle\nabla\rangle^s z\|_{L^q L^r([-T,T]\times\mathbb{R}^3)} < R, \tag{48}$$

and for any solution v to (6) we have

$$\|v\|_{L^{10}L^{10}([-T,T]\times\mathbb{R}^3)} < R. \tag{49}$$

Now assume that ω is indeed in the set $\Omega_{T,R}$. Note that by the local well-posedness theory a solution exists on some short time interval $(-t, t)$. Suppose for sake of contradiction there is a pair of times $-T < T_{\min} < 0 < T_{\max} < T$ such that the solution $v(t)$ cannot be extended in H^{1+c} past $(T_{\min}, T_{\max})$.

We know that $\|v\|_{L^{10}L^{10}((T_{\min},T_{\max})\times\mathbb{R}^3)} < R$ and we have

$$\|\langle\nabla\rangle^s z\|_{L^q L^r([-T,T]\times\mathbb{R}^3)} < R$$

for each necessary pair (q, r). Therefore, we can split $[T_{\min}, T_{\max}]$ into a finite number of subintervals I on which $\|v\|_{L^{10}L^{10}(I\times\mathbb{R}^3)} < \epsilon$ and $\|\langle\nabla\rangle^s z\|_{L^q L^r(I\times\mathbb{R}^3)} < \epsilon$ for $(q, r) \in \left\{(10, 10), \left(\frac{15}{2}, \frac{15}{7}\right), \left(\frac{30}{7}, 15\right)\right\}$.

This means that on each subinterval $[t_i, t_{i+1}]$, the conditions of Proposition 6.1 are met, and therefore the $\|v\|_{S^{1+c}([t_i,t_{i+1}]\times\mathbb{R}^3)}$ norm is finite. Therefore, there exists a solution in the space $S^{1+c}([t_i, t_{i+1}] \times \mathbb{R}^3)$ on each successive interval $[t_i, t_{i+1}]$, which implies that the $\|v\|_{L^\infty H^{1+c}}$ norm is bounded at each endpoint. This means the S^{1+c} norm is bounded on the next interval. Iterating this argument over each subinterval, this implies the S^{1+c} norm of the nonlinear solution $v(t)$ is bounded on the whole interval $[T_{\min}, T_{\max}]$. In addition, $\|v(T_{\min})\|_{H^{1+c}}$ and $\|v(T_{\max})\|_{H^{1+c}}$ are both finite. Therefore, one can apply the local well-posedness theory to extend the solution beyond $[T_{\min}, T_{\max}]$, which is a contradiction. □

This concludes the proof of Theorem 1.2. It remains to prove Proposition 6.1.

Proof of Proposition 6.1. The nonlinear part of the solution v satisfies the differential equation

$$iv_t + \Delta v = (v + z)|v + z|^4 = v|v|^4 + f(v, z) \tag{50}$$

for the function $f(v, z) = (v + z)|v + z|^4 - v|v|^4 \lesssim |z|^5 + |z| \cdot |v|^4$.

By the Strichartz estimates (21), we have the bound

$$\|v\|_{S^{1+c}([t_1,t_2]\times\mathbb{R}^3)} \lesssim \|v(t_1)\|_{H^{1+c}} + \left\|v|v|^4\right\|_{N^{1+c}([t_1,t_2]\times\mathbb{R}^3)} + \|f(v, z)\|_{N^{1+c}([t_1,t_2]\times\mathbb{R}^3)}. \tag{51}$$

So we need to bound the two remaining terms.

Lemma 6.2. *If v is a solution to (50), then*

$$\left\|v|v|^4\right\|_{N^{1+c}([t_1,t_2]\times\mathbb{R}^3)} \lesssim \epsilon^4 \|v\|_{S^{1+c}([t_1,t_2]\times\mathbb{R}^3)}.$$

Proof. Note that the pair $\left(\frac{10}{3}, \frac{10}{3}\right)$ is Schrödinger-admissible and has Hölder conjugate $\left(\frac{10}{7}, \frac{10}{7}\right)$. Therefore, by (21) we have

$$\begin{aligned}\left\|v \cdot |v|^4\right\|_{N^{1+c}([t_1,t_2]\times\mathbb{R}^3)} &\lesssim \left\|\langle\nabla\rangle^{1+c} v \cdot |v|^4\right\|_{L^{10/7}L^{10/7}([t_1,t_2]\times\mathbb{R}^3)} \\ &\lesssim \|\langle\nabla\rangle^{1+c} v\|_{L^{10/3}L^{10/3}([t_1,t_2]\times\mathbb{R}^3)} \|v\|^4_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)} \\ &\lesssim \|v\|_{S^{1+c}([t_1,t_2]\times\mathbb{R}^3)} \|v\|^4_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)} \\ &\lesssim \epsilon^4 \|v\|_{S^{1+c}([t_1,t_2]\times\mathbb{R}^3)}.\end{aligned}$$

□

Proposition 6.3. *Assume* $0 < c < \frac{1}{8}$, $f(v,z) = (v+z)|v+z|^4 - v|v|^4$ *and that* z *and* v *satisfy the* R *and* ϵ *bounds in the proposition, where* z *is the linear solution and* v *is the solution to* (6). *Then we have*

$$\begin{aligned}\|f\|_{N^{1+c}([t_1,t_2]\times\mathbb{R}^3)} &\lesssim \|\langle\nabla\rangle^{1+c} f\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)} \\ &\lesssim \sqrt{\epsilon^7}R\big(\sqrt{\epsilon R} + \|v_1(t_1)\|_{H^1(\mathbb{R}^3)} \\ &\qquad + \|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big). \qquad (52)\end{aligned}$$

Proof. Observing that $\big(3, \frac{18}{5}\big)$ is Schrödinger-admissible, we have

$$\begin{aligned}\|f\|_{N^{1+c}([t_1,t_2]\times\mathbb{R}^3)} &\le \|f\|_{L^{3/2}W^{1+c,18/13}([t_1,t_2]\times\mathbb{R}^3)} \\ &\lesssim \sup_{\|w_6\|_{L^3L^{18/5}([t_1,t_2]\times\mathbb{R}^3)}\le 1} \int_{t_1}^{t_2}\int_x \langle\nabla\rangle^{1+c}[f]w_6\,dw. \qquad (53)\end{aligned}$$

The function $f(v,z)$ is a sum of terms of the form $w_1w_2w_3w_4z_5$, where each w_i is either a v or z term. We dyadically decompose these first five terms (not w_6), refer to $P_{N_i}w_i$ as w_i, and sum over all frequencies N_1–N_5 and combinations of v, z in integrals of the form

$$\begin{aligned}&\|\langle\nabla\rangle^{1+c} f\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)} \\ &\qquad\lesssim \sup_{\|w_6\|_{L^3L^{18/5}([t_1,t_2]\times\mathbb{R}^3)}\le 1} \int_{t_1}^{t_2}\int_x \langle\nabla\rangle^{1+c}[w_1w_2w_3w_4v_5]w_6\,dw. \qquad (54)\end{aligned}$$

We can assume that the $1+c$ derivatives fall on the term with highest frequency. Before going through cases, we prove the following lemmas that combine interpolation with the bilinear estimate, Theorem 4.6.

Lemma 6.4. *If* $N_1 \le N_2$, *then for any pair of dyadic components* $v_1 = P_{N_1}v$, $z_2 = P_{N_2}z$ *we have the bound*

$$\begin{aligned}&\|v_1z_5\|_{L^{30/11}L^{15/8}([t_1,t_2]\times\mathbb{R}^3)} \\ &\quad\lesssim N_2^{-1/4-s+\delta/2}\|v_1\|^{1/2}_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)} \\ &\qquad\times\big(\|v_1(t_1)\|_{H^1(\mathbb{R}^3)} + \|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big)^{1/2}\sqrt{\epsilon R}. \qquad (55)\end{aligned}$$

Proof. First note that for $N_1 \le N_2$ we have the bilinear estimate Theorem 4.6:

$$\begin{aligned}&\|v_1z_2\|_{L^2L^2([t_1,t_2]\times\mathbb{R}^3)} \\ &\quad\lesssim N_2^{-1/2+\delta}N_1^{1-\delta}\big(\|v_1(t_1)\|_{L^2(\mathbb{R}^3)} + \|u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big)\|z(t_1)\|_{L^2(\mathbb{R}^3)},\end{aligned}$$

$$\begin{aligned}&\|v_1 z_2\|_{L^2L^2([t_1,t_2]\times\mathbb{R}^3)}\\&\qquad\lesssim N_2^{-1/2-s+\delta}\big(\|v_1(t_1)\|_{H^1(\mathbb{R}^3)}+\|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big)R.\end{aligned}\tag{56}$$

Also, by Hölder's inequality

$$\begin{aligned}\|v_1 z_2\|_{L^{30/7}L^{30/17}([t_1,t_2]\times\mathbb{R}^3)}&\le\|v_1\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\|z_2\|_{L^{30/4}L^{30/14}([t_1,t_2]\times\mathbb{R}^3)}\\&\le N_2^{-s}\|v_1\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\epsilon.\end{aligned}\tag{57}$$

Now note that

$$\frac{1}{30/11}=\frac{1/2}{2}+\frac{1/2}{30/7},\qquad\frac{1}{15/8}=\frac{1/2}{2}+\frac{1/2}{30/17}.\tag{58}$$

Interpolating with exponents $\frac{1}{2}$, $\frac{1}{2}$ yields

$$\begin{aligned}&\|v_1 z_5\|_{L^{30/11}L^{15/8}([t_1,t_2]\times\mathbb{R}^3)}\\&\qquad\lesssim N_2^{-1/4-s+\delta/2}\|v_1\|^{1/2}_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\\&\qquad\qquad\times\big(\|v_1(t_1)\|_{H^1(\mathbb{R}^3)}+\|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big)^{1/2}\sqrt{\epsilon R}.\end{aligned}\tag{59}$$

This completes the proof. □

Lemma 6.5. *If $N_2\le N_1$, then for any pair of dyadic components $v_1=P_{N_1}v$, $z_2=P_{n_2}z$ we have the bound*

$$\begin{aligned}&\|v_1 z_2\|_{L^{30/11}L^{15/8}([t_1,t_2]\times\mathbb{R}^3)}\\&\qquad\lesssim N_1^{-3/4+\delta/2}N_2^{1/2-s}\\&\qquad\qquad\times\big(\|v_1(t_1)\|_{H^1(\mathbb{R}^3)}+\|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big)\sqrt{\epsilon R}.\end{aligned}\tag{60}$$

Proof. The bilinear estimate Theorem 4.6 tells us that

$$\begin{aligned}&\|v_1 z_2\|_{L^2L^2([t_1,t_2]\times\mathbb{R}^3)}\\&\qquad\lesssim N_1^{-1/2+\delta}N_2^{1-\delta}\\&\qquad\qquad\times\big(\|v_1(t_1)\|_{L^2(\mathbb{R}^3)}+\|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big)\|z(t_1)\|_{L^2(\mathbb{R}^3)}\\&\|v_1 z_2\|_{L^2L^2([t_1,t_2]\times\mathbb{R}^3)}\\&\qquad\le N_1^{-3/2+\delta}N_2^{1-s}\big(\|v_1\|_{H^1(\mathbb{R}^3)}+\|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big)R.\end{aligned}\tag{61}$$

Also, by Hölder's inequality we have

$$\begin{aligned}\|v_1 z_2\|_{L^{30/7}L^{30/17}([t_1,t_2]\times\mathbb{R}^3)}&\lesssim\|v_1\|_{L^\infty L^2([t_1,t_2]\times\mathbb{R}^3)}\|z_2\|_{L^{30/7}L^{15}([t_1,t_2]\times\mathbb{R}^3)}\\&\lesssim N_1^{-1}N_2^{-s}\|v_1\|_{S^1([t_1,t_2]\times\mathbb{R}^3)}\epsilon.\end{aligned}\tag{62}$$

So with exponents $\frac{1}{2}, \frac{1}{2}$ we interpolate between the L^2L^2 and $L^{30/7}L^{30/17}$ bounds, and apply the Strichartz estimate to get

$$\begin{aligned}\|v_1 z_2\|_{L^{30/11}L^{15/8}([t_1,t_2]\times\mathbb{R}^3)} \\ &\lesssim N_1^{-3/4+\delta/2} N_2^{1/2-s} \\ &\qquad \times \left(\|v_1(t_1)\|_{H^1(\mathbb{R}^3)} + \|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\right)\sqrt{\epsilon R}, \quad (63)\end{aligned}$$

completing the proof. □

Lemma 6.6. *If $N_1 \leq N_2$ then for $z_1 = P_{N_1}z$, $z_2 = P_{n_2}z$ we have*

$$\|z_1 z_2\|_{L^{30/11}L^{15/8}([t_1,t_2]\times\mathbb{R}^3)} \leq N_2^{-1/4-s+\delta/2} N_1^{1/2-s} R\epsilon. \quad (64)$$

Proof. The proof is identical to that of Lemma 6.4 except that v_1 has been replaced with z_1, which is put in an $L^{10}L^{10}$ norm. □

In analyzing terms of the form $w_1w_2w_3w_4v_5$, there are two cases for where the highest frequencies occur:

- Case 1: The highest frequency is on a z term.
- Case 2: The highest frequency is on a v term.

Throughout these cases we utilize the facts that $\|v\|_{L^qL^r} \lesssim \|v\|_{S^1}$ for $\frac{2}{q} + \frac{3}{r} = \frac{1}{2}$ and $\|v\|_{L^qL^r} \lesssim \|v\|_{S^0}$ for (q,r) Schrödinger admissible. We also use the above three lemmas. Now we begin the analysis of cases.

Case 1: In this case the highest frequency is on z_5. We have all the derivatives falling on z_5.

(1a) $v_1w_2w_3w_4z_5$ case.

Applying Hölder's inequality, Lemma 6.4 and our assumptions about ϵ we have

$$\begin{aligned}I &= \int_{t_1}^{t_2}\int_x v_1w_2w_3w_4\langle\nabla\rangle^{1+c} z_5 w_6\, dw \\ &\leq N_5^{1+c}\|w_2\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\|w_3\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\|w_4\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)} \\ &\qquad \times \|v_1z_5\|_{L^{30/11}L^{15/8}([t_1,t_2]\times\mathbb{R}^3)}\|w_6\|_{L^3L^{18/5}} \\ &\leq N_5^{3/4+c-s+\delta/2}\|w_2\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\|w_3\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)} \\ &\qquad \times \|w_4\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\|v_1\|^{1/2}_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)} \\ &\qquad \times \left(\|v_1(t_1)\|_{H^1(\mathbb{R}^3)} + \|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\right)^{1/2} \\ &\qquad \times \sqrt{\epsilon R}\|w_6\|_{L^3L^{18/5}}. \quad (65)\end{aligned}$$

For $c < \frac{1}{8}$, $s > \frac{7}{8}$ and $\delta = 0+$, the power of N_5 is negative, and the sum converges. The w_i terms are all bounded by $\|v_i\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}$ or ϵ. So this is bounded by $\epsilon^4 R^{1/2}\big(\|v_1(t_1)\|_{H^1(\mathbb{R}^3)} + \|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big)^{1/2}$.

(1b) $z_1z_2z_3z_4z_5$ case.

Applying Hölder's inequality, Lemma 6.6, and our ϵ bounds we have

$$\begin{aligned}
I &= \int_{t_1}^{t_2}\int_x z_1z_2z_3z_4\langle\nabla\rangle^{1+c}z_5w_6\,dw\\
&\le N_5^{1+c}\|z_2\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\|z_3\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\\
&\qquad\times\|z_4\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\|z_1z_5\|_{L^{30/11}L^{15/8}([t_1,t_2]\times\mathbb{R}^3)}\|w_6\|_{L^3L^{18/5}}\\
&\le N_1^{1/2-s}N_5^{3/4-s+c+\delta}\|z_2\|_{L^{10}L^{10}}\|z_3\|_{L^{10}L^{10}}\\
&\qquad\times\|z_4\|_{L^{10}L^{10}}\epsilon R\|w_6\|_{L^3L^{18/5}}.
\end{aligned}\tag{66}$$

For $s > \frac{7}{8}$ and $c < \frac{1}{8}$ and $\delta = 0+$, both powers are negative and this is bounded by $\epsilon^4 R$.

Case 2: In this case the highest frequency falls on v, meaning $N_1 \ge N_2, \dots, N_5$. Applying Hölder's inequality, Lemma 6.5, and our ϵ bounds, for $N_5 \le N_1$ we have

$$\begin{aligned}
I &= \int_{t_1}^{t_2}\int_x \langle\nabla\rangle^{1+c}v_1w_2w_3w_4z_5w_6\,dw\\
&\le N_1^{1+c}\|w_2\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\|w_3\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\\
&\qquad\times\|w_4\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\|v_1z_5\|_{L^{30/11}L^{15/8}([t_1,t_2]\times\mathbb{R}^3)}\|w_6\|_{L^3L^{18/5}}\\
&\le N_1^{-1/4+c+\delta/2}N_5^{1/2-s}\|w_2\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\|w_3\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\\
&\qquad\times\|w_4\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}\big(\|v_1(t_1)\|_{H^1(\mathbb{R}^3)} + \|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big)\\
&\qquad\times\sqrt{\epsilon R}\|w_6\|_{L^3L^{18/5}}.
\end{aligned}\tag{67}$$

As in case (1a), the $\|w_i\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}$ terms are all bounded by either ϵ or $\|v\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}$. Thus, for $c < \frac{1}{8}$ the sum over frequencies is bounded by

$$\sqrt{\epsilon^7 R}\big(\|v_1(t_1)\|_{H^1(\mathbb{R}^3)} + \|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big).$$

So in all cases the integral is bounded by

$$\sqrt{\epsilon^7 R}\big(\sqrt{\epsilon R} + \|v_1(t_1)\|_{H^1(\mathbb{R}^3)} + \|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big).$$

This completes the proof of Proposition 6.3. □

So combining Lemma 6.2, Proposition 6.3 and the fact that $\epsilon \ll 1$ we arrive at the following pair of inequalities:

$$\begin{aligned}
&\|v\|_{S^{1+c}([t_1,t_2]\times\mathbb{R}^3)} \\
&\qquad\lesssim \|v(t_1)\|_{H^{1+c}} + \|\langle\nabla\rangle^{1+c} f\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}, \\
&\|\langle\nabla\rangle^{1+c} f\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)} \\
&\qquad\lesssim \sqrt{\epsilon^7 R}\big(\sqrt{\epsilon R} + \|v_1(t_1)\|_{H^1(\mathbb{R}^3)} + \|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big).
\end{aligned}$$

So all that remains is to bound

$$\begin{aligned}
&\|\langle\nabla\rangle u^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)} \\
&\qquad\le \|\langle\nabla\rangle v^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)} + \|\langle\nabla\rangle f\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}. \qquad (68)
\end{aligned}$$

First observe that

$$\begin{aligned}
&\|\langle\nabla\rangle v^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)} \\
&\qquad\le \|\langle\nabla\rangle v\cdot v^4\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)} \\
&\qquad\le \|\langle\nabla\rangle v\|_{L^{15/4}L^{90/29}([t_1,t_2]\times\mathbb{R}^3)}\|v\|^4_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)} \\
&\qquad\le \|v\|_{S^1([t_1,t_2]\times\mathbb{R}^3)}\|v\|^4_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)} \\
&\qquad\lesssim \|\langle\nabla\rangle v^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\|v\|^4_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)} \\
&\qquad\qquad + \big(\|v(t_1)\|_{H^1(\mathbb{R}^3)} + \|f\|_{N^1([t_1,t_2]\times\mathbb{R}^3)}\big)\|v\|^4_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}
\end{aligned}$$

and for $\|v\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}$ less than ϵ we have

$$\|\langle\nabla\rangle v^5\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)} \lesssim \epsilon^4\big(\|v(t_1)\|_{H^1(\mathbb{R}^3)} + \|f\|_{N^1([t_1,t_2]\times\mathbb{R}^3)}\big). \qquad (69)$$

Noting that $\|v\|_{L^{10}L^{10}([t_1,t_2]\times\mathbb{R}^3)}$ is small and combining Proposition 6.3, (68) and (69), we have

$$\begin{aligned}
&\|\langle\nabla\rangle^{1+c} f\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)} \\
&\quad\lesssim \sqrt{\epsilon^7 R}\big(\sqrt{\epsilon R} + \|v(t_1)\|_{H^1(\mathbb{R}^3)} + \|\langle\nabla\rangle f\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big) \\
&\quad\lesssim \sqrt{\epsilon^7 R}\big(\sqrt{\epsilon R} + \|v(t_1)\|_{H^1(\mathbb{R}^3)} + \|\langle\nabla\rangle^{1+c} f\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)}\big). \qquad (70)
\end{aligned}$$

For $\epsilon \ll \frac{1}{R}$ this implies that

$$\|\langle\nabla\rangle^{1+c} f\|_{L^{3/2}L^{18/13}([t_1,t_2]\times\mathbb{R}^3)} \lesssim \sqrt{\epsilon^7 R}\big(\sqrt{\epsilon R} + \|v(t_1)\|_{H^1(\mathbb{R}^3)}\big).$$

This gives us the necessary bound on f.

Combining this result with Lemma 6.2, we have

$$\|v\|_{S^{1+c}([t_1,t_2]\times\mathbb{R}^3)} \lesssim \|v(t_1)\|_{H^{1+c}(\mathbb{R}^3)} + C(\epsilon) \qquad (71)$$

for sufficiently small $\epsilon \ll \frac{1}{R}$, which completes the proof of Proposition 6.1. □

References

[Alazard and Carles 2009] T. Alazard and R. Carles, "Loss of regularity for supercritical nonlinear Schrödinger equations", *Math. Ann.* **343**:2 (2009), 397–420. MR Zbl

[Bényi et al. 2015a] Á. Bényi, T. Oh, and O. Pocovnicu, "Wiener randomization on unbounded domains and an application to almost sure well-posedness of NLS", pp. 3–25 in *Excursions in harmonic analysis* (College Park, MD, 2013), vol. 4, edited by R. Balan et al., Springer, 2015. MR Zbl

[Bényi et al. 2015b] Á. Bényi, T. Oh, and O. Pocovnicu, "On the probabilistic Cauchy theory of the cubic nonlinear Schrödinger equation on $\mathbb{R}^d$, $d \geq 3$", *Trans. Amer. Math. Soc. Ser. B* **2** (2015), 1–50. MR Zbl

[Bourgain 1998] J. Bourgain, "Refinements of Strichartz' inequality and applications to 2D-NLS with critical nonlinearity", *Internat. Math. Res. Notices* **1998**:5 (1998), 253–283. MR Zbl

[Burq and Tzvetkov 2008] N. Burq and N. Tzvetkov, "Random data Cauchy theory for supercritical wave equations, I: Local theory", *Invent. Math.* **173**:3 (2008), 449–475. MR Zbl

[Christ et al. 2003] M. Christ, J. Colliander, and T. Tao, "Ill-posedness for nonlinear Schrödinger and wave equations", preprint, 2003. arXiv

[Hadac et al. 2009] M. Hadac, S. Herr, and H. Koch, "Well-posedness and scattering for the KP-II equation in a critical space", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **26**:3 (2009), 917–941. MR Zbl

[Herr et al. 2011] S. Herr, D. Tataru, and N. Tzvetkov, "Global well-posedness of the energy-critical nonlinear Schrödinger equation with small initial data in $H^1(\mathbb{T}^3)$", *Duke Math. J.* **159**:2 (2011), 329–349. MR Zbl

[Ozawa and Tsutsumi 1998] T. Ozawa and Y. Tsutsumi, "Space-time estimates for null gauge forms and nonlinear Schrödinger equations", *Differential Integral Equations* **11**:2 (1998), 201–222. MR Zbl

[Strichartz 1977] R. S. Strichartz, "Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations", *Duke Math. J.* **44**:3 (1977), 705–714. MR Zbl

[Tao 2006] T. Tao, *Nonlinear dispersive equations: Local and global analysis*, CBMS Regional Conference Series in Mathematics **106**, American Mathematical Society, Providence, RI, 2006. MR Zbl

[Visan 2006] M. Visan, *The defocusing energy-critical nonlinear Schrödinger equation in dimensions five and higher*, Ph.D. thesis, University of California, Los Angeles, 2006, available at https://search.proquest.com/docview/305342414/.

[Yajima 1987] K. Yajima, "Existence of solutions for Schrödinger evolution equations", *Comm. Math. Phys.* **110**:3 (1987), 415–426. MR Zbl

[Zhang and Fang 2012] T. Zhang and D. Fang, "Random data Cauchy theory for the generalized incompressible Navier–Stokes equations", *J. Math. Fluid Mech.* **14**:2 (2012), 311–324. MR Zbl

Received 9 Apr 2018.

JUSTIN T. BRERETON:
jbrere@berkeley.edu
Department of Mathematics, University of California, Berkeley, CA, United States

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the submission page.

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 are usually in English or French, 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 refer to bibliography keys. 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 a Mathematics Subject Classification for the article, and, for each author, affiliation (if appropriate) and email address.

Format. Authors are encouraged to use LaTeX and the standard amsart class, but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should normally 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 as many details as you can about how your graphics were generated.

Bundle your figure files into a single archive (using zip, tar, rar or other format of your choice) and upload on the link you been provided at acceptance time. Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables.

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.

Tunisian Journal of Mathematics

2019 vol. 1 no. 3