Pacific Journal of Mathematics

Volume 278 No. 2 December 2015

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

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

The subscription price for 2015 is US $420/year for the electronic version, and $570/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.278.257

DIFFERENTIAL HARNACK AND LOGARITHMIC SOBOLEV INEQUALITIES ALONG RICCI-HARMONIC MAP FLOW

ABIMBOLA ABOLARINWA

This paper introduces a new family of entropy functionals which is proved to be monotonically nondecreasing along the Ricci-harmonic map heat flow. Some of the consequences of the monotonicity are combined to derive gradient estimates and Harnack inequalities for all positive solutions to the associated conjugate heat equation. We relate the entropy monotonicity and the ultracontractivity property of the heat semigroup, and as a result we obtain the equivalence of logarithmic Sobolev inequalities, conjugate heat kernel upper bounds and uniform Sobolev inequalities under Ricci-harmonic map heat flow.

1. Introduction

Let (M, g) and (N, ξ) be compact Riemannian manifolds (without boundary) of dimensions m and n respectively. Let a smooth map $u : M \to N$ be a critical point of the Dirichlet energy integral $E(u) = \int_M |\nabla u|^2 \, d\mu_g$, where N is isometrically embedded in $\mathbb{R}^d$, $d \geq n$, by the Nash embedding theorem. The configuration $(g(x,t), u(x,t))$, $t \in [0, T)$, of a one-parameter family of Riemannian metrics $g(x,t)$ and a family of smooth maps $u(x,t)$ is defined to be a Ricci-harmonic map flow if it satisfies the coupled system of nonlinear parabolic equations denoted by $(\mathrm{RH})_\alpha$

$$
\begin{cases} \dfrac{\partial}{\partial t} g(x,t) = -2\,\mathrm{Rc}(x,t) + 2\alpha \nabla u(x,t) \otimes \nabla u(x,t), \\ \dfrac{\partial}{\partial t} u(x,t) = \tau_g u(x,t), \end{cases} \tag{1-1}
$$

where $\mathrm{Rc}(x,t)$ is the Ricci curvature tensor for the metric g, $\alpha(t) \equiv \alpha > 0$ is a time-dependent coupling constant and $\tau_g u$ is the intrinsic Laplacian of u which denotes the tension field of the map u. The system (1-1) was first studied by B. List [2008] in a special case, $N \subseteq \mathbb{R}$ and $\alpha = 2$, where the flow was modified by the Lie derivative of g with respect to a gradient vector field to give a gradient flow

MSC2010: 35J05, 53C44, 58J35, 58J60.
Keywords: Ricci-harmonic map heat flow, monotonicity formula, Harnack inequalities, ultracontractivity, heat semigroup, logarithmic Sobolev inequalities.

of an energy functional whose stationary points are solutions to the static Einstein vacuum equations arising in general relativity. This has since been generalised by R. Müller [2012] to the general case $N \hookrightarrow \mathbb{R}^d$, for sufficiently large d. Precisely, the system couples together the Ricci flow of Hamilton [1982] and the heat flow for harmonic maps of Eells and Sampson [1964]. The system $(\mathrm{RH})_\alpha$ is closer to the former in behaviours, such as in existence and singularities, though may be less singular than both. Hence, the analysis of the flow is usually done along the line of Ricci flow and for this, Perelman's works [2002; 2003b; 2003a] on Ricci flow are very applicable to the theory and applications of the Ricci-harmonic map flow.

In this paper we study the behaviour of all positive solutions to the associated conjugate heat equation along the Ricci-harmonic map flow. Let $h,\ H : M \times [0, T) \to (0, \infty)$ satisfy

$$\Big(\frac{\partial}{\partial t} - \Delta_g\Big)h = 0 \quad \text{and} \quad \Big(-\frac{\partial}{\partial t} - \Delta_g + R - \alpha|\nabla u|_g^2\Big)H = 0,$$

with

$$\int_0^T \int_M \Big(\frac{\partial}{\partial t} - \Delta_g\Big)hH \, d\mu_g \, dt = \int_0^T \int_M h\Big(-\frac{\partial}{\partial t} - \Delta_g + R - \alpha|\nabla u|_g^2\Big)H \, d\mu_g \, dt,$$

where Δ_g is the usual Laplace–Beltrami operator and $\Box^* := -\partial/\partial t - \Delta_g + R - \alpha|\nabla u|_g^2$ is the standard conjugate to the heat operator $\Box := \partial/\partial t - \Delta_g$. We say h and H are respectively solutions to the heat equation and conjugate heat equation. The main idea here is to solve the Ricci-harmonic map flow forward in time and solve the conjugate heat equation backward in time. Fixing the coordinate (y, s), $H = H(x, t; y, s)$ will be called the conjugate heat kernel (the positive minimal solution) if it tends to a δ-function as $t \to T$.

Our main results in the first part of this paper are Perelman's differential Harnack estimates for $f \in C^\infty(M \times [0, T))$ satisfying $H(x, \tau; y, s) = (4\pi\tau)^{-m/2} e^{-f(x,\tau)}$, $\tau = T - t$,

$$-\frac{d}{dt} f(\gamma(\tau), \tau) \le \frac{1}{2}\Big(|\gamma'(\tau)|^2 + S_g(\gamma(\tau), \tau) - \frac{m}{2\tau}\Big), \tag{1-2}$$

and Li–Yau Harnack estimates for all positive solutions to the conjugate heat equation

$$\frac{H(x_2, t_2)}{H(x_1, t_1)} \le \Big(\frac{\tau_1}{\tau_2}\Big)^{m/4} \exp\Big(\frac{1}{2}\int_{t_1}^{t_2} (|\gamma'(t)|^2 + S_g(\gamma(t), t)\, dt\Big), \tag{1-3}$$

where $S_g = R_g - \alpha|\nabla u|_g^2$. (The proofs of (1-2) and (1-3) are delayed until Section 4). Both results stated above are consequences of a monotonicity formula for a new entropy functional $\mathcal{W}_{\alpha,\epsilon}$ introduced in Section 3, where we obtain the Harnack inequality for $0 < \epsilon^2 \leq 4\pi$,

$$\frac{\epsilon^2 \tau}{4\pi}(2\Delta f - |\nabla f|^2 + S_g) + f - \frac{m\epsilon^2}{4\pi} \leq 0 \tag{1-4}$$

for all time t and prove that

$$\lim_{\tau \to 0} \int_M \frac{\epsilon^2 \tau}{4\pi}(2\Delta f - |\nabla f|^2 + S_g) + f - \frac{m\epsilon^2}{4\pi} hH \, d\mu_g \longrightarrow 0, \tag{1-5}$$

with the condition that $\epsilon^2 \to 4\pi$ as $t \to T$. Monotonicity formulas are generally useful in controlling solutions of evolution equations. This entropy is also intimately related to the logarithmic Sobolev inequality of Gross [1975]. Perelman used this property to obtain upper bounds for the fundamental solution to the adjoint heat equation via his reduced length. This leads to the proof of the noncollapsing theorem on Riemannian manifolds and, consequently, to the completion of R. Hamilton's program on the Poincaré conjecture. See [Perelman 2002; 2003b; 2003a; Cao et al. 2003]. Among several examples, Perelman's entropy and the gradient estimates of Li and Yau [1986] are important ones that show close relations between entropy monotonicity and the gradient estimate for the heat equation (forward or backward in time). Lei Ni [2004] has also considered a case for the linear heat equation on a static manifold with nonnegative Ricci curvature. We notice that coupling a heat-type equation with geometric flow began with [Hamilton 1993] and it has since become a very active research area and has led to numerous physical and geometric applications; for examples, see [Băileşteanu et al. 2010; Băileşteanu and Tran 2013; Cao and Zhang 2011; Kuang and Zhang 2008; List 2008; Müller 2012; Ni 2006; Zhang 2007] and the references therein.

Another important application of Perelman's $\mathcal{W}$-entropy monotonicity is in the derivation of uniform Sobolev inequalities by Q. Zhang [2007]; see also [Hsu 2008; Ye 2007]. In the second part of this paper, we relate the entropy monotonicity and the ultracontractivity property of the heat semigroup, and as a result we establish the equivalence of logarithmic Sobolev inequalities, conjugate heat kernel upper bounds and uniform Sobolev inequalities under Ricci-harmonic map heat flow. Precisely, let A_0 and B_0 be finite positive constants depending only on m, g_0, the lower bound for the Ricci curvature and the injectivity radius of M. For any $v \in W^{1,2}(M, g_0)$ such that

$$\|v\|_{2m/(m-2)} \leq A_0 \|\nabla v\|_2 + B_0 \|v\|_2, \tag{1-6}$$

where $m \geq 3$ and $\|\cdot\|_q = \left(\int_M |\cdot|^q \, d\mu_g\right)^{1/q}$, $1 \leq p < \infty$, we have the following result.

Theorem. *Let M be a compact Riemannian manifold of dimension $m \geq 3$. Let the solution to the* $(\mathrm{RH})_\alpha$*-flow exist for all times $t \in [0, T)$. Assume the Sobolev* *embedding (1-6) holds; then for finite positive constants A and B depending on m, A_0, B_0, the lower bound for R_{g_0} and T,*

$$\left(\int_M v^{2m/(m-2)}\,d\mu_g\right)^{(m-2)/2} \le A\int_M\left(|\nabla v|^2+\tfrac{1}{4}S_g v^2\right)d\mu_g + B\int_A v^2\,d\mu_g \tag{1-7}$$

and

$$\int_M v^2\ln v^2\,d\mu_{g(t)} \le \sigma^2\int_M(4|\nabla v|^2+S_g v^2)\,d\mu_{g(t)} - \frac{m}{2}\ln\sigma^2+(t+\sigma^2)\beta_1+\frac{m}{2}\ln\frac{mA}{2e}, \tag{1-8}$$

hold for each $t\in[0,T)$ and $v\in W^{1,2}(M)$ if $\lambda_\alpha=\inf_{\|v\|_2=1}\int_M(4|\nabla v|^2+S_g v^2)\,d\mu_{g_0}$; that is, $\lambda_{\alpha 0}$ is the first eigenvalue of the operator $-\Delta+\frac{1}{4}S_g$.

Finally, for some constant C depending on m, t, T, A_0, B_0 and $\sup S_g(\,\cdot\,,0)$, *the estimate*

$$H(x,T;y)\le CT^{-m/2} \tag{1-9}$$

for the positive solution to the conjugate heat equation associated to $(\mathrm{RH})_\alpha$ *holds.*

The three results in the above theorem are essentially equivalent, and their proofs occupy Sections 5 – 7. The approach to the proof here is *Sobolev inequality* (1-7) $\Longrightarrow$ *log-Sobolev inequality* (1-8) $\Longrightarrow$ *heat kernel upper bound* (1-9) $\Longrightarrow$ *Sobolev inequality* (1-7). Indeed, any of them can be derived from the other. The results of the above form [Hsu 2008; Ye 2007; Zhang 2007] yield a long time κ-noncollapsing estimate which generalises Perelman's short time result [2002] along the Ricci flow.

We recall that the nonnegativity of the scalar curvature R_g is preserved along Ricci flow [Chow and Knopf 2004], so the nonnegativity of S_g is also preserved as long as $(\mathrm{RH})_\alpha$ exists. Indeed, S_g evolves by a reaction-diffusion equation which helps to visualise its behaviour up to singular time (we discuss this in the next section). The condition $S_g=R_g-\alpha|\nabla u|_g^2\ge 0$ at the starting time $t=0$ must now be considered. The assumption is not necessary for the derivation of (1-7) since additional geometric data are not usually required to derive a Sobolev inequality from either a log-Sobolev inequality or the heat kernel bound. The assumption is required for the condition that a certain eigenvalue λ_α for the initial metric is positive, which is required to pass to (1-8). The class of manifold (M,g_0) with $\lambda_\alpha 0>0$ is a very large one and significant from a geometric point of view. Moreover, if $\lambda_\alpha 0>0$ for $S_g(0)>0$ (i.e., $R_g(0)>\alpha(0)|\nabla u(0)|^2$) then A, B are independent of time and $B=0$. Corollary 7.5 below presents corresponding versions of (1-7) and (1-8) in this case.

In the next section we discuss necessary background on Perelman–Müller entropy monotonicity formulas for $(\mathrm{RH})_\alpha$.

2. Background on entropy formulas for $(\mathrm{RH})_\alpha$-flow

Let (M, g) be a compact Riemannian manifold. For the metric g, any smooth functions $u \in C^\infty(M, N)$, $u(M) \subseteq N \hookrightarrow \mathbb{R}^n$, $f \in C^\infty(M)$ and constant $\alpha > 0$, Perelman and Müller's energy functional [Müller 2012] is defined on the triple (g, u, f) by

$$\mathcal{F}_\alpha(g, u, f) := \int_M (R_g + |\nabla f|_g^2 - \alpha|\nabla u|_g^2) e^{-f} \, d\mu_g, \tag{2-1}$$

which can also be written in two other ways,

$$\begin{aligned}\mathcal{F}_\alpha(g, u, f) &= \int_M (S_g + \Delta_g f) e^{-f} \, d\mu_g \\ &= \int_M (2\Delta_g f - |\nabla f|_g^2 + S_g) e^{-f} \, d\mu_g,\end{aligned}$$

since $\int_M \Delta(e^{-f}) = 0 = \int_M (-\Delta f + |\nabla f|_g^2) e^{-f} \, d\mu_g$. For any diffeomorphism $\phi : M \to M$, we have $\mathcal{F}_\alpha(\phi^* g, \phi^* u, \phi^* f) = \mathcal{F}_\alpha(g, u, f)$. If (g, u) is a solution to the system (1-1), Müller [2012] proved that the $\mathcal{F}_\alpha$-functional is nondecreasing under the flow and showed that the system is equivalent (after pulling back with a diffeomorphism generated by a vector field) to the gradient flow system for the energy functional $\mathcal{F}_\alpha$, locally written as,

$$\begin{cases} \dfrac{\partial}{\partial t} g_{ij} = -2R_{ij} + 2\alpha \nabla_i u \otimes \nabla_j u + 2\nabla_i \nabla_j f, \\ \dfrac{\partial}{\partial t} u = \tau_g u - \langle \nabla u, \nabla f \rangle, \\ \dfrac{\partial}{\partial t} f = -R + \alpha |\nabla u|^2 - \Delta f. \end{cases} \tag{2-2}$$

More precisely,

$$\begin{aligned}&\frac{d}{dt}\mathcal{F}_\alpha(g, u, f) \\ &\quad = 2\int_M \big(|\operatorname{Rc} - \alpha \nabla u \otimes \nabla u + \nabla\nabla f|^2 + \alpha|\tau_g u - \langle \nabla u, \nabla f \rangle|^2\big) e^{-f} \, d\mu_g \geq 0.\end{aligned} \tag{2-3}$$

An application of this is that $\mathcal{F}_\alpha$ is constant if and only if (g, u) is a steady gradient soliton.

Define

$$\lambda_\alpha(g) = \inf\left\{\mathcal{F}_\alpha(g, u, f) : f \in C^\infty(M), \int_M e^{-f} \, d\mu_g = 1\right\}.$$

Then $\lambda_\alpha(g)$ is the first eigenvalue of the operator $-4\Delta + S_g$, where the nondecreasing property of $\mathcal{F}_\alpha$ implies $\lambda_\alpha(g)$ is nondecreasing and we have, by setting $v = e^{-f/2}$, the corresponding normalised eigenvector,

$$-4\Delta v + S_g v = \lambda_\alpha(g)v.$$

Hence

$$\lambda_\alpha(g,u) = \inf\left\{\int_M (4|\nabla v|^2 + S_g v^2)\,d\mu_g : \int_M v^2\,d\mu_g = 1\right\}.$$

Similar to the case of Hamilton's Ricci flow, all geometric quantities associated with the source manifold evolve along $(\mathrm{RH})_\alpha$. For instance, we consider those quantities that are directly relevant at the present; the metric inverse, volume element, Laplace–Beltrami operator and S_g evolve as follows:

$$\frac{\partial}{\partial t}g^{ij} = 2S^{ij}, \qquad \frac{\partial}{\partial t}\Delta_g = 2S^{ij}\nabla_i\nabla_j - 2\alpha\tau_g u\langle\nabla u, \nabla\cdot\rangle,$$
$$\frac{\partial}{\partial t}d\mu_g = -S_g\,d\mu_g, \quad \frac{\partial}{\partial t}S_g = \Delta S_g + 2|S_{ij}|^2 + 2\alpha|\tau_g u|_g^2,$$

where $S_{ij} = \mathrm{Rc} - \alpha\nabla u\otimes\nabla u$ and $g^{ij}S_{ij} = S_g$. The nonnegativity of the curvature operator and S_g are preserved during the flow; for example, the evolution of $S_g = R_g - \alpha|\nabla u|^2$ is governed by the differential inequality

$$\frac{\partial}{\partial t}S_g \geq \Delta S_g + \frac{2}{m}S_g^2,$$

since $\alpha \geq 0$ and $|S_{ij}|^2 \geq (1/m)S_g^2$. Suppose $S_{g_0} \geq \rho$. We can use the maximum principle by comparing the solution of the above inequality with that of the ODE

$$\begin{cases} \dfrac{d\psi(t)}{dt} = \dfrac{2}{m}(\psi(t))^2, \\ \psi(0) = \rho, \end{cases} \tag{2-4}$$

solving to

$$\psi(t) = \frac{1}{\frac{1}{\rho} - \frac{2}{m}t}.$$

Therefore,

$$S_{g(t)} \geq \psi(t) = \frac{1}{\frac{1}{\rho} - \frac{2}{m}t} \tag{2-5}$$

for all $t \geq 0$ as long as the flow exists. We remark that (2-5) implies

$$S_{g(t)\min} \geq \frac{S_{g(0)\min}}{1 - (2t/m)S_{g(0)\min}}.$$

Clearly, $S_{g(0)\min} > 0$ implies $S_{g(t)\min} \to \infty$ in finite time $T_\epsilon \leq m/(2S_{g(0)\min}) < \infty$. This also implies that $R_{g(t)\min} \to \infty$ as $t \to T_\epsilon$, and thus $g(t)$ becomes singular in finite time $T_{\text{singular}} \leq T_\epsilon < \infty$.

Recall the Perelman–Müller $\mathcal{W}_\alpha$-entropy functional also introduced in [Müller 2012] as

$$\mathcal{W}_\alpha(g,u,f,\tau) := \int_M \big(\tau(S_g + |\nabla f|_g^2) + f - m\big)\frac{e^{-f}}{(4\pi\tau)^{m/2}}\, d\mu_g, \tag{2-6}$$

where $\tau \in \mathbb{R}$ is a real number, $f \in C^\infty(M \times [0,T))$, $\alpha > 0$ is a constant and $u \in C^\infty(M,N)$ is a harmonic map between the m-dimensional manifold M and the n-dimensional manifold N, which by the Nash embedding theorem is isometrically embedded in $\mathbb{R}^d$ for sufficiently large d. The above entropy functional is analogous to Perelman's $\mathcal{W}$-entropy for shrinkers [2002] under the Ricci flow. $\mathcal{W}_\alpha$ is equally used for shrinkers under Ricci-harmonic map flow as can be traced back to List [2008]. As pointed out in [Perelman 2002], such an entropy is invariant and monotone. In fact, given a constant $\lambda > 0$ and a diffeomorphism ϕ of M, under simultaneous scaling of g and τ, we have

$$\mathcal{W}_\alpha(\lambda g, u, f, \lambda\tau) = \mathcal{W}_\alpha(g,u,f,\tau),$$

and under the pullback of g, u and f, we have

$$\mathcal{W}_\alpha(\phi^* g, \phi^* u, \phi^* f, \tau) = \mathcal{W}_\alpha(g,u,f,\tau).$$

More importantly, we have the following monotonicity formula.

Proposition 2.1 [List 2008; Müller 2012]. *Let $\big(g(t), u(t), f(t), \tau(t)\big)$, $t \in [0,T)$ be a solution of the system*

$$\begin{cases} \dfrac{\partial}{\partial t} g = -2\,\mathrm{Rc} + 2\alpha \nabla u \otimes \nabla u, \\ \dfrac{\partial}{\partial t} u = \tau_g u, \\ \Big(-\dfrac{\partial}{\partial t} - \Delta_g + R - \alpha|\nabla u|_g^2\Big)\dfrac{e^{-f}}{(4\pi\tau)^{m/2}} = 0, \\ \dfrac{\partial}{\partial t}\tau = -1. \end{cases} \tag{2-7}$$

Then the $\mathcal{W}_\alpha$-entropy is nondecreasing with

$$\begin{aligned} \frac{d}{dt}\mathcal{W}_\alpha(g,u,f,\tau) = 2\tau \int_M \Big|\mathrm{Rc} - \alpha\nabla u \otimes \nabla u + \nabla\nabla f - \frac{1}{2\tau} g\Big|^2 \frac{e^{-f}}{(4\pi\tau)^{m/2}}\, d\mu_g \\ + 2\tau \int_M \alpha |\tau_g u - \langle \nabla u, \nabla f\rangle|^2 \frac{e^{-f}}{(4\pi\tau)^{m/2}}\, d\mu_g. \end{aligned} \tag{2-8}$$

Notice that the third equation in the above system is equivalent to the following backward heat equation

$$\frac{\partial f}{\partial t} = -\Delta_g f + |\nabla f|_g^2 - R + \alpha|\nabla u|_g^2 + \frac{m}{2\tau}, \tag{2-9}$$

A monotonicity formula of the above type is used to rule out any periodic solution to the $(\mathrm{RH})_\alpha$-flow other than those that are striking and Einstein [List 2008; Müller 2012; Perelman 2002].

Similar to $\lambda_\alpha(g,u)$ above, define the minimizing problem

$$\mu_\alpha(g,u,\tau) := \inf\left\{\mathcal{W}_\alpha(g,u,f,\tau) : f \in C^\infty(M),\ \int_M (4\pi\tau)^{-m/2} e^{-f}\, d\mu_g = 1\right\},$$

replacing f by $v = e^{-f/2}$. We have an equivalent minimizing integral

$$\mathcal{W}_\alpha(g,u,v,\tau) = \int_M \bigl(\tau(4|\nabla v|^2 + S_g v^2) - v^2 \ln v^2 - mv^2\bigr)(4\pi\tau)^{-m/2}\, d\mu_g$$

for functions $v \in H^1(M)$ with $\int_M v^2 (4\pi\tau)^{-m/2}\, d\mu_g = 1$. Then v satisfies the Euler–Lagrange equation, and it follows that $\mu_\alpha(g,u,\tau)$ is achieved by a minimizer f_τ satisfying

$$\tau(2\Delta f_\tau - |\nabla f_\tau|^2 + S_g) + f_\tau - n = \mu(g,\tau).$$

By the result of Perelman, it is well understood that for any metric g on a compact manifold M and $\tau > 0$, we have $\mu(g,u,\tau) > -\infty$ and it approaches zero as $\tau \to 0$.

3. A new entropy monotonicity formula

In this section we introduce a new family of dual entropy formulas, which are dual in the sense that they generalise Ni's entropy formula [2004] for the forward heat equation on the one hand and generalise Perelman and Müller's $\mathcal{W}_\alpha$-entropy on the other hand. A similar family of entropy functionals was constructed by Kuang and Zhang [2008]. The monotonicity property discussed here is very crucial to the derivation of our results in the rest of this paper.

Definition 3.1. Let $f : M \times [0,T] \to \mathbb{R}$ be smoothly defined with normalisation condition

$$\int_M \frac{e^{-f}}{(4\pi\tau)^{m/2}}\, d\mu_g = 1.$$

We define a generalised family of entropy by

(3-1)
$$\mathcal{W}_{\alpha,\epsilon}(g,u,f,\tau) = \int_M \left(\frac{\epsilon^2\tau}{4\pi}(S_g + |\nabla f|_g^2) + f - \frac{m\epsilon^2}{4\pi} + \frac{m}{2}\ln\frac{4\pi}{\epsilon^2}\right)\frac{e^{-f}}{(4\pi\tau)^{m/2}}\, d\mu_g,$$

where $\tau(t) = T - t > 0$, $0 < \epsilon^2 \le 4\pi$ and $S_g = S_g(x,t) = (R_g - \alpha|\nabla u|_g^2)(x,t)$.

Let $H = H(x,t)$ be a positive solution to the conjugate heat equation on a complete compact manifold with metric $g = g(x,t)$, evolving by the $(\mathrm{RH})_\alpha$. Let

$H=(4\pi\tau)^{-m/2}e^{-f}$ and $\int_M H\,d\mu_g=1$. Then

$$\Big(-\frac{\partial}{\partial t}-\Delta_g+S_g\Big)H=0. \tag{3-2}$$

Theorem 3.2. *Suppose that* $(g(t),u(t))$, $t\in[0,T)$, *solves* $(\mathrm{RH})_\alpha$ *with* $\alpha(t)\equiv\alpha>0$ *and* τ *is a backward time with* $\partial\tau/\partial t=-1$. *Suppose that* $H:M\times[0,T)\to(0,\infty)$ *solves the conjugate heat equation* $(-\partial/\partial t-\Delta_g+S_g)H=0$. *The entropy functional* $\mathcal{W}_{\alpha,\epsilon}$ *is nondecreasing by the formula*

$$\begin{aligned}\frac{d}{dt}&\mathcal{W}_{\alpha,\epsilon}(g,u,f,\tau)\\ &\geq\frac{\epsilon^2\tau}{2\pi}\int_M\Big(\Big|\mathrm{Rc}-\alpha\nabla u\otimes\nabla u+\nabla\nabla f-\frac{1}{2\tau}g\Big|^2+\alpha|\tau_g u-\langle\nabla u,\nabla f\rangle|^2\Big)H\,d\mu_g\end{aligned}\tag{3-3}$$

for $0<\epsilon^2\leq 4\pi$.

Remark 3.3. We remark that if $\epsilon^2=4\pi$, we recover Perelman and Müller's $\mathcal{W}_\alpha$-entropy.

Scaling and diffeomorphism invariance of $\mathcal{W}_{\alpha,\epsilon}$. Before we prove the monotonicity formula (3-3), we shall first establish the invariance of our new entropy with respect to dilation and diffeomorphism.

Lemma 3.4 [Chow and Knopf 2004, Lemma 6.57]. *If* g *and* h *are two Riemannian metrics on an* n*-dimensional Riemannian manifold and they are related by the time-scale factor* λ (*i.e.*, $g=\lambda h$), *then the various geometric quantities scale as follows*:

$$\begin{gathered} g^{ij}=\frac{1}{\lambda}h^{ij},\qquad \Gamma^k_{ij\,(g)}=\Gamma^k_{ij\,(h)},\\ R^l{}_{ijk}(g)=R^l{}_{ijk}(h),\qquad R_{ijkl}(g)=\phi R_{ijkl}(h),\\ R_{ij}(g)=R_{ij}(h),\qquad R_{(g)}=\frac{1}{\lambda}R_{(h)},\qquad d\mu_{(g)}=\lambda^{n/2}\,d\mu_{(h)}.\end{gathered}$$

Lemma 3.5. *Let* $\lambda>0$ *be any constant and* $\phi:M\to M$ *be a one-parameter family of diffeomorphisms. Then*

$$\begin{aligned}\mathcal{W}_{\alpha,\epsilon}(\lambda g,u,f,\lambda\tau)&=\mathcal{W}_{\alpha,\epsilon}(g,u,f,\tau),\\ \mathcal{W}_{\alpha,\epsilon}(\phi^*g,\phi^*u,\phi^*f,\tau)&=\mathcal{W}_{\alpha,\epsilon}(g,u,f,\tau).\end{aligned}$$

Proof. By a straightforward computation, we have

$$\begin{aligned}\mathcal{W}_{\alpha,\epsilon}(\lambda g,u,f,\lambda\tau)=\int_M\bigg(&\frac{\epsilon^2\lambda\tau}{4\pi}(R(\lambda g)-\alpha(\lambda g)^{ij}\nabla_i u\otimes\nabla_j u+(\lambda g)^{ij}\nabla_i f\nabla_j f)\\ &+f-\frac{m\epsilon^2}{4\pi}+\frac{m}{2}\ln\frac{4\pi}{\epsilon^2}\bigg)\frac{e^{-f}}{(4\pi\lambda\tau)^{m/2}}\sqrt{\det(\lambda g)}\,dx\end{aligned}$$

$$= \int_M \biggl(\frac{\epsilon^2 \lambda \tau}{4\pi} (\lambda^{-1} R(g) - \alpha \lambda^{-1} g^{ij} \nabla_i u \otimes \nabla_j u + \lambda^{-1} g^{ij} \nabla_i f \nabla_j f) \\ + f - \frac{m\epsilon^2}{4\pi} + \frac{m}{2} \ln \frac{4\pi}{\epsilon^2} \biggr) \frac{e^{-f}}{\lambda^{m/2}(4\pi\tau)^{m/2}} \sqrt{\lambda^m \det(g)}\, dx$$

$$= \int_M \biggl(\frac{\epsilon^2 \tau}{4\pi} (R_g - \alpha |\nabla u|_g^2 + |\nabla f|_g^2) \\ + f - \frac{m\epsilon^2}{4\pi} + \frac{m}{2} \ln \frac{4\pi}{\epsilon^2} \biggr) \frac{e^{-f}}{(4\pi\tau)^{m/2}}\, d\mu_g$$

$$= \mathcal{W}_{\alpha,\epsilon}(g, u, f, \tau).$$

The invariance under diffeomorphisms is trivial since $(\mathrm{RH})_\alpha$-flow is equivalent to the flow modified by the time-dependent diffeomorphism ϕ generated by the gradient of f, where $\phi^* g$ is the pulled-back metric and $\phi^* f = f \circ \phi$. For the harmonic map u, the invariance holds if we combine the following facts: ϕ is a C^∞-diffeomorphism and $u \in C^\infty(M, N)$ is a harmonic map with respect to (M, g); then $\phi^* u = u \circ \phi \in C^\infty(M, N)$ is a harmonic map with respect to $(M, \phi^* g)$ with the identity

$$\int_M |\nabla u|_g^2\, d\mu_g = \int_M |\nabla (u \circ \phi)|_{\phi^* g}^2\, d\mu_{\phi^* g}.$$

Then, all the geometric quantities are invariant under $(\mathrm{RH})_\alpha$-flow and the diffeomorphism invariance of $\mathcal{W}_{\alpha,\epsilon}$ follows. □

Proof of Theorem 3.2 (the monotonicity formula for $\mathcal{W}_{\alpha,\epsilon}$).

Proof. The entropy functional can be rewritten as

$$\mathcal{W}_{\alpha,\epsilon}(g, u, f, \tau) \\ = \frac{\epsilon^2}{4\pi} \int_M (\tau(S_g + |\nabla f|^2) + f - m) H\, d\mu_g + \left(1 - \frac{\epsilon^2}{4\pi}\right) \int_M f H\, d\mu_H + \frac{m}{2} \ln \frac{4\pi}{\epsilon^2}.$$

By direct computation we obtain the evolution equation

$$\frac{d}{dt} \mathcal{W}_{\alpha,\epsilon}(g, u, f, \tau) = \frac{\epsilon^2}{4\pi} \frac{\partial}{\partial t} \left(\int_M V\, d\mu_g \right) + \left(1 - \frac{\epsilon^2}{4\pi}\right) \frac{\partial}{\partial t} \left(\int_M f H\, d\mu_g \right), \tag{3-4}$$

where

$$V := (\tau(2\Delta_g f + S_g - |\nabla f|^2) + f - m) H \tag{3-5}$$

since $\int_M (\Delta_g f - |\nabla f|_g^2) e^{-f}\, d\mu_g = 0$ on a closed manifold M. We make two claims here, which we shall prove in the next two propositions, namely,

$$\frac{\partial}{\partial t} \left(\int_M V\, d\mu_g \right) = \int_M -\square^* V\, d\mu_g = \frac{d}{dt} \mathcal{W}_\alpha(g, u, f, \tau) \tag{3-6}$$

and

$$\frac{\partial}{\partial t}\left(\int_M fH\,d\mu_g\right) = -\mathcal{F}_\alpha(g,u,f) + \frac{m}{2\tau} \geq 0. \tag{3-7}$$

With the above two claims, we arrive at

$$\frac{d}{dt}\mathcal{W}_{\alpha,\epsilon}(g,u,f,\tau) = \frac{\epsilon^2}{4\pi}\frac{d}{dt}\mathcal{W}_\alpha(g,u,f,\tau) + \frac{m}{2\tau} - \mathcal{F}_\alpha(g,u,f), \tag{3-8}$$

which proves the monotonicity formula (3-3) for $0 < \epsilon^2 \leq 4\pi$. □

Proposition 3.6. *With the assumptions of Theorem 3.2, the quantity*

$$V := \big(\tau(2\Delta_g f + S_g - |\nabla f|^2) + f - m\big)H$$

satisfies

$$\Box^* V = -2\tau\left(\left|\mathrm{Rc} - \alpha\nabla u\otimes\nabla u + \nabla\nabla f - \frac{1}{2\tau}g\right|^2 + \alpha|\tau_g u - \langle\nabla u, \nabla f\rangle|^2\right)H \tag{3-9}$$

and

$$\frac{d}{dt}\mathcal{W}_\alpha(g,u,f,\tau) = -\int_M \Box^* V\,d\mu_g. \tag{3-10}$$

Moreover if H tends to a δ-function as $t \to T$, then $V \leq 0$ for all $t < T$ with $H(x,\tau)$ replaced with $H(x,\tau;y,\sigma)$, the fundamental solution.

Proof. Let $P = \tau(2\Delta f - |\nabla f|^2 + S_g) + f - n$, and $\partial_t\tau = -1$ since $\tau = T - t$. Thus,

$$\begin{aligned}\Box^* V &= (-\partial_t - \Delta + S_g)(PH)\\ &= -(\partial_t + \Delta)PH - 2\langle\nabla P, \nabla H\rangle\end{aligned}$$

and

$$H^{-1}\Box^* V = -(\partial_t + \Delta)P + 2\langle\nabla P, \nabla f\rangle$$

since $f = -\ln H - (m/2)\ln(4\pi\tau)$ implies that $\nabla f = -H^{-1}\nabla H$. Let us compute $(\partial_t + \Delta)P$ as follows:

$$\frac{\partial P}{\partial t} = -(2\Delta f - |\nabla f|^2 + S_g) + \tau\frac{\partial}{\partial t}(2\Delta f - |\nabla f|^2) + \tau\frac{\partial}{\partial t}S_g + \frac{\partial}{\partial t}f. \tag{3-11}$$

Note that

$$\begin{aligned}\partial_t f &= -\Delta_g f - S_g + |\nabla f|_g^2 + \frac{m}{2\tau},\\ \partial_t S_g &= \Delta S_g + 2|S_{ij}|_g^2 + 2\alpha|\tau_g u|_g^2.\end{aligned}$$

Then a straightforward computations yields

$$2\frac{\partial}{\partial t}(\Delta f) = 4S_{ij}\nabla_i\nabla_j f - 4\alpha\tau_g u\langle\nabla u, \nabla f\rangle + 2\Delta(-\Delta f + |\nabla f|^2 - S_g), \tag{3-12}$$

$$\frac{\partial}{\partial t}|\nabla f|^2 = \frac{\partial}{\partial t}(g^{ij}\nabla_i f\nabla_j f) = 2S_{ij}\nabla_i f\nabla_j f + 2\left\langle\nabla f, \nabla\frac{\partial}{\partial t}f\right\rangle. \tag{3-13}$$

Combining (3-11)–(3-13) with the identity $\Delta P=\tau\big(2\Delta(\Delta f)-\Delta|\nabla f|^2+\Delta S_g\big)+\Delta f$, we have

$$\begin{aligned}\Big(\frac{\partial}{\partial t}+\Delta\Big)P &= -2\Delta f-2|\nabla f|^2-2S_g+\frac{m}{2\tau}\\ &\quad+\tau\big(4S_{ij}\nabla_i\nabla_j f-4\alpha\tau_g u\langle\nabla u,\nabla f\rangle+\Delta|\nabla f|^2+2\langle\nabla f,\nabla\Delta f\rangle\\ &\qquad-2S_{ij}\nabla_i f\nabla_j f+2|S_{ij}|^2+2\alpha|\tau_g u|^2-2\langle\nabla f,\nabla|\nabla f|^2\rangle+2\langle\nabla f,\nabla S_g\rangle\big).\end{aligned}$$

Similarly,

$$\begin{aligned}2\langle\nabla P,\nabla f\rangle &= 2\langle\nabla(\tau(2\Delta f-|\nabla f|^2+S-g)+f),\nabla f\rangle\\ &= 2\tau\big(2\langle\nabla\Delta f,\nabla f\rangle-\langle\nabla|\nabla f|^2,\nabla f\rangle+\langle\nabla S_g,\nabla f\rangle\big)+2|\nabla f|^2.\end{aligned}$$

Therefore,

$$\begin{aligned}&-\Big(\frac{\partial}{\partial t}+\Delta\Big)P+2\langle\nabla P,\nabla f\rangle\\ &\quad=\Big(2\Delta f+2S_g-\frac{m}{2\tau}\Big)-\tau\big(4S_{ij}\nabla_i\nabla_j f-4\alpha\tau_g u\langle\nabla u,\nabla f\rangle+\Delta|\nabla f|^2\\ &\qquad\qquad-2\langle\nabla f,\nabla\Delta f\rangle-2S_{ij}\nabla_i f\nabla_j f+2|S_{ij}|^2+2\alpha|\tau_g u|^2\big)\\ &\quad=\Big(2\Delta f+2S_g-\frac{m}{2\tau}\Big)-\tau\big(4S_{ij}\nabla_i\nabla_j f+2|\nabla\nabla f|^2+2|S_{ij}|^2\big)\\ &\qquad\qquad-2\tau\alpha(|\tau_g u+\langle\nabla u,\nabla f\rangle^2-2\tau_g u\langle\nabla u,\nabla f\rangle)\\ &\quad=-2\tau\Big(2S_{ij}\nabla_i\nabla_j f+|\nabla\nabla f|^2+|S_{ij}|^2-\frac{1}{\tau}\Big(\Delta f+R-\frac{m}{4\tau}\Big)\Big)\\ &\qquad\qquad-2\tau\alpha(|\tau_g u-\langle\nabla u,\nabla f\rangle|^2)\\ &\quad=-2\tau\Big((S_{ij}+\nabla_i\nabla_j f)^2-\frac{1}{\tau}\Big(\Delta f+R-\frac{m}{4\tau}\Big)\Big)-2\tau\alpha(|\tau_g u-\langle\nabla u,\nabla f\rangle|^2)\\ &\quad=-2\tau\Big|S_{ij}+\nabla_i\nabla_j f-\frac{1}{2\tau}g_{ij}\Big|^2-2\tau\alpha(|\tau_g u-\langle\nabla u,\nabla f\rangle|^2),\end{aligned}$$

where we have used the following calculation by Bochner's identity:

$$\begin{aligned}\Delta|\nabla f|^2-2\langle\nabla f,\nabla\Delta f\rangle-2S_{ij}\nabla_i f\nabla_j f &= 2|\nabla\nabla f|^2+2(R_{ij}-S_{ij})\nabla_i f\nabla_j f\\ &= 2|\nabla\nabla f|^2+2\alpha\langle\nabla u,\nabla f\rangle^2.\end{aligned}$$

Hence,

$$H^{-1}\Box^* V=-2\tau\Big|S_{ij}+\nabla_i\nabla_j f-\frac{1}{2\tau}g_{ij}\Big|^2-2\tau\alpha(|\tau_g u-\langle\nabla u,\nabla f\rangle|^2)$$

and

$$\Box^* V=-2\tau\Big|S_{ij}+\nabla_i\nabla_j f-\frac{1}{2\tau}g_{ij}\Big|^2 H-2\tau\alpha(|\tau_g u-\langle\nabla u,\nabla f\rangle|^2)H.$$

The consequence of which is a localised version of Perelman's $\mathcal{W}$-entropy monotonicity formula. Thus,

$$
\begin{aligned}
\frac{d\mathcal{W}}{dt} = \frac{\partial}{\partial t}\int_M V\,d\mu_g &= \int_M (\partial_t V - RV + \alpha|\nabla u|_g^2 V)\,d\mu \\
&= \int_M (-\Box^* V - \Delta_g V)\,d\mu_g = \int_M -\Box^* V\,d\mu_g \\
&= 2(T-t)\int_M \Big(\Big|S_{ij} + \nabla_i\nabla_j f - \frac{1}{2(T-t)} g_{ij}\Big|^2 \\
&\qquad + \alpha(|\tau_g u - \langle \nabla u, \nabla f\rangle|^2)\Big)\frac{e^{-f}}{(4\pi\tau)^{-m/2}}\,d\mu_g. \qquad \square
\end{aligned}
$$

Proposition 3.7. *With the assumptions of Theorem 3.2, we have*

$$
\frac{\partial}{\partial t}\Big(\int_M fH\,d\mu_g\Big) \geq 0. \tag{3-14}
$$

Proof. By direct computation,

$$
\begin{aligned}
\frac{\partial}{\partial t}\Big(\int_M fH\,d\mu\Big) &= \int_M \Big(\frac{\partial}{\partial t} fH + f\frac{\partial}{\partial t}H - S_g fH\Big)\,d\mu_g \\
&= \int_M \Big(-\Delta_g f + |\nabla f|_g^2 - S_g + \frac{m}{2\tau}\Big) H\,d\mu_g \\
&\qquad\qquad + \int_M f(-\Delta_g H + S_g H)\,d\mu - \int_M S_g f H\,d\mu_g \\
&= \int_M (-2\Delta_g f + |\nabla f|_g^2) H\,d\mu + \int_M \Big(\frac{m}{2\tau} - S_g\Big) H\,d\mu_g,
\end{aligned}
$$

where we used integration by parts on $-\int_M \Delta_g f H = -\int_M f\Delta_g H$. Rearranging the above, we have

$$
\begin{aligned}
\frac{\partial}{\partial t}\Big(\int_M fH\,d\mu_g\Big) &= \int_M (-S_g - 2\Delta_g f + |\nabla f|_g^2) H d\mu_g + \frac{m}{2\tau}\int_M H\,d\mu_g \\
&= -\int_M (S_g + |\nabla f|_g^2) H\,d\mu_g + \frac{m}{2\tau} \\
&= -\mathcal{F}_\alpha + \frac{m}{2\tau},
\end{aligned}
$$

where $\mathcal{F}_\alpha = \int_M (S_g + |\nabla f|_g^2) H\,d\mu_g$ is the Perelman energy functional introduced in [Müller 2012], which we discussed in Section 2. Next is to show that

$$
\frac{\partial}{\partial t}\Big(\int_M fu\,d\mu\Big) = -\mathcal{F}_\alpha + \frac{m}{2\tau} \geq 0. \tag{3-15}
$$

Recall the evolution of $\mathcal{F}_\alpha$:

$$
\frac{d}{dt}\mathcal{F}_\alpha(g,f) = 2\int_M \big(|\operatorname{Rc} - \alpha\nabla u\otimes\nabla u + \nabla\nabla f|^2 + \alpha|\tau_g u - \langle\nabla u, \nabla f\rangle|^2\big) H\,d\mu_g. \tag{3-16}
$$

Straightforward analysis, using an elementary inequality and the Cauchy–Schwarz inequality, gives

$$|\operatorname{Rc} - \alpha \nabla u \otimes \nabla u + \nabla\nabla f|^2 \geq \frac{1}{m}(R_g - \alpha|\nabla u|_g^2 + \Delta_g f)^2 \tag{3-17}$$

so that

$$\int (S_g + \Delta_g f) H \, d\mu_g \leq \left(\int (S_g + \Delta_g f)^2 H \, d\mu_g \right)^{1/2} \left(\int H \, d\mu_g \right)^{1/2},$$

which implies

$$\left(\int_M (S_g + \Delta_g f) H \, d\mu \right)^2 \leq \int_M (S_g + \Delta_g f)^2 H \, d\mu_g.$$

Hence by (3-16) and (3-17), we obtain

$$\frac{d}{dt}\mathcal{F}_\alpha \geq \frac{2}{m} \int_M (S_g + \Delta_g f)^2 H \, d\mu_g + \int_M 2\alpha |\tau_g u - \langle \nabla u, \nabla f \rangle|^2 H \, d\mu_g. \tag{3-18}$$

We can then solve

$$\frac{d}{dt}\mathcal{F}_\alpha \geq \frac{2}{m}\mathcal{F}_\alpha^2, \quad \mathcal{F}_\alpha \geq 0.$$

This implies

$$\frac{d\mathcal{F}_\alpha}{\mathcal{F}_\alpha^2} \geq \frac{2}{m}\, dt \Longrightarrow -\frac{1}{\mathcal{F}_\alpha}\bigg|_t^T \geq \frac{2}{m}(T - t) \Longrightarrow \frac{1}{\mathcal{F}_\alpha(t)} - \frac{1}{\mathcal{F}_\alpha(T)} \geq \frac{2}{m}\tau$$

$$\Longrightarrow \frac{1}{\mathcal{F}_\alpha(t)} \geq \frac{2}{m}\tau + \frac{1}{\mathcal{F}_\alpha(T)}.$$

From here we can conclude as follows:

(i) Suppose $\mathcal{F}_\alpha(T) > 0$. Then

$$\frac{1}{\mathcal{F}(t)} \geq \frac{2\tau}{m}; \quad \text{i.e, } \mathcal{F}_\alpha(t) \leq \frac{m}{2\tau}.$$

(ii) Suppose $\mathcal{F}_\alpha(T) \leq 0$. Then $\mathcal{F}_\alpha(t) \leq 0 \leq m/(2\tau)$ for all $t \in [0, T)$, since we know that $d\mathcal{F}_\alpha/dt \geq 0$.

Hence

$$\mathcal{F}_\alpha(t) \leq \frac{m}{2\tau} \quad \text{for } t \in [0, T),$$

which proves the claim (3-15). □

4. Differential Harnack estimates

In this section we obtain Perelman's differential Harnack-type estimate which holds for the fundamental solution and, of course, all positive solutions to the conjugate heat equation coupled to the Ricci-harmonic map flow. There is an improvement over some known results as there is no explicit restriction on the curvature and no recourse to Perelman's reduced distance. In what follows, we want to show that the local entropy satisfies a pointwise differential inequality for the positive minimal solution. Define a differential Harnack quantity

$$P_\epsilon := \frac{\epsilon^2\tau}{4\pi}(2\Delta f - |\nabla f|^2 + S_g) + f + \frac{m}{2}\ln\frac{4\pi}{\epsilon^2} - \frac{m\epsilon^2}{4\pi}.$$

Theorem 4.1. *Let M be a closed manifold with bounded Ricci curvature and $H(x, y, t) = H = (4\pi t)^{-n/2}e^{-f}$ satisfy $\Box^* H = 0$, where H tends to a δ-function as $t \to T$ and satisfies $\int_M H\,d\mu_g = 1$. Then for all $t < T$ and $\epsilon^2 \to 4\pi$ as $t \to T$, we have*

$$\frac{\epsilon^2\tau}{4\pi}(2\Delta f - |\nabla f|^2 + S_g) + f - \frac{m\epsilon^2}{4\pi} \le 0. \tag{4-1}$$

Proof. Let h be any compactly supported smooth function for all $t_0 > 0$. Suppose $h(\cdot\,, t)$ is a positive solution to the ordinary heat equation $(\partial_t - \Delta)h = 0$ (this is Perelman's argument in [2002, Corollary 9.3]). Then, it is clear that

$$\frac{\partial}{\partial t}\int_M H h\,dV = 0$$

and we have by direct calculation that

$$\begin{aligned}\frac{\partial}{\partial t}\int_M hP_\epsilon H\,d\mu_g &= \int_M \big(\partial_t h(P_\epsilon H) + h\partial_t(P_\epsilon H) - S_g P_\epsilon H\big)\,d\mu_g\\ &= \int_M \big((\partial_t - \Delta)h(P_\epsilon H) + h(\partial_t + \Delta - S_g)P_\epsilon H)\big)\,d\mu_g\\ &= -\int_M h\Box^*(P_\epsilon H)\,d\mu_g\\ &= -\frac{\epsilon^2}{4\pi}\int_M h\Box^* V_\epsilon\,d\mu_g \ge 0.\end{aligned}$$

The inequality is due to Proposition 3.6. We are left to show that for the everywhere positive function $h(\cdot\,, t)$, the limit of $\int_M hV_\epsilon\,d\mu_g$ is nonpositive as $t \to T$. We assume the claim a priori, i.e, $\lim_{t\to T}\int_M hV_\epsilon\,d\mu_g = 0$, with

$$V_\epsilon = \Big(\tau(2\Delta f - |\nabla f|^2 + S_g) + \frac{4\pi}{\epsilon^2}f - m\Big)H,$$

and conclude the result. □

For completeness, we devote the next effort to justifying the claim

$$\lim_{t\to T}\int_M hV_\epsilon\,d\mu_g \le 0 \iff \lim_{t\to T}\int_M hP_\epsilon H\,d\mu_g \le \frac{m}{2}\lim_{t\to T}\Big(\ln\frac{4\pi}{\epsilon^2}\int_M hH\,d\mu_g\Big). \tag{4-2}$$

Our argument follows from [Ni 2006; Perelman 2002] and can be compared with the recent preprint [Băileşteanu and Tran 2013, Proposition 4.2] (see also [Chow et al. 2008, Section 16.4]), where we know that $\lim_{t\to T}\int_M Vh\,d\mu_g \le 0$ (where V is as defined in Proposition 3.6). To see this clearly, we write

$$P_\epsilon H = \frac{\epsilon^2}{4\pi}V_\epsilon + \frac{m}{2}\ln\frac{4\pi}{\epsilon^2}H,$$

which implies

$$\lim_{t\to T}\int_M hP_\epsilon H\,d\mu_g = \frac{\epsilon^2}{4\pi}\lim_{t\to T}\int_M V_\epsilon h\,d\mu_g + \frac{m}{2}\ln\frac{4\pi}{\epsilon^2}\lim_{t\to T}\int_M hH\,d\mu_g. \tag{4-3}$$

If H tends to a Dirac δ-function, say at a point $p\in M$, for $\tau\to T$, then f satisfies $f(x,\tau)\to d^2(p,\ x)/4\tau$. This is in relation to the l-length of Perelman. This yields

$$\lim_{\tau\to 0}\int_M fhH\,d\mu_g \le \limsup_{\tau\to 0}\int_M \frac{d^2(p,x)}{4\tau}hH\,d\mu_g = \frac{m}{2}h(p,T). \tag{4-4}$$

Meanwhile, by the strong maximum principle, we have $h(x,T)>0$ and

$$\lim_{\tau\to 0}\int_M hH\,d\mu_g = h(x,T).$$

Hence by a scaling argument, we assume that $h(x,T)=1$. Rewriting P_ϵ and using integration by parts, we have

$$\begin{aligned}\int_M P_\epsilon hH\,d\mu_g = \int_M \frac{\epsilon^2\tau}{4\pi}\Big(|\nabla f|^2 + S_g\,d\mu_g - \frac{m}{2\tau}\Big)hH\,d\mu_g - \frac{\epsilon^2\tau}{2\pi}\int_M \langle\nabla f,\nabla h\rangle H\,dV\\ + \int_M fHh\,d\mu_g + \frac{m}{2}\Big(\ln\frac{4\pi}{\epsilon^2} - \frac{\epsilon^2}{4\pi}\Big)\int_M Hh\,d\mu_g.\end{aligned}$$

We should also note that since $h(\,\cdot\,,t_0)$ is compactly supported and by the strong maximum principle, we have that $h(\,\cdot\,,t_0)$, $|\nabla h(\,\cdot\,,t_0)|$ and $|\Delta h(\,\cdot\,,t_0)|$ are bounded on M. This implies that there exists a bounded solution $h(\,\cdot\,,t_0)$. Now we claim that the first three terms on the right-hand side of the last equation vanish as $\tau\to 0$. We can see this, for instance, in the following argument: By integration by parts and the fact that $\nabla H = -H\nabla f$, we have

$$-\tau\int_M \langle\nabla f,\nabla h\rangle H\mu_g = \tau\int_M \langle\nabla H,\nabla h\rangle\mu_g = -\tau\int_M H\Delta h\mu_g \tag{4-5}$$

is bounded since $|\Delta h|$ is bounded as stated earlier. Thus, the second term in right-hand side of the preceding equation is bounded and goes to zero as $\tau\to 0$, so the same is true for first terms (which follows from gradient estimates [Chow et al. 2008, Lemma 16.47]). Thus the analysis is reduced to showing that

$$\lim_{\tau\to 0}\int V_\epsilon h\,d\mu_g < C(m) \le 0. \tag{4-6}$$

By the monotonicity formula for $W_{\alpha,\epsilon}$, we have

$$\frac{\partial}{\partial t}\int_M P_\epsilon hH\,d\mu_g = \frac{\epsilon^2\tau}{4\pi}\frac{\partial}{\partial t}\int_M V_\epsilon h\,d\mu_g \geq 0.$$

By the mean value theorem, there exists a sequence $\tau_k \to 0$ such that

$$\lim_{\tau_k\to 0}\tau_k\int_M\Big(\Big|\mathrm{Rc}-\alpha\nabla u\otimes\nabla u+\nabla\nabla f-\frac{n}{2\tau_k}g\Big|^2+\alpha|\tau_g u-\langle\nabla u,\nabla f\rangle|^2\Big)Hh\,d\mu_g=0.$$

Applying the Cauchy–Schwarz and Hölder inequalities, we have

$$\Big|\mathrm{Rc}-\alpha\nabla u\otimes\nabla u+\nabla\nabla f-\frac{1}{2\tau_k}g\Big|^2\geq\frac{1}{m}\Big(R_g-\alpha|\nabla u|_g^2+\Delta_g f-\frac{n}{2\tau_k}g\Big)^2$$

and

$$\begin{aligned}&\int_M\tau_k\Big(S_g+\Delta f-\frac{n}{2\tau_k}\Big)Hh\,d\mu_g\\ &\leq\Big(\tau_k^2\int_M\Big(S_g+\Delta f-\frac{n}{2\tau_k}\Big)^2Hh\,d\mu_g\Big)^{1/2}\Big(\int_M Hh\,d\mu_g\Big)^{1/2}\\ &\leq\sqrt{m}\Big(\tau_k^2\int_M\Big|\mathrm{Rc}-\alpha\nabla u\otimes\nabla u+\nabla\nabla f-\frac{1}{2\tau_k}\Big|^2Hh\,d\mu_{g(\tau_k)}\Big)^{1/2}\Big(\int_M Hh\,d\mu_g\Big)^{1/2}\to 0\end{aligned}$$

as $\tau_k\to 0$, since $\alpha|\tau_g u-\langle\nabla u,\nabla f\rangle|^2\geq 0$ and $\lim_{\tau_k\to 0}\int_M Hh\,d\mu_{g(\tau_k)}$ is finite.

Then we have

$$\begin{aligned}&\lim_{t\to T}\int V_\epsilon h\,d\mu_g\\ &\quad=\lim_{t\to T}\int_M\Big(\frac{\epsilon^2}{4\pi}\tau_k(2\Delta f-|\nabla f|^2+S_g)+\frac{4\pi}{\epsilon^2\tau}f-m\Big)Hh\,d\mu_g\\ &\quad=\lim_{t\to T}\int_M\Big(\frac{\epsilon^2\tau_k}{4\pi}\Big(\Delta f+S_g-\frac{n}{2\tau_k}\Big)\Big)Hh\,d\mu_{g(\tau_k)}\\ &\qquad+\lim_{t\to T}\int_M\Big(\frac{\epsilon^2\tau_k}{4\pi}(\Delta f-|\nabla f|^2)\Big)Hh\,d\mu_g+\lim_{t\to T}\int_M\Big(f-\frac{m\epsilon^2}{8\pi}\Big)Hh\,d\mu_g\\ &\quad=\lim_{t\to T}\int_M\Big(f-\frac{m\epsilon^2}{8\pi}\Big)Hh\,d\mu_g,\end{aligned}$$

where we have used the identity

$$\int_M(\Delta f-|\nabla f|^2)H\,d\mu=-\int_M\Delta H\,d\mu=0$$

for any positive solution H and the fact that each quantity in (4-5) is bounded to obtain $\lim_{t\to T}\tau_k\int_M(\Delta f-|\nabla f|^2)Hh\,d\mu_g=0$.

By (4-4) and the asymptotic behaviour of the heat kernel, i.e, $f \approx d^2/(4\tau)$ as $\tau \to 0$, we have (see [Ni 2006, Theorem 2.1])

$$H(x, y, \tau) \sim (4\pi\tau)^{-m/2} \exp\left(\frac{d^2(x, y)}{4\tau}\right) \sum_{j=0}^{\infty} u_j(x, y, \tau)\tau^j := w_k(x, y, \tau)$$

as $\tau \to 0$, where $d^2(x, y)$ is the distance function and $w_k(x, y, t)$ satisfies

$$w_k(x, y, \tau) = O\left(\tau^{k+1-m/2} \exp\left(\frac{\delta d^2(x, y)}{4\tau}\right)\right)$$

uniformly for all $x, y \in M$ and δ is just a number depending only on the geometry of (M, g). The function can be chosen such that $u_0(x, y, 0) = 1$. Though, the above asymptotic result does not require any curvature assumption, a result due to Cheeger and Yau [1981] states that on a manifold with bounded Ricci curvature (which is our case), the heat kernel satisfies

$$H(x, y, \tau) \geq (4\pi\tau)^{-m/2} \exp\left(\frac{d^2(x, y)}{4\tau}\right),$$

which implies

$$f(x, \tau) \leq \frac{d^2(x, y)}{4\tau}.$$

Therefore,

$$\begin{aligned}\lim_{\tau\to 0} \int_M \left(f - \frac{m\epsilon^2}{8\pi}\right) hH \, d\mu_g &\leq \limsup_{\tau\to 0} \int_M \left(\frac{d^2(x, y)}{4\tau} - \frac{m\epsilon^2}{8\pi}\right) h(y, t) H(x, y, \tau) \, d\mu_g \\ &= \limsup_{\tau\to 0} \int_M \left(\frac{d^2(x, y)}{4\tau} - \frac{m\epsilon^2}{8\pi}\right) \frac{e^{-d^2(x,y)/4\tau}}{(4\pi\tau)^{m/2}} h(y, t) \, d\mu_g.\end{aligned}$$

It is easy to see that for any $\delta > 0$, the integration of the above integrand in the domain $d(x, y) \geq \delta$ converges to zero. Therefore,

$$\begin{aligned}\text{(4-7)} \quad &\lim_{t\to 0} \int_M \left(f - \frac{m\epsilon^2}{8\pi}\right) hH \, d\mu_g \\ &\qquad \leq \lim_{t\to 0} \int_{d(x,y)\leq\delta} \left(\frac{d^2(x, y)}{4t} - \frac{m\epsilon^2}{8\pi}\right) \frac{e^{-d^2(x,y)/(4t)}}{(4\pi t)^{m/2}} h(y, t) \, d\mu_g.\end{aligned}$$

Whenever δ is chosen sufficiently small, $d(x, y)$ is asymptotically sufficiently close to the Euclidean distance. Then by a standard approximation using local coordinates, we have

$$\begin{aligned}\text{(4-8)} \quad &\lim_{t\to 0} \int_M \left(\frac{d^2(x, y)}{4t} - \frac{m\epsilon^2}{8\pi}\right) hH \, d\mu_g \\ &\qquad = \lim_{t\to 0} \int_{\mathbb{R}^m} \left(\frac{|x - y|^2}{4\tau} - \frac{m\epsilon^2}{8\pi}\right) \frac{e^{-|x-y|^2/(4\tau)}}{(4\pi\tau)^{m/2}} h_p(y) \, dy,\end{aligned}$$

where h_p is the pullback of $h(\cdot, 0)$ from the region $d(x, y) \le \delta$ to the Euclidean space.

Splitting the last integrand as in [Kuang and Zhang 2008], we are left with

$$\lim_{t\to 0}\int_M \left(f - \frac{m\epsilon^2}{8\pi}\right) hH\, d\mu_g \le h_p(x) \lim_{t\to 0}\int_{\mathbb{R}^m} \left(\frac{|x-y|^2}{4\tau} - \frac{m\epsilon^2}{8\pi}\right) \frac{e^{-|x-y|^2/(4t)}}{(4\pi\tau)^{m/2}}\, dy$$
$$= h_p(\cdot) \lim_{\tau\to 0}\int_{\mathbb{R}^m} \left(\frac{|y|^2}{4\tau} \frac{e^{-|y|^2/(4\tau)}}{(4\pi\tau)^{m/2}}\right) dy - \frac{m\epsilon^2}{8\pi} h_p(\cdot).$$

Lastly, we have that the right-hand side evaluates to a constant $C(m) \le 0$ by using the standard Gauss integral

$$\int_{\mathbb{R}^m} \left(\frac{|y|^2}{4\tau} \frac{e^{-|y|^2/(4\tau)}}{(4\pi\tau)^{n/2}}\right) dy = \frac{m}{2}$$

and the condition $\epsilon \to 2\sqrt{\pi}$ as $\tau \to 0$. The claim then follows.

Finally in this section we prove Perelman's differential Harnack estimates for f as an application of Theorem 4.1 and the monotonicity of $\mathcal{W}_{\epsilon,\alpha}$. A corollary to this gives estimates of Li–Yau type for all positive solutions $H(x, \tau)$.

Proposition 4.2. *Let the assumptions of Theorem 4.1 hold. Then for any smooth curve $\gamma(\tau)$ in M, we have the estimate*

$$-\frac{d}{dt} f(\gamma(\tau), \tau) \le \frac{1}{2}\left(|\gamma'(\tau)|^2 + S_g(\gamma(\tau), \tau) - \frac{m}{2\tau}\right). \tag{4-9}$$

After the usual integration of (4-9) along the path $\gamma(\tau)$ and exponentiation, we have the following result.

Corollary 4.3. *With the notation and assumptions of Theorem 4.1, the following Li–Yau Harnack estimate holds*:

$$\frac{H(x_2, t_2)}{H(x_1, t_1)} \le \left(\frac{T - t_1}{T - t_2}\right)^{m/4} \exp\left(\frac{1}{2}\int_{t_1}^{t_2} (|\gamma'(t)|^2 + S_g(\gamma(t), t)\, dt\right). \tag{4-10}$$

Proof of Proposition 4.2. Precisely from (4-1), we have

$$f \le \frac{m\epsilon^2}{4\pi} - \frac{\epsilon^2\tau}{4\pi}(2\Delta f - |\nabla f|^2 + S_g) \le \frac{m\epsilon^2}{8\pi}$$

since $\Delta f + S_g - m/(2\tau) \ge 0$ by the monotonicity formula (3-3). Now multiplying (4-1) through by $2\pi/(\epsilon^2\tau)$, we have

$$\Delta f - \tfrac{1}{2}|\nabla f|^2 + \tfrac{1}{2} S_g + \frac{2\pi}{\epsilon^2 t} f - \frac{m}{2\tau} \le 0.$$

Using $\Delta f = -\partial_t f + |\nabla f|^2 - S_g + m/(2\tau)$ from (2-9), we obtain

$$-\partial_t f + \tfrac{1}{2}|\nabla f|^2 \le \tfrac{1}{2} S_g - \frac{2\pi}{\epsilon^2 t} f. \tag{4-11}$$

By Young's inequality, we have

$$\begin{aligned}-\frac{d}{dt}f(\gamma(\tau),\tau) &= -\partial_t f(\gamma(\tau),\tau)-\langle\nabla f(\gamma(\tau),\tau),\gamma'(\tau)\rangle\\ &\le -\partial_t f+\tfrac12|\nabla f|^2+\tfrac12|\gamma'(\tau)|^2\\ &= \tfrac12|\gamma'(\tau)|^2+\tfrac12 S_g(\gamma(\tau),\tau)-\frac{2\pi}{\epsilon^2\tau}f(\gamma(\tau)\tau)\end{aligned}$$

on the path $\gamma(\tau)$. The result follows by using the fact that $f\le m\epsilon^2/(8\pi)$. □

5. Log-Sobolev inequalities along (RH)$_\alpha$-flow

By the results of Aubin [1976] and Hebey [1996] for complete manifolds whose Ricci curvature is bounded from below and injectivity radius is positive and bounded from above, we can assume the Sobolev embedding on the initial metric, since $(M, g(0))$ is a compact Riemannian manifold. Let A_0, $B_0<\infty$ be positive constants such that for all $v\in W^{1,2}(M, g_0)$,

$$\|v\|_{2m/(m-2)}\le A_0\|\nabla v\|_2+B_0\|v\|_2, \tag{5-1}$$

where A_0 and B_0 depend only on m, g_0, the lower bound for the Ricci curvature and the injectivity radius. We can then write (5-1) as

(5-2)
$$\left(\int_M v^{2m/(m-2)}\,d\mu_{g_0}\right)^{(m-2)/m}\le A\int_M(4|\nabla v|^2+S_g v^2)\,d\mu_{g_0}+B\int_M v^2\,d\mu_{g_0},$$

where

$$A=\tfrac14 A_0 \quad\text{and}\quad B=\tfrac14 A_0\sup S_g^-(\,\cdot\,,0)+B_0$$

since $S_g(x,0)+\sup S_g^-(\,\cdot\,,0)=S_g^+(x,0)-S_g^-(x,0)$. We will assume that (5-2) holds uniformly for $g(t)$, $t>0$, and different A and B in order to prove the logarithmic Sobolev inequalities.

The usual way of deriving logarithmic Sobolev inequalities follows from a careful application of Hölder's and Jensen's inequalities since $\log v$ is a concave function, in which case

$$\int v^2\ln v^{q-2}\,d\mu\le\ln\int v^q\,d\mu$$

with the assumption that $\int v^2\,d\mu=1$. Then

$$\int v^2\ln v\,d\mu\le\frac{q}{q-2}\ln\left(\int v^q\,d\mu\right)^{1/q}.$$

Taking $q=2m/(m-2)$, we have

$$\int v^2\ln v\,d\mu\le\frac{m}{2}\ln\left(\int v^{2m/(m-2)}\,d\mu\right)^{(m-2)/2m},$$

and by multiplying both sides by 2 we obtain the following result.

Lemma 5.1. *For any* $v \in W^{1,2}(M, g_0)$ *with* $\|v\|_2 = 1$,

$$\int_M v^2 \ln v^2 \, d\mu_{g_0} \leq \frac{m}{2} \ln\left(A \int_M (4|\nabla v|^2 + S_g v^2) \, d\mu_{g_0} + B \right). \tag{5-3}$$

See [Hsu 2008; Ye 2007; Zhang 2007] for similar proofs. Inequalities of the form (5-3) are usually estimated further by the application of an elementary inequality of the form $\ln y \leq \theta y - \ln\theta - 1$, where $\theta, y \geq 0$. Precisely, taking

$$y = A \int_M (4|\nabla v|^2 + S_g v^2) \, d\mu_{g_0} + B$$

in (5-3) gives us

$$\begin{aligned} \int_M v^2 \ln v^2 \, d\mu_{g_0} &\leq \frac{m\theta}{2}\left(A \int_M (4|\nabla v|^2 + S_g v^2) \, d\mu_{g_0} + B \right) - \frac{m}{2}(1 + \ln\alpha) \\ &= \frac{m\theta A}{2} \int_M (4|\nabla v|^2 + S_g v^2) \, d\mu_{g_0} + \frac{m\theta B}{2} - \frac{m}{2} - \frac{m}{2}\ln\alpha. \end{aligned} \tag{5-4}$$

We will now modify the arguments in both [Ye 2007] and [Zhang 2007] to prove the following result which says the monotonicity of the $\mathcal{W}_{\alpha,\epsilon}$-entropy implies a logarithmic Sobolev inequality (not with the best constant). Here we assume the flow exists for all time.

Theorem 5.2. *Let* (M, g) *be a compact Riemannian manifold of dimension* $m \geq 3$ *and the metric* $g(t)$ *evolved by the* $(\mathrm{RH})_\alpha$*-flow. Assume that an* L^2*-Sobolev embedding* (5-2) *holds true with respect to the initial metric* $g(0) = g_0$. *Then, we have*

$$\int_M v^2 \ln v^2 \, d\mu_{g(t)} \leq \int_M \sigma^2 (4|\nabla v|^2 + S_g v^2) \, d\mu_{g(t)} - \frac{m}{2}\ln\sigma^2 + (t + \sigma^2)\beta_1 + \frac{m}{2}\ln\frac{mA}{2e}, \tag{5-5}$$

where $\sigma > 0$, $\beta_1 = 4A_0^{-1}B_0 + \sup S_g^-(\,\cdot\,, 0)$ *and*

$$\lambda_{\alpha 0} = \inf_{\|v\|_2 = 1} \int_M (4|\nabla v|^2 + S_g v^2) \, d\mu_{g_0};$$

that is, $\lambda_{\alpha 0}$ *is the first eigenvalue of the operator* $-4\Delta + S_g$.

Moreover, if $\lambda_{\alpha 0}$ *is strictly positive for* $S_g(\,\cdot\,, 0) > 0$ *(i.e.,* $R(\,\cdot\,, 0) > \alpha(0)|\nabla u(0)|^2$*), then*

$$\int_M v^2 \ln v^2 \, d\mu_{g(t)} \leq \int_M \sigma^2 (4|\nabla v|^2 + S_g v^2) \, d\mu_{g(t)} - \frac{m}{2}\ln\sigma^2 + (t + \sigma^2)\beta_2 + \frac{m}{2}\ln\frac{mA}{2e} \tag{5-6}$$

holds with $B_0 = 0$, *i.e.,* $\beta_2 = \sup S_g^-(\,\cdot\,, 0)$.

We first discuss some vital issues that will help put the proof of the above theorem in perspective. Now take an L^2-solution $H = H(x,t)$ of the conjugate heat equation

$$\partial_t H = -\Delta H + S_g H \tag{5-7}$$

to be $H = (4\pi\tau)^{m/2}e^{-f}$. Relating the entropy $\mathcal{W}_{\alpha,\epsilon}$ with the idea of logarithmic Sobolev inequalities, we consider a function

$$v = \sqrt{H} = \frac{e^{-f/2}}{(4\pi\tau)^{m/4}} \tag{5-8}$$

such that $\int_M v^2\,d\mu = 1$. We also notice that (5-8) implies $f = -\ln v^2 - (m/2)\ln\tau - (m/2)\ln(4\pi)$; hence the entropy (3-1) is rewritten as

$$\begin{aligned}\mathcal{W}_\epsilon(g,v,\tau) = \frac{\epsilon^2}{4\pi}\int_M (\tau(4|\nabla v|^2 + S_g v^2) - v^2\ln v^2)\,d\mu - \frac{\epsilon^2}{4\pi}\frac{m}{2}\ln\tau \\ - \frac{\epsilon^2}{4\pi}\frac{m}{2}\ln(4\pi) + \Big(1 - \frac{\epsilon^2}{4\pi}\Big)\int_M f v^2\,d\mu - \frac{m\epsilon^2}{4\pi} + \frac{m}{2}\ln\frac{4\pi}{\epsilon^2}.\end{aligned} \tag{5-9}$$

Define

$$\mathcal{W}^*_\epsilon(g,v,\tau) = \frac{\epsilon^2}{4\pi}\int_M (\tau(4|\nabla v|^2 + S_g v^2) - v^2\ln v^2)\,d\mu \tag{5-10}$$

and

$$\mu^*_\epsilon(g,v,\tau) = \inf\Big\{\mathcal{W}^*_\epsilon(g,v,\tau) : \int_M v^2\,d\mu = 1\Big\}. \tag{5-11}$$

Set $T^* = t^* + \sigma^2$ and $\tau(t) = T^* - t$ for $0 \le t \le t^*$, $\sigma > 0$. Then

$$\begin{aligned}&\frac{d}{dt}\mathcal{W}_\epsilon(g,v,\tau)\\ &\quad = \frac{d}{dt}\mathcal{W}^*_\epsilon(g,v,\tau) - \frac{m\epsilon^2}{8\pi}\frac{d}{dt}\ln\tau + \Big(1 - \frac{\epsilon^2}{4\pi}\Big)\frac{\partial}{\partial t}\int_M f v^2\,d\mu + \frac{m}{2}\ln\frac{4\pi}{\epsilon^2} \ge 0,\end{aligned}$$

where the last inequality is due to the monotonicity of $\mathcal{W}_\epsilon(g,f,\tau)$, the proof of which also reveals that

$$\frac{\partial}{\partial t}\int_M f v^2\,d\mu = -\mathcal{F}_\alpha + \frac{m}{2\tau},$$

where $\mathcal{F}_\alpha = \int_M (S_g + |\nabla f|^2)v^2\,d\mu$ is Perelman and Müller's energy functional. Let $\lambda_{\alpha 0}$ be the first eigenvalue of the operator $-4\Delta + S_g$. Then, we know that $\lambda_{\alpha 0} = \inf_{\|u\|_2=1}\mathcal{F}_\alpha$. Therefore we arrive at

$$\frac{d}{dt}\mathcal{W}^*_\epsilon \ge \frac{n\epsilon^2}{8\pi}\frac{d}{dt}\ln\tau + \Big(1 - \frac{\epsilon^2}{4\pi}\Big)\lambda_{\alpha 0}.$$

To continue this argument, we should note that either (5-7) or (5-8) implies that the function $f = f(t)$ satisfies the following backward heat equation

$$\frac{\partial f}{\partial t} = -\Delta f + |\nabla f|^2 - S_g + \frac{m}{2\tau}, \tag{5-12}$$

with $v = v(x,t)$ satisfying

$$\frac{\partial v}{\partial t} = -\Delta v + \frac{|\nabla v|^2}{v} + \frac{S_g}{2} v \tag{5-13}$$

on $[0, t^*]$ with a given terminal value at $t + t^*$ with $g = g(t^*)$.

Let v_0 be a minimizer of the entropy $\mathcal{W}_\epsilon(g, f, \tau_0)$ for all v with $\int_M v_0^2 \, d\mu_{g(t_0)} = 1$. We can then solve heat equation (5-12) backward in time with initial data $f(t_0) = f_0$ and v_0 chosen at $t = t_0$. Let u_j be the value of the conjugate heat equation (5-13) at $t = t_j$. We can define functions f_j, $j = 1, 2$, by

$$u_j = \frac{e^{-f_j/2}}{(4\pi\tau_j)^{n/4}}, \qquad j = 1, 2.$$

Then by the monotonicity of $\mathcal{W}_{\alpha\epsilon}(g, f, \tau)$-entropy, using Perelman's approach we have

$$\begin{aligned}\mu_\epsilon(g(t_1), \tau(t_1)) &= \inf_{\|v_0\|_{g(t_1)}=1} \mathcal{W}_\epsilon(g(t_1), f_0, \tau_1) \le \mathcal{W}_\epsilon(g(t_1), f_1, \tau_1) \\ &\le \mathcal{W}_\epsilon(g(t_2), f_2, \tau_2) = \inf_{\|v_0\|_{g(t_2)}=1} \mathcal{W}_\epsilon(g(t_2), f, \tau_2) = \mu_\epsilon(g(t_2), \tau(t_2)).\end{aligned}$$

It follows from the above that

$$\mu_\epsilon^*(g(t_1), \tau(t_1)) \le \mu_\epsilon^*(g(t_2), \tau(t_2)) + \frac{n\epsilon^2}{8\pi} \ln \frac{\tau_1}{\tau_2}$$

for any $t_1 < t_2$, where $\tau_j = \tau(t_j)$, $j = 1, 2$. Choosing $t_1 = 0$ and $t_2 = t^*$, we then obtain

$$\mu_\epsilon^*(g(0), t^* + \sigma^2) \le \mu_\epsilon^*(g(t^*), \sigma^2) + \frac{n\epsilon^2}{8\pi} \ln \frac{t^* + \sigma^2}{\sigma}. \tag{5-14}$$

Since $0 < t^* < T$ is arbitrary, we can write (5-14) as

$$\mu_\epsilon^*(g(t), \sigma^2) \ge \mu_\epsilon^*(g(0), t + \sigma^2) + \frac{n\epsilon^2}{8\pi} \ln \frac{\sigma^2}{t + \sigma^2} \tag{5-15}$$

for all $t \in [0, T)$. [1] We now state the proof.

Proof. We now apply (5-4) with $g = g_0$ to estimate $\mu_\epsilon^*(g(0), t + \sigma^2)$. For any function $v \in W^{1,2}(M, g)$ with $\|v\|_2 = 1$ and using

$$\frac{m\theta A}{2} = t + \sigma^2 \quad \Longrightarrow \quad \theta = \frac{8(t + \sigma^2)}{nA_0},$$

[1] The case $t = 0$ is optimal, as equality is attained.

the inequality in (5-4) becomes

$$\int_M v^2 \ln v^2 \, d\mu_{g_0} \le (t+\sigma^2) \int_M (4|\nabla v|^2 + S_g v^2)\, d\mu_{g_0} + \frac{m}{2}\frac{8(t+\sigma^2)}{mA_0} B - \frac{m}{2} \ln \frac{8(t+\sigma^2)}{nA_0} - \frac{m}{2}$$

$$= (t+\sigma^2) \int_M (4|\nabla v|^2 + S_g v^2)\, d\mu_{g_0} + 4(t+\sigma^2) B A_0^{-1} - \frac{m}{2}\ln(t+\sigma^2) + \frac{m}{2}(\ln A_0 + \ln m - 3\ln 2 - 1).$$

Choosing $\epsilon^2 \le 4\pi$ as before, it then follows that

$$\mu_\epsilon^*(g(0), t+\sigma^2) \ge \frac{m\epsilon^2}{4\pi}\Big(\tfrac{1}{2}\ln(t+\sigma^2) - \frac{4}{m}(t+\sigma)BA_0^{-1} - \tfrac{1}{2}(\ln A_0 + \ln m - 3\ln 2 - 1)\Big). \tag{5-16}$$

Combining (5-15) and (5-16), we obtain

$$\mu_\epsilon^*(g(t), \sigma^2) \ge \frac{m\epsilon^2}{8\pi}\ln\sigma^2 - \frac{m\epsilon^2}{\pi}(t+\sigma^2)BA_0^{-1} - \frac{m\epsilon^2}{8\pi}(\ln A_0 + \ln m - 3\ln 2 - 1), \tag{5-17}$$

which implies

$$\frac{\epsilon^2}{4\pi}\int_M \big(\sigma^2(4|\nabla v|^2 + S_g v^2) - v^2 \ln v^2\big)\, d\mu \ge \frac{m\epsilon^2}{8\pi}\ln\sigma^2 - \frac{m\epsilon^2}{\pi}(t+\sigma^2)BA_0^{-1} - \frac{m\epsilon^2}{8\pi}\ln\frac{mA_0}{8e}.$$

Therefore,

$$\int_M v^2 \ln v^2 \, d\mu \le \int_M \sigma^2(4|\nabla v|^2 + S_g v^2)\, d\mu - \frac{m}{2}\ln\sigma^2 + 4(t+\sigma^2)BA_0^{-1} - \frac{m}{2}\ln\frac{nA_0}{8e}. \tag{5-18}$$

Choosing $\beta_1 = 4BA_0^{-1} = 4A_0^{-1}(B_0 + A \sup S_g^-(x,0))$ and $A = A_0/4$, we obtain the result. We can also derive (5-6) in a similar manner. □

6. Heat kernel bound via log-Sobolev inequalities

We apply the logarithmic Sobolev inequality obtained in the last section to derive an upper bound for the conjugate heat kernel along the Ricci flow, demonstrating that there is a lot of geometric information embedded in such inequalities. The basic ideas, due to Davies and Simon [1984], relate Nelson's hypercontractivity (see [Gross 1975]) to ultracontractivity (see also [Davies 1989]). These ideas always yield estimates with sharp constants. We modify the argument in [Zhang 2007] (see also [Lieb and Loss 1997; Zhang 2011]) to prove our result.

Theorem 6.1. *Suppose there exists a solution to the* $(\mathrm{RH})_\alpha$*-flow with* $m \geq 2$ *and let* $H(x,t;y)$ *be the fundamental solution to the conjugate heat equation*

$$\big(-\partial_t - \Delta + S_g(x,\tau)\big) w(x,\tau) = 0. \tag{6-1}$$

Then, for some nonnegative finite constant C *depending on* n, t, T, A_0, B_0 *and* $\sup S_g^-(\,\cdot\,, 0)$, *the estimate*

$$H(x,T;y) \leq CT^{-m/2} \tag{6-2}$$

holds, where $\partial_t \tau = -1$ *and* A_0, B_0 *are as defined in the last section.*

Without loss of generality, we may assume $w = w(x,t)$ to be a nonnegative solution of the conjugate heat equation (6-1) on the interval $[0, T]$, where $\partial_t \tau = -1$. Let $T > 0$ and $r(\tau) : [0, T] \to [1, \infty]$ be a continuously differentiable increasing function such that $r(0) = \infty$ and $r(T) = 1$. The function $r(\tau) = T/\tau$ gives a perfect example as we shall see below.

The idea here follows from the fact that if

$$w(x,t) = \int H(x,t;y) w_0(y)\, d\mu(y)$$

solves the heat equation, where $H(x,t;y)$ is the heat kernel, then

$$\sup_{w \neq 0} \frac{\|w(\,\cdot\,,t)\|_\infty}{\|w(\,\cdot\,,0)\|_1} = \sup_{x,y} H(x,t;y).$$

We may obtain an estimation of the time derivative for the logarithms of the quantity

$$\|w\|_{r(t)} = \left(\int_M |w|^{r(t)}\, d\mu_{g(t)}\right)^{1/r(t)}$$

as follows:

$$\int_0^T \frac{\partial}{\partial t} \ln \|w\|_{r(t)}\, dt = \ln \frac{\|w(\,\cdot\,,t)\|_\infty}{\|w(\,\cdot\,,0)\|_1}.$$

Proof. By routine computation,

$$\begin{aligned}
\partial_t \|w\|_{r(t)} &= \partial_t \left(\int_M |w|^{r(t)}\, d\mu_{g(\tau)}\right)^{1/r(\tau)} \\
&= -\frac{\dot r(\tau)}{r^2(\tau)} \|w\|_{r(\tau)} \ln \|w\|_{r(\tau)}^{r(\tau)} + \frac{\|w\|_{r(\tau)}^{1-r(\tau)}}{r(\tau)} \bigg(\dot r(\tau) \int_M w^{r(\tau)} \ln w\, d\mu_{g(\tau)} \\
&\qquad + r(\tau) \int_M \big(w^{r(\tau)-1}(-\Delta w + S_g w) + w^{r(\tau)}(-S_g) \big)\, d\mu_{g(\tau)} \bigg).
\end{aligned}$$

Multiplying both sides by $r^2(\tau)\|w\|_{r(\tau)}^{r(\tau)}$, we have

$$\begin{aligned}
&r^2(\tau)\|w\|_{r(\tau)}^{r(\tau)}\,\partial_t\|w\|_{r(t)}\\
&\qquad= -\dot r(\tau)\|w\|_{r(\tau)}^{r(\tau)+1}\ln\|w\|_{r(\tau)}^{r(\tau)} + r(\tau)\dot r(\tau)\|w\|_{r(\tau)}\int_M w^{r(\tau)}\ln u\,d\mu_{g(\tau)}\\
&\qquad\quad + r^2(\tau)\|w\|_{r(\tau)}\int_M w^{r(\tau)-1}(-\Delta w)\,d\mu_{g(\tau)} + r^2(\tau)\|w\|_{r(\tau)}\\
&\qquad\quad \times\int_M w^{r(\tau)-1}(S_g w)\,d\mu_{g(\tau)} - r(\tau)\|w\|_{r(\tau)}\int_M w^{r(\tau)}S_g\,d\mu_{g(\tau)}.
\end{aligned}$$

By the application of integration by parts, we have

$$\begin{aligned}
&r^2(\tau)\|w\|_{r(\tau)}\int_M w^{r(\tau)-1}(-\Delta w)\,d\mu_{g(\tau)}\\
&\qquad= r^2(\tau)\|u\|_{r(\tau)}\int_M \nabla(u^{r(\tau)-1})\nabla w\,d\mu_{g(\tau)}\\
&\qquad= r^2(\tau)(r(\tau)-1)\|w\|_{r(\tau)}\int_M w^{r(\tau)-2}|\nabla w|^2\,d\mu_{g(\tau)}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&r^2(\tau)\|w\|_{r(\tau)}^{r(\tau)}\,\partial_t\|w\|_{r(t)}\\
&\qquad= -\dot r(\tau)\|w\|_{r(\tau)}^{r(\tau)+1}\ln\|w\|_{r(\tau)}^{r(\tau)} + r(\tau)\dot r(\tau)\|w\|_{r(\tau)}\int_M w^{r(\tau)}\ln w\,d\mu_{g(\tau)}\\
&\qquad\quad + r^2(\tau)(r(\tau)-1)\|w\|_{r(\tau)}\int_M w^{r(\tau)-2}|\nabla w|^2\,d\mu_{g(\tau)}\\
&\qquad\quad + r(\tau)(r(\tau)-1)\|w\|_{r(\tau)}\int_M S_g\,w^{r(\tau)}\,d\mu_{g(\tau)}.
\end{aligned}$$

Further dividing both sides by $\|w\|_{r(\tau)}$, we obtain

$$\begin{aligned}
\text{(6-3)}\quad &r^2(\tau)\|w\|_{r(\tau)}^{r(\tau)}\,\partial_t(\ln\|w\|_{r(t)})\\
&\qquad= -\dot r(\tau)\|u\|_{r(\tau)}^{r(\tau)}\ln\|w\|_{r(\tau)}^{r(\tau)} + r(\tau)\dot r(\tau)\int_M w^{r(\tau)}\ln w\,d\mu_{g(\tau)}\\
&\qquad\quad + r^2(\tau)(r(\tau)-1)\int_M w^{r(\tau)-2}|\nabla w|^2\,d\mu_{g(\tau)}\\
&\qquad\quad + r(\tau)(r(\tau)-1)\int_M S_g\,w^{r(\tau)}\,d\mu_{g(\tau)}.
\end{aligned}$$

Using

$$v=\frac{w^{r(\tau)/2}}{\|w^{r(\tau)/2}\|_2}\quad\Longrightarrow\quad v^2=\frac{w^{r(\tau)}}{\|w\|_{r(\tau)}^{r(\tau)}},$$

we have

$$|\nabla v|^2 = \frac{r^2(\tau)}{4\|w\|_{r(\tau}^{r(\tau)}} w^{r(\tau)-2}|\nabla w|^2$$

and

$$\ln v^2 = \ln w^{r(\tau)} - \ln \|w\|_{r(\tau)}^{r(\tau)}.$$

Therefore,

$$\begin{aligned}\dot r(\tau)\int_M v^2\ln v^2\,d\mu_{g(\tau)} &= \dot r(\tau)\int_M \frac{w^{r(\tau)}}{\|u\|_{r(\tau)}^{r(\tau)}}(\ln w^{r(\tau)} - \ln\|w\|_{r(\tau)}^{r(\tau)} g)\,d\mu_{g(\tau)}\\ &= \frac{\dot r(\tau)r(\tau)}{\|w\|_{r(\tau)}^{r(\tau)}}\int_M w^{r(\tau)}\ln w^{r(\tau)}\,d\mu_{g(\tau)} - \dot r\ln\|w\|_{r(\tau)}^{r(\tau)}.\end{aligned}$$

Plugging these into (6-3), we arrive at

$$\begin{aligned} r^2(\tau)\partial_t(\ln\|w\|_{r(t)}) &= \dot r(\tau)\int_M v^2\ln v^2\,d\mu_{g(\tau)} + 4(r(\tau)-1)\int_M |\nabla v|^2\,d\mu_{g(\tau)}\\ &\qquad + r(\tau)(r(\tau)-1)\int_M Rv^2\,d\mu_{g(\tau)}\\ &= \dot r(\tau)\int_M v^2\ln v^2\,d\mu_{g(\tau)} + (r(\tau)-1)\int_M (4|\nabla v|^2 + S_g v^2)\,d\mu_{g(\tau)}\\ &\qquad + (r(\tau)-1)^2\int_M S_g v^2\,d\mu_{g(\tau)}.\end{aligned}$$

Using the choice $r(\tau) = T/\tau$, we have $\dot r(\tau) = -T/\tau^2$ and $r(\tau)-1 = (T-\tau)/\tau$ so that we write the last equality as

$$\begin{aligned} r^2(\tau)\partial_t(\ln\|w\|_{r(t)}) &= -\frac{T}{\tau^2}\int_M v^2\ln v^2\,d\mu_{g(\tau)} + \frac{T-\tau}{\tau}\int_M (4|\nabla v|^2 + S_g v^2)\,d\mu_{g(\tau)}\\ &\qquad + \Big(\frac{T-\tau}{\tau}\Big)^2\int_M S_g v^2\,d\mu_{g(\tau)}\\ &= \frac{T}{\tau^2}\bigg(\frac{\tau(T-\tau)}{T}\int_M (4|\nabla v|^2 + S_g v^2)\,d\mu_{g(\tau)} - \int_M v^2\ln v^2\,d\mu_{g(\tau)}\bigg)\\ &\qquad + \Big(\frac{T-\tau}{\tau}\Big)^2\int_M S_g v^2\,d\mu_{g(\tau)}.\end{aligned}$$

From the log-Sobolev inequality (5-5) point of view, we may choose

$$\sigma^2 = \frac{4\tau(T-\tau)}{T} \le \frac{T}{4},$$

and we get

$$\text{(6-4)}\quad r^2(\tau)\partial_t(\ln\|w\|_{r(t)}) \geq \frac{T}{\tau^2}\Big(\frac{m}{2}\ln\sigma^2-\frac{m}{2}\ln\frac{mA}{2e}-(t_0+\sigma^2)\beta_1\Big)+\Big(\frac{T-\tau}{\tau}\Big)^2\int_M S_g v^2\,d\mu_{g(\tau)}$$

and

$$\text{(6-5)}\quad \partial_t(\ln\|w\|_{r(t)}) \geq \frac{1}{T}\Big(\frac{m}{2}\ln\frac{4\pi\tau(T-\tau)}{T}-\frac{m}{2}\ln\frac{m\pi A}{2e}-(t_0+\sigma^2)\beta_1-T\sup S_g^-(\,\cdot\,,0)\Big).$$

Notice that (since $\sigma^2\leq T/4$)

$$\begin{aligned}(t+\sigma^2)\beta_1+T\sup R^-(\,\cdot\,,0)&=4(t_0+\sigma^2)\big(A_0^{-1}B_0+\tfrac14\sup S_g^-(\,\cdot\,,0)\big)+T\sup S_g^-(\,\cdot\,,0)\\&\leq(4t_0+T)A_0^{-1}B_0+\tfrac14(4t_0+5T)\sup S_g^-(\,\cdot\,,0).\end{aligned}$$

Denoting D by

$$D\equiv\frac{m}{2}\ln\frac{m\pi A}{2e}+(4t_0+T)A_0^{-1}B_0,$$

substituting into (6-5) and integrating the result from 0 to T, we have

$$\begin{aligned}\ln\frac{\|w(\,\cdot\,,T)\|_{r(T)}}{\|w(\,\cdot\,,T)\|_{r(0)}}&\geq\frac{m}{2T}\int_0^T\ln\frac{4\pi\tau(T-\tau)}{T}\,dt-D-\tfrac14(4t_0+5T)\sup R^-(\,\cdot\,,0)\\&=\frac{m}{2}\ln(4\pi)-\frac{n}{2}\ln T-n+n\ln T-D-\tfrac14(4t_0+5T)\sup S_g^-(\,\cdot\,,0)\\&=\frac{m}{2}\ln(4\pi T)-m-D-(4t_0+5T)\sup S_g^-(\,\cdot\,,0).\end{aligned}$$

This then yields

$$\ln\frac{\|w(\,\cdot\,,T)\|_1}{\|w(\,\cdot\,,T)\|_\infty}\geq\frac{m}{2}\ln(4\pi T)-m-D-\tfrac14(4t_0+5T)\sup S_g^-(\,\cdot\,,0),$$

which implies

$$\|w(\,\cdot\,,T)\|_\infty\leq\|w(\,\cdot\,,T)\|_1\,\frac{\exp\big(\frac14(4t_0+5T)\sup S_g^-(\,\cdot\,,0)+D+m\big)}{(4\pi T)^{m/2}}.$$

Because

$$w(x,T)=\int_M H(x,T;y)w(y,0)\,d\mu(y)_{g(\tau)},$$

where $H(x,T;y)$ is the conjugate heat kernel,

$$H(x,T;y)\leq\frac{\exp(mD)}{(4\pi T)^{m/2}}\dot{\exp}\big(\tfrac14(4t_0+5T)\sup S_g^-(\,\cdot\,,0)\big).$$

This ends the proof of the estimate (6-2). ∎

7. The Sobolev inequality along $(\mathrm{RH})_\alpha$-flow

In this section we show that global bounds on the heat kernel to the conjugate heat equation imply a uniform Sobolev inequality under Ricci-harmonic map flow. This type of proof is standard as contained in [Davies 1989, Chapter 2]. The same procedures have been adapted in [Zhang 2007] for Kähler–Ricci flow; see also [Ye 2007; Hsu 2008]. For completeness we give the summary of the approach.

For any $t \in [0, T)$, we define the operator

$$L := -\Delta_g + \frac{S_g + \sup_M S_g^-}{4}. \tag{7-1}$$

Since $R_g(\cdot\,,\tau) \geq -\sup_M R_g(\cdot\,,\tau)$, we know that $\Phi = \frac{1}{4}(S_g + \sup_M S_g^-) \geq 0$, $\Phi \in L^\infty(M)$. Then $L \geq 0$ and is essentially a self-adjoint operator on $L^2(M)$ with the associated quadratic form

$$\boldsymbol{Q}(v) = \int_M (|\nabla v|^2 + \Phi v^2)\, d\mu_g \quad \forall v \in W^{1,2}(M). \tag{7-2}$$

By the heat kernel convolution property, we have

$$e^{-tL} w_0 = \int_M H(x, t; y) w_0(y)\, d\mu_g(y), \tag{7-3}$$

where e^{-tL} is a self-adjoint positivity preserving semigroup for all $t \geq 0$. It is also a contraction on $L^\infty(M)$ and $L^1(M)$ for all $t \geq 0$. Then

$$\|e^{-tL} w_0\|_\infty \leq C_0 t^{-m/2} \|w_0\|_1. \tag{7-4}$$

The next step is to apply a theorem in [Davies 1989], which we state below as a lemma.

Lemma 7.1. *If $m \geq 2$, then a bound of the form*

$$\|e^{-tL} w_0\|_\infty \leq C_1 t^{-m/4} \|w_0\|_2 \tag{7-5}$$

for all $t > 0$ and all $w_0 \in L^2(M)$ is equivalent to a bound of the form

$$\|w_0\|^2_{2m/(m-2)} \leq C_2 \boldsymbol{Q}(w_0) \quad \forall w_0 \in W^{1,2}(M). \tag{7-6}$$

By Lemma 7.1 we can prove that

$$\left(\int_M v^{2m/(m-2)}\, d\mu_g\right)^{(m-2)/2} \leq A_0 \int_M \big(|\nabla v|^2 + \tfrac{1}{4}(S_g + \sup_M S_g^-) v^2\big)\, d\mu_g \tag{7-7}$$

using an estimate of the form (1-9). The only thing remaining for us to show is that estimates (7-4) and (7-5) are equivalent. We do this via the following lemma and the Hölder inequality.

Lemma 7.2. *Suppose $m \geq 2$ and $T < \infty$. Let $C_1 > 0$ be the same as C_1 in (7-5). Then we have*

$$\|e^{-tL} w_0\|_2 \leq C_1 t^{-m/4} \|w_0\|_1 \quad \forall w_0 \in L^1(M). \tag{7-8}$$

Now write $e^{-tL} w_0 = e^{-1/2tL} e^{-1/2tL} w_0$ and by assuming (7-5), we have

$$\|e^{-tL} w_0\|_\infty \leq C_1 t^{-m/4} \|e^{-1/2tL} w_0\|_2 \leq C_1^2 t^{-m/2} \|w_0\|_1.$$

Similarly, combining the fact that e^{-tL} is a contraction on $L^\infty(M)$ with bound (7-4) gives us (7-5). Indeed,

$$\begin{aligned}\|e^{-tL} w_0\|_\infty &= \left| \int_M H(x, t; y) w_0(y)\, d\mu_g(y) \right| \\ &\leq \left(\int_M H^{q'}(x, t; y) \mu_g(y) \right)^{1/q'} \left(\int_M w_0^q \mu_g(y) \right)^{1/q} \leq C t^{-m/2q} \|w_0\|_q,\end{aligned}$$

for all $w_0 \in L^q(M)$ with $1/q = 1 - 1/q'$ and $\int_M H(x, t : y)\, d\mu_g \leq 1$. Here we take q to satisfy $1 \leq q < m$ for obvious reason. (Though, by the Riez–Thorin interpolation theorem, the above holds for any $1 \leq q < \infty$ since e^{-tL} is a contraction on $L^1(M)$ and $L^\infty(M)$.)

The main result of this section is as follows.

Theorem 7.3. *With the conditions of the theorem in the introduction, we claim that estimate (1-8) implies the uniform Sobolev inequality (1-7).*

Proof. Based on the previous argument and a modification of the calculation in [Zhang 2007], we define the operator $\widetilde{L} = L + 1$, which also has all the properties of L, ($\widetilde{L} \geq 0$ and generates a symmetric Markov semigroup). Then for any positive constant c depending on m, T, a lower bound for R_{g_0} and an upper bound for A_0 such that for all $t \in [0, T)$ and $v \in \mathrm{Dom}(\widetilde{L}) \subseteq W^{1,q}(M)$,

$$\|\widetilde{L}^{-1/2} w\|_{mq/(m-q)} \leq c \|w\|_q \quad \forall w \in W_0^{1,2}(M) \tag{7-9}$$

holds for $m \geq 3$. Since $\widetilde{L}^{-1/2}$ is of weak type (p, q), $p = mq/(m-q)$ for any $1 < q < m$. A simple analysis and the Marcinkiewicz interpolation theorem tell us that $\widetilde{L}^{-1/2}$ is a bounded operator from L^q to L^p and that (7-9) holds.

Define $v(x, t) = \widetilde{L}^{-1/2} w(x, t)$, which implies $w(x, t) = \widetilde{L}^{1/2} v(x, t)$. Taking $q = 2$, we have

$$\|w\|_2^2 = \int_M \widetilde{L}^{1/2} v \widetilde{L}^{1/2} v\, d\mu_g = \int_M (\widetilde{L} v) v\, d\mu_g = \int_M ((L+1)v) v\, d\mu_g.$$

Combining with (7-9) and (7-6), we obtain the Sobolev inequality

$$\|v\|_{2m/(m-2)}^2 \leq c C_2 \left(\boldsymbol{Q}(v) + \int_M v^2 \mu_g \right), \tag{7-10}$$

whereby (1-7) follows with $A = c C_2$ and $B = \frac{1}{4} c C_2 (\sup_M S_g + 1)$. □

Remark 7.4. Fixing t_0 during $(\mathrm{RH})_\alpha$-flow, it is clear that $\widetilde{H} = e^{-1}H$ is the heat kernel of $\widetilde{L}$ and that

$$\int_M \widetilde{H}(x,t;\,y)\,d\mu_g(y) \le \int_M H(x,t;\,y)\,d\mu_g(y) \le 1.$$

By the upper bound for H, we are sure that $\widetilde{H}$ obeys the global upper bound

$$\widetilde{H}(x,t;\,y)\,d\mu_g(y) \le \widetilde{C}t^{-m/2}, \quad t > 0,$$

where $\widetilde{C}$ depends on m, A_0, B_0, t_0 and T. Similarly,

$$\|e^{-t\widetilde{L}}w\|_\infty = \|e^{-t}e^{-t\widetilde{L}}\|_\infty \le e^{-t}Ct^{-m/2}\|w\|_1 = \widetilde{C}t^{-m/2}\|w\|_1.$$

As a corollary, suppose

$$\lambda_{\alpha 0} = \inf_{\|v\|_2=1}\int_M (4|\nabla v|^2 + S_g v^2)\,d\mu_{g_0} > 0.$$

It can be proved by following [Zhang 2007] that Sobolev inequality (5-2) holds with $B = B(t=0) = 0$ on a compact manifold (M, g_0); i.e.,

$$\left(\int_M v^{2m/(m-2)}\,d\mu_{g_0}\right)^{(m-2)/m} \le \widetilde{A}_0\int_M (4|\nabla v|^2 + S_g v^2)\,d\mu_{g_0}, \tag{7-11}$$

where $\widetilde{A}$ depends only on m, g_0 and $\lambda_{\alpha 0}$. Therefore, we have the following result.

Corollary 7.5. *Let (M, g) be a compact Riemannian manifold of dimension $m \ge 3$ and the metric $g(t)$ evolved by the $(\mathrm{RH})_\alpha$-flow. Assume that L^2-Sobolev embedding (7-11) holds true with respect to the initial metric $g(0) = g_0$. Then, there exists a positive constant $\widetilde{A}$ depending on $\widetilde{A}_0$ such that for all $v \in W^{1,2}(M, g(t))$, $t \in [0, T)$,*

$$\left(\int_M v^{2m/(m-2)}\,d\mu_{g(t)}\right)^{(m-2)/m} \le \widetilde{A}\int_M (4|\nabla v|^2 + S_g v^2)\,d\mu_{g(t)}, \tag{7-12}$$

and

$$\int_M v^2\ln v^2\,d\mu_{g(t)} \le \sigma^2\int_M (4|\nabla v|^2 + S_g v^2)\,d\mu_{g(t)} - \frac{m}{2}\ln\sigma^2 + \frac{m}{2}\ln\frac{mA}{2e}, \tag{7-13}$$

where $\sigma > 0$.

Remark 7.6. The smallest eigenvalue is an important quantity that gives a better understanding of the geometric nature of the underlying manifold. For instance, consider the operator semigroup e^{-tL} generated by $L := -\Delta + \Phi$, with $\Phi \in L^\infty(M, g)$. By spectral decomposition, we write a positive solution on M as

$$U = e^{-tL}u = \sum_{j=1}^{\infty} e^{-\lambda_j t}\psi_j\langle u, \psi_j\rangle_{L^2(M)}$$

for $u \in L^2(M)$ satisfying the Cauchy problem

$$\frac{\partial}{\partial t}(e^{-tL}u) = -Le^{-tL}u,$$
$$U|_{t=0} = u,$$

and the eigenvalue problem $L\psi = \lambda\psi$, where $\{\psi_j\}_{j=1}^{\infty}$ forms a complete set of L^2-orthonormal eigenfunctions of L and the corresponding eigenvalues can be arranged in a nondecreasing order $\lambda_1 \le \lambda_2 \le \cdots$, with $\lambda_j \to \infty$. An interested reader will find the books [Davies 1989] and [Schoen and Yau 1994] useful in this respect.

Acknowledgements

The author wishes to thank the anonymous referee and the editor for their useful comments. He acknowledges the support he received from his supervisor, Professor Ali Taheri, University of Sussex, UK. He is also indebted to Dr Mihai Băileşteanu, University of Rochester, NY, for the useful discussions he had with him when he visited the University of Rochester in February 2014. His research is supported by the Tertiary Education Trust Fund, Federal Republic of Nigeria.

References

[Aubin 1976] T. Aubin, "Problèmes isopérimétriques et espaces de Sobolev", *J. Differential Geometry* **11**:4 (1976), 573–598. MR 56 #6711 Zbl 0371.46011

[Băileşteanu et al. 2010] M. Băileşteanu, X. Cao, and A. Pulemotov, "Gradient estimates for the heat equation under the Ricci flow", *J. Funct. Anal.* **258**:10 (2010), 3517–3542. MR 2011b:53153 Zbl 1193.53139

[Băileşteanu and Tran 2013] M. Băileşteanu and H. Tran, "Heat kernel estimates under the Ricci–Harmonic map flow", preprint, 2013. arXiv 1310.1619

[Cao and Zhang 2011] X. Cao and Q. S. Zhang, "The conjugate heat equation and ancient solutions of the Ricci flow", *Adv. Math.* **228**:5 (2011), 2891–2919. MR 2012k:53127 Zbl 1238.53038

[Cao et al. 2003] H. D. Cao, B. Chow, S. C. Chu, and S. T. Yau (editors), *Collected papers on Ricci flow*, Series in Geometry and Topology **37**, International Press, Somerville, MA, 2003. MR 2006e:53002 Zbl 1108.53002

[Cheeger and Yau 1981] J. Cheeger and S. T. Yau, "A lower bound for the heat kernel", *Comm. Pure Appl. Math.* **34**:4 (1981), 465–480. MR 82i:58065 Zbl 0481.35003

[Chow and Knopf 2004] B. Chow and D. Knopf, *The Ricci flow: An introduction*, Mathematical Surveys and Monographs **110**, Amer. Math. Soc., Providence, RI, 2004. MR 2005e:53101 Zbl 1086.53085

[Chow et al. 2008] B. Chow, S.-C. Chu, D. Glickenstein, C. Guenther, J. Isenberg, T. Ivey, D. Knopf, P. Lu, F. Luo, and L. Ni, *The Ricci flow: Techniques and applications, II: Analytic aspects*, Mathematical Surveys and Monographs **144**, Amer. Math. Soc., Providence, RI, 2008. MR 2008j:53114 Zbl 1157.53035

[Davies 1989] E. B. Davies, *Heat kernels and spectral theory*, Cambridge Tracts in Mathematics **92**, Cambridge Univ. Press, 1989. MR 90e:35123 Zbl 0699.35006

[Davies and Simon 1984] E. B. Davies and B. Simon, "Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians", *J. Funct. Anal.* **59** (1984), 335–395. MR 86e:47054 Zbl 0568.47034

[Eells and Sampson 1964] J. Eells, Jr. and J. H. Sampson, "Harmonic mappings of Riemannian manifolds", *Amer. J. Math.* **86** (1964), 109–160. MR 29 #1603 Zbl 0122.40102

[Gross 1975] L. Gross, "Logarithmic Sobolev inequalities", *Amer. J. Math.* **97**:4 (1975), 1061–1083. MR 54 #8263 Zbl 0318.46049

[Hamilton 1982] R. S. Hamilton, "Three-manifolds with positive Ricci curvature", *J. Differential Geom.* **17**:2 (1982), 255–306. MR 84a:53050 Zbl 0504.53034

[Hamilton 1993] R. S. Hamilton, "The Harnack estimate for the Ricci flow", *J. Differential Geom.* **37**:1 (1993), 225–243. MR 93k:58052 Zbl 0804.53023

[Hebey 1996] E. Hebey, "Optimal Sobolev inequalities on complete Riemannian manifolds with Ricci curvature bounded below and positive injectivity radius", *Amer. J. Math.* **118**:2 (1996), 291–300. MR 97c:53058 Zbl 0863.53031

[Hsu 2008] S.-Y. Hsu, "Sobolev inequalities for manifolds evolving by Ricci flow", *Adv. Math. Sci. Appl.* **18**:2 (2008), 453–462. MR 2010f:53112 Zbl 1187.53070

[Kuang and Zhang 2008] S. Kuang and Q. S. Zhang, "A gradient estimate for all positive solutions of the conjugate heat equation under Ricci flow", *J. Funct. Anal.* **255**:4 (2008), 1008–1023. MR 2009h:53150 Zbl 1146.58017

[Li and Yau 1986] P. Li and S.-T. Yau, "On the parabolic kernel of the Schrödinger operator", *Acta Math.* **156**:3-4 (1986), 153–201. MR 87f:58156 Zbl 0611.58045

[Lieb and Loss 1997] E. H. Lieb and M. Loss, *Analysis*, Graduate Studies in Mathematics **14**, Amer. Math. Soc., Providence, RI, 1997. MR 98b:00004 Zbl 0873.26002

[List 2008] B. List, "Evolution of an extended Ricci flow system", *Comm. Anal. Geom.* **16**:5 (2008), 1007–1048. MR 2010i:53126 Zbl 1166.53044

[Müller 2012] R. Müller, "Ricci flow coupled with harmonic map flow", *Ann. Sci. Éc. Norm. Supér.* (4) **45**:1 (2012), 101–142. MR 2961788 Zbl 1247.53082

[Ni 2004] L. Ni, "The entropy formula for linear heat equation", *J. Geom. Anal.* **14**:1 (2004), 87–100. MR 2004m:53118a Zbl 1044.58030

[Ni 2006] L. Ni, "A note on Perelman's LYH-type inequality", *Comm. Anal. Geom.* **14**:5 (2006), 883–905. MR 2007h:53103 Zbl 1116.58031

[Perelman 2002] G. Perelman, "The entropy formula for the Ricci flow and its geometric application", preprint, 2002. arXiv math.DG/0211159v1

[Perelman 2003a] G. Perelman, "Finite extinction time for the solutions to the Ricci flow on certain three-manifolds", preprint, 2003. arXiv math.DG/0307245v1

[Perelman 2003b] G. Perelman, "Ricci flow with surgery on three manifolds", preprint, 2003. arXiv math.DG/0303109v1

[Schoen and Yau 1994] R. Schoen and S.-T. Yau, *Lectures on differential geometry*, Conference Proceedings and Lecture Notes in Geometry and Topology **1**, International Press, Cambridge, MA, 1994. MR 97d:53001 Zbl 0830.53001

[Ye 2007] R. Ye, "The logarithmic Sobolev inequalities along the Ricci flow", preprint, 2007. arXiv 0707.2424v4

[Zhang 2007] Q. S. Zhang, "A uniform Sobolev inequality under Ricci flow", *Int. Math. Res. Not.* **2007**:17 (2007), Art. ID rnm056. MR 2008g:53083 Zbl 1141.53064

[Zhang 2011] Q. S. Zhang, *Sobolev inequalities, heat kernels under Ricci flow, and the Poincaré conjecture*, CRC Press, Boca Raton, FL, 2011. MR 2011m:53127 Zbl 1208.58032

Received March 7, 2014. Revised January 14, 2015.

ABIMBOLA ABOLARINWA
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SUSSEX
BRIGHTON
BN1 9QH
UNITED KINGDOM
a.abolarinwa@sussex.ac.uk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.278.291

ON *J*-HOLOMORPHIC CURVES IN ALMOST COMPLEX MANIFOLDS WITH ASYMPTOTICALLY CYLINDRICAL ENDS

ERKAO BAO

Symplectic field theory is the study of *J*-holomorphic curves in almost complex manifolds with cylindrical ends. One natural generalization is to replace "cylindrical" by "asymptotically cylindrical". We generalize a number of asymptotic results about the behavior of *J*-holomorphic curves near infinity to the asymptotically cylindrical setting. We also sketch how these asymptotic results allow compactness theorems in symplectic field theory to be extended to the asymptotically cylindrical case.

1. Introduction

Introduced by Gromov [1985], J-holomorphic curves have been studied intensively in closed symplectic manifolds. Hofer [1993] studied the behaviors of J-holomorphic curves in symplectizations of contact manifolds, which are noncompact. Shortly after that, Eliashberg, Givental and Hofer [2000] invented symplectic field theory, which greatly helps us understand symplectic manifolds and contact manifolds. In most of the previous literature, the almost complex structure J is cylindrical near the ends of the noncompact symplectic manifolds. Here cylindrical means that J is independent of the radial direction. In [Bourgeois et al. 2003] the notion was introduced of an asymptotically cylindrical almost complex structure, which is a natural generalization of a cylindrical almost complex structure. However, no results corresponding to the notion of asymptotically cylindrical almost

MSC2010: primary 53D05, 53D10, 53D12, 53D40, 53D42; secondary 58J05.

Keywords: asymptotically cylindrical almost complex structure, symplectic field theory, compactness, Hofer energy, J-holomorphic curve, Morse–Bott, stable Hamiltonian structure.

complex structures in that paper have been proven. Intuitively, we expect similar results as in the cylindrical case. However, the original proofs rely heavily on the cylindrical nature of the almost complex structure, which prevents us from a direct generalization to the asymptotically cylindrical case. In this paper, we give a modified definition of asymptotically cylindrical almost complex structure, which includes an exponential decay condition that is satisfied in all interesting examples, and prove some parallel analytical results as in the cylindrical case. Based on these results we can compactify the moduli space of J-holomorphic curves in almost complex manifolds with asymptotically cylindrical ends by adding the holomorphic buildings introduced by [Bourgeois et al. 2003].

This generalization is needed for application purposes, since in many cases the natural almost complex structure is only asymptotically cylindrical (see Examples 2.5 and 4.1). For instance, we can use the generalized results to prove Gromov's monotonicity theorem with multiplicity (see [Bao 2014]). We also take this chance to fill in some gaps in the literature.

In the asymptotically cylindrical case, the proofs of some theorems are significantly different and more sophisticated than the proofs in the cylindrical case (see the proofs of Proposition 3.4, Theorem 2.8 and Theorem 3.7, for example). The extra difficulties mainly come from the following two facts: (1) the translations in the cylindrical almost complex manifold are not J-holomorphic anymore; (2) the unmodified Hofer energy is not positive when restricted to J-complex planes, and the modified Hofer energy is not closed. Crucial uses of Gromov's monotonicity theorem are the main ingredients to overcoming these extra difficulties.

In Section 2, we give the definition of asymptotically cylindrical almost complex manifolds and the definition of Hofer energy of J-holomorphic curves in this context.

In Section 3, we give the proofs of the main results listed in Section 2. The proofs follow the schemes of [Hofer 1993; Hofer et al. 2001; Hofer et al. 2002; Bourgeois 2002; Bourgeois et al. 2003].

In Section 4, we give the definition of almost complex manifolds with asymptotically cylindrical ends and the definition of Hofer energy in this context. Finally we state and outline the proof of the compactness result in this context.

2. Asymptotically cylindrical almost complex structures

2A. *Definitions.* Let V be a smooth closed oriented manifold of dimension $2n+1$, and let J be a smooth almost complex structure in $W := \mathbb{R}^+ \times V$. Assume that the orientation of W determined by J is the same as the orientation coming from the standard orientation of $\mathbb{R}^+$ and the orientation of V. Let $\boldsymbol{R} := J(\partial/\partial r)$ be a smooth vector field on W, and let ξ be a subbundle of the tangent bundle TW

defined by $\xi_{(r,v)} = (0 \times T_v V) \cap J(0 \times T_v V) \subset T_{(r,v)} W$, for $(r, v) \in W$. The tangent bundle TW splits as $TW = \mathbb{R}(\partial/\partial r) \oplus \mathbb{R}(\boldsymbol{R}) \oplus \xi$.

Define a 1-form λ on W by $\lambda(\xi) = 0,\ \lambda(\partial/\partial r) = 0,\ \lambda(\boldsymbol{R}) = 1$, and a 1-form σ on W by $\sigma(\xi) = 0,\ \sigma(\partial/\partial r) = 1,\ \sigma(\boldsymbol{R}) = 0$.

We call a tensor on W translationally invariant if it is independent of the r-coordinate. Let $f_s : W \to W$ be the translation along the $\mathbb{R}^+$-direction defined by $f_s(r, v) := (r + s, v)$.

Definition 2.1. Under the above notation, J is called asymptotically cylindrical at positive infinity if, for all $l \in \mathbb{Z}_{\geq 0}$, the following five conditions are satisfied:

(AC1) There exists a smooth translationally invariant almost complex structure J_∞ on W and constants $K_l^+, \delta_l > 0$ such that

$$\left\| \nabla^l (J - J_\infty)|_{[r,+\infty)\times V} \right\|_{C^0} \leq K_l^+ e^{-\delta_l r} \tag{1}$$

for all $r \geq 0$, where $\|\cdot\|_{C^0}$ is computed using a translationally invariant metric g_W on W (for example, $g_W = dr^2 + g_V$), and ∇ is the corresponding Levi-Civita connection. We further require that K_l^+ is sufficiently small such that the ω defined in Equation (2) satisfies requirements (a) and (b) in Section 2B. (Remark 2.2 explains that K_l^+ being small is not restrictive.)

(AC2) $i(\boldsymbol{R}_\infty)\, d\lambda_\infty = 0$, where $\boldsymbol{R}_\infty := \lim_{s\to\infty} f_s^* \boldsymbol{R}$, $\lambda_\infty := \lim_{s\to\infty} f_s^* \lambda$, and both limits exist by (AC1).

(AC3) $\boldsymbol{R}_\infty(r, v) = J_\infty(\partial/\partial r) \in 0 \times T_v V$.

(AC4) There exists a closed 2-form ω_∞ on V such that $i(\boldsymbol{R}_\infty)\omega_\infty = 0$.

(AC5) $\omega_\infty(\cdot\,, J_\infty\,\cdot)$ is a metric on ξ_∞, where $\xi_\infty = \lim_{s\to\infty} f_s^* \xi$.

Remark 2.2. The definition we use is slightly different from the one in [Bourgeois et al. 2003]. We require that J converges to J_∞ exponentially fast in condition (AC1). This is the accurate condition to guarantee that the J-holomorphic curve converges to the periodic orbits of $\boldsymbol{R}_\infty$ exponentially fast by the footnote of formula (35). If we are only interested in the behavior of a J-holomorphic curve near infinity, then the requirement that K_l^+ is small can be achieved by restricting W to $r \geq r_0$ for some large r_0.

We can restate the above conditions using the notion of hamiltonian structure as in [Eliashberg 2007]. That the 2-form ω_∞ has rank $2n$ says that (V, ω_∞) is a hamiltonian structure. The conditions (AC3), $i(\boldsymbol{R}_\infty)\omega_\infty = 0 = i(\boldsymbol{R}_\infty)\, d\lambda_\infty$ and $\lambda_\infty(\boldsymbol{R}_\infty) = 1$ say that (V, ω_∞) is a stable hamiltonian structure. The condition $\xi_\infty = \ker \lambda_\infty$, that J_∞ is an almost complex structure on ξ_∞, and that J_∞ is compatible with ω_∞ (by (AC5)) imply that $(\lambda_\infty, J_\infty)$ is a framing of (V, ω_∞). If in addition $\omega_\infty = d\lambda_\infty$, then we say (V, ω_∞) is of contact type.

We call (λ, J) defined as above an asymptotically cylindrical framing of the stable hamiltonian structure (V, ω_∞).

Similarly, we can define the notion of J being asymptotically cylindrical on $\mathbb{R}^- \times V$ at $-\infty$. When we say J is asymptotically cylindrical, we choose $\omega_{\pm\infty}$ without mention.

The following definition is the case considered in [Hofer 1993; Hofer et al. 2001; Hofer et al. 2002; Bourgeois 2002; Bourgeois et al. 2003].

Definition 2.3. An almost complex structure J on $\mathbb{R}^\pm \times V$ is said to be a cylindrical almost complex structure at $\pm\infty$ if J is an asymptotically cylindrical almost complex structure at $\pm\infty$ and J is translationally invariant near $\pm\infty$.

An almost complex structure J on $\mathbb{R} \times V$ is said to be a cylindrical almost complex structure if J is asymptotically cylindrical at both ∞ and $-\infty$ and J is translationally invariant.

Example 2.4 (Symplectization). Assume (V, ξ) is a contact manifold with contact 1-form λ and Reeb vector field $\boldsymbol{R}$, i.e., $\xi = \ker \lambda$, $\lambda \wedge (d\lambda)^n \neq 0$, $i_{\boldsymbol{R}}\, d\lambda = 0$, and $\lambda(\boldsymbol{R}) = 1$. Let $\omega_\infty = d\lambda$ and let J_ξ be an almost complex structure in ξ such that it is compatible with $\omega_\infty|_\xi$, i.e., $d\lambda(\,\cdot\,, J_\xi\,\cdot)$ is a metric on ξ. We extend J_ξ to $\mathbb{R} \times V$ by setting $J(\partial/\partial r) = \boldsymbol{R}$. Then J is a cylindrical almost complex structure and, in particular, an asymptotically cylindrical almost complex structure at $\pm\infty$.

Refer to [Bourgeois et al. 2003] for other interesting examples of cylindrical almost complex structures.

Example 2.5. Assume J is a smooth almost complex structure on $\mathbb{R}^{2n+2}$ with $J(0) = J_0(0)$, where J_0 is the standard complex structure on $\mathbb{R}^{2n+2}$. Consider $\mathbb{R}^{2n+2} \backslash \{0\}$ and pick a polar coordinate chart

$$\varphi : \mathbb{R}^- \times S^{2n+1} \to \mathbb{R}^{2n+2} \backslash \{0\}, \qquad (r, \Theta) \mapsto e^r \Theta,$$

where we view S^{2n+1} as the unit sphere inside $\mathbb{R}^{2n+2}$. Let $\lambda_{-\infty}$ be the standard contact form on S^{2n+1}. Define the 2-form $\omega_{-\infty}$ on $\mathbb{R}^- \times S^{2n+1}$ by $\omega_{-\infty} = d\lambda_{-\infty}$. Now it is clear that $J|_{\mathbb{R}^- \times S^{2n+1}}$ is an asymptotically cylindrical almost complex structure near $-\infty$.

By (AC1) and (AC3) we can see that $\boldsymbol{R}_\infty$ is a translationally invariant vector field on W and that it is tangent to each level set $\{r\} \times V$, so we can view $\boldsymbol{R}_\infty$ as a vector field on V. Let ϕ^t be the flow of $\boldsymbol{R}_\infty$ on V, i.e., let $\phi^t : V \to V$ satisfy $(d/dt)\phi^t = \boldsymbol{R}_\infty \circ \phi^t$. Then we have

$$\frac{d}{dt}[(\phi^t)^* \lambda_\infty] = (\phi^t)^* (i(\boldsymbol{R}_\infty)\, d\lambda_\infty + d i(\boldsymbol{R}_\infty)\lambda_\infty) = 0.$$

Hence ϕ^t preserves λ_∞ and thus also ξ_∞. Similarly, ϕ^t preserves ω_∞.

Let's denote by $\mathcal{P}$ the set of periodic trajectories, counting their multiples, of the vector field $\boldsymbol{R}_\infty$ restricting to V. Notice that any smooth family of periodic trajectories from $\mathcal{P}$ has the same period by Stokes' theorem.

Definition 2.6. A T-periodic orbit γ of $\boldsymbol{R}_\infty$ is called nondegenerate if $d\phi^T|_{\xi_\infty(\gamma(0))}$ does not have 1 as an eigenvalue, where ϕ^t is the flow of $\boldsymbol{R}_\infty$. We say that J is nondegenerate if all the periodic solutions of $\boldsymbol{R}_\infty$ are nondegenerate.

A weaker requirement for J than nondegenerate is Morse–Bott.

Definition 2.7. We say that J is of the Morse–Bott type if, for every $T > 0$, the subset $N_T \subset V$ formed by the closed trajectories from $\mathcal{P}$ of period T is a smooth closed submanifold of V such that the rank of $\omega_\infty|_{N_T}$ is locally constant and $T_pN_T = \ker(d\phi^T - \mathrm{Id})_p$.

We always assume J is of Morse–Bott type in this paper.

2B. *Energy of J-holomorphic curves.* Let J be an asymptotically cylindrical almost complex structure on $W := \mathbb{R}^+ \times V$. Let's denote the projections from $TW = \mathbb{R}(\partial/\partial r) \oplus \mathbb{R}(\boldsymbol{R}) \oplus \xi$ to each subbundle by π_r, $\pi_{\boldsymbol{R}}$ and π_ξ. It is convenient to introduce a 2-form ω on W by

$$\omega(x, y) = \tfrac{1}{2}[\omega_\infty(\pi_\xi x, \pi_\xi y) + \omega_\infty(J\pi_\xi x, J\pi_\xi y)]. \tag{2}$$

It is easy to check that $i(\partial/\partial r)\omega = 0 = i(\boldsymbol{R})\omega$. We assume that K_l^+ in (AC1) is sufficiently small for all $l \in \mathbb{Z}_{\geq 0}$ such that ω satisfies the following two conditions:

(a) $\omega|_\xi(\cdot\,, J\,\cdot)$ is a metric on ξ.

(b) There exist constants $\varepsilon_l, \delta_l > 0$ such that, for all $r \geq 0$,

$$\left\| (\omega - \omega_\infty)|_{[r,+\infty)\times V} \right\|_{C^l} \leq \varepsilon_l e^{-\delta_l r}.$$

Let (Σ, j) be a punctured Riemann surface (with or without boundary) and let $\tilde{u} = (a, u) : (\Sigma, j) \to (W, J)$ be a J-holomorphic curve, i.e., $T\tilde{u} \circ j = J(\tilde{u}) \circ T\tilde{u}$. The following definition is a modification of Hofer energy in the cylindrical almost complex structure case. The ω-energy and λ-energy are defined, respectively, as

$$E_\omega(\tilde{u}) = \int_\Sigma \tilde{u}^*\omega, \qquad E_\lambda(\tilde{u}) = \sup_{\phi \in \mathcal{C}} \int_\Sigma \tilde{u}^*(\phi(r)\,\sigma \wedge \lambda),$$

where $\mathcal{C} = \{\phi \in C_c^\infty(\mathbb{R}, [0, 1]) : \int_{-\infty}^{+\infty} \phi(x)\,dx = 1\}$[1], and λ, σ are defined as in the beginning of Section 2A. Let's define the energy of $\tilde{u}$ by

$$E(\tilde{u}) = E_\omega(\tilde{u}) + E_\lambda(\tilde{u}).$$

[1]In [Bourgeois et al. 2003], the set $\mathcal{C}$ is given by $\mathcal{C} = \{\phi \in C_c^\infty(\mathbb{R}, \mathbb{R}^+) : \int_{-\infty}^{+\infty} \phi(x)\,dx = 1\}$. It is easier to get uniform energy bounds using the modified definition in the case when the almost complex structure is only asymptotically cylindrical.

Equip $\mathbb{R}^+\times S^1$ with the standard complex structure and coordinate (s,t), and consider a J-holomorphic map $\tilde{u}=(a,u):\mathbb{R}^+\times S^1\to W$. Here we view S^1 as $\mathbb{R}/\mathbb{Z}$. Notice that

$$\tilde{u}^*\omega=\omega(\pi_\xi\tilde{u}_s, J(\tilde{u})\pi_\xi\tilde{u}_s)\,ds\wedge dt, \tag{3}$$

$$\tilde{u}^*(\phi(r)\,\sigma\wedge\lambda)=\phi(a)[\sigma(\tilde{u}_s)^2+\lambda(\tilde{u}_s)^2]\,ds\wedge dt. \tag{4}$$

Thus, we have $E_\omega(\tilde{u})\ge 0$ and $E_\lambda(\tilde{u})\ge 0$.

2C. ***Main results.*** The next two theorems tell us the behaviors of J-holomorphic curves near infinity.

Theorem 2.8. *Suppose that J is an asymptotically cylindrical almost complex structure on $\mathbb{R}^\pm\times V$ at $\pm\infty$, and suppose that J is of the Morse–Bott type. Let $\tilde{u}=(a,u):\mathbb{R}^\pm\times\mathbb{R}/\mathbb{Z}\to\mathbb{R}^\pm\times V$ be a finite energy J-holomorphic curve. Suppose that the image of $\tilde{u}$ is unbounded in $\mathbb{R}^\pm\times V$. Then there exists a periodic orbit γ of $\boldsymbol{R}_\infty$ of period $|T|$ with $T\neq 0$ such that, in $C^\infty(S^1)$,*

$$\lim_{s\to\pm\infty}u(s,t)=\gamma(Tt)\quad\text{and}\quad\lim_{s\to\pm\infty}\frac{a(s,t)}{s}=T.$$

The above theorem tells us that when $|s|$ is large enough $u(s,t)$ lies inside a small neighborhood of γ. We will construct a coordinate chart for such a neighborhood $U\subset S^1\times\mathbb{R}^{2n}\to V$, and then we can view the map $\tilde{u}$ as

$$\tilde{u}(s,t)=(a(s,t),\vartheta(s,t),z(s,t))\in\mathbb{R}^\pm\times\mathbb{R}\times\mathbb{R}^{2n},$$

where ϑ is the coordinate of the universal cover of $S^1=\mathbb{R}/\mathbb{Z}$.

Theorem 2.9. *Under the same assumption as in Theorem 2.8, there exist constants $M_\beta, d_\beta, a_0, \vartheta_0, s_0>0$ such that*

$$\begin{aligned}|D^\beta\{a(s,t)-Ts-a_0\}|&\le M_\beta e^{\mp d_\beta s},\\ |D^\beta\{\vartheta(s,t)-Tt-\vartheta_0\}|&\le M_\beta e^{\mp d_\beta s},\\ |D^\beta z(s,t)|&\le M_\beta e^{\mp d_\beta s},\end{aligned}$$

for all $s>s_0$ and $\beta=(\beta_1,\beta_2)\in\mathbb{Z}_{\ge 0}\times\mathbb{Z}_{\ge 0}$.

3. Proof of main results

The proofs for $\mathbb{R}^+\times V$ and $\mathbb{R}^-\times V$ are almost the same, so we will focus on the $\mathbb{R}^+\times V$ case. The proof is done in three steps. The first step is to show that the gradient of a finite Hofer energy J-holomorphic curve $\tilde{u}=(a,u)$ is bounded. The second step is to show “subsequence convergence”: briefly, given a sequence of numbers R_k converging to infinity, we want to show that there exists a subsequence R_{k_n} such that $u(R_{k_n}, t)$ converges to a periodic solution of the vector field $\boldsymbol{R}_\infty$. The third step is to get an exponential decay estimate and then prove Theorems 2.8 and 2.9.

3A. *Gradient bounds.* We cite the following two lemmata for later use.

Lemma 3.1 [Hofer 1993]. *Let (X, d) be a metric space. The following statements are equivalent*:

(a) *(X, d) is complete.*

(b) *For every continuous map $\phi : X \to [0, +\infty)$ and a given $x \in X$, $\varepsilon > 0$ there exist $x' \in X$, $\varepsilon' > 0$ such that*
   - $\varepsilon' \le \varepsilon$, $\phi(x')\varepsilon' \ge \phi(x)\varepsilon$,
   - $d(x, x') \le 2\varepsilon$,
   - $2\phi(x') \ge \phi(y)$ *for all $y \in X$ with $d(y, x') \le \varepsilon'$.*

Let J be an asymptotically cylindrical almost complex structure on $W = \mathbb{R}^+ \times V$ at ∞, and let $\tilde{u} = (a, u)$ be a J-holomorphic map from $B(0, R)$ to W, where $B(z_0, R) := z = \{s + \sqrt{-1}t \in \mathbb{C} : |z - z_0| < R\}$. Define

$$\|\nabla\tilde{u}\| := \sup_{(s,t)\in B(0,R)} |\nabla\tilde{u}(s, t)| \tag{5}$$

and

$$\|\tilde{u}\|_{C^k(B(0,R),W)} := \sup_{x\in B(0,R)} \sum_{|l|=0}^{k} |\nabla^l\tilde{u}(x)|,$$

where the norm $|\cdot|$ is taken with respect to the standard metric ds^2+dt^2 on $B(z_0, R)$ and to a translationally invariant metric g_W on W (for example, $g_W = g_V + dr^2$), and ∇ is the Levi-Civita connection with respect to g_W on W. The following lemma says that the gradient bound implies a C^∞ bound.

Lemma 3.2 (Gromov–Schwarz). *Fix $0 < \varepsilon < 1$ and $k \in \mathbb{N}$. If $\|\nabla\tilde{u}\| < C' < +\infty$, then there exists a $C(k, C') > 0$ such that*

$$\|\tilde{u}\|_{C^k(B(0,R-\varepsilon),W)} \le C(k, C'),$$

where $C(k, C')$ does not depend on $\tilde{u}$.

Proof. This is a standard result. Using the gradient bound of $\tilde{u}$, we can find uniform coordinate charts both in domain and in target, then we can apply Proposition 2.36 in [Audin and Lafontaine 1994]. □

The following proposition, whose proof reveals the relation between the ω-energy and trajectory of $\boldsymbol{R}_\infty$, is one of the key steps in [Hofer 1993].

Proposition 3.3 [Hofer 1993]. *Suppose J is a cylindrical almost complex structure on $\mathbb{R} \times V$ and let $\tilde{u} = (a, u) : \mathbb{C} \to \mathbb{R} \times V$ be a finite Hofer energy J-holomorphic* *plane (i.e.,* $E(\tilde{u}) = E_\lambda(\tilde{u}) + E_\omega(\tilde{u}) < +\infty$*). If* $E_\omega(\tilde{u}) = 0$ *and* $\|\nabla\tilde{u}\| \le C$ *for some* $C > 0$*, then* $\tilde{u}$ *is constant.*

Proof. Suppose $\tilde{u}$ is not constant. By (3), $\pi_\xi \tilde{u}_s = 0 = \pi_\xi \tilde{u}_t$. Hence $\pi_\xi \circ T\tilde{u}$ is the zero section of $\tilde{u}^*\xi \to \mathbb{C}$. Therefore we have $u(s,t) = x \circ f(s,t)$, where $x : \mathbb{R} \to V$ satisfies $\dot{x} = \boldsymbol{R}(x)$ and $f : \mathbb{C} \to \mathbb{R}$ is a smooth function. Consequently, $f_s = -a_t$ and $f_t = a_s$. Hence $\Phi := f + ia$ is a holomorphic function on $\mathbb{C}$. Since $\|\nabla\tilde{u}\|$ is bounded, $\|\nabla\Phi\|$ is bounded; thus Φ is a linear function. By (4),

$$E_\lambda(\tilde{u}) = \sup_{\phi\in\mathcal{C}} \int_{\mathbb{C}} \phi(a)(a_s^2 + a_t^2)\, ds \wedge dt = +\infty,$$

via a linear change of variables. □

The proposition below generalizes Proposition 27 in [Hofer 1993] to the asymptotically cylindrical case.

Proposition 3.4. *If* J *is an asymptotically cylindrical almost complex structure on* $W = \mathbb{R}^+ \times V$ *at* ∞*, and* $\tilde{u}$ *is a* J*-holomorphic map from* $\mathbb{C}$ *to* W *satisfying* $E(\tilde{u}) < +\infty$*, then* $\|\nabla\tilde{u}\| < +\infty$*.*

Proof. Suppose to the contrary that there exists a sequence of points $z_k \in \mathbb{C}$ satisfying $|z_k| \to \infty$, $R_k := \|\nabla\tilde{u}(z_k)\| \to \infty$, as $k \to \infty$. By Lemma 3.1, we can modify z_k such that there exists a sequence of $\varepsilon_k > 0$ satisfying $\varepsilon_k \to 0$, $\varepsilon_k R_k \to +\infty$, and $|\nabla\tilde{u}(z)| \le 2R_k$ for $z \in B(z_k, \varepsilon_k)$. Now there are two cases.

Case 1: $\{a(z_k)\}_{k\in\mathbb{Z}}$ is unbounded.

Then there exists a subsequence of z_k, still denoted by z_k, such that $a(z_k) \to +\infty$ or $a(z_k) \to -\infty$. Without loss of generality, let's assume $a(z_k) \to +\infty$. Pick a further subsequence of z_k such that $a(z_k) \ge 2^{k+2}$. Let $\varepsilon'_k := \min\{\varepsilon_k, 2^k/R_k\}$. Then we have $\varepsilon'_k \to 0$, $\varepsilon'_k R_k \to +\infty$, and $|a(z) - a(z_k)| \le 2\varepsilon'_k R_k \le 2(2^k/R_k)R_k = 2^{k+1}$, for $|z - z_k| \le \varepsilon'_k$. Thus, $a(z) \ge a(z_k) - 2^{k+1} \ge 2^{k+2} - 2^{k+1} = 2^{k+1}$, for $|z - z_k| \le \varepsilon'_k$.

Since $\tilde{u}$ is J-holomorphic, we have

$$J(\tilde{u}) \circ T\tilde{u} = T\tilde{u} \circ i. \tag{6}$$

Thus

$$J_\infty(\tilde{u}) \circ T\tilde{u} = T\tilde{u} \circ i + (J_\infty - J)(\tilde{u}) \circ T\tilde{u}. \tag{7}$$

By (AC1), we have, as $k \to +\infty$,[2]

$$\sup_{z\in B(z_k,\varepsilon'_k)} \|(J_\infty - J)(\tilde{u}(z))\| \to 0.$$

[2]Actually, to prove Proposition 3.4, Proposition 3.5 and Theorem 3.7 we only need $f_s^* J \to J_\infty$ in C^1_{loc} as $s \to \infty$. We need the stronger condition (AC1) to prove exponential decay in Section 3C and thus the main theorems.

Define maps $\tilde{u}_k(z) = (a(z_k + z/R_k) - a(z_k), u(z_k + z/R_k))$ from $\mathbb{C}$ to $\mathbb{R} \times V$. For any $R' > 0$, when k is large, $\|\nabla \tilde{u}_k(z)\| \le 2$ for $z \in B(0, R')$. By Lemma 3.2, for any $n \in \mathbb{Z}_{\ge 0}$, there exists a $C(n, R')$ satisfying

$$\|\tilde{u}_k\|_{C^n(B(0,R'-1),W)} \le C(n, R'). \tag{8}$$

We also have

$$|\nabla \tilde{u}_k(0)| = 1, \tag{9}$$

$$|\nabla \tilde{u}_k(z)| \le 2 \quad \text{for all } |z| \le \varepsilon_k' R_k. \tag{10}$$

We apply the Ascoli–Arzela theorem to get a subsequence, still called $\tilde{u}_k$, satisfying $\tilde{u}_k \to \tilde{u}_\infty$ in C^∞_{loc} as $k \to \infty$. Here $\tilde{u}_\infty : \mathbb{C} \to \mathbb{R} \times V$ is a J_∞-holomorphic map satisfying

$$|\nabla \tilde{u}_\infty(0)| = 1 \quad \text{and} \quad \|\nabla \tilde{u}_\infty\| \le 2.$$

Indeed, $\tilde{u}_k$ satisfies

$$J_\infty(\tilde{u}_k) T\tilde{u}_k = T\tilde{u}_k i + o_k, \tag{11}$$

where $\|o_k\|_{C^0(B(0,\varepsilon_k' R_k))} \to 0$ as $k \to \infty$. Therefore, $\tilde{u}_\infty$ is J_∞-holomorphic.

Now let's look at its energy:

$$\int_{B(0,R')} \tilde{u}_k^* \omega_\infty = \int_{B(z_k,R'/R_k)} \tilde{u}^* \omega + \int_{B(z_k,R'/R_k)} \tilde{u}^*(\omega - \omega_\infty). \tag{12}$$

From $E(\tilde{u}) < +\infty$ we see that $\int_{B(z_k,R'/R_k)} \tilde{u}^* \omega \to 0$ as $k \to +\infty$. We also have

$$\begin{aligned}\left| \int_{B(z_k,R'/R_k)} \tilde{u}^*(\omega_\infty - \omega) \right| &\le \int_{B(z_k,R'/R_k)} (2R_k)^2 \left| (\omega_\infty - \omega)\left(\frac{\tilde{u}_s}{2R_k}, \frac{\tilde{u}_t}{2R_k} \right) \right| ds \wedge dt \\ &\le \pi \left(\frac{R'}{R_k} \right) (2R_k)^2 c_k \to 0,\end{aligned}$$

where

$$c_k := \sup_{z \in B(z_k, \varepsilon_k')} \left| (\omega_\infty - \omega)\left(\frac{\tilde{u}_s}{2R_k}, \frac{\tilde{u}_t}{2R_k} \right) \right|,$$

and by (AC4) $c_k \to 0$ as $k \to \infty$. Therefore,

$$E_{\omega_\infty}(\tilde{u}_\infty) = \int_{\mathbb{C}} \tilde{u}_\infty^* \omega_\infty = 0.$$

Moreover, we have $E_{\lambda_\infty}(\tilde{u}_\infty) < +\infty$. Given $\phi \in \mathcal{C}$, define $\phi_k(r) := \phi(r - a(z_k)) \in \mathcal{C}$. Then we have

$$\begin{aligned}(13)\quad & \Big| \int_{B(0,R')} \tilde{u}_k^*(\phi(r)\, dr \wedge \lambda_\infty) \Big| \\ &\le \Big| \int_{B(z_k,R'/R_k)} \phi_k(a) \tilde{u}^*(\sigma \wedge \lambda) \Big| + \Big| \int_{B(z_k,R'/R_k)} \phi_k(a) \tilde{u}^*(dr \wedge \lambda_\infty - \sigma \wedge \lambda) \Big|.\end{aligned}$$

We also have

$$(14)\qquad \Big| \int_{B(z_k,R'/R_k)} \phi_k(a) \tilde{u}^*(\sigma \wedge \lambda) \Big| \le \Big| \int_{\mathbb{C}} \phi_k(a) \tilde{u}^*(\sigma \wedge \lambda) \Big| \le E_\lambda(\tilde{u})$$

and

$$\begin{aligned}(15)\quad & \Big| \int_{B(z_k,R'/R_k)} \phi_k(a) \tilde{u}^*(dr \wedge \lambda_\infty - \sigma \wedge \lambda) \Big| \\ &\le \int_{B(z_k,R'/R_k)} \phi_k(a)(2R_k)^2 \Big| (dr \wedge \lambda_\infty - \sigma \wedge \lambda)\Big(\frac{\tilde{u}_s}{2R_k}, \frac{\tilde{u}_t}{2R_k}\Big) \Big|\, ds \wedge dt \\ &\le \big(\sup_{x \in \mathbb{R}} \phi(x)\big)(2R_k)^2 r_k \pi \Big(\frac{R'}{R_k}\Big)^2 \to 0,\end{aligned}$$

where

$$r_k := \sup_{z \in B(z_k,R'/R_k)} \Big| (dr \wedge \lambda_\infty - \sigma \wedge \lambda)\Big(\frac{\tilde{u}_s}{2R_k}, \frac{\tilde{u}_t}{2R_k}\Big) \Big| \to 0$$

as $k \to \infty$. Combining (13), (14) and (15), we get the following result: given $R' > 0$ and $\phi \in \mathcal{C}$, there exists a constant K such that, for all $k > K$,

$$\Big| \int_{B(0,R')} \tilde{u}_k^*(\phi(r)\, dr \wedge \lambda_\infty) \Big| \le E_\lambda(\tilde{u}) + 1.$$

Therefore, $E_{\lambda_\infty}(\tilde{u}_\infty) \le E_\lambda(\tilde{u}) + 1$. Altogether, we get a J_∞-holomorphic map $\tilde{u}_\infty : \mathbb{C} \to W$ satisfying

$$\|\nabla \tilde{u}_\infty\| \le 2, \quad |\nabla \tilde{u}_\infty(0)| = 1, \quad E_{\omega_\infty}(\tilde{u}_\infty) = 0, \quad E(\tilde{u}_\infty) < +\infty.$$

By Proposition 3.3, we get a contradiction, which finishes the proof for Case 1.

Case 2: $\{a(z_k)\}_{k \in \mathbb{Z}}$ is bounded.

Now let us define $\tilde{u}_k$, differently from Case 1, by

$$\tilde{u}_k(z) := \tilde{u} \circ l_k = (a(z_k + z/R_k), u(z_k + z/R_k)).$$

Then $\tilde{u}_k$ satisfies

$$|\nabla \tilde{u}_k(z)| \le 2 \quad \text{for } z \in B(0, \varepsilon_k R_k),$$
$$\{\tilde{u}_k(0)\}_{k\in\mathbb{Z}^+} \text{ is bounded,}$$
$$|\nabla \tilde{u}(0)| = 1.$$

Similar to Case 1, by applying the Ascoli–Arzela theorem we get a subsequence, still called $\tilde{u}_k$, converging to $\tilde{u}_\infty = (a_\infty, u_\infty) : \mathbb{C} \to W$ in the C^∞_{loc} sense. Here $\tilde{u}_\infty$ is J-holomorphic, satisfying

$$|\nabla \tilde{u}_\infty(0)| = 1, \tag{16}$$
$$\|\nabla \tilde{u}_\infty\| \le 2, \tag{17}$$
$$\int_{B(0,\varepsilon_k R_k)} \tilde{u}_k^* \omega = \int_{B(z_k,\varepsilon_k)} \tilde{u}^* \omega \to 0 \quad \text{as } k \to +\infty. \tag{18}$$

Thus, $E_\omega(\tilde{u}_\infty) = \int_{\mathbb{C}} \tilde{u}_\infty^* \omega = 0$. Moreover, given $R' > 0$ and $\phi \in \mathcal{C}$, we have

$$\int_{B(0,R')} \tilde{u}_k^*[\phi(r)\,\sigma \wedge \lambda] = \int_{B(z_k, R'/R_k)} \tilde{u}^*[\phi(r)\,\sigma \wedge \lambda] \to 0$$

as $k \to +\infty$. This means $\int_{B(0,R')} \tilde{u}_\infty^*[\phi(r)\,\sigma \wedge \lambda] = 0$, and so $E_\lambda(\tilde{u}_\infty) = 0$. Hence, $\tilde{u}_\infty$ is constant, contradicting (16). □

Proposition 3.5. *Suppose J is a cylindrical almost complex structure on $\mathbb{R} \times V$. Let $\tilde{v} : \mathbb{R}^+ \times S^1 \to W$ be a J-holomorphic map with respect to the standard complex structure on $\mathbb{R}^+ \times S^1$, and assume $E(\tilde{v}) < +\infty$. Then we have*

$$\|\nabla \tilde{v}\| < +\infty, \quad \textit{where } \|\nabla \tilde{v}\| := \sup_{(s,t)\in\mathbb{R}^+\times S^1} |\nabla \tilde{v}(s,t)|,$$

and the norm $|\cdot|$ is taken with respect to the standard metric ds^2+dt^2 on $\mathbb{R}^+\times S^1$ and to a translationally invariant metric g_W on W, and ∇ is the Levi-Civita connection with respect to g_W.

Proof. The proof is almost the same as the proof of Proposition 3.4. □

Remark 3.6. Actually, we can see that we can get a gradient bound with respect to a metric g_D on the domain and a translationally invariant metric g_W on W, as long as the injectivity radius of g_D is bounded away from 0.

3B. *Subsequence convergence.*

Theorem 3.7. *Let J be an asymptotically cylindrical almost complex structure on $\mathbb{R}^\pm \times V$, and let $\tilde{v} = (a, v) : \mathbb{R}^\pm \times S^1 \to \mathbb{R}^\pm \times V$ be a J-holomorphic curve with $E(\tilde{v}) < +\infty$. Suppose that $\tilde{v}(\mathbb{R}^\pm \times S^1)$ is unbounded. Then for any sequence*

$k_n \to +\infty$, there exists a subsequence k_{n_i} such that $v(k_{n_i}, \cdot)$ converges in $C^\infty(S^1)$ to a map $S^1 \to V$ given by $t \mapsto x(tT)$, where $x : \mathbb{R} \to V$ is a $|T|$-periodic solution of $\dot{x} = \boldsymbol{R}_\infty(x)$.

Proof. We prove this theorem for the case $\mathbb{R}^+ \times V$. The proof for the $\mathbb{R}^- \times V$ case can be carried out similarly, and hence is omitted. By Proposition 3.5 we have $\|\nabla\tilde{v}\| \le C$ for some $C > 0$. Since $\tilde{v}(\mathbb{R}^+ \times S^1)$ is not bounded, there exists a sequence of points $(s_k, t_k) \in \mathbb{R}^+ \times S^1$ such that $|a(s_k, t_k)| \to +\infty$. Now there are two cases.

Case 1: $a(s_k, t_k) \to +\infty$.

Suppose that there exists a sequence of points $(s'_k, t'_k) \in \mathbb{R}^+ \times S^1$ such that $a(s'_k, t'_k) < Q$ for some constant Q. Pick a subsequence of (s_k, t_k), still called (s_k, t_k), and a subsequence of (s'_k, t'_k), still called (s'_k, t'_k), so that they satisfy $s'_k < s_k < s'_{k+1}$ for all k. This is possible because $s_k \to +\infty$. Since $\|\nabla\tilde{v}\| \le C$, we have $a(s'_k, t) < Q+C$ for $t \in S^1$. Consider the compact manifold $N := [Q, Q+2C] \times M \subset W = \mathbb{R}^+ \times V$. Pick a $\phi \in \mathcal{C}$ such that $\phi|_{[Q,Q+2C]} > 0$. By Gromov's monotonicity theorem (see for example Theorem 1.3 in [Hummel 1997]), there exists an $\iota > 0$ such that

$$\int_{\tilde{v}([s'_k, s_k] \times S^1)} \omega + \phi(r)\,\sigma \wedge \lambda \ge \iota > 0$$

for all k. This contradicts the fact that $E(\tilde{v}) < +\infty$. Thus $a(s,t) \to +\infty$ uniformly in t as $s \to +\infty$.

Define

$$\tilde{v}_n(s,t) = (a(s+k_n, t) - a(k_n, 0),\, v(s+k_n, t)).$$

Then the sequence $\tilde{v}_n(0,0) = (0, v(k_n, 0))$ is bounded. Since $\tilde{v}$ is J-holomorphic, by Lemma 3.2 and the Ascoli–Arzela theorem, there exists a subsequence, still called $\tilde{v}_n$, converging to $\tilde{v}_\infty = (b, v_\infty) : \mathbb{R} \times S^1 \to W$ in C^∞_{loc}. We know $\tilde{v}_\infty$ is J_∞-holomorphic. Define the translation map $\tau_n : \mathbb{R} \times S^1 \to \mathbb{R} \times S^1$ by $\tau_n(s,t) = (s+k_n, t)$. Now observe that

$$\int_{[-R,R]\times S^1} \tilde{v}_n^*\omega_\infty = \int_{[-R+k_n, R+k_n]\times S^1} \tilde{v}^*\omega + \int_{[-R+k_n, R+k_n]\times S^1} \tilde{v}^*(\omega_\infty - \omega). \tag{19}$$

For the first term on the right-hand side we have

$$\int_{[-R+k_n, R+k_n]\times S^1} \tilde{v}^*\omega \to 0 \tag{20}$$

as $n \to \infty$, since $E_\omega(\tilde{v})$ is finite. By (AC4), the second term satisfies, as $n \to +\infty$,

$$\int_{[-R+k_n, R+k_n]\times S^1} \tilde{v}^*(\omega_\infty - \omega) \le \int_{[-R+k_n, R+k_n]\times S^1} |(\omega_\infty - \omega)(\tilde{v}_s, \tilde{v}_t)|\, ds \wedge dt \to 0. \tag{21}$$

Combining (19), (20) and (21), we can see that $\int_{[-R,R]\times S^1} \tilde{v}_\infty^* \omega_\infty = 0$ and hence $E_{\omega_\infty}(\tilde{v}_\infty) = 0$, so there exists a smooth map $f : \mathbb{R}^2 \to \mathbb{R}$ such that $\tilde{v}_\infty = (b, x \circ f)$, where $x : \mathbb{R} \to V$ is the solution of $\dot{x} = \boldsymbol{R}_\infty(x)$. Let Φ be the holomorphic function defined by $\Phi = b + if$. Since $\|\nabla\Phi\| \leq C$, we know that Φ is linear. Thus, $\Phi(s,t) = \alpha(s+it) + \beta$, where $\alpha = T + il$, $\beta = m + in \in \mathbb{C}$ are constants. But $b(s,t) - b(s,t+1) = 0$ implies $l = 0$, and $b(0,0) = 0$ implies $m = 0$. Thus,

$$f = Tt + n, \tag{22}$$

$$b = Ts. \tag{23}$$

Therefore, $a_s(k_n, t) \to T$ uniformly in t as $n \to +\infty$ (recall the notation $\tilde{v} = (a, v)$, $\tilde{v}_\infty = (b, v_\infty)$). Moreover, we have

$$\int_{\{0\}\times S^1} \tilde{v}_\infty^* \lambda_\infty = \int_{\{0\}\times S^1} \lambda_\infty[(\tilde{v}_\infty)_t]\, dt = \int_{\{0\}\times S^1} b_s\, dt = T. \tag{24}$$

Claim: $T \neq 0$.

It follows from the claim and (22) that $\tilde{v}_\infty$ is not constant. Indeed, by (22), $f(s, t+1) = T(t+1) + n$, so $x(T(t+1)+n) = x(Tt+n)$. Hence, x is T-periodic.

Proof. Suppose $T = 0$. Since $a(s,t) \to +\infty$ uniformly in t as $s \to +\infty$, we can choose a subsequence k_{n_m} of k_n and a sequence $t_m \in S^1$ so that we have $a(k_{n_{m+1}}, t_{m+1}) - a(k_{n_m}, t_m) \geq 4C$. Denote $a(k_{n_m}, t_m)$ by a_m. Then from $\|\nabla\tilde{u}\| \leq C$ we get

$$a(k_{n_m}, t) \in [a_m - C, a_m + C], \tag{25}$$

$$a(k_{n_{m+1}}, t) \geq a_m + 3C. \tag{26}$$

Let $\psi_m : \mathbb{R} \to [0,1]$ be a smooth map, satisfying $\psi_m(r) = \frac{1}{7C}(r - a_m + \frac{3}{2}C)$ for $r \in [a_m - C, a_m + 5C]$ and $\phi_m = \psi_m' \in \mathcal{C}$. If we further require $C > 1$, then $\phi_m(r) \leq \frac{1}{7C} < 1$. Observe that

$$\int_{[k_{n_m}, k_{n_{m+1}}]\times S^1} \tilde{v}^*\, d(\psi_m(r)\lambda) = \int_{\{k_{n_{m+1}}\}\times S^1} \tilde{v}^*(\psi_m(r)\lambda) - \int_{\{k_{n_m}\}\times S^1} \tilde{v}^*(\psi_m(r)\lambda).$$

We also have, as $m \to +\infty$,

$$\left| \int_{\{k_{n_{m+1}}\}\times S^1} \tilde{v}^*(\psi_m(r)\lambda) \right| = \left| \int_{\{k_{n_{m+1}}\}\times S^1} \psi_m(\tilde{v})\lambda(\tilde{v}_t)\, dt \right| \leq \int_{\{k_{n_{m+1}}\}\times S^1} |\lambda(\tilde{v}_t)|\, dt \to T = 0.$$

Similarly, $\int_{\{k_{n_m}\}\times S^1} \tilde{v}^*(\psi_m(r)\lambda) \to 0$. Thus, by Stokes' theorem,

$$\int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*\, d(\psi_m(r)\lambda) \to 0. \tag{27}$$

Observe that

$$\begin{aligned}\int_{[k_{n_m},k_{n_{m+1}}]\times S^1} & \tilde{v}^*(\phi_m(r)\, \sigma \wedge \lambda) \\ &= \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*(\phi_m(r)\, dr \wedge \lambda) + \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*[\phi_m(r)\, (\sigma - dr) \wedge \lambda].\end{aligned} \tag{28}$$

For the first term on the right-hand side, we have, for some $c > 0,\ c_m > 0$,

$$\begin{aligned}\Bigg| \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} & \tilde{v}^*(\phi_m(r)\, dr \wedge \lambda)\Bigg| \\ &\leq \Bigg| \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*\, d(\psi_m(r)\lambda)\Bigg| + \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} |\tilde{v}^*(\psi_m(r)\, d\lambda)| \\ &\leq \Bigg| \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*\, d(\psi_m(r)\lambda)\Bigg| + \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*(c\omega + c_m\, \sigma \wedge \lambda).\end{aligned} \tag{29}$$

The second inequality is due to the fact that $c\omega + c_m\, \sigma \wedge \lambda$ is positive on all J-complex planes; also since $d\lambda \to d\lambda_\infty$ and $i(\partial/\partial r)\, d\lambda_\infty = 0 = i(\boldsymbol{R}_\infty)\, d\lambda_\infty$, we can require that c is independent of m and c_m goes to 0 as $m \to +\infty$. Similarly, we have

$$\Bigg| \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*[\phi_m(r)\, (\sigma - dr) \wedge \lambda]\Bigg| \leq \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*[c\omega + c_m\, \sigma \wedge \lambda]. \tag{30}$$

When k is large, from (28), (29) and (30) we get

$$\int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*(\phi_m(r)\, \sigma \wedge \lambda) \leq D\Bigg\{\Bigg| \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*\, d(\psi_m(r)\lambda)\Bigg| + \int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*\omega\Bigg\}, \tag{31}$$

for some constant $D > 0$ which does not depend on m and $\tilde{v}$. The term

$$\int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*(c_m\, \sigma \wedge \lambda)$$

does not show up on the right-hand side of (31) because it is absorbed by the left-hand side, since $\phi_m|_{[k_{n_m},k_{n_{m+1}}]\times S^1} = 1/7$. Since $E_\omega(\tilde{v})$ is finite,

$$\int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*\omega \to 0.$$

Together with (27), we get

$$\int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*(\phi_m(r)\,dr \wedge \lambda) \to 0.$$

Summing up, we have, as $m \to +\infty$,

$$\int_{[k_{n_m},k_{n_{m+1}}]\times S^1} \tilde{v}^*(\omega + \phi_m(r)\,dr \wedge \lambda) \to 0. \tag{32}$$

Now consider $N_m = [a_m + C, a_m + 3C] \times V \subseteq W$ with an almost complex structure $J_m := J|_{N_m}$ and a nondegenerate 2-form $\Omega_m := \omega + \phi_m(r)\,\sigma \wedge \lambda|_{N_m}$. Because of the asymptotic condition, we can find uniform constants $C_0, r_0 > 0$ such that by Gromov's monotonicity theorem, for any J_m-holomorphic curve $h_m : (S, j) \to (N_m, J_m)$, where (S, j) is a Riemann surface with boundary, if the boundary $h_m(\partial S)$ is contained in the complement of the ball $B(h_m(s_0), r)$, where $s_0 \in \operatorname{Int} S_m$ and $r < r_0$, then we have

$$\int_{h_m(S)\cap B(h_m(s_0),r))} \Omega_m \geq C_0 r^2.$$

By (25) and (26) we can see $\tilde{u}(k_{n_m}, S^1) \cap \operatorname{Int} N_m = \varnothing$ and $\tilde{u}(k_{n_{m+1}}, S^1) \cap \operatorname{Int} N_m = \varnothing$. This contradicts (32). Thus, $T \neq 0$. □

Case 2: $a(s_k, t_k) \to -\infty$.

We deal with this case similarly. □

Corollary 3.8. *Under the assumptions of Theorem 3.7, there exists a number $T > 0$ such that, as $s \to \pm\infty$,*

$$\partial^\beta[a(s,t) - Ts] \to 0 \tag{33}$$

uniformly in t, provided $\beta = (\beta_1, \beta_2) \in \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ and $|\beta| = \beta_1 + \beta_2 \geq 1$.

Proof. By Theorem 3.7, there exist a number $T > 0$ and a sequence of numbers s'_k such that $s'_k \to +\infty$ and $v(s'_k, \cdot) \to x(T\,\cdot)$, for some T-periodic orbit x of $\boldsymbol{R}_\infty$. Suppose (33) is not true for this T. Then there exists a sequence of points (s_k, t_k) such that $s_k \to +\infty$ and $\partial^\beta[a(s,t) - Ts]|_{(s_k,t_k)} \to c$ as $k \to +\infty$ for some $|\beta| \geq 1$, where c is a nonzero constant (or $\pm\infty$). Define $\bar{a}_k(s,t) := a(s+s_k, t+t_k) - a(s_k, t_k)$

and then $\bar{a}_k(0,0)=0$. From the proof of Theorem 3.7 we get a subsequence of k, still called k, and a T'-periodic orbit x' of $\boldsymbol{R}_\infty$ such that $\bar{a}_k \to T's$ in $C^\infty_{\mathrm{loc}}(\mathbb{R}^+\times S^1,\mathbb{R})$. By a straightforward modification of the proof of Proposition 2.1 in [Hofer et al. 2001] to the Morse–Bott case, we can show that x' and x lie in the same component of N_T (see Definition 2.7) and in particular $T'=T$. Thus,

$$\begin{aligned}\partial^\beta[a(s,t)-Ts]|_{(s_k,t_k)} &= \partial^\beta[a(s+s_k,t+t_k)-a(s_k,t_k)-Ts]|_{(0,0)}\\ &= \partial^\beta(\bar{a}_k(s,t)-Ts)|_{(0,0)}\\ &\to 0,\end{aligned}$$

which contradicts the assumption. □

To prove Theorems 2.8 and 2.9, we need to obtain exponential decay estimates.

3C. *Exponential decay estimates.* In this subsection, we will follow the schemes in [Bourgeois 2002] to prove Theorems 2.8 and 2.9. The strategy is as follows: firstly, we pick a neighborhood U of the orbit γ, restrict the J-holomorphic curve $\tilde{u}$ to a sequence of cylinders inside the domain so that the images lie in the neighborhood and satisfy certain inequalities, and estimate the behaviors of each finite cylinder by the behaviors of boundaries of the cylinder. Secondly, since we have a sequence of circles in the domain whose images lie in U, we get that the cylinders bounded by the circles also lie in U, based on the estimates. We also show that near the end of the domain $\tilde{u}$ satisfies the inequalities. Once these are achieved, Theorems 2.8 and 2.9 follow easily.

In order to study the J-holomorphic curve equation around γ, we need to introduce a good coordinate chart around a neighborhood of γ.

Lemma 3.9 [Bourgeois et al. 2003]. *Suppose that J_∞ is a cylindrical almost complex structure of the Morse–Bott type on $\mathbb{R}^+\times V$ at ∞. Let N be a component of the set $N_T\subset V$ (see Definition 2.7), and let γ be one of the orbits from N.*

(a) *If T is the minimal period of γ then there exists a neighborhood $U\supset\gamma$ in V such that $U\cap N$ is invariant under the flow of $\boldsymbol{R}_\infty$, and one finds coordinates $(\vartheta,x_1,\dots,x_n,y_1,\dots,y_n)$ of U such that*

$$N=\{x_1,\dots,x_p=0,\ y_1,\dots,y_q=0\},$$

for $0\le p,\ q\le n$, and

$$\boldsymbol{R}_\infty|_N=\frac{\partial}{\partial\vartheta},\qquad \omega_\infty|_N=\omega_0|_N,$$

where $\omega_0=\sum_{i=1}^n dx_i\wedge dy_i$.

(b) *If γ is an m-multiple of a trajectory $\bar{\gamma}$ of a minimal period T/m then there exists a tubular neighborhood $\bar{U}$ of $\bar{\gamma}$ such that its m-multiple cover U together* *with all the structures induced by the covering map from $U \to \bar{U}$ from the corresponding objects on $\bar{U}$ satisfy the properties of part (a).*

Proof. Refer to Lemma A.1 in [Bourgeois et al. 2003]. □

Using this coordinate chart, we can work locally in $U \subset (\mathbb{R}/T\mathbb{Z}) \times \mathbb{R}^{2n}$ and make T the minimal period of γ. Denote by z_{in} the coordinate $(x_1, \dots, x_p, y_1, \dots, y_q)$ and by z_{out} the coordinate $(x_{n-p+1}, \dots, x_n, y_{n-p+1}, \dots, y_n)$. We easily obtain the following lemma about the behavior of a J-holomorphic curve in the z_{out} direction.

Lemma 3.10. *Let J be an asymptotically cylindrical almost complex structure on $W = \mathbb{R}^+ \times V$, and let $\tilde{u}$ be a finite Hofer energy J-holomorphic curve from $\mathbb{R}^+ \times S^1$ to W. Suppose $[m_k, n_k]$ is a sequence of intervals in $\mathbb{R}^+$ with $m_k \to +\infty$ and $\tilde{u}([m_k, n_k] \times S^1) \subset U$. Then we have, as $k \to +\infty$,*

$$\sup_{(s,t)\in[m_k,n_k]\times S^1} |\partial^\beta z_{\text{out}}(s,t)| \to 0$$

for all $\beta \in \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$.

Proof. The proof is very similar to the proof of Corollary 3.8, so we omit it here. □

Let's study the J-holomorphic curve equation in $\mathbb{R}^+ \times U \subset \mathbb{R}^+ \times (\mathbb{R}/T\mathbb{Z}) \times \mathbb{R}^{2n}$. Define $\theta := [s_0, s_1] \times S^1$ for some $s_0 < s_1$ and let $\tilde{u} = (a, \vartheta, z) : \theta \to \mathbb{R} \times U$ be a J-holomorphic curve. Then we have

$$(a_s, \vartheta_s, z_s) + J(\tilde{u})(a_t, \vartheta_t, z_t) = 0. \tag{34}$$

Rewriting this equation according to its z-, ϑ-, and a-components we get[3]

$$z_s + M z_t + S z_{\text{out}} + L = 0, \tag{35}$$

$$a_s - \vartheta_t + B z_{\text{out}} + B' z_t + N = 0, \tag{36}$$

$$a_t + \vartheta_s + C z_{\text{out}} + C' z_s + O = 0, \tag{37}$$

where M, S, B, B', C, C' depend on $a(s,t)$, $\vartheta(s,t)$, $z(s,t)$ and are bounded by a constant C_0, and L, N, O depend on $a(s,t)$, $\vartheta(s,t)$, $z(s,t)$ and are bounded by $C_0 e^{-\delta a}$.

Define an operator $A(s) : W^{1,2}(S^1, \mathbb{R}^{2n}) \to L^2(S^1, \mathbb{R}^{2n})$ by

$$(A(s)w)(t) = -M(\tilde{u}(s,t)) w_t(t) - S(\tilde{u}(s,t)) w_{\text{out}}(t).$$

Then by (35) we get

$$A(s) z(s, \cdot) = z_s + L. \tag{38}$$

[3]From (35) we can see that if we require z, z_s and z_t to decay exponentially, L must decay exponentially. The condition $f_s^* J \to J_\infty$ in C^∞_{loc} is not enough to guarantee that L decays exponentially.

Notice that $A(s)$ depends on the map $\tilde{u} = (a, \vartheta, z_{\text{in}}, z_{\text{out}})$. If we do not use the original J-holomorphic curve $\tilde{u}$ and instead we substitute $\vartheta(s,t) = \vartheta(s_0, 0) + Tt$, $a(s,t) = Ts$, $z_{\text{out}}(s,t) = 0$, and $z_{\text{in}}(s,t) = z_{\text{in}}(s_0, t)$, then we get another operator denoted by $\tilde{A}(s)$. We can easily see that $\lim_{s\to+\infty} \tilde{A}(s)$ exists and denote the limiting operator by A_0. Similarly, we get two matrices $M_0(t)$ and $S_0(t)$, and then we have

$$M_0(t)^2 = -\text{id},$$

and

$$(A_0 w)(t) = -M_0(t) w_t(t) - S_0(t) w_{\text{out}}. \tag{39}$$

Consider an inner product on $L^2(S^1, \mathbb{R}^{2n})$ defined by

$$\langle u, v\rangle_0 = \int_0^1 \langle u, -J_0 M_0 v\rangle \, dt, \tag{40}$$

where the inner product is given by $\langle\,\cdot\,,\,\cdot\,\rangle = \omega_0(\,\cdot\,, J_0\,\cdot\,)$, and J_0 is the standard complex structure on $\mathbb{R}^{2n}$. With respect to the inner product $\langle\,\cdot\,,\,\cdot\,\rangle_0$, one can check directly that M_0 is antisymmetric and that A_0 is self-adjoint.

Remark 3.11. A_0 is injective if and only if γ is nondegenerate.

It is not hard to see that $\ker A_0$ consists of the constant vector fields in N along γ_0. Denote by P_0 the projection onto $\ker A_0$ with respect to $\langle\,\cdot\,,\,\cdot\,\rangle_0$, and let $Q_0 := I - P_0$. It is easy to check the following lemma.

Lemma 3.12. *Q_0 satisfies*

$$(Q_0 w)_t = w_t, \quad (Q_0 w)_s = Q_0 w_s, \quad (Q_0 w)_{\text{out}} = w_{\text{out}}, \quad Q_0 A_0 = A_0 Q_0.$$

The following lemma will be needed in proving Lemma 3.14.

Lemma 3.13. *There exists a constant $C > 0$ such that*

$$\|A_0 Q_0 w\|_0 \geq C(\|Q_0 w\|_0 + \|(Q_0 w)_t\|_0)$$

for $w \in W^{1,2}(S^1, \mathbb{R}^{2n})$, where $\|\cdot\|_0$ is defined using the inner product $\langle\,\cdot\,,\,\cdot\,\rangle_0$.

Proof. To prove the lemma we only need to prove that $\|A_0 Q_0 w\|_0 \geq C' \|Q_0 w\|_0$ for some $C' > 0$, because by definition we have

$$A_0 Q_0 w = -M_0 (Q_0 w)_t - S_0 Q_0 w. \tag{41}$$

Suppose to the contrary that there exist an $\varepsilon_n \to 0$ and $w_n \in W^{1,2}(S^1, \mathbb{R}^{2n})$ satisfying $\|Q_0 w_n\|_0 = 1$ and $\|A_0 Q_0 w_n\|_0 \leq \varepsilon_n$. Then we have

$$\|(Q_0 w_n)_t\|_0 \leq \|M_0 A_0 Q_0 w_n\|_0 + \|M_0 S_0 Q_0 w_n\|_0 \leq \varepsilon_n + C''.$$

Therefore, $Q_0 w_n$ is bounded in $W^{1,2}(S^1, \mathbb{R}^{2n})$. Since $W^{1,2}(S^1, \mathbb{R}^{2n})$ embeds compactly in $L^2(S^1, \mathbb{R}^{2n})$ we get a subsequence of w_n, still denoted by w_n, such that $Q_0 w_n$ is a Cauchy sequence in $L^2(S^1, \mathbb{R}^{2n})$. But it is easy to see that $(Q_0 w_n)_t$ is also a Cauchy sequence in $L^2(S^1, \mathbb{R}^{2n})$. Therefore, $Q_0 w_n$ converges to some η in $W^{1,2}(S^1, \mathbb{R}^{2n})$, so η is an element of $\ker A_0$. Because η also lies in the orthogonal complement of $\ker A_0$, we must have $\eta = 0$, which contradicts the fact that $\|\eta\|_0 = \lim_{n\to 0} \|Q_0 w_n\|_0 = 1$. □

Define $\kappa_0(s) := (\vartheta(s_0, 0) - \vartheta(s, 0), z_{\text{in}}(s_0, 0) - z_{\text{in}}(s, 0))$, $g_0(s) := \frac{1}{2}\|Q_0 z(s)\|_0^2$.

Lemma 3.14. *There exist $\delta = \delta(\beta) > 0$, $\flat = \flat(\beta) > 0$ and $\bar\kappa = \bar\kappa(\beta) > 0$ such that if, for any multi-indices β,*

$$a(s_0, 0) \geq \flat, \quad |\kappa_0(s_0)| \leq \bar\kappa, \quad \sup_{(s,t)\in\theta} |\partial^\beta z_{\text{out}}(s, t)| \leq \delta,$$

and, for any multi-indices β with $|\beta| > 0$,

$$\sup_{(s,t)\in\theta} |\partial^\beta (a(s, t) - Ts)| \leq \delta, \quad \sup_{(s,t)\in\theta} |\partial^\beta (\vartheta(s, t) - Tt)| \leq \delta, \quad \sup_{(s,t)\in\theta} |\partial^\beta z_{\text{in}}(s, t)| \leq \delta,$$

then we have, for $s \in [s_0, \mathfrak{s}]$,

$$g_0''(s) \geq c^2 g_0(s) - c_2 e^{-c_1(s-s_0)},$$

where

$$\mathfrak{s} := \sup\{s \in [s_0, s_1] : |\kappa_0(s')| \leq \bar\kappa \textit{ for all } s' \in [s_0, s]\},$$

and $c, c_1, c_2 > 0$ are constants independent of s_0 and s_1.

Proof. All constants in the proof may depend on β. Notice that from the assumption we have

$$\sup_{(s,t)\in\theta} |\partial^\beta (\vartheta(s, t) - \vartheta(s, 0) - Tt)| \leq \delta, \quad \sup_{(s,t)\in\theta} |\partial^\beta (z_{\text{in}}(s, t) - z_{\text{in}}(s, 0))| \leq \delta,$$

for all multi-indices β.

Define an operator $\bar A(s) w = -\bar M(\tilde u(s, t)) w_t(t) - \bar S(\tilde u(s, t)) w_{\text{out}}(t)$ in the same way as $A(s)$ but using J_∞ instead of J.

From (38) we get

$$z_s = A_0 z + (\Delta_0 + \tilde\Delta_0 \kappa_0) z_t + (\hat\Delta_0 + \bar\Delta_0 \kappa_0) z_{\text{out}} + [A(s) - \bar A(s)] z - L. \tag{42}$$

Applying Q_0 to (42) gives us

$$\begin{aligned}(Q_0 z)_s = A_0 Q_0 z &+ Q_0(\Delta_0 + \tilde\Delta_0 \kappa_0)(Q_0 z)_t \\ &+ Q_0(\hat\Delta_0 + \bar\Delta_0 \kappa_0)(Q_0 z)_{\text{out}} + Q_0[A(s) - \bar A(s)] z - Q_0 L,\end{aligned} \tag{43}$$

where $\Delta_0 = \bar{M}_0 - \bar{M}$ and $\hat{\Delta}_0 = \bar{S}_0 - \bar{S}$, satisfying for any multi-indices β

$$\sup_{(s,t)\in\theta} |\partial^\beta \Delta_0(s,t)| \le C\delta, \qquad \sup_{(s,t)\in\theta} |\partial^\beta \hat{\Delta}_0(s,t)| \le C\delta,$$

and $\tilde{\Delta}_0\kappa_0 = M_0 - \bar{M}_0$ and $\bar{\Delta}_0\kappa_0 = S_0 - \bar{S}_0$, satisfying for any multi-indices β

$$\sup_{(s,t)\in\theta} |\partial^\beta \tilde{\Delta}_0(s,t)| \le C, \qquad \sup_{(s,t)\in\theta} |\partial^\beta \bar{\Delta}_0(s,t)| \le C.$$

We can require $0 < \delta < T/2$, and then we get

$$a(s,t) \ge a(s_0,0) + \frac{T}{2}(s-s_0) - \delta \ge (\flat - \delta) + \frac{T}{2}(s-s_0).$$

Because J is an asymptotically cylindrical almost complex structure, we get

$$\|Q_0 L\|_0 \le c_0 e^{-c_0'(\flat-\delta)} e^{-c_0' \frac{T}{2}(s-s_0)}$$

for some constants $c_0, c_0' > 0$. Define $c_1 := c_0' T/2$ and $c_2 := c_0 e^{-c_0'(\flat-\delta)}$. Then we have

$$\|Q_0 L\|_0 \le c_2 e^{-c_1(s-s_0)}.$$

We also have

$$\|\{\partial^\beta[A(s) - \bar{A}(s)]\}z\|_0 \le c_2 e^{-c_1(s-s_0)} \|Q_0 z\|_{0,W^{1,2}} \tag{44}$$

for multi-indices β, by picking c_0 larger if necessary.

Now we are ready to estimate $g_0''(s)$. Obviously we have

$$g_0''(s) \ge \langle Q_0 z_{ss}, Q_0 z\rangle_0.$$

Now let's compute the right-hand side of the above inequality. Differentiating (43) with respect to s, we obtain

$$\begin{aligned}(Q_0 z)_{ss} = A_0 Q_0 z_s &+ Q_0(\Delta_0 + \tilde{\Delta}_0\kappa_0)(Q_0 z)_{st} + Q_0(\Delta_0 + \tilde{\Delta}_0\kappa_0)_s (Q_0 z)_t \\ &+ Q_0(\hat{\Delta}_0 + \bar{\Delta}_0\kappa_0)(Q_0 z_s)_{\mathrm{out}} + Q_0(\hat{\Delta}_0 + \bar{\Delta}_0\kappa_0)_s (Q_0 z)_{\mathrm{out}} \\ &+ Q_0[A(s) - \bar{A}(s)]_s z + Q_0[A(s) - \bar{A}(s)] z_s - Q_0 L_s.\end{aligned}$$

Thus we see that $\langle Q_0 z_{ss}, Q_0 z\rangle_0$ contains 8 terms. When we are estimating these terms, each time we see $Q_0 z_s$, we replace it using (43). A straightforward calculation using Lemma 3.13 and the fact that

$$-c_2 e^{-c_1(s-s_0)} \|Q_0 z\|_{0,W^{1,2}} \ge -c_2 e^{-c_1(s-s_0)} - c_2 e^{-c_1(s-s_0)} \|Q_0 z\|^2_{0,W^{1,2}}$$

gives us

$$g_0''(s) \ge (1 - 10C\delta - 10C|\kappa_0| - 10Cc_2 e^{-c_1(s-s_0)}) g_0(s) - c_2 e^{-c_1(s-s_0)}.$$

From the definition of c_2 we can see that if $\flat$ is large enough, c_2 can be very close to 0. Therefore,

$$g_0''(s) \geq c^2 g_0(s) - c_2 e^{-c_1(s-s_0)}.$$

We can require further that $c_1 > c > 0$. □

From Lemma 3.14 we easily obtain the following lemma.

Lemma 3.15. *Under the same assumption as in Lemma 3.14, we have for $s_0 \leq s \leq \mathfrak{s}$,*

$$g_0(s) \leq \max\{g_0(s_0), g_0(\mathfrak{s})\} \frac{\cosh\left[c\left(s - \frac{s_0+\mathfrak{s}}{2}\right)\right]}{\cosh\left(c\frac{\mathfrak{s}-s_0}{2}\right)} + \frac{c_2}{c_1^2 - c^2} \frac{\sinh(c(\mathfrak{s}-s))}{\sinh(c(\mathfrak{s}-s_0))}.$$

Proof. Let

$$\begin{aligned} h(s) := \max\{g_0(s_0), g_0(\mathfrak{s})\} \frac{\cosh\left[c\left(s - \frac{s_0+\mathfrak{s}}{2}\right)\right]}{\cosh\left(c\frac{\mathfrak{s}-s_0}{2}\right)} + \frac{c_2}{c_1^2 - c^2} \frac{1}{\sinh(c(\mathfrak{s}-s_0))} \\ \times\{\sinh(c(\mathfrak{s}-s)) + e^{-c_1(\mathfrak{s}-s_0)} \sinh(c(s-s_0)) - e^{-c_1(s-s_0)} \sinh(c(\mathfrak{s}-s_0))\}. \end{aligned}$$

Then $h(s)$ satisfies

$$\begin{cases} h''(s) - c^2 h(s) = -c_2 e^{-c_1(s-s_0)}, \\ h(s_0) = \max\{g_0(s_0), g_0(\mathfrak{s})\}, \\ h(\mathfrak{s}) = \max\{g_0(s_0), g_0(\mathfrak{s})\}. \end{cases} \tag{45}$$

Let $l(s) := g_0(s) - h(s)$. Then $l(s)$ satisfies

$$\begin{cases} l''(s) - c^2 l(s) \geq 0, \\ l(s_0) \leq 0, \\ l(\mathfrak{s}) \leq 0. \end{cases} \tag{46}$$

Then by the maximal principle we get $l(s) \leq 0$ for $s_0 \leq s \leq \mathfrak{s}$. Now the lemma follows from the fact that

$$e^{-c_1(\mathfrak{s}-s_0)} \sinh(c(s-s_0)) - e^{-c_1(s-s_0)} \sinh(c(\mathfrak{s}-s_0)) \leq 0. \qquad \square$$

Now let's study the component z_{in}.

Lemma 3.16. *Let e be a unit vector in $\mathbb{R}^{2n}$ with $e_{\text{out}} = 0$. Under the assumption of Lemma 3.14 and for $s \in [s_0, \mathfrak{s}]$, we have*

$$|\langle z(s), e\rangle_0 - \langle z(s_0), e\rangle_0| \leq \frac{8C}{c} \max(\|Q_0 z(s_0)\|_0, \|Q_0 z(\mathfrak{s})\|_0) + o(c_2),$$

where $o(c_2)$ satisfies $\lim_{c_2\to 0} o(c_2) = 0$, and C is a constant independent of s_0, s_1.

Proof. The inner product of the Cauchy–Riemann equation (35) with e gives

$$\frac{d}{ds}\langle z, e\rangle_0 + \langle M z_t, e\rangle_0 + \langle S z_{\text{out}}, e\rangle_0 + \langle L, e\rangle_0 = 0.$$

From

$$\begin{aligned}\langle M z_t, e\rangle_0 &= \int_0^1 \omega_0(M(Q_0 z)_t, M_0 e)\,dt \\ &= -\int_0^1 \omega_0(M_t Q_0 z, M_0 e)\,dt - \int_0^1 \omega_0(M Q_0 z, (M_0)_t e)\,dt\end{aligned}$$

we can see that

$$|\langle M z_t, e\rangle_0| \le C\|Q_0 z\|_0.$$

Together with the facts $|\langle S z_{\mathrm{out}}, e\rangle_0| \le C\|Q_0 z\|_0$ and $|\langle L, e\rangle_0| \le c_2 e^{-c_1(s-s_0)}$ we get

$$\begin{aligned}\langle z(s), e\rangle_0 - \langle z(s_0), e\rangle_0 &\le \int_{s_0}^{s} [2C\|Q_0 z(\mathfrak{r})\|_0 + c_2 e^{-c_1(\mathfrak{r}-s_0)}]\,d\mathfrak{r} \\ &\le 2C\int_{s_0}^{s} \sqrt{2g_0(\mathfrak{r})}\,d\mathfrak{r} + \frac{c_2}{c_1}.\end{aligned}$$

The proof is finished with a straightforward calculation using Lemma 3.15 and the fact that $\sqrt{\cosh u} < \sqrt{2}\cosh(u/2)$. □

Remark 3.17. By requiring $\flat$ to be sufficiently large, we can make c_2 sufficiently small.

Now let's estimate the derivatives of z.

Lemma 3.18. *There exist $\delta = \delta(\beta) > 0$, $\flat = \flat(\beta) > 0$ and $\bar\kappa = \bar\kappa(\beta) > 0$ such that if, for any multi-indices β,*

$$\sup_{(s,t)\in\theta} |\partial^\beta z_{\mathrm{out}}(s,t)| \le \delta, \quad a(s_0, 0) \ge \flat,$$

and, for any multi-indices β with $|\beta| > 0$,

$$\sup_{(s,t)\in\theta} |\partial^\beta(a(s,t) - Ts)| \le \delta, \quad \sup_{(s,t)\in\theta} |\partial^\beta(\vartheta(s,t) - t)| \le \delta, \quad \sup_{(s,t)\in\theta} |\partial^\beta z_{\mathrm{in}}(s,t)| \le \delta,$$

then we have, for $s \in [s_0, \mathfrak{s}]$,

$$\begin{aligned}\|\partial^\beta z(s)\|_0 &\le C_\beta \max_{|\beta'|\le|\beta|}\{\|Q_0\partial^{\beta'} z(s_0)\|_0, \|Q_0\partial^{\beta'} z(\mathfrak{s})\|_0\}\sqrt{\frac{\cosh\bigl(c_1\bigl(s - \frac{s_0+\mathfrak{s}}{2}\bigr)\bigr)}{\cosh\bigl(c_1\bigl(\frac{s_0-\mathfrak{s}}{2}\bigr)\bigr)}} \\ &\quad + D_\beta(c_2)\sqrt{\frac{\sinh(c(\mathfrak{s}-s))}{\sinh(c(\mathfrak{s}-s_0))}} + c_2 e^{-c_1(s-s_0)},\end{aligned}$$

where

$$\mathfrak{s} := \sup\{s \in [s_0, s_1] : |\kappa_0(s')| \le \bar\kappa \textit{ for all } s' \in [s_0, s]\},$$

and $C_\beta, c_1 > 0$ are constants independent of s_0 and s_1, and $D_\beta(c_2)$ is a function of c_2 independent of s_0 and s_1, satisfying $\lim_{c_2\to 0} C^\beta(c_2) = 0$, *and l is the integer in Definition 2.1.*

Proof. Let's prove the estimate for $|\beta| = 1$. The proof of the estimates of the higher derivatives is almost the same. Refer to Lemma A.6 in [Bourgeois et al. 2003] for the estimates for all derivatives in the cylindrical case.

Equation (42) can be rewritten as

$$z_s = A_0 z + \dot{\Delta} z_t + \ddot{\Delta} z_{\text{out}} + \dddot{\Delta} z - L, \tag{47}$$

with $\dot{\Delta} = \Delta_0 + \tilde{\Delta}_0\kappa_0$, $\ddot{\Delta} = \hat{\Delta}_0 + \bar{\Delta}_0\kappa_0$, and $\dddot{\Delta} = [A(s) - \bar{A}(s)]$. If we define $\mathcal{W} := (Q_0 z, \partial/\partial s(Q_0 z), A_0 Q_0 z, \partial/\partial s(A_0 Q_0 z))$, then $\mathcal{W}$ satisfies

$$\mathcal{W}_s = \mathcal{A}_0\mathcal{W} + \mathcal{Q}_0 \dot{\mathbf{\Delta}} \mathcal{W}_t + \mathcal{Q}_0 \ddot{\mathbf{\Delta}} \mathcal{W}_{\text{out}} + \dddot{\mathbf{\Delta}} \mathcal{W} - \boldsymbol{\mathcal{L}},$$

where $\mathcal{A}_0 = \operatorname{diag}(A_0, A_0, A_0, A_0)$, $\mathcal{Q}_0 = \operatorname{diag}(Q_0, Q_0, Q_0, Q_0)$, and $\dot{\mathbf{\Delta}}, \ddot{\mathbf{\Delta}}, \dddot{\mathbf{\Delta}}, \boldsymbol{\mathcal{L}}$ satisfy similar estimates as $\dot{\Delta}, \ddot{\Delta}, \dddot{\Delta}, L$ respectively. Indeed, for $|\beta| = 1$ we can derive this equation by direct computation. For general β, we can derive it by induction on $|\beta|$. This equation is of the same type as (47). Copying the proofs of Lemmata 3.14, 3.15 and 3.16, we can get the desired estimate for $\mathcal{W}$. In particular, we get the estimates for $(Q_0 z)_s$ and $A_0 Q_0 z$.

From the equation $z_t = M_0 A_0 Q_0 z + M_0 Q_0 S_0 z_{\text{out}}$ we get the estimate for z_t. Applying P_0 to (47), we get

$$(P_0 z)_s = P_0 \dot{\Delta} z_t + P_0 \ddot{\Delta} z_{\text{out}} + P_0 \dddot{\Delta} z - P_0 L.$$

This equation together with the estimate of $\dddot{\Delta} z$ (see (44)) gives us the desired estimate for $P_0 z_s$. Then the estimate for z_s follows from $z_s = P_0 z_s + Q_0 z_s$. □

Lemma 3.19. *Define*

$$\vartheta_0 = \int_0^1 \left[\vartheta\left(\tfrac{s_0+\mathfrak{s}}{2}, t\right) - Tt\right] dt, \quad a_0 = \int_0^1 \left[a\left(\tfrac{s_0+\mathfrak{s}}{2}, t\right) - Ts_0\right] dt,$$

and define $\tilde{a} = a(s,t) - Ts - a_0$ and $\tilde{\vartheta} = \vartheta(s,t) - Tt - \vartheta_0$. Under the assumptions of Lemma 3.18, we have, for $s \in [s_0, \mathfrak{s}]$ and every multi-index β,

$$\begin{aligned} &\|\partial^\beta(\tilde{a}(s,t)\|^2, \|\partial^\beta(\tilde{\vartheta}(s,t))\|^2 \\ &\qquad \le C_1 \max_{|\beta'| \le |\beta|+3} \{\|Q_0 \partial^{\beta'} z(s_0)\|_0^2, \|Q_0 \partial^{\beta'} z(\mathfrak{s})\|_0^2\} \\ &\qquad\qquad + C_1 \max\{\|\tilde{a}(s_0,\cdot\,)\|^2 + \|\tilde{\vartheta}(s_0,\cdot\,)\|^2, \|\tilde{a}(\mathfrak{s},\cdot\,)\|^2 + \|\tilde{\vartheta}(\mathfrak{s},\cdot\,)\|^2\} + o(c_2), \end{aligned}$$

where $\|\cdot\|$ is the L^2-norm, $o(c_2)$ satisfies $\lim_{c_2\to 0} o(c_2) = 0$, *and C_1 is a constant independent of $\tilde{u}$.*

Proof. We can modify the proofs of Lemmata 3.8–3.13 in [Hofer et al. 2002] in the obvious way, similar to what we did in the proof of Lemma 3.14, and then use Lemma 3.18 to prove this lemma. We omit the proof here, since it is essentially not new.[4] □

Remark 3.20. When $\mathfrak{s}$ is infinity, we can get a better exponential decay estimate using the same proof, and in that case the term $o(c_2)$ can be replaced by $c_2 e^{-(s-s_0)}$.

Proof of Theorem 2.8. Let's follow the proof in [Bourgeois 2002]. By Theorem 3.7, we can find a sequence $s_{0m} \to \infty$ such that

$$\lim_{m\to\infty} u(s_{0m}, t) = \gamma(Tt), \qquad \lim_{m\to\infty} a(s_{0m}, t) = \pm\infty$$

for some T-periodic orbit γ of $\boldsymbol{R}_\infty$. From the proof of Theorem 3.7, we can further require for any multi-indices α with $|\alpha| > 0$ we have $\sup_{t\in S^1} \|\partial^\alpha z(s_{0m}, t)\| \to 0$ as $m \to +\infty$.

Given $\sigma > 0$, let $\zeta_m > 0$ be the largest number such that $u(s, t) \in S^1 \times [-\sigma, \sigma]^{2n}$ for all $s \in [s_{0m}, s_{0m} + \zeta_m]$. Let $\theta_m := [s_{0m}, s_{0m} + \zeta_m] \times S^1$ and let $\kappa_{0m}(s) := (\vartheta(s_{0m}, 0) - \vartheta(s, 0), z_{\text{in}}(s_{0m}, 0) - z_{\text{in}}(s, 0))$. Now we can define the operator A_{0m}, similar to how it was defined before, in the obvious way.

By Corollary 3.8, given $\delta > 0$ we have

$$\sup_{(s,t)\in\theta_m} |\partial^\beta (a(s, t) - Ts)| \le \delta$$

for those multi-indices β with $|\beta| > 0$, when m is large. This implies that $a(s_{0m}, 0) \to +\infty$ as $m \to +\infty$. Notice that the other requirements in Lemmata 3.14 and 3.18 are also satisfied; i.e., given $\delta > 0$, there exists an m_δ such that for $m > m_\delta$ we have

$$\sup_{(s,t)\in\theta_m} |\partial^\beta z_{\text{out}}(s, t)| \le \delta$$

for multi-indices β, and

$$\sup_{(s,t)\in\theta_m} |\partial^\beta (\vartheta(s, t) - Tt)| \le \delta, \tag{48}$$

$$\sup_{(s,t)\in\theta_m} |\partial^\beta z_{\text{in}}(s, t)| \le \delta$$

for those multi-indices β with $|\beta| > 0$. Indeed, if $\{(s_{m_k}, t_{m_k})\}$ violates one of these properties, we can define

$$\tilde{u}_{m_k}(s, t) := (a(s - s_{m_k}, t - t_{m_k}) - a(s_{m_k}, t_{m_k}), u(s - s_{m_k}, t - t_{m_k})).$$

[4] The proof of Proposition 3.4 in [Bourgeois 2002] is inaccurate, and this lemma fills in the gap.

By Ascoli–Arzela, we can extract a subsequence, still called $\tilde{u}_{m_k}(s,t)$, such that $\tilde{u}_{m_k}(s,t)$ converges in C^∞_{loc} to a J_∞-holomorphic cylinder $\tilde{u}_\infty$ over a periodic orbit $\gamma' \in N$. Since $\tilde{u}_\infty$ must satisfy those three properties, we get a contradiction.

By construction, $|\langle z(s_{0m}), e\rangle_{0m}| \to 0$ and $\|Q_{0m}\partial^\alpha z(s_{0m})\| \to 0$, for all multi-indices α with $|\alpha| \geq 0$. Let $\bar{\kappa}_m$ be the "$\bar{\kappa}$" in Lemmata 3.14 and 3.18 applied to $\tilde{u}|_{\theta_m}$ and let $\mathfrak{s}_m := \sup\{s \in [s_{0m}, s_{0m}+\zeta_m] : |\kappa_{0m}(s')| \leq \bar{\kappa}_m \text{ for all } s' \in [s_0, s]\}$, and notice that $\bar{\kappa}_m$ can actually be chosen independent of m. We can extract a subsequence so that $u(\mathfrak{s}_m, t)$ converges to a closed Reeb orbit $\gamma'' \in N$. Therefore, $\|Q_{0m}\partial^\alpha z(\mathfrak{s}_m)\| \to 0$, for all multi-indices α with $|\alpha| \geq 0$. Since $\langle z(\mathfrak{s}_m), e\rangle_0 \to 0$ and $\sup_{t\in S^1} |(\partial/\partial t)z_{\text{in}}(\mathfrak{s}_m, t)| \to 0$, we obtain $\sup_{t\in S^1} |z_{\text{in}}(\mathfrak{s}_m, t)| \to 0$. By Lemmata 3.14 and 3.18, we have

$$\sup_{s\in[s_{0m},\mathfrak{s}_m]} \|\partial^\beta z(s)\|_{0m} \to 0 \tag{49}$$

for $|\beta| \leq k$. Therefore,

$$\begin{aligned}
\sup_{(s,t)\in[s_{0m},\mathfrak{s}_m]\times S^1} |z_{\text{in}}(s,t)| &\leq \sup_{s\in[s_{0m},\mathfrak{s}_m]} \|z_{\text{in}}(s,\cdot)\|_{C^0(S^1)} \\
&\leq C \sup_{s\in[s_{0m},\mathfrak{s}_m]} \|z_{\text{in}}(s,\cdot)\|_{W^{1,2}(S^1)} \\
&\leq C_1\Big\{\sup_{s\in[s_{0m},\mathfrak{s}_m]} \|(\partial/\partial t)z_{\text{in}}(s,\cdot)\|_{0l} + \sup_{s\in[s_{0m},\mathfrak{s}_m]} \|z_{\text{in}}(s,\cdot)\|_{0m}\Big\} \\
&\to 0.
\end{aligned}$$

Lemma 3.19 and formula (48) imply that $|\vartheta(\mathfrak{s}_m, 0) - \vartheta(s_{0m}, 0)| \to 0$ as $m \to \infty$. Thus, we have $\mathfrak{s}_m = s_{0m} + \zeta_m$ for m large enough, and

$$\sup_{(s,t)\in[s_{0m},s_{0m}+\zeta_m]\times S^1} |z(s,t)| \to 0$$

as $m \to \infty$. Therefore, $\zeta_m = +\infty$ for m large. □

Furthermore, we can show that the convergence of a J-holomorphic curve is exponentially fast.

Proof of Theorem 2.9. Now with the help of the previous lemmata, the proof of the third inequality is almost evident. Indeed, since $\mathfrak{s} = +\infty$, Lemma 3.15 becomes

$$g_0(s) \leq \Big(g_0(s_0) + \tfrac{c_2}{c_1^2 - c^2}\Big)e^{-c(s-s_0)}.$$

Consequently, in the proof Lemma 3.16, we can get

$$|\langle z(s), e\rangle_0| \leq \int_s^{+\infty} [2C\|Q_0 z(\mathfrak{r})\|_0 + c_2 e^{-c_1(\mathfrak{r}-s_0)}]\, d\mathfrak{r} \leq C' e^{-c(s-s_0)},$$

where C' is independent of s. Similarly, we can get the corresponding statement of Lemma 3.18 for $\mathfrak{s} = +\infty$.

The proof for the rest is a straightforward modification of the original proof in [Hofer et al. 2001]. □

So far we have studied the behaviors of a finite energy J-holomorphic curve whose domain is an infinite cylinder. In order to compactify the moduli space of holomorphic curves, we also need to understand the behavior of a finite energy J-holomorphic curve whose domain is a long but finite interval and whose ω-energy is small. To do that, we need the following lemma.

Lemma 3.21 (bubbling lemma [Bourgeois et al. 2003; Hofer and Viterbo 1992]). *Let J^0 be a cylindrical almost complex structure on $W = \mathbb{R}^+ \times V$. There exists a constant $\hbar > 0$ depending only on (W, J^0, ω^0) where $J^0 = J^0_\infty$ and $\omega^0 = \omega^0_\infty$ (see Definitions 2.1 and 2.3 and Section 2B), such that the following holds true. Let (J^n, ω^n_∞) be a sequence of pairs satisfying (AC1)–(AC5) on W and converging to (J^0, ω^0) in the C^∞_{loc} sense. Consider a sequence of J^n-holomorphic maps $\tilde{u}_n = (a_n, u_n)$ from the unit disc $B(0, 1)$ to W satisfying $E_n(\tilde{u}_n) = E_{\omega^n}(\tilde{u}_n) + E_{\lambda_n}(\tilde{u}_n) \le C$ (see Section 2B) for some constant C, such that the sequence $a_n(0)$ is bounded, and such that $\|\nabla \tilde{u}_n(0)\| \to +\infty$ as $n \to +\infty$. Then there exists a sequence of points $z_n \in B(0, 1)$ converging to 0, and sequences of positive numbers ε_n and R_n satisfying*

$$\varepsilon_n \to 0, \quad R_n \to +\infty, \quad \varepsilon_n R_n \to +\infty, \quad |z_n| + \varepsilon_n < 1,$$

such that the rescaled maps

$$\tilde{u}^0_n : B(0, \varepsilon_n R_n) \to W, \quad z \mapsto \tilde{u}_n(z_n + R_n^{-1} z)$$

converge in C^1_{loc} to a J_0-holomorphic map $\tilde{u}^0 : \mathbb{C} \to W$ which satisfies $E(\tilde{u}^0) \le C$ and $E_{\omega^0}(\tilde{u}^0) > \hbar$.

Moreover, this map is either a J_0-holomorphic plane asymptotic as $|z| \to \infty$ to a periodic orbit of the vector field $\boldsymbol{R}^0$ defined by $\boldsymbol{R}^0 = J_0(\partial/\partial r)$, or extendable to a J_0-holomorphic sphere $\mathbb{P}^1 \to W$ by Gromov's removal of singularity theorem.

A similar statement is also true for $\mathbb{R}^- \times V$.

Proof. See [Hofer and Viterbo 1992]. □

The following theorem studies the behavior of a long cylinder having small ω-area. It is needed in order to prove the compactness results for the moduli space of J-holomorphic curves in symplectic field theory. Refer to [Hofer et al. 2002; Bourgeois et al. 2003] for the cylindrical case.

Theorem 3.22. *Suppose that J is an asymptotically cylindrical almost complex structure on $W = \mathbb{R}^\pm \times V$ at $\pm\infty$, and suppose that J is of the Morse–Bott type.*

Given $E_0 > 0$ and $\varepsilon > 0$, there exist constants $\sigma, c > 0$ such that for every $R > c$ and every J-holomorphic cylinder $\tilde{u} = (a, u) : [-R, R] \times S^1 \to W$ satisfying the inequalities $E_\omega(\tilde{u}) < \sigma$ and $E(\tilde{u}) < E_0$, we have $u(s, t) \in B_\varepsilon(u(0, t))$ for all $s \in [-R + c, R - c]$ and all $t \in S^1$.

Proof. The proof follows the scheme in [Bourgeois et al. 2003] with some modification.

By contradiction, assume that there exist sequences $c_n \to +\infty$, $R_n > c_n$ and $\tilde{u}_n = (a_n, u_n) : [-R_n, R_n] \times S^1 \to W$. The sequence $\tilde{u}_n$ is J-holomorphic, satisfying $E(\tilde{u}_n) \le E_0$, $E_\omega(\tilde{u}_n) \to 0$, and $u_n(s_n, t_n) \notin B(u_n(0, t_n), \varepsilon)$ for some $s_n \in [-k_n, k_n]$, $k_n = R_n - c_n$ and $t_n \in S^1$. By the proof of Proposition 3.4 together with the bubbling lemma (Lemma 3.21), $\|\nabla \tilde{u}_n\|$ is uniformly bounded on each compact subset. We can extract a subsequence of n, still denoted by n, such that $a_n(s_n, t_n) \to \pm\infty$. This is because, otherwise, we can get a contradiction as in the proof of Proposition 3.4. Now define $\tilde{u}^0_n(s, t) := (a^0_n, u^0_n) = (a_n(s, t) - a_n(s_n, t_n), u_n(s, t))$. By Ascoli–Arzela, we can extract a subsequence, still called $\tilde{u}^0_n$, converging to a J_∞-holomorphic cylinder $\tilde{u} : \mathbb{R} \times S^1 \to \mathbb{R} \times V$. Since $\tilde{u}$ satisfies $E_\omega(\tilde{u}) = 0$ and $E(\tilde{u}) \le E_0$, we have that $\tilde{u}$ is a trivial cylinder over some periodic orbit γ. Let's choose a neighborhood around γ, and pick the coordinate as in Lemma 3.9, and show that

$$\sup_{(s,t)\in[-k_n,k_n]\times S^1} |\partial^\beta z_{\mathrm{out},n}(s, t)| \to 0 \tag{50}$$

for multi-indices β and

$$\sup_{(s,t)\in[-k_n,k_n]\times S^1} |\partial^\beta (a_n(s, t) - Ts)| \to 0, \tag{51}$$

$$\sup_{(s,t)\in[-k_n,k_n]\times S^1} |\partial^\beta z_{\mathrm{in},n}(s, t)| \to 0, \tag{52}$$

$$\sup_{(s,t)\in[-k_n,k_n]\times S^1} |\partial^\beta (\vartheta_n(s, t) - Tt)| \to 0 \tag{53}$$

for multi-indices β with $|\beta| > 0$, when $n \to +\infty$.

If this were not true, suppose there exists a subsequence of $\{n\}$, still denoted by $\{n\}$, such that (s'_n, t'_n) violates one of these properties. Then we can make the same argument using (s'_n, t'_n) instead of (s_n, t_n) as above and get a trivial cylinder contradicting the fact that (s'_n, t'_n) violates one of these properties.

Define A_{0n} and Q_{0n} in the obvious way using γ and $s_{0n} = 0$. Then apply Lemmata 3.14–3.16 and 3.18 to each $\tilde{u}_n|_{[-k_n,k_n]}$ to get $\sup_{s\in[-k_n,k_n]} \|Q_{0n} z_n(s)\|_{0,n} \to 0$. Then the Sobolev embedding theorem tells us that $\kappa_{0n} \to 0$ as $n \to +\infty$. This contradicts the assumption that $u_n(s_n, t_n) \notin B(u_n(0, t), \varepsilon)$. □

We need the following theorem later to prove the surjectivity of the gluing map in the subsequent paper. After proving all the previous lemmata and theorems, the proof of the following theorem is standard. For the case when J is cylindrical and nondegenerate and V is a contact manifold, the proof is given in [Hofer et al. 2002].

Theorem 3.23. *Suppose that J is an asymptotically cylindrical almost complex structure on $W = \mathbb{R}^+ \times V$ at ∞, and suppose that J is of the Morse–Bott type. Given $E_0 > 0$ and sufficiently small $\varepsilon > 0$, there exist constants $\sigma, c, \flat, \nu > 0$ such that, for every $R > c$ and every J-holomorphic cylinder $\tilde{u} = (a, u) : [-R, R] \times S^1 \to (\flat, \infty) \times V$ satisfying the inequalities $E_\omega(\tilde{u}) < \sigma$ and $E(\tilde{u}) < E_0$, there exists either a point $w \in W$ such that $\tilde{u}(s, t) \in B_\varepsilon(w)$ for $s \in [-R + c, R - c]$ and $t \in S^1$, or a T-periodic orbit γ of $\boldsymbol{R}_\infty$ such that $u(s, t) \in B_\varepsilon(\gamma(Tt))$ for $s \in [-R + c, R - c]$ and $t \in S^1$. In the second case, we have a coordinate around γ as in Lemma 3.9 such that*

$$|D^\beta\{a(s,t) - Ts - a_0\}|^2 \le \varepsilon^2 M_\beta \frac{\cosh(2\nu s)}{\cosh(2\nu(R-c))} + C_\beta e^{-c_\beta(s+R-c)},$$

$$|D^\beta\{\vartheta(s,t) - Tt - \vartheta_0\}|^2 \le \varepsilon^2 M_\beta \frac{\cosh(2\nu s)}{\cosh(2\nu(R-c))} + C_\beta e^{-c_\beta(s+R-c)},$$

$$|D^\beta z(s,t)|^2 \le \varepsilon^2 M_\beta \frac{\cosh(2\nu s)}{\cosh(2\nu(R-c))} + C_\beta e^{-c_\beta(s+R-c)},$$

for $s \in [-R + c, R - c]$, $t \in S^1$, and $\beta \in \mathbb{N} \times \mathbb{N}$ such that $|\beta| \le l - 3$, where M_β, C_β, c_β are constants independent of $\tilde{u}$ and ε, and C_β converges to 0 as $\flat$ converges to $+\infty$, and M_β and c_β are independent of $\flat$.

A similar statement is also true for $\mathbb{R}^- \times V$.

4. Almost complex manifolds with asymptotically cylindrical ends

In this section, we introduce the notion of almost complex manifolds with asymptotically cylindrical ends.

4A. ***Definitions.*** Let (W_0, ω') be a closed symplectic manifold with boundary $\partial W_0 = V_+ \sqcup V_-$, where $V_\pm$ is an oriented closed manifold. Let W be the noncompact smooth manifold obtained by attaching $E_\pm := \mathbb{R}^\pm \times V_\pm$ to W_0 along $\{0\} \times V_\pm$ and $V_\pm$. Suppose that there exists an almost complex structure J on W such that $J|_{W_0}$ is compatible with ω' and $(E_\pm, J|_{E_\pm})$ is asymptotically cylindrical at $\pm\infty$. We assume that the orientation of $E_\pm$ determined by $J|_{E_\pm}$ coincides with the orientation coming from the standard orientation of $\mathbb{R}^\pm$ and the orientation of $V_\pm$. This assumption distinguishes V_+ from V_-. Furthermore, we assume $\omega'|_{V_\pm} = \omega_{\pm\infty}$, where $\omega_{\pm\infty}$ is the 2-form on $V_\pm$ from Definition 2.1. In this case, we say (W, J) is an almost complex manifold with asymptotically cylindrical ends.

Example 4.1 [Bourgeois et al. 2003]. Let (X, ω', J) be an almost Kähler manifold, and let $Y \subset X$ be an embedded closed almost Kähler submanifold. We claim that $(X\backslash Y, J|_{X\backslash Y})$ has an asymptotically cylindrical negative end. Let N be the normal bundle of Y in X with the metric $\omega'(\cdot, J\cdot)|_Y$, let V be the associated unit sphere bundle of N defined by $V = \{u \in N : |u| = 1\}$, and let U_ε be the disc bundle of N defined by $U_\varepsilon = \{u \in N : |u| \le \varepsilon\}$. For small enough $\varepsilon > 0$, we have that U_ε is diffeomorphic to a tubular neighborhood of Y in X via the exponential map with respect to the metric $\omega'(\cdot, J\cdot)$. Since $U_\varepsilon\backslash Y$ is also diffeomorphic to $(-\infty, \log\varepsilon] \times V$ via the map $u \mapsto (\log|u|, u/|u|)$, one can check that this makes $(X\backslash Y, J|_{X\backslash Y})$ an almost complex manifold with an asymptotically cylindrical negative end.

In particular, if we pick Y to be a point in X, we get Example 2.5 as a special case.

4B. ***Energy of J-holomorphic curves.*** Let w be a J-holomorphic map from a punctured Riemann surface (Σ, j) to (W, J), and define

$$\begin{aligned}
E_{\text{symp}}(w) &= \int_{w^{-1}(W_0)} w^*\omega', \\
E_\omega(w) &= \int_{w^{-1}(E_+)} w^*\omega + \int_{w^{-1}(E_-)} w^*\omega, \\
E_\lambda(w) &= \sup_{\phi\in\mathcal{C}_+} \int_{w^{-1}(E_+)} w^*(\phi\,\sigma\wedge\lambda) + \sup_{\phi\in\mathcal{C}_-} \int_{w^{-1}(E_-)} w^*(\phi\,\sigma\wedge\lambda),
\end{aligned}$$

where

$$\mathcal{C}_+ = \Big\{\phi \in C_c^\infty(\mathbb{R}^+, [0,1]) : \int \phi = 1\Big\}, \quad \mathcal{C}_- = \Big\{\phi \in C_c^\infty(\mathbb{R}^-, [0,1]) : \int \phi = 1\Big\},$$

and

$$E(w) = E_{\text{symp}}(w) + E_\omega(w) + E_\lambda(w).$$

Theorem 4.2. *Suppose (W, J) is an almost complex manifold with asymptotically cylindrical ends, and suppose that J is of the Morse–Bott type. Let w be a J-holomorphic curve from a punctured Riemann surface Σ to W with $E(w) < \infty$. Then around each puncture, either w can be extended holomorphically over the puncture, or one can choose a holomorphic coordinate chart $\mathbb{R}^+ \times S^1$ or $\mathbb{R}^- \times S^1$ in S around the puncture such that w converges to a Reeb orbit in E_+ or E_- in the sense of Theorems 2.8 and 2.9.*

Proof. If w is bounded around a puncture, then Gromov's removal of singularity theorem implies that w can be extended holomorphically over the puncture.

Suppose that w is not bounded around a puncture. We pick a holomorphic cylindrical coordinate $\mathbb{R}^+ \times S^1$ around the puncture of Σ. By Proposition 3.5,

$|\nabla w| < C$ with respect to the standard metric on $\mathbb{R}^+ \times S^1$. If w keeps coming back to a compact region of W and also escaping to the positive (or negative) end of W, we can find an r_0 such that w touches $\{r_0\} \times V_\pm$ and $\{r_0 \pm 3C\} \times V_\pm$ infinitely many times. Then we can apply Gromov's monotonicity theorem to w in the region $[r_0 \pm C, r_0 \pm 2C] \times V_\pm$ as in the argument of Case 1 in the proof of Theorem 3.7, and get $E(w) = \infty$, which contradicts the assumption. Therefore, near the puncture, w converges to ∞ or $-\infty$ in E_+ or E_-. Then Theorem 4.2 follows from Theorems 2.8 and 2.9. □

Proposition 4.3. *Suppose (W, J) is an almost complex manifold with asymptotically cylindrical ends, and suppose that J is of the Morse–Bott type. Then there exists a constant $\varepsilon_0 > 0$ such that if $K_0^\pm < \varepsilon_0$, where $K_0^\pm$ is the constant in (AC1), the following holds.*

Let w be a J-holomorphic curve from a punctured Riemann surface Σ to W such that, around punctures of Σ, we have that w converges to the periodic orbits $\gamma_1^+, \dots, \gamma_p^+$ inside V_+ and $\gamma_1^-, \dots, \gamma_q^-$ inside V_-. Then

$$E(w) \le C_1 \sum_{i=1}^{p} \int_{\gamma_i^+} \lambda_\infty - C_2 \sum_{j=1}^{q} \int_{\gamma_j^-} \lambda_{-\infty} + C_3 \Big\{ \int_{w^{-1}(E_+)} w^*\omega_\infty + \int_{w^{-1}(W_0)} w^*\omega' + \int_{w^{-1}(E_-)} w^*\omega_{-\infty} \Big\},$$

where C_1, C_2, C_3 are positive constants that are independent of w. In particular, $E(w)$ only depends on the homology class of w in $H_2\big(W, \big(\bigcup_{i=1}^p \gamma_i^+\big) \cup \big(\bigcup_{j=1}^q \gamma_j^-\big)\big)$.

Proposition 4.3 is the asymptotically cylindrical version of Proposition 6.13 in [Bourgeois et al. 2003]. The extra work to prove it for the asymptotically cylindrical case is essentially carried out in the Appendix of [Bao 2014] where we assume $\omega_{\pm\infty} = d\lambda_\infty$. For the sake of completeness, we reproduce the proof here.

Proof. First, we restrict ourselves to E_+ and denote $w_\pm := w|_{w^{-1}(E_\pm)}$. Note that, when restricted to J-complex planes, we have

$$|\omega - \omega_\infty| \le \varepsilon e^{-\delta s}(\omega + \sigma \wedge \lambda), \tag{54}$$

$$|d\lambda_\infty| \le C\omega + \varepsilon e^{-\delta s} \sigma \wedge \lambda, \tag{55}$$

$$|\sigma \wedge \lambda - dr \wedge \lambda_\infty| \le \varepsilon e^{-\delta s}(\sigma \wedge \lambda + \omega), \tag{56}$$

where C is a positive constant and the constant $\varepsilon > 0$ can be chosen to be small if K_0^+ is small. Since $\int_0^\infty \delta e^{-\delta s}\, ds = 1$, we get

$$\int_{w^{-1}(E_+)} w^*\omega \le \int_{w^{-1}(E_+)} w^*\omega_\infty + \varepsilon \int_{w^{-1}(E_+)} w^*\omega + \frac{\varepsilon}{\delta} E_\lambda(w_+),$$

where

$$E_\lambda(w_\pm) := \sup_{\phi\in\mathcal{C}_\pm} \int_{w^{-1}(E_\pm)} w^*(\phi\,\sigma\wedge\lambda).$$

Absorbing the second term on the right-hand side to the left-hand side, we get

$$E_\omega(w_+) \le C_1 \int_{w^{-1}(E_+)} w^*\omega_\infty + C_2\varepsilon E_\lambda(w_+), \tag{57}$$

for some constants C_1, C_2, where $E_\omega(w_\pm) := \int_{w^{-1}(E_\pm)} w^*\omega$.

For any $\phi \in \mathcal{C}_+$, let $\Phi(s) = \int_0^s \phi(l)\,dl$. Then using (55) and (56) we have

$$\begin{aligned}
&\int_{w^{-1}(E_+)} w^*\phi\,\sigma\wedge\lambda \\
&\quad= \int_{w^{-1}(E_+)} w^*\phi\,dr\wedge\lambda_\infty + \int_{w^{-1}(E_+)} w^*\phi(\sigma\wedge\lambda - dr\wedge d\lambda_\infty) \\
&\quad\le \int_{w^{-1}(E_+)} w^*\,d(\Phi\lambda_\infty) - \int_{w^{-1}(E_+)} w^*\Phi\,d\lambda_\infty + \int_{w^{-1}(E_+)} w^*\varepsilon e^{-\delta s}\phi(\sigma\wedge\lambda+\omega) \\
&\quad\le \sum_{i=1}^{p}\int_{\gamma_i^+}\lambda_\infty + \int_{w^{-1}(E_+)} w^*(C\omega + \varepsilon e^{-\delta s}\sigma\wedge\lambda) + \int_{w^{-1}(E_+)} w^*\varepsilon e^{-\delta s}\phi(\sigma\wedge\lambda+\omega) \\
&\quad\le \sum_{i=1}^{p}\int_{\gamma_i^+}\lambda_\infty + CE_\omega(w_+) + \varepsilon E_\lambda(w_+),
\end{aligned}$$

where in the last inequality we get the constants C and ε by slightly abusing the notations, but we can still have ε small. Taking the sup over ϕ, we get

$$E_\lambda(w_+) \le \sum_{i=1}^{p}\int_{\gamma_i^+}\lambda_\infty + CE_\omega(w_+) + \varepsilon E_\lambda(w_+). \tag{58}$$

Therefore, by (57) and (58) we have

$$E_\omega(w_+) + E_\lambda(w_+) \le C_1\int_{\gamma_+}\lambda_\infty + C_2\int_{w^{-1}(E_+)} w^*\omega_\infty, \tag{59}$$

where constants C_1 and C_2 are not necessarily the same as before.

For E_-, by the proof of Theorem 10 in [Bao 2014], if K_0^- is small we have

$$E_\omega(w_-)+E_\lambda(w_-)\le C_1' E_{\text{symp}}(w)+C_2'\int_{w^{-1}(E_-)} w^*\omega_\infty - C_3'\sum_{j=1}^{q}\int_{\gamma_j^-}\lambda_{-\infty}, \tag{60}$$

where C_1', C_2', C_3' are positive constants independent of w. Here we recall that

$$E_{\text{symp}}(w)=\int_{w^{-1}(W_0)} w^*\omega'. \tag{61}$$

Now by (59) and (60) we have

$$\begin{aligned}E(w)&=E_\omega(w_+)+E_\lambda(w_+)+E_\omega(w_-)+E_\lambda(w_+)+E_{\text{symp}}(w)\\ &\le a_1(E_\omega(w_+)+E_\lambda(w_+))+a_2(E_\omega(w_-)+E_\lambda(w_+))+a_3E_{\text{symp}}(w_0)\\ &\le C_1\int_{\gamma_+}\lambda_\infty - C_2\int_{\gamma_-}\lambda_{-\infty}+C_3\biggl\{\int_{w^{-1}(E_+)} w^*\omega_\infty+\int_{w^{-1}(W_0)} w^*\omega'+\int_{w^{-1}(E_-)} w^*\omega_\infty\biggr\},\end{aligned}$$

where $a_1, a_2, a_3 \ge 1$ are positive constants chosen in a way such that the last inequality holds for some positive constants C_1, C_2 and C_3. $\square$

Let $\mathcal{M}^A_{g,p+q}(\gamma_1^+,\dots,\gamma_p^+,\gamma_1^-,\dots,\gamma_q^-;J)$ be the moduli space of J-holomorphic curves of genus g in W that converge to periodic orbits $\gamma_1^+,\dots,\gamma_p^+$ inside V_+ and $\gamma_1^-,\dots,\gamma_q^-$ inside V_- and represent the homology class A, which is an element of $H_2\bigl(W,\bigl(\bigcup_{i=1}^p\gamma_i^+\bigr)\cup\bigl(\bigcup_{j=1}^q\gamma_j^-\bigr)\bigr)$. Let $\bar{\mathcal{M}}^A_{g,p+q}(\gamma_1^+,\dots,\gamma_p^+,\gamma_1^-,\dots,\gamma_q^-;J)$ be the compactification of the space $\mathcal{M}^A_{g,p+q}(\gamma_1^+,\dots,\gamma_p^+,\gamma_1^-,\dots,\gamma_q^-;J)$ by allowing stable holomorphic buildings. See Theorems 8.1 and 8.2 in [Bourgeois et al. 2003] for the definition of stable holomorphic buildings in manifolds with cylindrical ends and the topology of the moduli space of holomorphic buildings. Finally, let us state the compactness results.

Theorem 4.4. *Suppose (W, J) is an almost complex manifold with asymptotically cylindrical ends, and suppose that J is of the Morse–Bott type. Then $\bar{\mathcal{M}}^A_{g,p+q}(\gamma_1^+,\dots,\gamma_p^+,\gamma_1^-,\dots,\gamma_q^-;J)$ is compact.*

Proof. The extra difficulty of proof that comes from J being asymptotically cylindrical is taken care of by Theorem 4.2; the rest of the proof is a straightforward modification of [Bourgeois et al. 2003]. For the sake of completeness, we outline the proof as follows.

Suppose that (Σ_n, w_n) is a sequence of J-holomorphic maps from a punctured Riemann surface Σ_n, with $E(w_n) < C$.

First, we add additional marked points to Σ_n to stabilize Σ_n, and we use the unique hyperbolic metric on Σ_n to decompose Σ_n into ε-thick part $\Sigma_n^{\varepsilon\text{-thick}}$ and

ε-thin part $\Sigma_n^{\varepsilon\text{-thin}}$ according to the injectivity radius, for $\varepsilon > 0$. Take a subsequence of Σ_n, still called Σ_n, such that Σ_n converges to a nodal surface Σ_∞ in the Deligne–Mumford sense. By continuing to add marked points to Σ_n, if necessary, one can keep track of all the sphere bubbles of w_n as $n \to \infty$. Eventually, for fixed $\varepsilon > 0$, we achieve that $w_n|_{\Sigma_n^{\varepsilon\text{-thick}}}$ has a uniformly gradient bound. By Ascoli–Arzela and elliptic estimates, we can extract a convergent subsequence of w_n, still called w_n. Now letting ε tend to 0 and picking a diagonal subsequence, we get a convergent subsequence of w_n, still called w_n, with the limit $(\Sigma_\infty, w_\infty|_{\Sigma_\infty})$. By Theorem 4.2, we know that, around a puncture, the limit $w_\infty|_{\Sigma_\infty}$ either has a removable singularity or converges to a Reeb orbit. But at the current stage, w_∞ may not be defined around the nodal points.

Secondly, for ε sufficiently small, the ε-thin part is a disjoint union of finite cylinders or half-finite cylinders. If $E_\omega(w_n|_{\Sigma_n^{\varepsilon\text{-thin}}}) \to 0$ as $n \to \infty$, then the behavior of $w_n|_{\Sigma_n^{\varepsilon\text{-thin}}}$ is controlled by Theorem 3.22. In this case, the convergence of w_n in the thick part can be continuously extended over Σ_n. Otherwise, $w_n|_{\Sigma_n^{\varepsilon\text{-thin}}}$ can have the additional broken trajectory degeneration. By adding more marked points to keep track of all of the broken trajectory, one has that $E_\omega(w_n|_{\Sigma_n^{\varepsilon\text{-thin}}}) \to 0$ as $n \to \infty$. □

Acknowledgements

Firstly, I would like to express my deepest gratitude to my advisor Yong-Geun Oh for suggesting such an interesting project. Secondly, I would like to thank Conan Leung for very patiently answering my questions about symplectic geometry. Thirdly, I would like to thank Lino Amorim, Garrett Alston, Dongning Wang, Rui Wang and Ke Zhu for all the fruitful discussions and valuable suggestions. Last but not least, I thank the anonymous referee for all the critiques and suggestions, which largely improved the quality of this paper.

References

[Audin and Lafontaine 1994] M. Audin and J. Lafontaine (editors), *Holomorphic curves in symplectic geometry*, Progress in Mathematics **117**, Birkhäuser, Basel, 1994. MR 95i:58005 Zbl 0802.53001

[Bao 2014] E. Bao, “On Hofer energy for J-holomorphic curves for asymptotically cylindrical J”, 2014. To appear in *J. Symplectic Geom.* arXiv 1303.4430

[Bourgeois 2002] F. Bourgeois, *A Morse–Bott approach to contact homology*, Ph.D. thesis, Stanford University, 2002, available at http://search.proquest.com/docview/305591502. MR 2703292 Zbl 1046.57017

[Bourgeois et al. 2003] F. Bourgeois, Y. Eliashberg, H. Hofer, K. Wysocki, and E. Zehnder, “Compactness results in symplectic field theory”, *Geom. Topol.* **7** (2003), 799–888. MR 2004m:53152 Zbl 1131.53312

[Eliashberg 2007] Y. Eliashberg, “Symplectic field theory and its applications”, pp. 217–246 in *International Congress of Mathematicians, Vol. I*, edited by M. Sanz-Solé et al., Eur. Math. Soc., Zürich, 2007. MR 2008f:53119 Zbl 1128.53059

[Eliashberg et al. 2000] Y. Eliashberg, A. Givental, and H. Hofer, "Introduction to symplectic field theory", *Geom. Funct. Anal.* Special Volume, Part II (2000), 560–673. MR 2002e:53136 Zbl 0989.81114

[Gromov 1985] M. Gromov, "Pseudo holomorphic curves in symplectic manifolds", *Invent. Math.* **82**:2 (1985), 308–347. Zbl 0592.53025

[Hofer 1993] H. Hofer, "Pseudoholomorphic curves in symplectizations with applications to the Weinstein conjecture in dimension three", *Invent. Math.* **114**:3 (1993), 515–563. MR 94j:58064 Zbl 0797.58023

[Hofer and Viterbo 1992] H. Hofer and C. Viterbo, "The Weinstein conjecture in the presence of holomorphic spheres", *Comm. Pure Appl. Math.* **45**:5 (1992), 583–622. MR 93h:58055 Zbl 0773.58021

[Hofer et al. 2001] H. Hofer, K. Wysocki, and E. Zehnder, "The asymptotic behavior of a finite energy plane", preprint, 2001, available at http://e-collection.library.ethz.ch/eserv/eth:25385/eth-25385-01.pdf.

[Hofer et al. 2002] H. Hofer, K. Wysocki, and E. Zehnder, "Finite energy cylinders of small area", *Ergodic Theory Dynam. Systems* **22**:5 (2002), 1451–1486. MR 2003k:53121 Zbl 1041.57009

[Hummel 1997] C. Hummel, *Gromov's compactness theorem for pseudo-holomorphic curves*, Progress in Mathematics **151**, Birkhäuser, Basel, 1997. MR 98k:58032 Zbl 0870.53002

Received January 19, 2014. Revised April 21, 2015.

Erkao Bao
Department of Mathematics
University of California, Los Angeles
520 Portola Plaza, Math Science Building, UCLA
Los Angeles, CA 90095
United States
bao@math.ucla.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.278.325

INTEGRATION OF COUPLING DIRAC STRUCTURES

OLIVIER BRAHIC AND RUI LOJA FERNANDES

Coupling Dirac structures are Dirac structures defined on the total space of a fibration, generalizing hamiltonian fibrations from symplectic geometry, where one replaces the symplectic structure on the fibers by a Poisson structure. We study the associated Poisson gauge theory, in order to describe the presymplectic groupoid integrating coupling Dirac structures. We find the obstructions to integrability, and we give explicit geometric descriptions of the integration.

1. Introduction

A Dirac structure on a manifold M is a (possibly singular) foliation of M by presymplectic leaves. It is well known that Dirac structures can be expressed in terms of a Lagrangian subbundle L of the generalized tangent bundle $TM \oplus T^*M$. The bundle L inherits a Lie algebroid structure from the Courant bracket [1990], so Dirac structures are infinitesimal objects. Bursztyn et al. [2004] showed that the global object underlying a given Dirac structure L is a *presymplectic groupoid*, i.e., a Lie groupoid $\mathcal{G} \rightrightarrows M$ with a multiplicative closed 2-form $\Omega_{\mathcal{G}}$ satisfying a certain nondegeneracy condition. Not all Lie algebroids can be integrated to Lie groupoids, and Dirac structures are no exception: not all Dirac structures can be integrated to presymplectic groupoids. The obstructions to integrability follow from the general obstruction theory discovered by Crainic and Fernandes [2003; 2004].

Brahic was partially supported by CNPq grant 401253/2012-0. Fernandes was partially supported by NSF grants DMS 13-08472 and DMS 14-05671. Both authors acknowledge the support of the *Ciências sem Fronteiras* program.
MSC2010: 53D17, 58H05.
Keywords: Dirac structure, coupling, presymplectic integration.

The general methods presented in [Crainic and Fernandes 2003; 2004] allow one to decide if a given Lie algebroid is integrable or not, and to produce a canonical integration in terms of an abstract path space construction. While the obstructions to integrability can be computed explicitly in many examples, describing the canonical integration $\mathcal{G}(L) \rightrightarrows M$ of a given integrable Dirac structure (M, L) is, in general, a very difficult task. However, for a few classes of Dirac structures one does have explicit integrations and often in such cases the construction of the groupoid has a nice geometric flavor.

In this paper we discuss the integration of *coupling Dirac structures*. The simplest examples of such couplings arise in the context of a *symplectic fibration* $p : E \to B$: a *coupling form* is a closed 2-form $\omega \in \Omega^2(E)$ on the total space of the fibration whose pullback to each fiber F_b is the symplectic form ω_{F_b} on the fiber. The obstructions to the existence of such a coupling form are well known and we will recall them below. We are interested in the more general situation of a *Poisson fibration*: now one looks for a coupling Dirac structure on the total space of the fibration which glues the Poisson structures on the fibers. This idea of a coupling is only a rough approximation: Dirac structures are very flexible and extra care must be taken in defining precisely the notion of a coupling [Dufour and Wade 2008; Brahic and Fernandes 2008; Vaisman 2006; Wade 2008].

Coupling Dirac structures appear very naturally in Poisson and Dirac geometry. One reason is that tubular neighborhoods of symplectic and presymplectic leaves in arbitrary Poisson and Dirac manifolds are always coupling Dirac structures. Our first main result concerning the integration of couplings can be stated as follows:

Theorem 1.1. *Let L be a coupling Dirac structure on $p : E \to B$. If L is integrable and $(\mathcal{G}, \Omega) \rightrightarrows B$ is a source connected, presymplectic groupoid integrating L, then Ω is a coupling form for a fibration $\bar{p} : \mathcal{G} \to \Pi(B)$ obtained by integrating the algebroid morphism $p_* \circ \sharp : L \to TB$.*

In other words, coupling Dirac structures integrate to coupling forms. Moreover, one can express the geometric data of the integration in terms of the geometric data associated with the coupling Dirac structure L. As a consequence of this result, any presymplectic groupoid integrating a coupling Dirac structure on $p : E \to B$ is Morita equivalent to a symplectic groupoid integrating the induced vertical Poisson structure on a fiber E_b.

The previous result describes the symplectic geometry of the integration. One is also interested in the groupoid structure of the integration and the obstructions to integrability. Our inspiration to deal with this integration problem comes from a beautiful gauge construction, known as the *Yang–Mills setup*, which yields coupling Dirac structures [Brahic and Fernandes 2014; Guillemin et al. 1996; McDuff and Salamon 1998; Weinstein 1980; Wade 2008].

Start with a principal G-bundle $P \to B$, with a connection Γ, and a hamiltonian action $G \times F \to F$ on a Poisson manifold (F, π_F), and construct a coupling Dirac structure L on the associated bundle $E = P \times_G F$ extending the Poisson structures on the fibers. This construction can be further twisted by a closed 2-form on the base B, and it leads to many examples of coupling Dirac structures.

We show that one can integrate a Yang–Mills phase space as follows:

(i) First integrate the fiber (F, π_F) to a symplectic groupoid $\mathcal{F} \rightrightarrows F$, and the principal G-bundle $P \to B$ to the gauge groupoid $\mathcal{G}(P) \rightrightarrows B$.

(ii) Then integrate the vertical Poisson structure Ver^* to a *fibered symplectic groupoid* $\mathcal{G}_V = P \times_G \mathcal{F} \rightrightarrows E$.

(iii) The gauge groupoid $\mathcal{G}(P) \rightrightarrows B$ acts on the fibered groupoid $\mathcal{G}_V \rightrightarrows E \to B$, yielding a semidirect product groupoid $\mathcal{G}(P) \ltimes \mathcal{G}_V \rightrightarrows E$.

(iv) Finally, the integration of the Yang–Mills phase space is a quotient

$$\mathcal{G}(L) = \mathcal{G}(P) \ltimes \mathcal{G}_V / \mathcal{C},$$

where $\mathcal{C}$ is a certain *curvature groupoid*.

Along the way we obtain the obstructions to integrability of a Yang–Mills phase space. Our integration procedure does not uses the principal bundle connection; hence, all the different couplings obtained by varying the connection have the same integrating Lie groupoid $\mathcal{G} \rightrightarrows E$. On the other hand, we also provide a construction for the presymplectic form $\Omega_{\mathcal{G}}$, which obviously depends on the choice of principal connection.

We show that, provided one is willing to accept infinite dimensional principal bundles, every coupling on a locally trivial fibration arises as a Yang–Mills phase space. This observation turns out to be the clue to integrate arbitrary coupling Dirac structures:

Theorem 1.2. *Let L be a coupling Dirac structure on $E \to B$. The source 1-connected groupoid $\mathcal{G}(L)$ integrating L naturally identifies with equivalence classes in $P(TB) \ltimes_B \mathcal{G}(\mathrm{Ver}^*)$ under the equivalence relation*:

- $(\gamma_0, g_0) \sim (\gamma_1, g_1)$ *if and only if there exists a homotopy* $\gamma_B : I \times I \to B$, $(t, \epsilon) \mapsto \gamma_B^\epsilon(t)$ *between* γ_0 *and* γ_1, *such that* $g_1 = \partial(\gamma_B, \boldsymbol{t}(g_0)) \,.\, g_0$.

where $\partial : P(TB) \times_B E \to \mathcal{G}(\mathrm{Ver}^*)$ *is a certain "groupoid" homomorphism that can be computed explicitly.*

The quotes in "groupoid" are used here because the path space $P(TB)$ is not really a groupoid, since associativity only holds up to isomorphism.

Again, Theorem 1.2 should be viewed as an infinite dimensional version of the groupoid integrating the Yang–Mills phase space. It also gives rise to the integrability obstructions of coupling Dirac structure. Namely, one checks that the restriction of the map $\partial : P(TB) \times_B E \to \mathcal{G}(\mathrm{Ver}^*)$ to a sphere in B based at some $b \in B$ (seen as a map $\gamma_B : I^2 \to B$ such that $\gamma_B(\partial I^2) = \{b\}$) is independent of its homotopy class. Then, if we let $\mathcal{M} := \partial(\pi_2(B) \times_B E)$, which we call the *monodromy groupoid of the fibration*, then we have the theorem:

Theorem 1.3. *Let L be a coupling Dirac structure on $E \to B$ and assume that the associated connection Γ is complete. Then, L is an integrable Lie algebroid if and only if the following conditions hold*:

(i) *the typical Poisson fiber $(E_x, \pi_V|_{E_x})$ is integrable*;

(ii) *the injection $\mathcal{M} \hookrightarrow \mathcal{G}(\mathrm{Ver}^*)$ is an embedding.*

The transgression map $\partial : \pi_2(B) \times_B E \to \mathcal{G}(\mathrm{Ver}^*)$ is computable in many examples, and so are the integrability obstructions of Theorem 1.3. We refer to the last section of the paper, where we will discuss for example the trivial fibration $p : \mathbb{S}^2 \times \mathfrak{so}^*(3) \to \mathbb{S}^2$, with the usual Lie–Poisson structure on the fibers. Using Theorem 1.3 one can see that there is only a 2-parameter family of integrable Dirac couplings of rank 4, while there is an infinite dimensional family of nonintegrable Dirac couplings of rank 4.

2. Coupling Dirac structures

The notion of a *coupling* was first introduced in the context of Dirac geometry [Vaisman 2006] (see also [Brahic and Fernandes 2008; Dufour and Wade 2008; Wade 2008]) but their origins lie in the theory of symplectic and hamiltonian fibrations; see, e.g., [Guillemin et al. 1996; McDuff and Salamon 1998]. In this section we recall the definition of a *coupling Dirac structure* and study its first properties.

2A. ***Fiber nondegenerate Dirac structures.*** We shall use some standard notions from Dirac geometry; see, e.g., [Courant 1990]. So given a smooth manifold M, we denote by $\mathbb{T}M := TM \oplus T^*M$ its generalized tangent bundle. The space of sections $\Gamma(\mathbb{T}M)$ has two natural pairings:

$$\langle (X, \alpha), (Y, \beta) \rangle_\pm := \tfrac{1}{2}(i_Y \alpha \pm i_X \beta), \tag{1}$$

and a skew-symmetric bracket, called the *Courant bracket*, given by

$$[\![(X, \alpha), (Y, \beta)]\!] := \big([X, Y],\ \mathcal{L}_X \beta - \mathcal{L}_Y \alpha + \mathrm{d}\langle (X, \alpha), (Y, \beta) \rangle_- \big). \tag{2}$$

An *almost Dirac structure* L on M is a subbundle $L \subset \mathbb{T}M := TM \oplus T^*M$ of the generalized tangent bundle, which is maximally isotropic with respect to $\langle \cdot, \cdot \rangle_+$. An almost Dirac structure is said to be a *Dirac structure* if it is furthermore closed under the bracket $[\![\cdot, \cdot]\!]$.

In general, the Courant bracket *does not* satisfy the Jacobi identity. For a Dirac structure L, however, its restriction to $\Gamma(L)$ yields a Lie bracket, and if we let $\sharp : L \to TM$ be the restriction of the projection to TM, then $(L, [\![\cdot,\cdot]\!], \sharp)$ defines a Lie algebroid. Each leaf of the corresponding characteristic foliation, obtained by integrating the singular distribution $\mathrm{Im}\,\sharp$, carries a presymplectic form ω: if $X, Y \in \mathrm{Im}\,\sharp$, we can choose $\alpha, \beta \in T^*M$ such that $(X, \alpha), (Y, \beta) \in L$ and set

$$\omega(X, Y) := \langle (X, \alpha), (Y, \beta)\rangle_- = i_Y\alpha = -i_X\beta. \tag{3}$$

One can check that this definition is independent of choices and that ω is indeed closed. Thus we may think of a Dirac manifold as a manifold foliated by (possibly singular) presymplectic leaves.

In what follows, unless otherwise stated, by a *fibration* $p : E \to B$ we mean a surjective submersion.

Definition 2.1. Let $p : E \to B$ be a fibration. An almost Dirac structure L on E is called *fiber nondegenerate* if

$$(\mathrm{Ver} \oplus \mathrm{Ver}^0) \cap L = \{0\}. \tag{4}$$

Here, $\mathrm{Ver} := \ker p_* \subset TE$ denotes the vertical distribution, and $\mathrm{Ver}^0 \subset T^*E$ its annihilator. When L is both Dirac and fiber nondegenerate, we shall refer to L as a *coupling Dirac structure*.

In the terminology of [Mărcuţ and Frejlich 2013], when L is a Poisson structure, this condition means that the fibers of $p : E \to B$ are *Poisson transversals*.

In order to understand the geometric meaning of this definition, one needs to decompose a fiber nondegenerate almost structure L into its various components:

- First, L gives rise to an *Ehresmann connection* by setting:

$$\mathrm{Hor} := \{X \in TE : \exists\, \alpha \in (\mathrm{Ver})^0 \text{ such that } (X, \alpha) \in L\}. \tag{5}$$

  The fact that $\mathrm{Hor} \oplus \mathrm{Ver} = TE$ is an easy consequence of (4).

- Next, it follows from (5) that the horizontal distribution Hor is contained in the characteristic distribution of L. Hence, we obtain a *horizontal 2-form* $\omega_H \in \Omega^2(\mathrm{Hor})$ by restricting the natural 2-form on the characteristic distribution to Hor. More precisely, (4) and (5) together show that for each $X \in \mathrm{Hor}$, there exists a unique $\alpha \in \mathrm{Ver}^0$ such that $(X, \alpha) \in L$. The skew-symmetric bilinear form $\omega_H : \mathrm{Hor} \times \mathrm{Hor} \to \mathbb{R}$ is defined by

$$\omega_H(X_1, X_2) := \langle (X_1, \alpha_1), (X_2, \alpha_2)\rangle_-, \tag{6}$$

  where $\alpha_1, \alpha_2 \in \mathrm{Ver}^0$ are the unique elements with $(X_1, \alpha_1), (X_2, \alpha_2) \in L$. Since L is maximal isotropic, this 2-form can also be written $\omega_H(X_1, X_2) = \alpha_1(X_2) = -\alpha_2(X_1)$.

- Finally, we can associate to L a *vertical bivector field* $\pi_V \in \mathfrak{X}^2(\mathrm{Ver})$. To see this, first observe that the annihilator of the horizontal distribution is:

$$\mathrm{Hor}^0 = \{\alpha \in T^*E : \exists X \in \mathrm{Ver} \text{ such that } (X, \alpha) \in L\}.$$

This, together with (4), shows that for each $\alpha \in \mathrm{Hor}^0$, there exists a unique $X \in \mathrm{Ver}$ such that $(X, \alpha) \in L$. Then one can define a skew-symmetric bilinear form $\pi_V : \mathrm{Hor}^0 \times \mathrm{Hor}^0 \to \mathbb{R}$ by letting:

$$\pi_V(\alpha_1, \alpha_2) := \langle (X_1, \alpha_1), (X_2, \alpha_2) \rangle_-, \tag{7}$$

where $X_1, X_2 \in \mathrm{Ver}$ are the unique elements with $(X_1, \alpha_1), (X_2, \alpha_2) \in L$. Since L is maximal isotropic, the form $\pi_V : \mathrm{Hor}^0 \times \mathrm{Hor}^0 \to \mathbb{R}$ can also be written as $\pi_V(\alpha_1, \alpha_2) = \alpha_1(X_2) = -\alpha_2(X_1)$. Notice that the splitting $TE = \mathrm{Hor} \oplus \mathrm{Ver}$ allows us to identify $\mathrm{Hor}^0 = \mathrm{Ver}^*$; thus, π_V becomes a bivector field on the fibers of $p : E \to B$.

A more geometric interpretation of π_V is that it is formed by the pullback to each fiber of the Dirac structure L; an easy computation shows that for each fiber $F_b = p^{-1}(b)$, the pullback Dirac structure $i_b^* L$ under the inclusion $i_b : F_b \hookrightarrow E$ coincides with π_V:

$$i_b^* L := \{(X, \alpha|_{\mathrm{Ver}}) \in \mathrm{Ver} \oplus \mathrm{Ver}^* : (X, \alpha) \in L\} = \mathrm{Graph}(\pi_V).$$

The preceding discussion justifies the following definition:

Definition 2.2. A *geometric data* on a fibration $p : E \to B$ is a triple $(\pi_V, \Gamma, \omega_H)$, where

- $\pi_V \in \mathfrak{X}^2(\mathrm{Ver})$ is a vertical bivector field.
- Γ is an *Ehresmann* connection, whose horizontal distribution is denoted Hor,
- $\omega_H \in \Omega^2(\mathrm{Hor})$ is a horizontal 2-form,

Proposition 2.3. *Given a fibration $E \to B$, there is a one-to-one correspondence between fiber nondegenerate almost Dirac structures and geometric data on the fibration.*

Proof. We have seen above how to associate to a fiber nondegenerate almost Dirac structure L, a geometric data $(\pi_V, \Gamma, \omega_H)$. Conversely, given a geometric data $(\Gamma_L, \omega_H, \pi_V)$ on a fibration $p : E \to B$, define an almost Dirac structure L by

$$L := \{(X + \pi_V^\sharp(\alpha), i_X\omega_H + \alpha) : X \in \mathrm{Hor} \text{ and } \alpha \in \mathrm{Hor}^0\}. \tag{8}$$

Notice that by using the identifications $\mathrm{Hor}^0 = \mathrm{Ver}^*$ and $\mathrm{Ver}^0 = \mathrm{Hor}^*$, we obtain simply $L = \mathrm{Graph}\,\omega_H \oplus \mathrm{Graph}\,\pi_V$, which will prove to be a meaningful way of presenting L later on. ∎

Given a fiber nondegenerate almost Dirac structure L with associated geometric data $(\Gamma, \omega_H, \pi_V)$, we now express the conditions on this data which will guarantee that L is a Dirac structure, i.e., that it is closed under the Courant bracket.

Let us first introduce some notations associated with the connection Γ. For a vector field $v \in \mathfrak{X}(B)$, we denote by $h(v) \in \mathfrak{X}(E)$ its horizontal lift. The exterior covariant differential $\mathrm{d}_\Gamma : \Omega^k(B) \otimes C^\infty(E) \to \Omega^{k+1}(B) \otimes C^\infty(E)$ is given by

$$\mathrm{d}_\Gamma\omega(v_0, \dots, v_k) := \sum_{i=0}^{k} (-1)^i \, \mathcal{L}_{h(v_i)} \, \omega(v_0, \dots, \hat{v}_i, \dots, v_k) + \sum_{i<j} (-1)^{i+j} \omega([v_i, v_j], v_0, \dots, \hat{v}_i, \dots, \hat{v}_j, \dots, v_k)$$

The curvature of Γ will be denoted by $\mathrm{Curv}_\Gamma \in \Omega^2(B, \mathrm{Ver})$ and is defined by

$$\mathrm{Curv}_\Gamma(v, w) := [h(v), h(w)] - h([v, w]) \quad \text{for } v, w \in \mathfrak{X}(B).$$

The curvature measures the failure of Hor in being involutive or, equivalently, the failure of d_Γ being a differential since

$$\mathrm{d}_\Gamma^2 f(u, v) = \mathcal{L}_{\mathrm{Curv}(u,v)} \, f,$$

for any $f \in C^\infty(E)$ and $u, v \in \mathfrak{X}(B)$.

Proposition 2.4. *Let $(\pi_V, \Gamma, \omega_H)$ be the geometric data determined by a fiber nondegenerate almost Dirac structure L on a fiber bundle $p : E \to B$. Then, L is a Dirac structure if and only if the following conditions hold*:

(i) *π_V is a vertical Poisson structure*:

$$[\pi_V, \pi_V] = 0;$$

(ii) *parallel transport along Γ preserves the vertical Poisson structure*:

$$\mathcal{L}_{h(v)} \, \pi_V = 0, \quad \textit{for any } v \in \mathfrak{X}(B);$$

(iii) *the horizontal 2-form ω_H is Γ-closed*:

$$\mathrm{d}_\Gamma\omega_H = 0;$$

(iv) *the curvature is hamiltonian*:

$$\mathrm{Curv}(u, v) = \pi_V^\sharp(\mathrm{d}i_{h(u)}i_{h(v)}\omega_H), \quad \textit{for any } u, v \in \mathfrak{X}(B). \tag{9}$$

A proof of Proposition 2.4 can be found in [Brahic and Fernandes 2008]. We shall refer to (9) as the *curvature identity*.

2B. *Examples.* The notion of coupling Dirac structure contains as special cases the notion of *coupling form* for symplectic fibrations (see, e.g., [Guillemin et al. 1996]) and the *coupling Poisson tensor* considered by Vorobjev [2001]. We now recall these examples as well as other ones.

2B1. *Coupling forms.* Let ω be a closed 2-form on the total space of a fibration $p : E \to B$. The associated Dirac structure $L := \mathrm{Graph}(\omega)$ is fiber nondegenerate if and only if the pullback of ω to each fiber is a nondegenerate 2-form. In this case, the fibration with the restriction of ω to the fibers becomes a symplectic fibration. The geometric data $(\pi_V, \Gamma, \omega_H)$ associated to L has a nondegenerate vertical Poisson structure π_V which coincides with the inverse of the restriction of ω to the fibers.

The converse is also true: a fiber nondegenerate Dirac structure L for which the geometric data $(\pi_V, \Gamma, \omega_H)$ has a nondegenerate vertical Poisson structure π_V is determined by a presymplectic form ω. In fact, it follows from (8) that L is the graph of the closed 2-form

$$\omega = \omega_H \oplus (\pi_V)^{-1}.$$

Hence, fiber nondegenerate presymplectic forms are the same as coupling forms for symplectic fibrations [Guillemin et al. 1996].

2B2. *Coupling Poisson structures.* Let π be a Poisson structure on the total space of a fibration $p : E \to B$. The Dirac structure $L = \mathrm{Graph}(\pi)$ is fiber nondegenerate if and only if π is horizontal nondegenerate in the sense of [Vorobjev 2001], i.e., if the bilinear form $\pi|_{\mathrm{Ver}^0} : \mathrm{Ver}^0 \times \mathrm{Ver}^0 \to \mathbb{R}$ is nondegenerate. In terms of the associated geometric data $(\pi_V, \Gamma_L, \omega_H)$ the Poisson structure on the fibers is π_V and ω_H is nondegenerate; in fact, π induces an isomorphism $\mathrm{Ver}^0 \to \mathrm{Hor}$ under which ω_H coincides with the restriction $\pi|_{\mathrm{Ver}^0}$.

The converse is also true: a fiber nondegenerate Dirac structure L for which the horizontal 2-form ω_H is nondegenerate is given by a Poisson structure π. In fact, it follows from (8) that

$$\pi = (\omega_H)^{-1} \oplus \pi_V.$$

Hence, fiber nondegenerate Poisson structures are the same thing as the horizontal nondegenerate Poisson structures of Vorobjev.

2B3. *Neighborhood of a presymplectic leaf.* Let L be any Dirac structure on a manifold M and fix a presymplectic leaf (S, ω) of L. Then, the restriction of L to any sufficiently small tubular neighborhood $p : \nu(S) \to S$ of the leaf is a coupling Dirac structure. To see this, one observes that along S:

$$L_x \cap \big(\nu(S)_x \oplus \nu_x(S)^0\big) = \{0\}, \quad \text{for all } x \in S.$$

It follows that L is fiber nondegenerate on a sufficiently small neighborhood of the zero section.

This shows that in a neighborhood of a presymplectic leaf the Dirac structure takes a special form and we can associate to it the geometric data $(\pi_V, \Gamma, \omega_H)$. The Poisson structure π_V is the transverse Poisson structure along S, while S (viewed as the zero section) is an integral leaf of Hor and the 2-form ω_H restricted to this leaf coincides with ω. Note that, in general, the distribution Hor fails to be integrable at other points.

2B4. *Reduction of canonical bundles.* Let $P \to M$ be a principal G-bundle. The action of G naturally lifts to a hamiltonian action of G on $(T^*P, \omega_{\text{can}})$. Clearly, T^*P is itself a principal G-bundle, sometimes called a *canonical bundle*, and it follows that the base manifold T^*P/G has an induced Poisson structure π.

Each choice of a principal bundle connection $\theta : TP \to \mathfrak{g}$ induces a projection map $p_\theta : T^*P/G \to T^*M$. It is easy to check that, for any choice of connection, the Dirac structure $L = \text{Graph}(\pi)$ on $E = T^*P/G$ is a coupling Dirac structure over $B = T^*M$.

2B5. *Yang–Mills–Higgs phase spaces.* There is a construction using principal bundles and hamiltonian actions which leads to an important class of coupling Dirac structures.

Definition 2.5. A classical Yang–Mills–Higgs setting is a triple (P, G, F) where $P \to B$ is a principal G-bundle and (F, π_F) is a G-hamiltonian Poisson manifold with equivariant moment map $\mu_F : F \to \mathfrak{g}^*$.

Proposition 2.6. *Let (P, G, F) be a classical Yang–Mills–Higgs setting. Each choice of a principal bundle connection $\theta : TP \to \mathfrak{g}$ determines a coupling Dirac structure on the associated fiber bundle $E := P \times_G F \to B$.*

The construction is well known (see [Brahic and Fernandes 2014; Weinstein 1980; Wade 2008]), so it will only be sketched. First, the connection $\theta : TP \to \mathfrak{g}$ gives a G-equivariant embedding $i_\theta : (\ker \mathrm{d}p)^* \hookrightarrow T^*P$, where $p : P \to B$ is the principal bundle projection. This allows us to pullback the hamiltonian G-space $(T^*P, \omega_{\text{can}}, \mu_{\text{can}})$, where $\mu_{\text{can}} : T^*P \to \mathfrak{g}^*$ is the dual of the infinitesimal action $\mathfrak{g} \to TP$, to obtain a hamiltonian G-space $((\ker \mathrm{d}p)^*, L_\theta, \mu_\theta)$, where $L_\theta := \text{Graph}(i_\theta^* \omega_{\text{can}})$ and $\mu_\theta : (\ker \mathrm{d}p)^* \to \mathfrak{g}^*$ is the composition $\mu_{\text{can}} \circ i_\theta$.

Next, combine the hamiltonian G-spaces $((\ker \mathrm{d}p)^*, L_\theta, \mu_\theta)$ and (F, L_{π_F}, μ_F), where $L_{\pi_F} = \text{Graph}(\pi_F)$, to obtain a hamiltonian G-space

$$((\ker \mathrm{d}p)^* \times F, L_\theta \times L_{\pi_F}, \mu_\theta + \mu_F),$$

where G acts diagonally on $(\ker \mathrm{d}p)^* \times F$.

Finally, observe that the hamiltonian quotient

$$((\ker \mathrm{d}p)^* \times F)//G := \{(v, f) \in (\ker \mathrm{d}p)^* \times F : \mu_\theta(u) + \mu_F(f) = 0\}/G$$

is isomorphic to $E := P \times_G F$: the map $[(v, f)] \mapsto [(u, f)]$, where $v \in \ker \mathrm{d}_u p$, gives the desired isomorphism. It follows that E has a quotient Dirac structure L, and this is indeed a coupling Dirac structure for the fibration $E \to B$.

The associated coupling Dirac structure can be described as follows. Since G acts on F by Poisson automorphisms, the associated bundle $E := P \times_G F$ has an induced vertical Poisson structure π_V with typical fiber (F, π_F). The induced connection Γ on E is a Poisson connection. Denoting by $\omega_\theta \in \Omega^2(B, \mathfrak{g})$ the curvature of the principal connection $\theta : TP \to \mathfrak{g}$, one obtains a well defined horizontal 2-form ω_H on E by setting

$$\omega_H(h(v_1), h(v_2)) := \langle \mu_F, \omega_\theta(v_1, v_2) \rangle.$$

The triple $(\pi_V, \Gamma, \omega_H)$ is the geometric data associated with L. One can also easily check that this triple satisfies the conditions in proposition 2.4.

Dirac structures obtained in this way are sometimes called classical *Yang–Mills–Higgs phase spaces*. We will be interested in the problem of integrability of such structures. In particular, the integrability of (F, π_F) is not enough to ensure the integrability of the associated bundle, as shown in the following example.

Example 2.7. Consider the Hopf fibration $P = \mathbb{S}^3 \to \mathbb{S}^2$, seen as an $\mathbb{S}^1$-principal bundle. One can choose a principal connection θ whose curvature is given by $\omega_\theta = p^*\omega$, where ω is the standard symplectic form on $\mathbb{S}^2$. Consider, furthermore, $F = \mathbb{R}$ endowed with the trivial Poisson structure $\pi_F = 0$, and let $\mathbb{S}^1$ act trivially on F. Any smooth function $f : F \to \mathbb{R}$ serves as a momentum map.

The associated bundle is trivial: $E = P \times_{\mathbb{S}^1} F = \mathbb{S}^2 \times \mathbb{R}$. Moreover, it is easily checked that the induced coupling Dirac structure has presymplectic leaves $(\mathbb{S}^2 \times \{x\}, f(x)\omega)$. Here, the associated geometric data is given by $(\pi_V, \mathrm{Hor}, \omega_H)$ where $\pi_V = 0$, Hor is the flat connection given by the trivialization, and $\omega_H := f p^*\omega$. Although π_V is integrable, the coupling Dirac structure L is not integrable whenever f has some critical point; see [Crainic and Fernandes 2004].

Remark 2.8. General coupling Dirac structures can be seen as infinite dimensional Yang–Mills–Higgs phase spaces, provided one allows for infinite dimensional structure groups, as shown in [Brahic and Fernandes 2008]. A precise formulation requires the theory of Fréchet manifolds and Fréchet Lie groups. However, one can use this *Poisson gauge theory* heuristically, offering guidance on how to extend constructions which work for a Yang–Mills–Higgs phase space to a general coupling Dirac structure. We will use this principle later in our construction of the integration of general coupling Dirac structures.

2C. ***Coupling Dirac structures as extensions.*** In the Yang–Mills–Higgs approach to general coupling Dirac structures, one must allow for infinite dimensional structure groups. An alternative approach (see [Brahic 2010]) is to observe that coupling Dirac structures give rise to Lie algebroid extensions.

Proposition 2.9. *Let L be a coupling Dirac structure on $p : E \to B$. The morphism $p_* \circ \sharp : L \to TB$ induces a Lie algebroid extension*

$$\operatorname{Graph}(\pi_V) \hookrightarrow L \twoheadrightarrow TB \tag{10}$$

Moreover, the decomposition (8) induces an Ehresmann connection with horizontal space $\operatorname{Graph}(\omega_H)$, *namely*

$$L = \operatorname{Graph}(\pi_V) \oplus \operatorname{Graph}(\omega_H). \tag{11}$$

Proof. The map $p_* \circ \sharp : L \to TB$ is clearly a Lie algebroid morphism, being the composition of algebroid morphisms. It is a surjective morphism because $\operatorname{Hor} \subset \operatorname{Im} \sharp$ and, since it covers the surjective submersion $p : E \to B$, it defines a Lie algebroid extension. The fiber nondegeneracy condition also implies that the kernel of the extension is given by

$$\ker(p_* \circ \sharp) = \sharp^{-1}(\mathrm{Ver}) = \operatorname{Graph}(\pi_V).$$

Hence, the kernel is exactly $\operatorname{Graph}(\pi_V) \subset \mathrm{Ver}^* \oplus \mathrm{Ver}$. The decomposition (11) gives a complementary vector subbundle to this kernel, i.e., an Ehresmann connection in the sense of Brahic. □

For a Lie algebroid extension which is split, as in (11), there is a natural decomposition of its Lie bracket [Brahic 2010, Lemma 1.8].

First, we may identify $\operatorname{Graph}(\pi_V)$ with Ver^*, so sections of $\operatorname{Graph}(\pi_V)$ are identified with vertical forms. Vertical forms $\alpha,\ \beta \in \Gamma(\mathrm{Ver}^*)$ come naturally equipped with a bracket and an anchor inherited from π_V:

$$\sharp_V(\alpha) := \pi_V^\sharp(\alpha), \tag{12}$$

$$[\alpha, \beta]_V := \mathcal{L}_{\sharp_V(\alpha)}\, \beta - \mathcal{L}_{\sharp_V(\beta)}\, \alpha - \mathrm{d}_V \pi_V(\alpha, \beta), \tag{13}$$

where $\mathrm{d}_V : C^\infty(E) \to \Gamma(\mathrm{Ver}^*)$ denotes the vertical de Rham differential. Since π_V is Poisson, this makes Ver^* into a Lie algebroid.

Second, for each $v \in \mathfrak{X}(B)$ there is a unique section $h^*(v)$ of $\operatorname{Graph}(\omega_H)$ such that $\mathrm{d}p \circ \sharp(h^*(v)) = v$. In fact, we have an isomorphism $\sharp : \operatorname{Graph}(\omega_V) \to \operatorname{Hor}$, so we can first lift v to $h(v) \in \Gamma(\operatorname{Hor})$ and then apply $\sharp^{-1}$, which gives

$$h^*(v) = (h(v), i_{h(v)}\omega_H) \in \Gamma(\operatorname{Graph}(\omega_H)).$$

We refer to $h^* : \mathfrak{X}(B) \to \Gamma(\operatorname{Graph}(\omega_H))$ as the *cohorizontal lifting map*.

Sections of L are generated by sections $\alpha \in \Gamma(\mathrm{Ver}^*)$ and $h^*(v)$, for $v \in \mathfrak{X}(B)$, so the Lie bracket on L is entirely determined by its value on these two types of sections.

Proposition 2.10 (Splitting Brackets). *Let L be a coupling Dirac structure on $E \to B$. Under the decomposition (11) the Lie bracket of L satisfies*:

$$[\alpha, \beta]_L = [\alpha, \beta]_V, \quad [h^*(v), \alpha]_L = \mathcal{L}_{h(v)}\alpha,$$
$$[h^*(v), h^*(w)]_L = h^*([v, w]) + \mathrm{d}_V\omega_H(h(v), h(w)),$$

while the anchor takes the form:

$$\sharp(h^*(v) + \alpha) = h(v) + \sharp_V(\alpha) \tag{14}$$

for any elements $v, w \in \mathfrak{X}(B)$ and $\alpha, \beta \in \Gamma(\mathrm{Ver}^*)$.

Proof. The proposition follows from straightforward computation using (2) and the identifications $\mathrm{Hor}^* \simeq \mathrm{Ver}^0$ and $\mathrm{Ver}^* \simeq \mathrm{Hor}^0$. □

In particular, we see that the curvature of $\mathrm{Graph}(\omega_H)$, as an Ehresmann connection on the Lie algebroid extension L, is given by

$$(\pi_V^\sharp \mathrm{d}_V\omega_H, \mathrm{d}_V\omega_H) \in \Omega^2\big(B, \Gamma(\mathrm{Graph}(\pi_V))\big).$$

Notice that this is just another way of expressing the curvature identity (9).

3. Integration of coupling Dirac structures I

As stated in the introduction, our main aim is to understand the integration of coupling Dirac structures. We now take care of the symplectic geometry, showing that an $\boldsymbol{s}$-connected groupoid integrating a coupling Dirac structure has a presymplectic 2-form which is itself a coupling form.

3A. *Presymplectic groupoids.* In the sequel, we will denote by $\mathcal{G} \rightrightarrows M$ a Lie groupoid, with source and target maps $\boldsymbol{s}, \boldsymbol{t} : \mathcal{G} \to M$, identity section $\iota : M \to \mathcal{G}$, $m \mapsto \mathbf{1}_m$, and inversion $i : \mathcal{G} \to \mathcal{G}$, $x \mapsto x^{-1}$. The composition of two arrows, denoted by $x \cdot y$, is only defined provided $\boldsymbol{s}(x) = \boldsymbol{t}(y)$.

We will denote by $p_A : A \to M$ a Lie algebroid with Lie bracket $[\cdot, \cdot]_A$ and anchor $\sharp : A \to TM$. Given a Lie groupoid $\mathcal{G}$, the corresponding Lie algebroid has underlying vector bundle $A(\mathcal{G}) := \ker \mathrm{d}_{\iota(M)}\boldsymbol{s}$ and anchor $\sharp := \mathrm{d}_{\iota(M)}\boldsymbol{t}$. The sections of $A(\mathcal{G})$ can be identified with the right invariant vector fields on $\mathcal{G}$, and this determines the Lie bracket on sections of $A(\mathcal{G})$. A groupoid that arises in this way is called *integrable*.

Not every Lie algebroid $p_A : A \to M$ is integrable. However, there always exists a topological groupoid $\mathcal{G}(A)$ with source 1-connected fibers, that formally "integrates" A, called the *Weinstein groupoid* of A. Moreover, A is integrable if and only if $\mathcal{G}(A)$ is smooth, in which case $A(\mathcal{G}(A))$ is canonically isomorphic to A; see [Crainic and Fernandes 2003; 2011].

Let us recall briefly the construction of $\mathcal{G}(A)$. More details can be found in [loc. cit.]. An *A-path* is a path $a : I \to A$ such that:

$$\sharp a(t) = \frac{\mathrm{d}}{\mathrm{d}t} p_A(a(t)).$$

We denote by $P(A)$ the space of A-paths (up to reparametrization), and we set $\boldsymbol{s}(a) := p_A \circ a(0)$ and $\boldsymbol{t}(a) := p_A \circ a(1)$. On the space $P(A)$, there is an equivalence relation $\sim$, called A-homotopy, that preserves the multiplication. The Weinstein groupoid is the quotient of $P(A)$ by A-homotopies:

$$\mathcal{G}(A) := P(A)/\!\sim.$$

Given an A-path a, we denote its A-homotopy class by $[a]_A$, or simply $[a]$ when no confusion seems possible.

When a Lie algebroid arises from a geometric structure, the corresponding Lie groupoid usually inherits some extra geometric structure. In the case of Dirac structures L, the Weinstein groupoid $\mathcal{G}(L)$ comes equipped with a multiplicative presymplectic form; see [Bursztyn et al. 2004].

Definition 3.1. A 2-form $\Omega \in \Omega^2(\mathcal{G})$ is multiplicative if

$$m^*\Omega = \mathrm{pr}_1^*\,\Omega + \mathrm{pr}_2^*\,\Omega,$$

where $m : \mathcal{G}^{(2)} \to \mathcal{G}$ is the multiplication of composable arrows and $\mathrm{pr}_i : \mathcal{G}^{(2)} \to \mathcal{G}$ is the projections onto factor i. A *presymplectic groupoid* is a Lie groupoid endowed with a multiplicative 2-form Ω such that

$$\ker \Omega_x \cap \ker(\mathrm{d}\boldsymbol{s})_x \cap \ker(\mathrm{d}\boldsymbol{t})_x = \{0\}, \quad \text{for all } x \in M. \tag{15}$$

Roughly speaking, Dirac structures integrate to presymplectic groupoids.

Theorem 3.2 [Bursztyn et al. 2004]. *Let L be a Dirac structure on a manifold M. If L is integrable, then $\mathcal{G}(L)$ has a naturally induced multiplicative presymplectic form such that the map $(\boldsymbol{t}, \boldsymbol{s}) : (\mathcal{G}(L), \Omega) \to (M \times M, L \times L^{\mathrm{op}})$ is f-Dirac.*

The aforementioned multiplicative presymplectic form Ω on $\mathcal{G}(L)$ is related to sections of L in the following way: for any $X \in T\mathcal{G}$ and any pair of sections $\eta = (v, \alpha)$, $\xi = (w, \beta)$ in $\Gamma(L)$,

$$\Omega(\overleftarrow{\eta}, X) = -\alpha(\boldsymbol{s}_* X), \quad \Omega(\vec{\xi}, X) = \beta(\boldsymbol{t}_* X),$$

where we denoted by $\overleftarrow{\eta}$ the left invariant vector field on $\mathcal{G}(L)$ associated to η and by $\vec{\xi}$ the right invariant vector field associated to ξ. Also, source and target fibers turn out to be presymplectically orthogonal:

$$\Omega(\overleftarrow{\eta}, \vec{\xi}) = 0.$$

Finally, if (S, ω_S) is the presymplectic leaf of (M, L) through $x \in M$, then the map $\boldsymbol{t} : \boldsymbol{s}^{-1}(x) \to S$ defines a principal G_x-bundle, and

$$i^*_{s^{-1}(x)}\Omega = \boldsymbol{t}|^*_{s^{-1}(x)}\omega_S,$$

where $i_{s^{-1}(x)} : s^{-1}(x) \hookrightarrow \mathcal{G}(L)$ denotes the inclusion.

Note that, given an integrable Dirac structure (M, L), there can be other presymplectic groupoids $(\mathcal{G}, \Omega_{\mathcal{G}})$ integrating (M, L) besides $(\mathcal{G}(L), \Omega)$. However, if $(\mathcal{G}, \Omega_{\mathcal{G}})$ has source connected fibers, then there is a covering Lie groupoid homomorphism $\Phi : (\mathcal{G}(L), \Omega) \to (\mathcal{G}, \Omega_{\mathcal{G}})$ with $\Phi^*\Omega_{\mathcal{G}} = \Omega$.

3B. ***Couplings integrate to couplings.*** Assume that L is an integrable coupling Dirac structure on a fibration $p : E \to B$. The anchor $\sharp : L \to TE$ is a Lie algebroid morphism that integrates to the groupoid morphism $\mathcal{G}(L) \to \Pi(E)$ which associates to the homotopy class of an L-path the homotopy class of its base path. On the other hand, $p_* : TE \to TB$ is a Lie algebroid morphism whose integration $\Pi(E) \to \Pi(B)$ is the morphism $[\gamma] \mapsto [p \circ \gamma]$. We will denote the composition of this two groupoid morphisms by $\tilde{p} : \mathcal{G}(L) \to \Pi(B)$.

Now the morphism $\tilde{p} : \mathcal{G}(L) \to \Pi(B)$ integrates the Lie algebroid morphism $p_* \circ \sharp : L \to TB$. Since $p_* \circ \sharp$ is surjective on the fibers, by the coupling condition, we see that $\tilde{p} : \mathcal{G}(L) \to \Pi(B)$ is a submersion, which is not necessarily surjective.

Proposition 3.3. *Suppose that L is an integrable coupling Dirac structure on a fibration $p : E \to B$ and the induced Ehresmann connection is complete. Then, $\tilde{p} : \mathcal{G}(L) \to \Pi(B)$ is surjective, so it is a fibration.*

Proof. Given $[\gamma] \in \Pi(B)$, where $\gamma : I \to B$ is a smooth path, completeness allows us to lift γ to a horizontal path $\tilde{\gamma} : I \to E$. Since $\tilde{\gamma}'(t) \in \operatorname{Im}\sharp$, we can find an L-path $a : I \to L$ with base path $\tilde{\gamma}$. Then, $\tilde{p}([a]) = [\gamma]$. □

Remark 3.4. One can show that if a locally trivial fibration $p : E \to M$ admits a complete Ehresmann connection, then $p_* : \Pi(E) \to \Pi(B)$ is also locally trivial and carries an induced Ehresmann connection. It follows then that if L is an integrable coupling Dirac structure on a fibration $p : E \to B$ and the induced Ehresmann connection is complete, then $\tilde{p} : \mathcal{G}(L) \to \Pi(B)$ is also a locally trivial fibration.

From now on, we will make the implicit assumption that our coupling Dirac structures have complete induced connections. This is the case, for example, if the fibers are compact.

Theorem 3.5. *Let L be a coupling Dirac structure on $p : E \to B$. If L is integrable, then the multiplicative presymplectic form Ω on $\mathcal{G}(L)$ is fiber nondegenerate for the fibration*

$$\tilde{p} : \mathcal{G}(L) \to \Pi(B), \tag{16}$$

obtained by integrating the Lie algebroid morphism $p_* \circ \sharp : L \to TB$.

Proof. Let us denote by $\mathrm{Ver}_{\mathcal{G}_L} := \ker \tilde{p}$. We only need to check that the nondegeneracy condition (4) holds:

$$(\mathrm{Ver}_{\mathcal{G}(L)} \oplus \mathrm{Ver}^0_{\mathcal{G}(L)}) \cap \mathrm{Graph}\,\Omega = \{0\}.$$

First notice that since $\tilde{p}$ is obtained by composing the groupoid maps

$$\mathcal{G}(L) \to \Pi(E) \to \Pi(B),$$

it follows that $(X, \alpha) \in T\mathcal{G}(L) \oplus T^*\mathcal{G}(L)$ lies in $\mathrm{Ver}_{\mathcal{G}(L)} \oplus \mathrm{Ver}^0_{\mathcal{G}(L)}$ if and only if it satisfies

$$(\boldsymbol{s}_* \times \boldsymbol{t}_*)(X) \in \mathrm{Ver} \times \mathrm{Ver} \quad \text{and} \quad \alpha \in (\boldsymbol{s}^* \times \boldsymbol{t}^*)(\mathrm{Ver}^0 \times \mathrm{Ver}^0).$$

Let $g \in \mathcal{G}(L)$ be the base point of (X, α), and set $x := \boldsymbol{s}(g) \in E$ and $y := \boldsymbol{t}(g) \in E$. The second condition shows that $\alpha \in \boldsymbol{s}^*(\mathrm{Ver}^0_x) + \boldsymbol{t}^*(\mathrm{Ver}^0_x)$, so there exists $a_0 \in \mathrm{Ver}^0_x$ and $a_1 \in \mathrm{Ver}^0_y$ such that $\alpha = \boldsymbol{s}^* a_0 - \boldsymbol{t}^* a_1$. It follows from the first condition that

$$(\boldsymbol{s}_* X, a_0) \in \mathrm{Ver}_x \oplus \mathrm{Ver}^0_x \quad \text{and} \quad (\boldsymbol{t}_* X, a_1) \in \mathrm{Ver}_y \oplus \mathrm{Ver}^0_y.$$

Thus, for any $(X, \alpha) \in (\mathrm{Ver}_{\mathcal{G}(L)} \oplus \mathrm{Ver}^0_{\mathcal{G}(L)}) \cap \mathrm{Graph}\,\Omega$, we must have $(\boldsymbol{s}_* X, a_0) \in L_x$ and $(\boldsymbol{t}_* X, a_1) \in L_y$, since $\boldsymbol{s} \times \boldsymbol{t}$ is a forward Dirac map. By the fiber nondegeneracy condition of L, we conclude that $(\boldsymbol{s}_* X, a_0) = 0$ and $(\boldsymbol{t}_* X, a_1) = 0$.

It follows that $\alpha = \boldsymbol{s}^* a_0 - \boldsymbol{t}^* a_1 = 0$ and that $X \in \ker \boldsymbol{s}_* \cap \ker \boldsymbol{t}_*$. Since $(X, \alpha) \in \mathrm{Graph}\,\Omega$, we conclude that $X \in \ker \boldsymbol{s}_* \cap \ker \boldsymbol{t}_* \cap \ker \Omega$. The nondegeneracy condition of Ω (see Definition 3.1) shows that we must also have $X = 0$. □

Remark 3.6. For each $b \in B$, the fiber $\tilde{p}^{-1}(\mathbf{1}_b)$ is a Lie subgroupoid of $\mathcal{G}(L)$ over the fiber $E_b := p^{-1}(b)$, and the restriction of Ω to the fiber is symplectic: it is a symplectic groupoid integrating the vertical Poisson structure (E_b, π_b); the fact that $\ker p_* \circ \sharp = \sharp^{-1}(\mathrm{Ver})$ is identified with $\mathrm{Graph}\,\pi_V = \mathrm{Ver}^*$ as a Lie algebroid is a consequence of Proposition 2.10. The kernel of $\tilde{p}$ is also a Lie subgroupoid of $\mathcal{G}(L)$ over E of a special kind, called a *fibered symplectic groupoid*, which we will study in Section 4A.

If $(\mathcal{G}, \Omega_{\mathcal{G}})$ is another presymplectic groupoid integrating (E, L) with source connected fibers, then we claim that there is also a fibration $\bar{p} : \mathcal{G} \to \mathcal{G}_B$, where $\mathcal{G}_B$ is a certain Lie groupoid integrating TB; in fact, since $\mathcal{G}$ has source connected fibers, there is a covering homomorphism $\Phi : \mathcal{G}(L) \to \mathcal{G}$, whose kernel $\mathcal{N} \subset \mathcal{G}(L)$

is an embedded bundle of normal subgroups. Its image $\tilde{p}(\mathcal{N}) \subset \pi_1(B)$ is also an embedded bundle of normal subgroups and so the quotient $\mathcal{G}_B := \pi_1(B)/\tilde{p}(\mathcal{N})$ is another Lie groupoid integrating TB. Moreover, we obtain a groupoid morphism $\bar{p} : \mathcal{G} \to \mathcal{G}_B$ from $\tilde{p} : \mathcal{G}(L) \to \pi(B)$ by passing to the quotient. We then obtain as a corollary of Theorem 3.5:

Corollary 3.7. *Let L be a coupling Dirac structure on $E \to B$. If $(\mathcal{G}, \Omega_{\mathcal{G}})$ is any source connected presymplectic groupoid integrating L, then Ω_G is a coupling form relative to the fibration $\bar{p} : \mathcal{G} \to \mathcal{G}_B$, the unique Lie groupoid homomorphism integrating the Lie algebroid morphism $\sharp \circ p_*$.*

3C. ***Integration of the geometric data.*** Let L be an integrable coupling Dirac structure on $p : E \to B$, with associated geometric data $(\pi_V, \Gamma, \omega_H)$, and $(\mathcal{G}, \Omega)$ a source connected presymplectic groupoid integrating L. According to the results of the previous section, Ω is a coupling form relative to a fibration $\bar{p} : \mathcal{G} \to \mathcal{G}_B$, which is a Lie groupoid homomorphism integrating the Lie algebroid morphism $p_* \circ \sharp$. We denote by $(\Omega_V, \tilde{\Gamma}, \Omega_H)$ the corresponding geometric data.

One can obtain the geometric data of the coupling multiplicative 2-form Ω in terms of the geometric data of the coupling Dirac structure L as follows:

Proposition 3.8 (integration of the geometric data). *The geometric data $(\Omega_V, \tilde{\Gamma}, \Omega_H)$ for Ω is related to the geometric data $(\pi_V, \Gamma, \omega_H)$ for L in the following way*:

(i) *$(\mathcal{G}_{E_b}, \Omega_{E_b}) := (\tilde{p}^{-1}(\mathbf{1}_b), i_b^* \Omega_V)$ is a symplectic Lie groupoid over E_b, which integrates $\pi_V|_{E_b}$, where $i_b : \tilde{p}^{-1}(\mathbf{1}_b) \hookrightarrow \mathcal{G}$ is the inclusion.*

(ii) *The connection $\tilde{\Gamma}$ has horizontal lift given by*

$$\mathrm{H}(v, w) = \overleftarrow{h}^*(v) - \overrightarrow{h}^*(w), \tag{17}$$

where h^* denotes the cohorizontal lift.

(iii) *Under the natural identification $T_g\mathcal{G}_B = T_{\boldsymbol{t}(g)}B \times T_{\boldsymbol{s}(g)}B$, the horizontal form Ω_H is given by*

$$\Omega_H(H(v_1, w_1), H(v_2, w_2)) = \omega_H(h(v_1), h(v_2)) \circ \boldsymbol{t} - \omega_H(h(w_1), h(w_2)) \circ \boldsymbol{s}. \tag{18}$$

Proof. Item (i) was already discussed in Section 3B (see Remark 3.6).

To prove item (ii), consider an element $(v, w) \in T_g\mathcal{G}_B = T_{\boldsymbol{t}(g)}B \times T_{\boldsymbol{s}(g)}B$. Using the expression $h^*(v) := (h(v), i_{h(v)}\omega_H)$ for the cohorizontal lifts, one checks that the right hand term in (17) projects onto (v, w) and lies in $\mathrm{Ver}_G^{\perp \Omega_L}$. By uniqueness, it must coincide with $\mathrm{H}(v, w)$.

Finally, expression (18) for the horizontal 2-form follows by straightforward computation, using the general properties of multiplicative 2-forms:

$$\begin{aligned}
\Omega_H\big(\mathrm{H}(v_1, w_1), \mathrm{H}(v_2, w_2)\big) \\
&= \Omega\big(\overleftarrow{h^*}(v_1), \overleftarrow{h^*}(v_2)\big) + \Omega\big(\overrightarrow{h^*}(w_1), \overrightarrow{h^*}(w_1)\big) \\
&= \big\langle (h(v_1), \eta_{v_1}), (h(v_2), \eta_{v_2}) \big\rangle_- \circ \boldsymbol{s} \,-\, \big\langle (h(w_1), \eta_{w_1}), (h(w_2), \eta_{w_2}) \big\rangle_- \circ \boldsymbol{t} \\
&= \omega_H(h(v_1), h(v_2)) \circ \boldsymbol{t} \,-\, \omega_H(w_1, w_2) \circ \boldsymbol{s}.
\end{aligned}$$

□

Remark 3.9. the groupoid geometric data $(\Omega_V, \tilde{\Gamma}, \Omega_H)$ has a *multiplicative* nature:

- The fiberwise symplectic forms are multiplicative 2-form on the vertical groupoids $\ker \tilde{p}$.
- The Ehresmann connection $\mathrm{H} \rightrightarrows \mathrm{Hor}$ is a *multiplicative distribution*, since it is a subgroupoid of $TG \rightrightarrows TE$ over $\mathrm{Hor} \subset TE$.
- Similarly, Equation (18) indicates that Ω_H is a multiplicative 2-form. There are several ways of expressing this multiplicativity. For example, one may say that for any pair of composable arrows $(v_1, w_1), (v_2, w_2) \in \mathrm{Hor}^{(2)}$, based at the same composable arrow $(g_1, g_2) \in \mathcal{G}^{(2)}$,

  $$\Omega_H(m_{\mathrm{Hor}}(v_1, w_1), m_{\mathrm{Hor}}(v_2, w_2)) = \Omega_H(v_1, v_2) + \Omega_H(w_1, w_2).$$

  One may also say that the composition $\mathrm{Hor} \to \mathrm{Hor}^* \to T^*G$ is a groupoid morphism, where the first map is contraction by Ω_H and the second one is the inclusion coming from the splitting $TG = \mathrm{Ver} \oplus \mathrm{Hor}$.

Observing that Ω is fiber nondegenerate for both $p \circ \boldsymbol{s}$ and $p \circ \boldsymbol{t}$, we obtain:

Corollary 3.10. *For each $b \in B$, the presymplectic groupoid $(\mathcal{G}, \Omega)$ and the symplectic groupoid $(\mathcal{G}_{E_b}, \Omega_{E_b})$ are Morita equivalent presymplectic groupoids*:

$$\begin{tikzcd}
& (\mathcal{P}, \Omega_{\mathcal{P}}) \arrow[dl, "\boldsymbol{t}|_{\mathcal{P}}"'] \arrow[dr, "\boldsymbol{s}|_{\mathcal{P}}"] & \\
(\mathcal{G}, \Omega) & & (\mathcal{G}_{E_b}, \Omega_{E_b})
\end{tikzcd}$$

where $\mathcal{P} := \boldsymbol{s}^{-1}(E_b)$ and $\Omega_{\mathcal{P}} := i_{\mathcal{P}}^*\Omega$, with $i_{\mathcal{P}} : \mathcal{P} \hookrightarrow \mathcal{G}$ denoting the inclusion.

4. Integration of the Yang–Mills–Higgs phase space

In [Brahic and Fernandes 2014], we have proposed an integration procedure for a Yang–Mills–Higgs phase space. This procedure consists in forming a certain hamiltonian quotient which is hard to make sense out of for arbitrary coupling Dirac structures, since it will involve an infinite dimensional reduction. In this section, we give a different approach to integrating a Yang–Mills–Higgs phase space.

This new construction of the integration of a Yang–Mills–Higgs phase space $E = P \times_G F$ associated with a triple (P, G, F) and a choice of connection $\theta : TP \to \mathfrak{g}$ involves the following steps:

(i) Integrate the Poisson structure on the fiber (F, π_F) to a symplectic groupoid $\mathcal{F} \rightrightarrows F$.

(ii) Integrate the vertical Poisson structure Ver* to a *fibered symplectic groupoid* $\mathcal{G}_V = P \times_G \mathcal{F} \rightrightarrows E$.

(iii) Integrate the principal G-bundle $P \to B$ to the gauge groupoid $\mathcal{G}(P) \rightrightarrows B$.

(iv) Let the gauge groupoid $\mathcal{G}(P) \rightrightarrows B$ act on the fibered groupoid $\mathcal{G}_V \rightrightarrows E \to B$, yielding a semidirect product groupoid $\mathcal{G}(P) \ltimes G_V \rightrightarrows E$.

(v) Finally, integrate the Yang–Mills phase space, forming a quotient

$$\mathcal{G}(L) = \mathcal{G}(P) \ltimes \mathcal{G}_V / \mathcal{C},$$

where $\mathcal{C}$ is a certain *curvature groupoid*.

The next paragraphs describe these constructions.

4A. *Fibered symplectic groupoids.* We discuss the first two integration steps above. For this, we recall briefly from [Brahic and Fernandes 2008] a few notions about fibered symplectic groupoids.

4A1. *Fibered groupoids.* Let us fix a base B. We have the category Fib of fibrations over B, where the objects are the fibrations $p : E \to B$ and the morphisms are the fiber preserving maps over the identity.

A *fibered groupoid* is an internal groupoid in Fib, i.e., an internal category where every morphism is an isomorphism. This means that both the total space $\mathcal{G}_V$ and the base E of a fibered groupoid are fibrations over B and all structure maps are fibered maps. For instance, the source and the target maps are fiber preserving maps over the identity:

$$\begin{tikzcd} \mathcal{G}_V \arrow[r, shift left] \arrow[r, shift right] \arrow[dr] & E \arrow[d, "p"] \\ & B \end{tikzcd}$$

It follows that any orbit of $\mathcal{G}_V$ lies in a fiber of $p : E \to B$. In fact, $\mathcal{G}_V|_{E_b} := (p \circ \boldsymbol{s})^{-1}(b) = (p \circ \boldsymbol{t})^{-1}(b)$ is a Lie groupoid over E_b.

The infinitesimal version of a fibered Lie groupoid $\mathcal{G}_V \rightrightarrows E \to B$ is a *fibered Lie algebroid* $A_V \to E \to B$. This means $\pi : A_V \to E$ is a Lie algebroid, the vector bundle projection is map of fibrations:

$$
\begin{tikzcd}
A_V \arrow[r] \arrow[dr] & E \arrow[d, "p"] \\
& B
\end{tikzcd}
$$

and the image of the anchor $\sharp_V$ takes value in the vertical bundle $\mathrm{Ver} \subset TE$. There is an obvious Lie functor from fibered groupoids to fibered algebroids.

A elementary way to obtain a fibered groupoid/algebroid is by using a principal bundle whose structure group acts on a Lie groupoid/algebroid by automorphisms. Then we have the following:

Proposition 4.1. *Given a principal G-bundle P and an action $\boldsymbol{A} : G \to \mathrm{Aut}(\mathcal{F})$ of G on a Lie groupoid $\mathcal{F} \rightrightarrows F$ by Lie groupoid automorphisms, the associated bundle $\mathcal{G}_V := P \times_G \mathcal{F}$ carries a natural structure of a fibered Lie groupoid over $E := P \times_G F$. The corresponding fibered Lie algebroid is $P \times_G A(\mathcal{F}) \to E$.*

Proof. The associated bundle $P \times_G \mathcal{F}$ is given by equivalence classes $[u : a]$ of couples $(u, a) \in P \times \mathcal{F}$ under the relation $[u : a] = [ug^{-1}, \boldsymbol{A}_g(a)]$ for all $g \in G$. The source and targets map $\boldsymbol{s}, \boldsymbol{t} : P \times_G \mathcal{F} \to P \times_G F$, given by

$$\boldsymbol{s}[u : a] := [u : \boldsymbol{s}(a)] \quad \text{and} \quad \boldsymbol{t}[u : a] := [u : \boldsymbol{t}(a)],$$

are easily checked to be well defined. Then, we define a composition by setting $[u' : a'] \cdot [u : a] = [u, \boldsymbol{A}_g(a') \cdot a]$, where g is the unique element of G such that $u' = ug$. Once we check that it is independent of g, we can write

$$[u : a'] \cdot [u : a] = [u : a \cdot a'],$$

which makes it straightforward to obtain a groupoid structure $P \times_G \mathcal{F} \rightrightarrows P \times_G F$, with inverse $[u : a]^{-1} = [u : a^{-1}]$ and units $\mathbf{1}_{[u:x]} = [u : \mathbf{1}_x]$. □

Remark 4.2. As a basic observation, note that each fiber of $P \times_G \mathcal{F}$ comes naturally equipped with the structure of a Lie groupoid over the corresponding fiber of $P \times_G F$, clearly isomorphic to the model $\mathcal{F} \rightrightarrows F$.

4A2. *Poisson fibrations.* We now apply these constructions to integrate Poisson fibrations into fibered symplectic groupoids.

Definition 4.3. A *Poisson fibration* $p : E \to B$ is a locally trivial fiber bundle, with fiber type a Poisson manifold (F, π_F) and with structure group a subgroup $G \subset \mathrm{Diff}_\pi(F)$. When π is symplectic the fibration is called a *symplectic fibration*.

The fibers $E_b := p^{-1}(b)$ of a Poisson fibration come with an induced Poisson structure π_{E_b} that glue to a Poisson structure π_V on the total space of the fibration, so that $\pi_{E_b} = \pi_V|_{E_b}$.

The bivector field π_V is *vertical*; that is, it takes values in $\bigwedge^2 \mathrm{Ver} \subset \bigwedge^2 TE$. Hence, the fibers (E_b, π_{E_b}) become Poisson submanifolds of (E, π_V).

It is important to distinguish π_V as a *vertical* Poisson structure from its underlying Poisson structure on E. In particular, the Lie algebroid structure associated to π_V as a vertical Poisson structure is defined on the covertical bundle Ver^*, rather than on T^*E. The corresponding bracket and anchor are given by (13) and (12). Clearly, this is a fibered version of the usual construction, which we formalize as follows:

Definition 4.4. A *fibered symplectic groupoid* is a fibered Lie groupoid $\mathcal{G}_V$ whose fiber type is a symplectic groupoid $(\mathcal{F}, \omega)$.

Therefore, if $\mathcal{G}_V$ is a fibered symplectic groupoid over B, then $p \circ \boldsymbol{s} = p \circ \boldsymbol{t} : \mathcal{G}_V \to B$ is a symplectic fibration, and each symplectic fiber $\mathcal{G}_V|_{E_b}$ is a symplectic groupoid over the corresponding fiber E_b.

Proposition 4.5. *The base $E \to B$ of a fibered symplectic groupoid $\mathcal{G}_V \rightrightarrows E \to B$ has a natural structure of a Poisson fibration.*

Conversely, a Poisson fibration whose fiber type is an integrable Poisson manifold, integrates to a fibered symplectic groupoid. In fact, standard facts about integration of Lie algebroids yield the following (see [Brahic and Fernandes 2008] for details):

Theorem 4.6. *Let $p : E \to B$ be a Poisson fibration with fiber type (F, π_F) an integrable Poisson manifold. There exists a unique (up to isomorphism) source 1-connected fibered symplectic groupoid integrating* Ver^*.

Remark 4.7. The integration of π_V as a Poisson fibration and as a Poisson structure differ since $\mathcal{G}(\mathrm{Ver}^*)$ has only dimension $2\dim(F) + \dim(B)$.

4B. *Action groupoids.*

Next we will discuss steps (iii) and (iv) in the integration of Yang–Mills phase spaces. We describe an action of the gauge groupoid of a principal bundle on an associated fibered groupoid, and the resulting action groupoid.

4B1. *Action of a Lie groupoid on a fibered Lie groupoid.*

Given a fibered groupoid $\mathcal{G}_V \rightrightarrows E \xrightarrow{p} B$, the *gauge groupoid* is the transitive (infinite dimensional) groupoid

$$\mathrm{Gau}(\mathcal{G}_V) := \{\mathcal{G}_V|_{E_b} \xrightarrow{g} \mathcal{G}_V|_{E_{b'}} : g \text{ is a Lie groupoid isomorphism}\},$$

with source $\boldsymbol{s}(g) = b$, target $\boldsymbol{t}(g) = b'$, and with the obvious composition.

Definition 4.8. An *action of a groupoid $\mathcal{G} \rightrightarrows B$ on a fibered groupoid $\mathcal{G}_V \rightrightarrows E \to B$* is a Lie groupoid homomorphism $\Phi : \mathcal{G} \to \mathrm{Gau}(\mathcal{G}_V)$.

There is an associated semidirect *action groupoid* $\mathcal{G} \ltimes \mathcal{G}_V \rightrightarrows E$ associated to such an action, whose space of arrows is defined as

$$\mathcal{G} \ltimes \mathcal{G}_V := \mathcal{G}\, {}_{\boldsymbol{s}}\!\times_{p\circ \boldsymbol{t}} \mathcal{G}_V = \{(g, a) \in \mathcal{G} \times \mathcal{G}_V : a \in \mathcal{G}_V|_{E_{\boldsymbol{s}(g)}}\}.$$

The source and target are given by $\boldsymbol{s}(g,a) := \boldsymbol{s}(a)$ and $\boldsymbol{t}(g,a) := \boldsymbol{t}(\Phi_g(a))$, the units by $\mathbf{1}_x = (\mathbf{1}_{p(x)}, \mathbf{1}_x)$, the inverses by $(g,a)^{-1} = (g^{-1}, \Phi_g(a)^{-1})$, and the composition by

$$(g_2, a_2) \cdot (g_1, a_1) = (g_2 \cdot g_1, \Phi_{g_1}^{-1}(a_2) \cdot a_1). \tag{19}$$

When both $\mathcal{G}$ and $\mathcal{G}_V$ are Lie groupoid, we say that the action is *smooth* whenever $\mathcal{G} \ltimes \mathcal{G}_V$ is a Lie groupoid for the obvious manifold structure.

In order to define the infinitesimal counterpart of this action groupoid, for a fibered Lie algebroid $A_V \to E$, define

$$\mathrm{Der}_B(A_V) := \{D \in \mathrm{Der}(A_V) : \text{the symbol of } D \text{ is } p_*\text{-projectable}\}$$

There is a well defined map $\rho : \mathrm{Der}_B(A_V) \to \mathfrak{X}(B)$, $D \mapsto p_* X_D$, where X_D denotes the symbol of D. Note that $\mathrm{Der}_B(A_V)$ is a $C^\infty(B)$-module by the formula $(f \,.\, D)(\alpha) := f D(\alpha)$. In fact, $\mathrm{Der}_B(A_V)$ is a Lie algebra over $C^\infty(B)$ and ρ a $C^\infty(B)$-linear morphism of Lie algebras. In this work, we will always assume that A_V is locally trivial, so that ρ is surjective.

Definition 4.9. An *action of a Lie algebroid $A \to B$ on a fibered Lie algebroid* $A_V \to E \to B$ is a $C^\infty(B)$-linear Lie algebra morphism $\mathcal{D} : \Gamma(A) \to \mathrm{Der}_B(A_V)$ covering the anchor map, that is, such that $\sharp_A = \rho \circ \mathcal{D}$.

Given such an action, $A \ltimes A_V := p^* A \oplus A_V$ comes naturally with the structure of a Lie algebroid over E. The bracket and anchor are given by

$$\sharp(v, \alpha) := X_{\mathcal{D}_v} + \sharp_V(\alpha),$$
$$[v, w] := [v, w]_A, \quad [\alpha, \beta] := [\alpha, \beta]_{A_V}, \quad [v, \alpha] := \mathcal{D}_v(\alpha).$$

for any sections $\alpha, \beta \in \Gamma(A_V)$ and $v, w \in \Gamma(A)$ seen as sections of $A \ltimes A_V$. The above brackets and anchor naturally extend to arbitrary sections of $A \ltimes A_V$ since $\gamma(A)$ generates $\Gamma(p^* A)$ as a $C^\infty(E)$-module.

Definition 4.10. Given an action $\mathcal{D} : \Gamma(A) \to \mathrm{Der}_B(A_V)$ of A on A_V, we call the Lie algebroid $A \ltimes A_V$ described above the *action Lie algebroid.*

Remark 4.11. If A_V is the Lie algebroid of a fibered Lie groupoid $\mathcal{G}_V$, then $\mathrm{Der}_B(A_V)$ can be thought of as the Lie algebroid of $\mathrm{Gau}(\mathcal{G}_V)$. Indeed, any smooth action $\Phi : \mathcal{G} \to \mathrm{Gau}(\mathcal{G}_V)$ differentiates to an action $\mathcal{D} : A \to \mathrm{Der}_B(A_V)$. Moreover, $A \ltimes A_V \to E$ is the Lie algebroid of $\mathcal{G} \ltimes \mathcal{G}_V \rightrightarrows E$.

In order to integrate an infinitesimal action $\mathcal{D} : \Gamma(A) \to \mathrm{Der}_B(A_V)$, note however that we need to assume that A acts by complete lifts, meaning that the symbol of $\mathcal{D}_\alpha$ is a complete vector field on E for any $\alpha \in \Gamma(A)$.

4B2. *Actions of principal bundles on fibered Lie groupoids.* Given a principal G-bundle P, recall that its *gauge groupoid* $\mathcal{G}(P) \rightrightarrows B$ has spaces of arrows the associated bundle $P \times_G P$.

Denoting by $[u_2 : u_1]$ the equivalence class of a couple $(u_2, u_1) \in P \times P$, the source and target of $\mathcal{G}(P)$ are defined by $\boldsymbol{s}([u_2 : u_1]) := q(u_1)$, $\boldsymbol{t}([u_2 : u_1]) := q(u_2)$, and the composition is well defined by setting:

$$[w : v] \cdot [v : u] = [w : u].$$

The inverses are given by $[v : u]^{-1} = [v : u]$ and identities by $\mathbf{1}_x = [u : u]$.

The groupoid $\mathcal{G}(P) \rightrightarrows B$ is transitive, and its isotropy groups fit into a Lie group bundle $\mathrm{Iso}_P \to B$ that canonically identifies with the associated bundle $P \times_G G$ by the injection $[u : h] \mapsto [uh : u]$. Note that the same goes for neutral components, namely $\mathrm{Iso}^\circ_P = P \times_G G^\circ$.

We will be interested in the $\boldsymbol{s}$-simply connected groupoid $\widetilde{\mathcal{G}}(P)$ corresponding to $\mathcal{G}(P)$ rather than $\mathcal{G}(P)$ itself. The principal bundle corresponding to $\widetilde{\mathcal{G}}(P)$ has total space the universal cover $\widetilde{P}$ of P. When G is connected, the structure group $\overline{G}$ of $\widetilde{P}$ fits into an exact sequence:

$$1 \to \mathrm{Im}\, \partial_2 \to \widetilde{G} \to \overline{G} \to 1,$$

where $\partial_2 : \pi_2(B) \to \pi_1(G)$ is the boundary operator in the homotopy sequence of the projection $P \to B$. This means one can always assume that $\pi_1(G) = \mathrm{Im}\, \partial_2$, provided one chooses to work with $\widetilde{\mathcal{G}}(P)$ instead of $\mathcal{G}(P)$.

Finally, the Lie algebroid associated to $\mathcal{G}(P)$ is usually denoted by TP/G. It is a vector bundle over B whose sections are the G-invariant vector fields on P, and who fits in the Atiyah sequence:

$$\ker \sharp \hookrightarrow TP/G \twoheadrightarrow TB.$$

Proposition 4.12. *Let P be a principal G-bundle and $\boldsymbol{A} : G \to \mathrm{Aut}(\mathcal{F})$ an action of G on a Lie groupoid $\mathcal{F} \rightrightarrows F$ by Lie groupoid automorphisms. Then, there is a natural action of the gauge groupoid $\mathcal{G}(P)$ on the associated fibered Lie groupoid $\mathcal{G}_V := P \times_G \mathcal{F} \rightrightarrows E := P \times_G F$.*

Proof. Define $\Phi : \mathcal{G}(P) \to \mathrm{Gau}(\mathcal{G}_V)$ by $\Phi_{[u_2:u_1]}([u : a]) := [u_2 g : a]$, where g is the unique element of G such that $u_1 g = u$. After checking that Φ is well defined, notice that a more convenient formula for Φ is simply:

$$\Phi_{[v:u]}([u : a]) = [v : a] \quad \text{for } u, v \in P,\ a \in \mathcal{F}.$$

This makes it straightforward to check that Φ indeed takes values in $\mathrm{Gau}(\mathcal{G}_V)$, and that it is a groupoid morphism. □

Proposition 4.13. *With the same assumptions as in Proposition 4.12, the action groupoid $\mathcal{G} \ltimes \mathcal{G}_V \rightrightarrows E$ identifies with the quotient $(P \times P \times \mathcal{F})/G$, where G acts on $P \times P \times F$ diagonally*: $[v : u : a] = [vg : ug : \boldsymbol{A}_g^{-1}(a)]$, *for any $g \in G$. Moreover, under this identification, the structure maps are given as follows*:

- *the source and target map are*

$$\boldsymbol{s}[v : u : a] = [u : \boldsymbol{s}(a)] \quad \textit{and} \quad \boldsymbol{t}[v : u : a] = [v : \boldsymbol{t}(a)],$$

- *the unit at a point $[u : x]$, where $x \in F$, $u \in P$, is*

$$\mathbf{1}_{[u:x]} = [u : u : \mathbf{1}_x],$$

- *the inverses are*

$$[v : u : a]^{-1} = [u : v : a^{-1}],$$

- *and the composition is*

$$[w : v : a'] \cdot [v : u : a] = [w : u : a' \cdot a].$$

Proof. By the construction of the Section 4B1, an arrow in $\mathcal{G}(P) \ltimes \mathcal{G}_V$ is a couple $([u_2 : u_1], [u : a])$, where $q(u) = q(u_1)$. Since there exists a unique $g \in G$ such that $u_1 g = u$, we can always assume that $u_1 = u$ and the identification easily follows. The formulas for the structure maps then come from Proposition 4.12 and the construction of the semidirect product. □

4B3. *Action groupoid of a Poisson fibration.* Let $E = P \times_G F \to B$ be a Poisson fibration associated with a principal G-bundle $p : P \to B$ and an action of G on an integrable Poisson manifold (F, π_F) by Poisson diffeomorphisms. The results above show that one obtains an action groupoid as follows.

First, we consider the source connected symplectic groupoid $\mathcal{F} \rightrightarrows F$ integrating (F, π_F). The G-action on F by Poisson diffeomorphisms lifts to Lie groupoid action $\boldsymbol{A} : G \to \mathrm{Aut}(\mathcal{F})$ by groupoid automorphisms; see, e.g., [Fernandes et al. 2009]. Therefore, according to Propositions 4.12 and 4.13, there is a natural action of the gauge groupoid $\mathcal{G}(P) \rightrightarrows B$ on the associated fibered Lie groupoid $\mathcal{G}_V := P \times_G \mathcal{F} \rightrightarrows E \to B$, giving rise to an action Lie groupoid $\mathcal{G}(P) \ltimes \mathcal{G}_V \rightrightarrows E$.

According to the preceding discussion (see Remark 4.11), the Lie algebroid of the action Lie groupoid $\mathcal{G}(P) \ltimes \mathcal{G}_V \rightrightarrows E$ has underlying vector bundle

$$p^* A \ltimes A_V = p^*(TP/G) \ltimes \mathrm{Ver}^*.$$

To determine the bracket and the anchor, we need to find the Lie algebra homomorphism $\mathcal{D} : \Gamma(TP/G) \to \mathrm{Der}_B(\mathrm{Ver}^*)$. Since Ver^* identifies naturally with the associated bundle $\mathrm{Ver}^* = P \times_G T^*F$ and since the action of G on T^*F is naturally lifted from the G-action on F, it follows that $\mathcal{D}$ associates to each G-invariant vector field X in P the Lie derivative of the vector field $X_E \in \mathfrak{X}(E)$, induced by the natural action on T^*P/G on E. In other words, $\mathcal{D}_v$ coincides with the Lie derivative of its own symbol:

$$\mathcal{D}_X(\alpha) = \mathcal{L}_{X_E}\alpha \quad \text{for } X \in \Gamma(TP/G),\ \alpha \in \Gamma(\text{Ver}^*),$$

where X_E is the projection on $E = P \times_G F$ of the vector field $(X, 0) \in \mathfrak{X}(P \times F)$. It follows that if X, Y denote G-invariant vector fields in P and $\alpha, \beta \in \Gamma(\text{Ver}^*)$, then the anchor of $p^*(TP/G) \ltimes \text{Ver}^*$ is given by

$$\sharp(X, \alpha) := X_E + \pi_V^\sharp(\alpha), \tag{20}$$

while the bracket takes the form

$$[X, Y]_{A \ltimes A_V} := [X, Y], \quad [\alpha, \beta]_{A \ltimes A_V} := [\alpha, \beta], \quad [X, \alpha]_{A \ltimes A_V} := \mathcal{L}_{X_E}(\alpha). \tag{21}$$

4C. *Integrability of Yang–Mills–Higgs phase spaces.* We consider now the last steps in the construction of the integration of Yang–Mills–Higgs phase space. So now we assume that we have

- $p : P \to B$ a principal G-bundle;
- (F, π_F) a Poisson manifold;
- $G \times F \to F$ a hamiltonian G-action on (F, π_F) with equivariant moment map $J_F : F \to \mathfrak{g}^*$.

Each choice of a principal connection $\theta : TP \to \mathfrak{g}$ yields a coupling Dirac structure on $E = P \times_G F$.

The fact that the action is hamiltonian implies that the G action on the algebroid T^*F is prehamiltonian, with premoment map (see):

$$\begin{aligned} \psi : \mathfrak{g} \ltimes F &\longrightarrow T^*F \\ (\xi, m) &\longmapsto \mathrm{d}_m\langle J, \xi\rangle. \end{aligned}$$

Therefore, by Theorem A.19, ψ integrates to a groupoid morphism

$$\Psi : G^\circ \ltimes F \to \mathcal{F},$$

where

$$\mathcal{F} := \Sigma(F)\,/\,\tilde{\Psi}(\pi_1(G) \ltimes F).$$

We will assume that $\tilde{\Psi}(\pi_1(G) \times F)$ is embedded in $\Sigma(F)$, so that $\mathcal{F}$ is smooth. Clearly, $\mathcal{G}_V := P \times_G \mathcal{F}$ is a symplectic groupoid integrating $\text{Ver}^* = P \times_G T^*F$.

The G-action on F lifts to a Lie groupoid action $\boldsymbol{A} : G \to \text{Aut}(\mathcal{F})$, so we can apply the construction of the previous subsection: we obtain an action groupoid $\mathcal{G}(P) \ltimes \mathcal{G}_V \rightrightarrows E$.

Definition 4.14. The *curvature subgroupoid*, denoted by $\mathcal{C} \rightrightarrows E$, is the subgroupoid $\mathcal{C} \subset \mathcal{G}(P) \ltimes \mathcal{G}_V$ given by:

$$\mathcal{C} := \mathrm{Graph}(\Psi_P \circ i) \subset \mathcal{G}(P) \ltimes \mathcal{G}_V,$$

where $\Psi_P : \mathrm{Iso}^\circ_P \times_B E \to \mathcal{G}_V$ is obtained by fibrating $\Psi : G^\circ \ltimes F \to \mathcal{F}$ along P and $i : \mathrm{Iso}^\circ \to \mathrm{Iso}^\circ$ denotes the inversion.

More explicitly, with the notations of Proposition 4.13, the curvature groupoid $\mathcal{C}$ is given by

$$\mathcal{C} := \{[uh^{-1} : u : \Psi(h, x)] \in \mathcal{G}(P) \ltimes \mathcal{G}_V \; : \; u \in P,\; h \in G^\circ,\; x \in F\}. \tag{22}$$

Proposition 4.15. *The curvature groupoid $\mathcal{C} \rightrightarrows E$ is a wide, normal, completely intransitive subgroupoid of $\mathcal{G}(P) \ltimes \mathcal{G}_V$.*

Proof. The result follows using the expression (22) for $\mathcal{C}$, the compositions rules in Proposition 4.13 and Equation (32) in Theorem A.19. The fact that $\mathcal{C}$ is a subgroupoid is rather straightforward. In order to see that it is normal, we pick any $[uh^{-1} : u : \Psi(h, x)] \in \mathcal{C}$ and $[v : u : a] \in \mathcal{G}(P) \ltimes \mathcal{G}_V$ which are composable, i.e., such that $x = \boldsymbol{s}(a)$, and we find that

$$[v : u : a] \cdot [uh^{-1} : u : \Psi(h, x)] \cdot [v : u : a]^{-1} = [vh^{-1} : v : \Psi(h, x)],$$

is an element in $\mathcal{C}$. □

Finally, putting all together, we conclude the following:

Theorem 4.16. *Suppose that (P, G, F) is a classical Yang–Mills–Higgs setting and $\theta : TP \to \mathfrak{g}$ is a principal connection. Let L be the corresponding coupling Dirac structure on $E = P \times_G F$ and assume that*

(i) *the Poisson manifold (F, π_F) is integrable, and*

(ii) *the groupoid $\tilde{\Psi}(\pi_1(G) \times F)$ is embedded in $\Sigma(F)$.*

Then, the quotient groupoid $\mathcal{G}(P) \ltimes \mathcal{G}_V / \mathcal{C}$ integrates (E, L).

Proof. As we saw above, the Lie algebroid of $\mathcal{G}(P) \ltimes \mathcal{G}_V$ is given by $A \ltimes A_V$, where $A = TP/G$ and $A_V = \mathrm{Ver}^*$. Furthermore, the principal connection induces a splitting of the Atiyah sequence, and we have an identification $TP/G \simeq TB \oplus \ker \sharp_A$. With this identification, the Lie algebroid $A_{\mathcal{C}}$ of $\mathcal{C}$ lies in

$$TP/G \ltimes \mathrm{Ver}^* \simeq (TB \oplus \ker \sharp_A) \ltimes \mathrm{Ver}^*$$

as

$$A_{\mathcal{C}} = \left\{(0, \xi, -\psi(\xi)) \in (TB \oplus \ker \sharp_A) \ltimes \mathrm{Ver}^* : \xi \in \ker \sharp_A\right\},$$

and the quotient $(TB \oplus \ker \sharp_A) \ltimes \mathrm{Ver}^* / A_{\mathcal{C}}$ identifies with $TB \times_B \mathrm{Ver}^*$, with canonical projection given by $\pi(X, \xi, \alpha) = (X, \alpha + \psi(\xi))$. It now follows from expressions

(20) and (21) for the anchor and the brackets that the Lie algebroid structure on $A \ltimes \mathrm{Ver}^*$ descends to a Lie algebroid structure on $TB \times_B A_V$ whose brackets and anchor are the same as those given in Proposition 2.10. Hence $A \ltimes \mathrm{Ver}^*/A_{\mathcal{C}}$ and L are isomorphic as Lie algebroids.

For the smoothness of the quotient, we observe that $\mathcal{G}(P) \ltimes \mathcal{G}_V/\mathcal{C}$ can also be thought of as an "associated bundle" $\mathcal{G}(P) \ltimes_{\mathrm{Iso}^\circ_P} \mathcal{G}_V$. Indeed, there is an action of the *bundle of* Lie groups Iso°_P on $\mathcal{G}(P) \ltimes \mathcal{G}_V$ which can be described as follows. On the one hand, the Lie groupoid morphism Ψ induces an action λ of G° on $\mathcal{F}$ by left multiplication: $\lambda_h(a) := \Psi(h, x) \cdot a$, where $h \in G^\circ$ and $x := \mathbf{s}(a)$. Fibering along P, we obtain an action of the bundle of Lie groups Iso°_P on $\mathcal{G}_V$:

$$\lambda^P_{[u:h]}([u : a]) := [u : \Psi(h, x) \cdot a].$$

Here we use the identification $\mathrm{Iso}^\circ_P \simeq P \times_G G^\circ$ to write an element of Iso°_P as a pair $[u : h]$. Note that this action is well defined by (33). On the other hand, Iso°_P acts on $\mathcal{G}(P)$ by right multiplication, which is a proper and free action. The two actions together give a proper and free action of Iso°_P on $\mathcal{G}(P) \ltimes \mathcal{G}_V$:

$$g \cdot (b, a) := \big(bg^{-1} : \lambda^P_g(a)\big).$$

and the quotient is the "associated bundle" $\mathcal{G}(P) \ltimes_{\mathrm{Iso}^\circ_P} \mathcal{G}_V$.

We claim that $\mathcal{G}(P) \ltimes_{\mathrm{Iso}^\circ_P} \mathcal{G}_V$ can be identified with $\mathcal{G}(P) \ltimes \mathcal{G}_V/\mathcal{C}$. This follows by observing that any $g \in \mathrm{Iso}^\circ_P$ can be written as $g = [u : uh] \in \mathcal{G}(P)$ so that (see Theorem A.19):

$$g \cdot (b, a) = \big(bg^{-1}, \lambda^P_g(a)\big) = (b, a) \cdot c$$

where $c := ([uh^{-1} : u : \Psi(h)]) \in \mathcal{C}$. Since the assignment $c \leftrightarrow g$ is one-to-one, the two quotients coincide. $\square$

Remark 4.17. Consider the Hopf fibration $P = \mathbb{S}^3 \to \mathbb{S}^2$, seen as an $\mathbb{S}^1$-principal bundle, and $F = \mathbb{R}$ acted upon trivially with momentum map $f : F \to \mathbb{R}$ any smooth function, as in Example 2.7. Then, the second condition in Theorem 4.16 fails if f has a critical point as explained in Example A.20.

Theorem 4.16 shows that the groupoid structure of $\mathcal{G}(L)$ does not depend on the choice of the principal G-bundle connection. In other words, two coupling Dirac structures associated with Yang–Mills data with the same principal G-bundle and hamiltonian G-action, but different principle bundle connections, give rise to the same Lie groupoid. Note, however, that the presymplectic forms will be distinct, as it is clear from their geometric data given in Proposition 3.8.

We can also give an explicit description of the presymplectic form Ω on $G(L)$, as follows. First, we use Proposition 4.13 to identify $\mathcal{G}(P) \ltimes \mathcal{G}_V \simeq (P \times P \times \mathcal{F})/G$. We then construct a presymplectic form $\widetilde{\Omega}$ on $\mathcal{G}(P) \ltimes \mathcal{G}_V$: we have a closed 2-form on $P \times P \times \mathcal{F}$ given by

$$\widetilde{\Omega} := p_{\mathcal{F}}^* \Omega_{\mathcal{F}} + \mathrm{d}\langle \theta, \bar{\mu} \rangle_1 - \mathrm{d}\langle \theta, \bar{\mu} \rangle_2,$$

where $p_{\mathcal{F}} : P \times P \times \mathcal{F} \to \mathcal{F}$ is the projection, and $\mathrm{d}\langle \theta, \bar{\mu} \rangle_i$ denotes the closed 2-form in $\Omega^2(P \times P \times \mathcal{F})$ obtained by differentiating the 1-form $\alpha_i \in \Omega^1(P \times P \times \mathcal{F})$ given by

$$\alpha_i|_{(u_1,u_2,g)}(v_1, v_2, w) := \langle \theta|_{u_i}(v_i), \bar{\mu}(g) \rangle.$$

Here, $\bar{\mu} : \mathcal{F} \to \mathfrak{g}^*$ denotes the moment map for the lifted G-action on $\mathcal{F}$, so that $\bar{\mu} = \mu \circ \boldsymbol{t} - \mu \circ \boldsymbol{s}$. One checks easily that the closed 2-form $\widetilde{\Omega}$ is basic for the G-action on $P \times P \times \mathcal{F}$, so it descends to a multiplicative 2-form in the quotient $(P \times P \times \mathcal{F})/G \simeq \mathcal{G}(P) \ltimes \mathcal{G}_V$.

Finally, one checks that resulting multiplicative 2-form on $\mathcal{G}(P) \ltimes \mathcal{G}_V$ further descends to the quotient $\mathcal{G}(P) \ltimes \mathcal{G}_V/\mathcal{C}$, giving a closed, multiplicative 2-form $\Omega_{\mathcal{G}}$ satisfying the nondegeneracy condition (15). A more-or-less tedious computation shows that the target map $\boldsymbol{t} : (\mathcal{G}(L), \Omega_{\mathcal{G}}) \to (E, L)$ is a forward Dirac map. Summarizing this discussion, we have:

Corollary 4.18. *Under the conditions of Theorem 4.16, the presymplectic form on the groupoid $G(L) = \mathcal{G}(P) \ltimes \mathcal{G}_V / \mathcal{C}$ is the quotient of the closed 2-form*

$$\widetilde{\Omega} := p_{\mathcal{F}}^* \Omega_{\mathcal{F}} + \mathrm{d}\langle \theta, \bar{\mu} \rangle_1 - \mathrm{d}\langle \theta, \bar{\mu} \rangle_2.$$

The integrability conditions in Theorem 4.16 can be made more explicit. On the one hand, the integrability of the fiber type (F, π_F) follows from general theory developed in [Crainic and Fernandes 2004] and can be expressed in terms of monodromy maps $\partial : \pi_2(S, x) \to \mathcal{G}(\mathfrak{g}_x)$, where S is the symplectic leaf of F through x and $\mathfrak{g}_x = \ker \pi_F^\sharp|_x$ is the isotropy Lie algebra at x. On the other hand, condition (ii) can be treated by the same methods as in [Brahic and Fernandes 2014, Section 4.3], and one gets another monodromy type map $\pi_1(G) \to \mathcal{F}_m$ controlling (ii). This will be treated elsewhere.

5. Integration of coupling Dirac structures II

A general coupling Dirac structure may not come from a principal bundle with structure group a finite dimensional Lie group. For instance, this is the case if the holonomy group induced by the connection (i.e., the group spanned by the holonomy along loops in the base) is not a finite dimensional subgroup of the Poisson automorphisms of the fiber. In such cases, one needs a formulation of the construction of Section 4C which avoids infinite dimensional reductions. In this section, we will take advantage of the fact that L fits into a Lie algebroid extension, to reformulate the construction given in Section 4C, without any mentioning to these infinite dimensional group quotients. We follow the ideas of [Brahic 2010] in order to describe L-paths and L-homotopies.

Recalling that $\mathrm{Ver}^* = \mathrm{Graph}(\pi_V)$, we know that L is a Lie algebroid extension

$$\mathrm{Ver}^* \hookrightarrow L \twoheadrightarrow TB,$$

which splits. The notion of holonomy makes sense for any Lie algebroid extension with a splitting (see [op. cit., Section 2.1]), and in our situation, given a TB-path $\dot{\gamma}_B \in P(TB)$, the holonomy is a Lie algebroid morphism

$$\Phi_{\gamma_B} : \mathrm{Ver}^*\big|_{E_{\gamma_B(0)}} \to \mathrm{Ver}^*\big|_{E_{\gamma_B(1)}}.$$

It will be useful to restrict γ_B to a path $[0, t] \to TB$, where $t \in [0, 1]$. The corresponding holonomy will then be

$$\Phi^{\gamma_B}_{t,0} : \mathrm{Ver}^*\big|_{E_{\gamma_B(0)}} \to \mathrm{Ver}^*\big|_{E_{\gamma_B(t)}}.$$

In the case of a coupling Dirac structure, there is another notion of holonomy to be taken into account, namely, the one induced by the *usual* Ehresmann connection Hor. Given a path $\gamma_B : [0, 1] \to B$ it gives rise to a holonomy map

$$\phi^{\gamma_B} : E_{\gamma_B(0)} \to E_{\gamma_B(1)}.$$

Again, restricting γ_B to a path $[0, t] \to B$ the corresponding holonomy will be denoted

$$\phi^{\gamma_B}_{t,0} : E_{\gamma_B(0)} \to E_{\gamma_B(t)}.$$

The two holonomies are related in a simple way:

Proposition 5.1. *The holonomy $\Phi^{\gamma_B}_{t,0}$ induced by the connection* Graph (ω_H) *on L is related to the holonomy $\phi^{\gamma_B}_{t,0}$ induced by* Hor *on TE by*

$$\Phi^{\gamma_B}_{t,0} = (\phi^{\gamma_B}_{0,t})^*.$$

Proof. The result follows directly from the identification $\mathrm{Ver}^* = \mathrm{Graph}(\pi_V)$ and from the particular form of the bracket given in Proposition 2.10. □

Recall that a Lie algebroid extension is called a *fibration* whenever the Ehresmann connection is complete [Brahic and Zhu 2011]. It follows from Proposition 5.1 that (10) is a fibration whenever the Ehresmann connection Hor is complete. In the sequel, we will always assume that this is the case.

5A. ***Splitting L-paths and L-homotopies.*** We see from Proposition 2.9 that any L-path a over $\gamma := p_L \circ a$ decomposes uniquely as a sum:

$$a(t) = h^*(\dot{\gamma}_B(t))_{\gamma(t)} + a_V(t) \tag{23}$$

where $\gamma_B := p \circ \gamma$. In this decomposition, neither $t \mapsto h^*(\dot{\gamma}_B(t))_{\gamma(t)}$, nor $t \mapsto a_V(t)$ is an L-path in general. However, it is possible to "split" the paths in $P(L)$ into horizontal and vertical parts, as follows:

Proposition 5.2 (splitting L-paths). *Let L be a coupling Dirac structure on a fibration $p : E \to B$. If the associated connection Γ is complete, then there is an isomorphism of Banach manifolds*:

$$\begin{array}{ccc} P(L) & \longrightarrow & P(TB)\ {}_{s}\times_{t\circ p} P(\mathrm{Ver}^*), \\ a & \longmapsto & (\dot{\gamma}_B,\, \tilde{a}), \end{array}$$

where the couple $(\dot{\gamma}_B, \tilde{a})$ is defined by

$$\dot{\gamma}_B := \mathrm{d}p \circ \sharp a, \quad \tilde{a}_t := a_V(t) \circ \mathrm{d}\phi^{\gamma_B}_{t,0}, \tag{24}$$

where $\phi^{\gamma_B}_{t,0} : E_{\gamma_B(0)} \to E_{\gamma_B(t)}$ denotes the holonomy along γ_B.

Proof. This follows from [Brahic 2010, Proposition 4.1] and Proposition 5.1. □

One should think of the couple $(\dot{\gamma}_B, \tilde{a})$ as a concatenation of L-paths of the form $h^*(\dot{\gamma}_B) \cdot \tilde{a}$. Here, $h^*(\dot{\gamma}_B)$ denotes the L-path defined by

$$h^*(\dot{\gamma}_B)(t) := h^*(\dot{\gamma}_B(t))_{\phi^{\gamma_B}_{t,1}(y)}, \tag{25}$$

where $y = s(a)$. Notice that the L-path (25) is different from the horizontal component appearing in (23) since the base paths are different. In particular, $h^*(\dot{\gamma}_B)$ as defined in (25) is *always* an L-path by construction.

Then Proposition 5.2 can be illustrated in a simple way as follows:

$$\begin{tikzcd}
\phi^{-1}_{\gamma_B}(y) \arrow[rr, bend left, "h^*(\dot{\gamma}_B)"] & & y & L \arrow[dd] \\
x \arrow[u, "\tilde{a}"] \arrow[urr, "a"'] & & & \\
\gamma_B(0) \arrow[rr, bend left, "\dot{\gamma}_B"] & & \gamma_B(1) & TB
\end{tikzcd}$$

In fact, it can be proved that a is L-homotopic to the concatenation $h^*(\dot{\gamma}_B) \cdot \tilde{a}$. However, for the sake of simplicity, in this work we shall simply think of the map $a \mapsto (\tilde{a}, \dot{\gamma}_B)$ as an mere identification.

Recall that for any A-path a, its *inverse path* is the A-path a^{-1} defined by $a^{-1}(t) := -a(1-t)$. Using Proposition 5.2, one can express the concatenation and inverses of L-paths as follows:

Proposition 5.3. *Under the isomorphism of Proposition 5.2, given two composable L-paths $a \simeq (\dot{\gamma}_B, \tilde{a})$ and $b \simeq (\dot{\delta}_B, \tilde{b})$, their concatenation is*

$$(\dot{\delta}_B, \tilde{b}) \cdot (\dot{\gamma}_B, \tilde{a}) := \big(\dot{\delta}_B \cdot \dot{\gamma}_B,\ \Phi^{-1}_{\gamma_B}(\tilde{b}) \cdot \tilde{a}\big).$$

Moreover, the inverse path a^{-1} of a is

$$(\dot{\gamma}_B, \tilde{a})^{-1} := \big(\dot{\gamma}_B^{-1},\ \Phi_{\gamma_B}(\tilde{a})^{-1}\big).$$

Proof. The result follows directly from (24) and from the fact that the holonomy commutes with taking concatenation and inverse of A-paths. □

Notice the analogy between the formula for concatenation in the previous proposition and formula (19) for the product in the action groupoid. In fact, if one thinks of $P(TB)$ as a groupoid over B, then the holonomy gives an action of $P(TB)$ on $P(\mathrm{Ver}^*)$ similar to the action of $\mathcal{G}(P)$ on $\mathcal{G}_V$ discussed in Section 4B1. For this reason, one may think of the fibered product $P(TB)\,{}_{s}\times_{t\circ p} P(\mathrm{Ver}^*)$ as a semidirect product $P(TB) \ltimes P(\mathrm{Ver}^*)$.

In general, the presence of curvature prevents the fundamental groupoid $\Pi(B)$ from acting on $P(\mathrm{Ver}^*)$. However, holonomy along a path $\gamma_B \in P(B)$ is a Lie algebroid morphism $\Phi_{\gamma_B} : \mathrm{Ver}^*|_{E_{\gamma_B(0)}} \to \mathrm{Ver}^*|_{E_{\gamma_B(1)}}$. Hence, it integrates to a groupoid morphism $\Phi_{\gamma_B} : \mathcal{G}(\mathrm{Ver}^*)|_{E_{\gamma_B(0)}} \to \mathcal{G}(\mathrm{Ver}^*)|_{E_{\gamma_B(1)}}$ that we still denote by Φ_{γ_B}. Here, $\mathcal{G}(\mathrm{Ver}^*)$ denotes the Weinstein groupoid of Ver^*. Finally, notice that the formulas in Proposition 5.3 still make sense when replacing Ver^*-paths by their homotopy classes; therefore, we will denote by $P(TB) \ltimes \mathcal{G}(\mathrm{Ver}^*)$ the fibered product $P(TB)\,{}_{s}\times_{t\circ p} \mathcal{G}(\mathrm{Ver}^*)$.

Theorem 5.4. *Suppose that L is a coupling Dirac structure on $E \to B$. The source 1-connected groupoid $\mathcal{G}(L)$ integrating L naturally identifies with equivalence classes in $P(TB) \ltimes_B \mathcal{G}(\mathrm{Ver}^*)$ under the following relation:*

- $(\gamma_0, g_0) \sim (\gamma_1, g_1)$ *if and only if there exists a homotopy $\gamma_B : I \times I \to B$, $(t, \epsilon) \mapsto \gamma_B^\epsilon(t)$ between γ_0 and γ_1, such that $g_1 = \partial(\gamma_B, \boldsymbol{t}(g_0)) \,.\, g_0$.*

Here, $\partial(\gamma_B, x_0)$ is the element in $\mathcal{G}(\mathrm{Ver}^*)$ represented by the Ver^*-path

$$\epsilon \longmapsto (\mathrm{d}_V)_{\tilde{\gamma}^\epsilon}\left(\int_0^1 (\phi^{\gamma_B^\epsilon}_{s,0})^* \omega_H(\gamma_B)_{s,\epsilon}\,\mathrm{d}s\right) \in \mathrm{Ver}^*_{\tilde{\gamma}(\epsilon)}, \tag{26}$$

where $\tilde{\gamma}(\epsilon) := \Phi^{-1}_{\gamma_B^\epsilon} \circ \Phi_{\gamma_B^0}(x_0)$ and

$$\omega_H(\gamma_B)_{s,\epsilon} := \omega_H\left(h\Big(\frac{\mathrm{d}\gamma_B}{\mathrm{d}t}(s,\epsilon)\Big),\ h\Big(\frac{\mathrm{d}\gamma_B}{\mathrm{d}\epsilon}(s,\epsilon)\Big)\right) \in C^\infty(E_{\gamma_B(s,\epsilon)})$$

One may illustrate the homotopy condition appearing in Theorem 5.4 in the following way:

$$\begin{tikzcd}
x_1 \arrow[rr, "h^*(\gamma_1)"] & & y \\
& x_0 \arrow[ul, dotted, "\partial(\gamma_B, x_0)"'] \arrow[ur, "h^*(\gamma_0)"'] & \\
& x \arrow[uul, bend left, "g_1"] \arrow[u, "g_0"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
b_0 \arrow[rr, bend left, "\gamma_1"] \arrow[rr, bend right, "\gamma_0"'] \arrow[rr, phantom, "\Uparrow \gamma_B"] & & b_1
\end{tikzcd}$$

Example 5.5. Let us consider the case where π_V is the trivial Poisson structure. This occurs, for instance, if L is the restriction of a regular Dirac structure to a tubular neighborhood $E \to B$ of one of its leaves B. Then $\mathcal{G}(\mathrm{Ver}^*)$ is a bundle of Lie groups that identifies with Ver^* with its additive structure. Furthermore, it follows from the curvature identity (9) that the connection is flat, so we have a genuine action of the fundamental groupoid $\Pi(B)$ on Ver^*. Up to a cover of B, we may assume that E is trivial as a representation of $\Pi(B)$. This means that E can be identified with $B \times F$ in such a way that the holonomy along any path is the identity:

$$\phi^{\gamma}_{s,0} = \mathrm{id}_F : \{\gamma(0)\} \times F \longrightarrow \{\gamma(1)\} \times F.$$

It follows that the horizontal and vertical distributions are respectively given by $\mathrm{Hor} = TB \times F$ and $\mathrm{Ver} = B \times TF$ in the decomposition $TE = TB \oplus TF$. Hence, ω_H can be seen as a family of 2-forms on B parametrized by F, and the leaves of L are of the form $B \times \{x\}$ with presymplectic form $\omega_H|_{B\times\{x\}}$, where $x \in F$.

The homotopy condition appearing in Theorem 5.4 can then be expressed as follows: two elements (γ_0, g_0) and (γ_1, g_1) in $P(TB) \times T^*_{x_0}F$ are homotopic if and only if there exists a TB-homotopy $\gamma_B : I^2 \mapsto B$ between γ_0 and γ_1 such that

$$g_1 - g_0 = (\mathrm{d}_V)_{x_0} \int_{\gamma_B} \omega_H,$$

where we integrate ω_H along γ_B as a 2-form with values in $C^\infty(F)$. In order to obtain the above formula, we simply replace $\phi^{\gamma_B^\epsilon}_{0,s}$ by id_F in (26) and then we use the fact that, T^*F being a bundle of abelian groups, any path in $T^*_{x_0}F$ can be represented by a constant paths. This amounts in (26) to average with respect to the ϵ variable. In particular, when $F = \mathbb{R}$, we recover the leafwise prequantization Lie algebroids and the homotopy condition appearing in [Crainic 2004].

Notice also that, applying the resulting 1-form to a vector $X_{x_0} \in TF$, one gets a geometrical interpretation of $g_1 - g_0$ as the variation of the presymplectic area of γ_B in the vertical directions:

$$\Big\langle \mathrm{d}_V \int_{\gamma_B} \omega_H,\ X_{x_0} \Big\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \int_{\gamma_B \times \{\phi_t^X(x_0)\}} \omega_H,$$

where ϕ_t^X is the flow of any vector field $X \in \mathfrak{X}(F)$ extending X_{x_0}.

Remark 5.6. The construction given in Theorem 5.4 can be interpreted as an infinite dimensional analogue of the construction given in Section 4 of the groupoid integrating a Yang–Mills phase space.

For this interpretation, one considers the Poisson frame bundle (see [Brahic and Fernandes 2008]), so that we can view our coupling as an infinite dimensional Yang–Mills phase space. One needs first to reduce the structure group from the group of Poisson diffeomorphisms between a fixed fiber E_{b_0} and any other fiber to the subgroup generated by the holonomy transformations Φ_{γ_B} along any path $\gamma_B \in P(B)$, with $\gamma_B(0) = b_0$. If $P \to B$ denotes the resulting principal bundle, then one can "identify" the corresponding gauge groupoid $\mathrm{Gau}(P) = P \times_B P$ with the "groupoid" $P(TB)$. Moreover, the equivalence relation $\sim$ of Theorem 5.4 can be viewed as the equivalence relation associated with the corresponding curvature groupoid.

5B. ***The monodromy groupoid.*** We now use the constructions of [Brahic 2010; Crainic and Fernandes 2004] in order to obtain the obstructions to integrability of a coupling Dirac structure L.

Consider the short exact sequence of Lie algebroids

$$\mathrm{Ver}^* \hookrightarrow L \twoheadrightarrow TB. \tag{27}$$

We obtain by integration the sequence of groupoid morphisms

$$\mathcal{G}(\mathrm{Ver}^*) \xrightarrow{j} \mathcal{G}(L) \xrightarrow{q} \Pi(B), \tag{28}$$

where $\Pi(B)$ denotes the fundamental groupoid of B. Recall that j and q are defined at the level of paths:

$$j([\tilde{a}]_V) := [i \circ a]_L \quad \text{and} \quad q([a]_L) := [p_* \circ \sharp(a)]_{TB}.$$

for any Ver^*-path $\tilde{a} : I \to \mathrm{Ver}^*$ and any L-path $a : I \to L$. Although the sequence (27) is exact, the sequence (28) might not be exact anymore, pointing out a lack of exactness of the integration functor. However, one can always ensure the right exactness:

Proposition 5.7. *Let L be a coupling Dirac structure whose induced Ehresmann connection is complete. Then, the sequence (28) is surjective at $\Pi(B)$ and exact at $\mathcal{G}(L)$.*

Proof. One may see that q is surjective, provided the connection is complete, by observing that, given $[\dot{\gamma}_B]_{TB} \in \Pi(B)$, the element $[h^*(\dot{\gamma}_B)] \in \mathcal{G}(L)$, defined by (25) maps to $[\dot{\gamma}_B]$.

For the exactness at $\mathcal{G}(L)$ one observes that, by the definition, the elements of $\ker q$ are represented by L-paths whose projection on TB is a contractile loop. Therefore, the inclusion $\operatorname{Im} j \subset \ker q$ is obvious. Conversely, given an element $[a]_L \in \ker\ q$, represented by some L-path a, we see that $a \sim (\tilde{a}, \dot{\gamma}_B)$, under the identifications of Proposition 5.2, where γ_B is a contractible loop based at some $b \in B$. Consider a contraction $\gamma_B^\epsilon : I^2 \to B$ between γ_B and the trivial path 0_b. Then, by Theorem 5.4, we see that $(\tilde{a}, \dot{\gamma}_B)$ is L-homotopic to $(\partial(\gamma_B) \cdot \tilde{a}, 0_b)$. Since $(\partial(\gamma_B) \cdot \tilde{a}, 0_b)$ represents a Ver*-path, we conclude that $[a]_L \in \operatorname{Im} j$, as claimed. □

It follows that (28) can only fail to be exact because of the lack of injectivity of j. In order to measure this failure, we introduce the following:

Definition 5.8. The *monodromy groupoid* associated with the fibration is the kernel of $j : \mathcal{G}(\mathrm{Ver}^*) \to \mathcal{G}(L)$, denoted by $\mathcal{M}$.

Obviously, by construction, we have an exact sequence of groupoids

$$\mathcal{M} \hookrightarrow \mathcal{G}(\mathrm{Ver}^*) \twoheadrightarrow \ker q,$$

and we can replace (28) by the exact sequence of groupoids:

$$\mathcal{G}(\mathrm{Ver}^*)/\mathcal{M} \hookrightarrow \mathcal{G}(\mathrm{Ver}^*) \twoheadrightarrow \Pi(B).$$

The kernel of this sequence $\ker q = \mathcal{G}(\mathrm{Ver}^*)/\mathcal{M}$ is a bundle of groupoids with typical fiber the neutral component of the restricted groupoid $\mathcal{G}(L)|_{E_{b_0}}$ to a fiber E_{b_0}. In particular, we see that if $\mathcal{G}(L)$ is integrable, then the monodromy groupoid $\mathcal{M}$ must be embedded in $\mathcal{G}(\mathrm{Ver}^*)$.

It remains to relate $\mathcal{M}$ to the global data associated with L on E.

Theorem 5.9. *Consider a coupling Dirac structure L on a fibration $E \to B$, and assume that the induced Eheresmann connection is complete. Then there exists a homomorphism*

$$\partial : \pi_2(B) \times_B E \to \mathcal{G}(\mathrm{Ver}^*),$$

that makes the following sequence exact:

$$\cdots \to \pi_2(B) \times_B E \to \mathcal{G}(\mathrm{Ver}^*) \to \mathcal{G}(L) \to \Pi(B).$$

In other words, Theorem 5.9 states that the monodromy groupoids of the fibration coincide with the image of the transgression map $\mathcal{M} = \operatorname{Im} \partial|_{\pi_2(B)}$.

Proof. The map ∂ in Theorem 5.4, when restricted to a sphere in B based at some $b \in B$ (seen as a map $\gamma_B : I^2 \to B$ such that $\gamma_B(\partial I^2) = \{b\}$) is independent of its homotopy class; see [Brahic 2010]. Then, it follows from that reference and [Brahic and Zhu 2011] that the restriction of the map ∂ to $\pi_2(B)$ corresponds precisely to the transgression map. □

Note the analogy between the monodromy groupoid described above and the monodromy groups that measure the integrability of an algebroid [Crainic and Fernandes 2003]. In fact, when E is a tubular neighborhood of a leaf $B \subset E$ in a Dirac structure, the restriction $\mathcal{M}|_B$ coincides, by construction, with the usual monodromy groups along B.

Finally, we can relate the monodromy groupoid of a coupling Dirac structure with the problem of integrability.

Theorem 5.10. *Let L be a coupling Dirac structure on $E \to B$ and assume that the associated connection Γ is complete. Then, L is an integrable Lie algebroid if and only if the following conditions hold*:

(i) *the typical Poisson fiber $(E_x, \pi_V|_{E_x})$ is integrable*;

(ii) *the injection $\mathcal{M} \hookrightarrow \mathcal{G}(\mathrm{Ver}^*)$ is an embedding.*

Proof. First, it is easily seen that since the associated Poisson fibration is locally trivial, Ver^* is integrable if and only if the typical Poisson fiber $(E_x, \pi_V|_{E_x})$ is integrable.

Assume now that L is integrable. Then, the projection $q : \mathcal{G}(E) \to \Pi(B)$ is a smooth surjective submersion. Therefore, $\ker q$ is a Lie groupoid integrating Ver^*; in particular, the typical Poisson fiber is integrable. Furthermore, since $\ker q = \mathcal{G}(\mathrm{Ver}^*)/\mathcal{M}$ is smooth, $\mathcal{M}$ is necessarily embedded in $\mathcal{G}(\mathrm{Ver}^*)$.

Conversely, suppose that $\mathcal{M}$ is embedded in $\mathcal{G}(\mathrm{Ver}^*)$ and consider a sequence $(\xi_n) \subset \mathcal{N}(L)$ of monodromy elements of L converging to a trivial path 0_x. Since $\ker \sharp \subset \mathrm{Ver}^*$, one can consider the sequence $[\xi_n]_V \in \mathcal{G}(\mathrm{Ver}^*)$, where ξ_n is considered as a constant path. By the definition [Crainic and Fernandes 2004] of the monodromy groups $\mathcal{N}(L)$ controlling the integrability of L, $[\xi_n]_L \in \mathcal{G}(L)$ is a sequence of units $[\xi_n]_L = \mathbf{1}_{x_n}$, therefore $[\xi_n]_V \in \mathcal{M}$. In other words, there exists a neighborhood U of the identity section in $\mathcal{G}(L)$ such that $\mathcal{N}(L) \cap U \subset \mathcal{M} \cap U$. Since $\mathcal{M}$ is embedded in $\mathcal{G}(\mathrm{Ver}^*)$, it follows that there exists a neighborhood $V \subset U$ of the identity section in $\mathcal{G}(L)$ such that $\mathcal{N}(L) \cap V$ coincides with the identity section. This shows that the obstructions to integrability of L vanish. □

Example 5.11 (hamiltonian symplectic fibrations). Assume L that is the graph of a presymplectic form. Then L identifies with TE as a Lie algebroid (using the anchor map). In particular L is integrable and $\mathcal{G}(L)$ identifies with the fundamental groupoid of E. Let us see how one can recover this using the previous construction.

In that case, π_V is the inverse of a symplectic (vertical) form. Thus, Ver^* identifies with Ver as a Lie algebroid, and $\mathcal{G}(\mathrm{Ver}^*) \simeq \mathcal{G}(\mathrm{Ver})$, which is just a fibered version of the fundamental groupoid. Therefore, the transgression map becomes $\partial : \pi_2(B) \ltimes E \to \mathcal{G}(\mathrm{Ver})$ and, as easily checked, corresponds to the usual transgression map in the homotopy long exact sequence associated to the fibration $E \to B$. It follows that $\mathcal{M}_x$ lies in the fundamental group $\pi_1(E_{p(x)})$ of the fiber through $x \in E$, and $\mathcal{M}$ is locally trivial over E. Now, Theorem 5.10 shows that L is integrable.

In fact, if the fibers are compact, one can even show that the transgression map vanishes. Indeed, given a sphere in B, it follows from (26) that the loops representing the image of the transgression map are the so-called hamiltonian loops; see [McDuff and Salamon 1998]. For compact symplectic manifolds, it is a well known fact that such hamiltonian loops are always contractile.

Example 5.12 (split Poisson structures). When a coupling Dirac structure L is the graph of a Poisson structure π, the decomposition (8) corresponds to a splitting $\pi = \pi_V + \pi_H$, where π_H is a bivector field; see [Vorobjev 2001].

One may check that the corresponding connection has vanishing curvature if and only if π_H is Poisson. The characteristic foliation of π_H is then given by the integrable distribution Hor. Moreover, it follows from the curvature identity (9) that the connection is flat if and only if ω_H takes values in the space of Casimirs of π_V.

Let us assume that π_H is indeed Poisson and, for the sake of simplicity, assume that $E = B \times F$ is a trivial fibration. Then, one can still interpret the elements of $\mathcal{M}$ in terms of variations of the symplectic area of spheres. First, notice that $\tilde{\gamma}^\epsilon$ is necessarily a trivial path since the connection is trivial. Furthermore, the integral in (26) involves

$$\omega_H\Big(\frac{\mathrm{d}\gamma_B}{\mathrm{d}t}, \frac{\mathrm{d}\gamma_B}{\mathrm{d}\epsilon}\Big),$$

which are Casimirs of the vertical Poisson structure on F. The resulting element in $\mathrm{Ver}^*_{x_0}$ lies in the center of the isotropy algebra at x_0. Thus, taking the corresponding Ver^*-homotopy class amounts to integrating along the ϵ variable.

Example 5.13. As a particular case of Example 5.12, consider the trivial Poisson fibration $E = \mathbb{S}^2 \times \mathfrak{so}_3^* \to \mathbb{S}^2$, where p is the projection onto the first factor. Let π_V be the linear Poisson structure on the fibers $\mathfrak{so}_3^*$ of the projection, and let $\mathrm{Hor}_{(b,x)} = T_b S^2 \times \{x\}$ be the trivial connection. Then, ω_H must necessarily be of the form $\omega_H = f \cdot \omega$, where ω denotes the standard symplectic form on S^2 and f is a Casimir of $\mathfrak{so}_3^*$, i.e., a smooth function of the radius $r \in C^\infty(\mathfrak{so}_3^*)$.

One knows (see, e.g., [Crainic and Fernandes 2003]) that the (usual) monodromy groups of the vertical Poisson structure at some $(b, x) \in B \times \mathfrak{so}_3^*$ are

$$\mathcal{N}(\mathrm{Ver}^*)_{(b,x)} = \begin{cases} 4\pi\mathbb{Z}\,.\,\mathrm{d}r & \text{if } x \neq 0, \\ \{0\} & \text{if } x = 0, \end{cases}$$

Applying the integrability criteria of Theorem 5.10, we see that L is integrable if and only if $4\pi f'(r)$ is a rational multiple of 4π, for any r. This means that f' must be constant, so $f(r) = \alpha r + \beta$, with $\alpha \in \mathbb{Q}$ and $\beta \in \mathbb{R}$.

One can also recover this result using the prequantization Lie algebroids of [Crainic 2004] associated with a product of presymplectic spheres. On each leaf, the restricted Lie algebroid $L|_{S^2 \times S^2 \times \{v\}}$ is the prequantization of a product of presymplectic spheres $(S^2 \times S^2,\ f'(v)\omega \times \omega)$. It is well known that leaf wise, $f'(v)$ must be a rational multiple of $\int_{S^2} \omega = 4\pi$.

This example shows how rigid the integrability conditions can be: in this example, the value and the derivative of f at a point entirely determines the structure.

Appendix

A1. ***Actions on Lie groupoids and Lie algebroids.*** We will have to look at various actions of Lie groups and algebras on Lie groupoids and Lie algebroids. The following diagram summarizes the various possibilities:

$$\begin{tikzcd}
\begin{array}{c}\text{Lie group action}\\ \text{on a groupoid:}\\ \boldsymbol{A} : G \to \mathrm{Aut}(\mathcal{G})\end{array} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \begin{array}{c}\text{Lie group action}\\ \text{on an algebroid}\\ \boldsymbol{a} : G \to \mathrm{Aut}(A)\end{array} \arrow[d, Rightarrow] \\
\begin{array}{c}\text{Lie algebra action}\\ \text{on a groupoid}\\ \mathfrak{A} : \mathfrak{g} \to \mathfrak{X}_{\mathrm{mult}}(\mathcal{G})\end{array} \arrow[r, Rightarrow] & \begin{array}{c}\text{Lie algebra action}\\ \text{on an algebroid}\\ \mathfrak{a}_* : \mathfrak{g} \to \mathrm{Der}(A)\end{array}
\end{tikzcd}$$

where the four corners have the following precise meaning:

- *Action of a Lie group G on a Lie groupoid $\mathcal{G}$:* This means a smooth action $\boldsymbol{A} : G \times \mathcal{G} \to \mathcal{G}$ such that for each $g \in G$ the map $\boldsymbol{A}_g : \mathcal{G} \to \mathcal{G}$, $x \mapsto gx$, is a Lie groupoid automorphism.

- *Action of a Lie group G on a Lie algebroid A:* This means a smooth action $\boldsymbol{a} : G \times A \to A$ such that for each $g \in G$ the map $\boldsymbol{a}_g : A \to A$, $a \mapsto ax$, is a Lie algebroid automorphism.

- *Action of a Lie algebra $\mathfrak{g}$ on a Lie groupoid $\mathcal{G}$:* This means a Lie algebra homomorphism $\mathfrak{A} : \mathfrak{g} \to \mathfrak{X}_{\mathrm{mult}}(\mathcal{G})$, where $\mathfrak{X}_{\mathrm{mult}}(\mathcal{G}) \subset \mathfrak{X}(\mathcal{G})$ denotes the multiplicative vector fields in $\mathcal{G}$.

- *Action of a Lie algebra* $\mathfrak{g}$ *on a Lie algebroid* A*:* This means a Lie algebra homomorphism $\mathfrak{a}_* : \mathfrak{g} \to \mathrm{Der}(A)$, where $\mathrm{Der}(A)$ is the space of derivations of the Lie algebroid A.

Clearly, Lie group actions on Lie groupoids and algebroids cover ordinary Lie group actions on the base manifold. Similarly, Lie algebra actions on Lie groupoids and algebroids cover ordinary Lie algebra actions on the base manifold.

The arrows in the diagram above represent natural differentiation operations, either along the group action or along the groupoid. The explicit description is left to the reader, and then the commutativity of the diagram becomes obvious.

Under appropriate assumptions one can also invert the arrows in the diagram above, namely:

- One can invert the horizontal arrows (integrate actions on Lie algebroids to actions on Lie groupoids) if $\mathcal{G} = \mathcal{G}(A)$, the source 1-connected Lie groupoid integrating A.
- One can invert the vertical arrows (integrate Lie algebra actions to Lie group actions) if $G = G(\mathfrak{g})$, the source 1-connected Lie group integrating $\mathfrak{g}$, and if the infinitesimal actions are complete (the flows are defined for all $t \in \mathbb{R}$).

The reader should be able to fill in the details.

A2. ***Inner actions.*** Recall that a bisection $b : M \to \mathcal{G}$ is a smooth section of the source map such that $\boldsymbol{t} \circ b$ is a diffeomorphism of M. The space $\mathrm{Bis}(\mathcal{G})$ of bisections has natural structure of a group, induced from the groupoid structure, and the map $\mathrm{Bis}(\mathcal{G}) \to \mathrm{Diff}(M)$, $b \mapsto \boldsymbol{t} \circ b$ is a morphism of groups.

The notion of inner action for Lie groupoids follows immediately from the following definitions:

- An *inner Lie groupoid automorphism* is a Lie groupoid automorphism $\Phi : \mathcal{G} \to \mathcal{G}$ of the form

$$\Phi(x) = b(\boldsymbol{t}(x)) \cdot x \cdot b(\boldsymbol{s}(x))^{-1}.$$

for some bisection $b : M \to \mathcal{G}$. They clearly form a subgroup $\mathrm{InnAut}(\mathcal{G}) \subset \mathrm{Aut}(\mathcal{G})$.

- A *inner Lie algebroid automorphism* is a Lie algebroid automorphism $\phi : A \to A$ of the form

$$\phi = \varphi^{D_\alpha}_{1,0},$$

for some time dependent section $\alpha_t \in \Gamma(A)$. Here, $t \mapsto \varphi^{D_\alpha}_{t,0}$ denotes the flow of the time dependent derivation $D_{\alpha_t} := [\alpha_t, \cdot\,])$. They generate a subgroup $\mathrm{InnAut}(A) \subset \mathrm{Aut}(A)$.

- A *multiplicative exact vector field* is a multiplicative vector field $X \in \mathfrak{X}_{\mathrm{mult}}(\mathcal{G})$ of the form

$$X = \vec{\alpha} - \overleftarrow{\alpha},$$

where α is a section of $A = A(\mathcal{G})$, and $\vec{\alpha}$ and $\overleftarrow{\alpha}$ are the right and left invariant vector fields in $\mathcal{G}$ determined by α. They form a Lie subalgebra $\mathfrak{X}_{\text{exact}}(\mathcal{G}) \subset \mathfrak{X}_{\text{mult}}(\mathcal{G})$.

- An *inner derivation* is a Lie algebroid derivation $D \in \operatorname{Der}(A)$ of the form

$$D = [\alpha, \cdot\,]_A,$$

for some section $\alpha \in \Gamma(A)$. They clearly form a Lie subalgebra $\operatorname{InnDer}(A) \subset \operatorname{Der}(A)$.

Now, one can define inner actions in a more-or-less obvious fashion. We obtain a diagram as above:

$$\begin{array}{ccc}
\begin{array}{c}\text{inner Lie group}\\ \text{action on a groupoid}\\ \boldsymbol{A} : G \to \operatorname{InnAut}(\mathcal{G})\end{array} & \Longrightarrow & \begin{array}{c}\text{inner Lie group}\\ \text{action on an algebroid}\\ \boldsymbol{a} : G \to \operatorname{InnAut}(A)\end{array}\\
\Big\Downarrow & & \Big\Downarrow\\
\begin{array}{c}\text{inner Lie algebra}\\ \text{action on a groupoid}\\ \mathfrak{A} : \mathfrak{g} \to \mathfrak{X}_{\text{exact}}(\mathcal{G})\end{array} & \Longrightarrow & \begin{array}{c}\text{inner Lie algebra}\\ \text{action on an algebroid}\\ \mathfrak{a}_* : \mathfrak{g} \to \operatorname{InnDer}(A)\end{array}
\end{array}$$

In this work, we will mainly consider inner actions associated with a Lie groupoid morphism $\Psi : G \times M \to \mathcal{G}$ given by

$$\boldsymbol{A}_g(x) = \Psi(g, \boldsymbol{t}(x)) \cdot x \cdot \Psi(g, \boldsymbol{s}(x))^{-1}, \quad \text{for } g \in G,\ x \in \mathcal{G}. \tag{29}$$

Notice that the map Ψ covers the ordinary action $G \times M \to M$ on the base. Furthermore, one may check that $\Psi : G \times M \to \mathcal{G}$ is a Lie groupoid morphism if and only if the map $G \to \operatorname{Bis}(\mathcal{G})$, $g \mapsto b^g(x) := \Psi(g, x)$ is a group morphism covering the usual Lie group action of G on M.

Similarly, the inner Lie algebra actions on a Lie algebroid $\mathfrak{a}_* : \mathfrak{g} \to \operatorname{InnDer}(A)$ will come associated with a Lie algebroid morphism $\psi : \mathfrak{g} \times M \to A$ (covering the identity on M) such that

$$(\mathfrak{a}_\xi)_* = [\psi_*(\xi), \cdot\,]_A, \quad \text{for } \xi \in \mathfrak{g}, \tag{30}$$

where $\psi_*(\xi) \in \Gamma(A)$ is defined by $\psi_*(\xi)_m = \psi(\xi, m)$, for any $m \in M$. The map $\psi_* : \mathfrak{g} \to \Gamma(A)$ covers the ordinary Lie algebra action $\mathfrak{g} \to \mathfrak{X}(M)$ on the base. Moreover, ψ_* is a Lie algebra morphism covering the infinitesimal action $\mathfrak{g} \to \mathfrak{X}(M)$ if and only if ψ is a Lie algebroid.

Proposition A.14. *Let $G \times M \to M$ be an action of a Lie group on a manifold. Then, any homomorphism $\Psi : G \ltimes M \to \mathcal{G}$ from the action Lie groupoid to a Lie groupoid $\mathcal{G}$ determines by formula (29) an inner action of G on $\mathcal{G}$ that covers the action on M.*

Similarly, let $\mathfrak{g} \to \mathfrak{X}(M)$ *be an action of a Lie algebra on a manifold. Then any homomorphism* $\psi_* : \mathfrak{g} \ltimes M \to A$ *from the action Lie algebroid to a Lie algebroid* A *determines by formula* (30) *an inner action of* $\mathfrak{g}$ *on* A *that covers the infinitesimal action on* M.

Remark A.15. Note that Ψ and ψ as above do not need to be morphisms in order for (29) and (30) to induce inner actions. In this paper though, we will always assume that it is the case.

The relevant notion for this work is the following:

Definition A.16. A prehamiltonian action of a Lie group G on a Lie algebroid A with *prehamiltonian moment map* $\psi_* : \mathfrak{g} \to \Gamma(A)$ is an action $\boldsymbol{a} : G \to \mathrm{Aut}(A)$ such that:

- $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{a}_{\exp(-t\xi)})_*(\beta)\big|_{t=0} = [\psi_*(\xi), \beta]_A, \quad$ for $\xi \in \mathfrak{g},\ \beta \in \Gamma(A)$,
- ψ_* is a G-equivariant morphism of Lie algebras.

Note that the G-equivariance is always satisfied when G is connected.

A3. ***Integration of inner actions.*** Let us now see in which circumstances one is able to invert arrows in the last diagram.

Proposition A.17. *Suppose that* $\mathfrak{g} \to \mathfrak{X}(M)$ *is a complete Lie algebra action and* $\psi : \mathfrak{g} \ltimes M \to A$ *is a Lie algebroid morphism from the action Lie algebroid to a Lie algebroid* A. *For any Lie groupoid* $\mathcal{G}$ *integrating* A, *the associated inner action* $\mathfrak{a}_* : \mathfrak{g} \to \mathrm{InnDer}(A)$ *integrates to an inner action* $\boldsymbol{A} : G(\mathfrak{g}) \to \mathrm{InnAut}(\mathcal{G})$, *where* $G(\mathfrak{g})$ *is the* 1*-connected Lie group integrating* $\mathfrak{g}$.

Proof. By the assumptions, we have a Lie group action $G(\mathfrak{g}) \times M \to M$, and the corresponding action groupoid $G(\mathfrak{g}) \ltimes M \rightrightarrows M$ is source 1-connected. Furthermore, the Lie algebroid morphism $\psi : \mathfrak{g} \ltimes M \to A$ integrates to a Lie groupoid morphism $\tilde{\Psi} : G(\mathfrak{g}) \ltimes M \to \mathcal{G}(A)$. Denote by Ψ the composition of $\tilde{\Psi}$ with the natural projection $\mathcal{G}(A) \to \mathcal{G}$. Then, one obtains an inner action $\boldsymbol{A} : G(\mathfrak{g}) \to \mathrm{InnAut}(\mathcal{G})$ by (29). As is easily checked, it integrates the inner action $\mathfrak{a}_* : \mathfrak{g} \to \mathrm{InnDer}(A)$. □

The above result is slightly better than the integration of non-inner actions we referred to in the end of the preceding subsection. In general, in order to integrate a Lie algebra action of $\mathfrak{g}$ on a Lie algebroid A to a Lie group action of G on a Lie groupoid $\mathcal{G}$, we need *both* G to be 1-connected and $\mathcal{G}$ to be source 1-connected (and the action to be complete).

This is important for our purposes, as G is the structure group of a principal bundle, thus its topology is imposed. So, we need to refine Proposition A.17 to groups that are neither simply connected nor connected.

Hence, assume that we want to integrate an inner action $\mathfrak{a} : \mathfrak{g} \to \mathrm{InnDer}(A)$ associated with $\psi : \mathfrak{g} \ltimes M \to A$ to an action of a connected (but not necessarily

1-connected) Lie group G with Lie algebra $\mathfrak{g}$. Recall that for a connected Lie group G, its fundamental group fits into the exact sequence

$$1 \longrightarrow \pi_1(G) \longrightarrow G(\mathfrak{g}) \longrightarrow G \longrightarrow 1. \tag{31}$$

Proposition A.18. *Suppose that G is a connected Lie group acting on a manifold M, $\mathcal{G}$ is a Lie groupoid over M, and $\overline{\Psi} : G(\mathfrak{g}) \ltimes M \to \mathcal{G}$ is a Lie groupoid morphism. Then, $\overline{\Psi}$ descends to a Lie groupoid morphism $\Psi : G \ltimes M \to \mathcal{G}$ if and only if it takes values in units when restricted to $\pi_1(G) \ltimes M \subset G(\mathfrak{g}) \ltimes M$, namely*

$$\overline{\Psi}(\pi_1(G) \ltimes M) = \mathbf{1}_M.$$

The proof is left to the reader. Note that although $\psi : \mathfrak{g} \ltimes M \to A$ may not integrate to $\Psi : G \ltimes M \to \mathcal{G}$, it may still integrate to a morphism $\Psi' : G \ltimes M \to \mathcal{G}'$ for a smaller Lie groupoid $\mathcal{G}'$ integrating A. This can be decided as follows. First we integrate $\psi : \mathfrak{g} \ltimes M \to A$ to a Lie groupoid morphism $\tilde{\Psi} : G(\mathfrak{g}) \ltimes M \to \mathcal{G}(A)$ with values in the source 1-connected Lie groupoid integrating A. Then, given any connected Lie group G integrating $\mathfrak{g}$, we introduce a bundle of groups Δ over M, defined in the following way:

$$\Delta := \tilde{\Psi}(\pi_1(G) \ltimes M) \subset \mathcal{G}(A)$$

Recall that if Δ is a totally disconnected wide *normal subgroupoid* of $\mathcal{G}(A)$, then the quotient $\mathcal{G}(A)/\Delta$ is a Lie groupoid; see, e.g., the discussion in [Gualtieri and Li 2014, Theorem 1.14]. Furthermore, it is easy to see that $\mathcal{G}(A)/\Delta$ integrates A. Therefore, we obtain a Lie groupoid morphism:

$$\Psi : G \ltimes M \to \mathcal{G}(A)/\Delta.$$

Of course, ψ integrates to a morphism $G \ltimes M \to \mathcal{G}$ whenever $\mathcal{G}$ is covered by $\mathcal{G}_0 := \mathcal{G}(A)/\Delta$.

Theorem A.19. *Let $\boldsymbol{a} : G \to \mathrm{Aut}(A)$ be a prehamiltonian action of a Lie group G on a Lie algebroid A with premoment map $\psi_* : \mathfrak{g} \to \Gamma(A)$ and:*

- *$\psi : \mathfrak{g} \ltimes M \to A$ the Lie algebroid associated with ψ_*,*
- *$\tilde{\Psi} : G(\mathfrak{g}) \ltimes M \to \mathcal{G}(A)$ the groupoid morphism integrating ψ,*
- *$\Delta \subset \mathcal{G}(A)$, the subset defined by $\Delta := \tilde{\Psi}(\pi_1(G) \ltimes M) \subset \mathcal{G}(A)$.*

Then, the following assertions hold:

(i) *Δ is a wide, normal, totally disconnected subgroupoid of $\mathcal{G}(A)$,*

(ii) *$\boldsymbol{a}$ integrates to a groupoid action $\boldsymbol{A} : G \to \mathrm{Aut}(\mathcal{G}(A)/\Delta)$.*

Moreover, $\mathcal{G}(A)/\Delta$ is a Lie groupoid if and only if $\Delta \subset \mathcal{G}(A)$ is an embedding. In this case, $\tilde{\Psi}$ descends to a Lie groupoid morphism $\Psi : G \times M \to \mathcal{G}(A)/\Delta$ and

$$\boldsymbol{A}_h(x) = \Psi_n(h) \cdot x \cdot \Psi_m(h)^{-1} \tag{32}$$

$$\Psi_{gm}(ghg^{-1}) = \boldsymbol{A}_g(\Psi_m(h)) \tag{33}$$

for any $h \in G^\circ$, $x \in \mathcal{G}(A)/\Delta$, $g \in G$, where $m := \boldsymbol{s}(x)$ and $n := \boldsymbol{t}(x)$.

Proof. Since G acts on A by Lie algebroid automorphisms, one can lift $\boldsymbol{a}$ into an action $\tilde{\boldsymbol{A}}$ of G on $\mathcal{G}(A)$ by groupoid automorphisms via

$$\tilde{\boldsymbol{A}}_g([q]_A) := [\boldsymbol{a}_g \circ q]_A, \quad \text{for } q \in P(A),\ g \in G.$$

Consider now an A-path q and an element h lying in the neutral component G° of G. We extend q into a time dependent section of A (which we still denote by q) and consider any $\mathfrak{g}$-path $\xi : \epsilon \mapsto \xi_\epsilon \in \mathfrak{g}$ that induces a path h_ϵ in G° between the identity and h. By the construction, $(\boldsymbol{a}_{h_\epsilon})_*(q_t)$ is a solution of the evolution equation

$$[(\boldsymbol{a}_{h_\epsilon})_*(q_t),\ \psi_*(\xi_\epsilon)]_A = \frac{\mathrm{d}}{\mathrm{d}\epsilon}(\boldsymbol{a}_{h_\epsilon})_*(q_t), \quad \text{for } \epsilon, t \in I.$$

Then, by [Brahic 2010, Proposition A.1], we obtain

$$\tilde{\boldsymbol{A}}_h([q]_A) = \tilde{\Psi}_y([\xi]_{\mathfrak{g}}) \cdot [q]_A \cdot \tilde{\Psi}_x([\xi]_{\mathfrak{g}})^{-1}. \tag{34}$$

In particular, if $h = 1$, that is, ξ_ϵ induces a loop in G°, then Δ is a normal subgroupoid of $\mathcal{G}(A)$.

Next, we have to make sure that $\tilde{\boldsymbol{A}}$ induces an action on $\mathcal{G}(A)/\Delta$, so we need to check that $\tilde{A}_h(\Delta) = \Delta$. For this, we apply Equation (34) with $q = \tilde{\Psi}[\eta]_{\mathfrak{g}}$, where η is a $\mathfrak{g}$-path inducing a loop in G, and we use successively the fact that $\tilde{\Psi}$ is a Lie groupoid homomorphism, then that $\pi_1(G)$ lies in $G(\mathfrak{g})$ as a normal subgroup. The first relation follows. The second one is obtained by using the equivariance condition in Definition A.16. ☐

Example A.20. Here is a basic example where the resulting groupoid $\mathcal{G}(A)/\Delta$ is not smooth. Consider the 1-dimensional (abelian) Lie algebra $\mathfrak{z} = \mathbb{R}$, its dual $\mathfrak{z}^*$ endowed with the trivial linear Poisson structure, and $A := T^*\mathfrak{z}^* \simeq \mathfrak{z} \ltimes \mathfrak{z}^*$ the corresponding Lie algebroid. Then, A integrates to a bundle of Lie groups $Z \ltimes \mathfrak{z}^*$, where Z is the Lie group $\mathbb{R}$.

Consider furthermore the trivial action of $G := S^1 \simeq \mathbb{R}/2\pi\mathbb{Z}$ on $\mathfrak{z}^*$. Then, any application $J : \mathfrak{z}^* \to \mathrm{Lie}(S^1)$ can be chosen to be a moment map. From the Lie algebroid point of view (see Definition A.16) we only have a prehamiltonian moment map $\psi_* : \mathrm{Lie}(S^1) \to \Gamma(\mathfrak{z} \ltimes \mathfrak{z}^*)$, $X \mapsto \mathrm{d}J(X)$. The corresponding Lie algebroid morphism $\psi : \mathrm{Lie}(S^1) \ltimes \mathfrak{z}^* \to \mathfrak{z} \ltimes \mathfrak{z}^*$ is given by $(X, z) \mapsto (\mathrm{d}J_z(X), z)$. It integrates to a Lie groupoid morphism $\mathbb{R} \ltimes \mathfrak{z}^* \to Z \ltimes \mathfrak{z}^*$ given by $(\theta, z) \mapsto (\mathrm{d}J_z(\theta), z)$. The exact sequence (31) reads

$$1 \longrightarrow 2 \longrightarrow \mathbb{R} \longrightarrow S^1 \longrightarrow 1,$$

Hence, we obtain $\Delta_z = \{(\mathrm{d}J_z(2k\pi), z) : k \in \mathbb{Z}\}$. Clearly, Δ defines a normal subgroupoid of $Z \ltimes \mathfrak{z}^*$; however, $Z \ltimes \mathfrak{z}^*/\Delta$ is not smooth if $\mathrm{d}J$ vanishes at some point z_0.

References

[Brahic 2010] O. Brahic, "Extensions of Lie brackets", *J. Geom. Phys.* **60**:2 (2010), 352–374. MR 2011b:58038 Zbl 1207.58018

[Brahic and Fernandes 2008] O. Brahic and R. L. Fernandes, "Poisson fibrations and fibered symplectic groupoids", pp. 41–59 in *Proceedings of the 5th International Conference* (Tokyo, 2006), edited by G. Dito et al., Contemp. Math. **450**, Amer. Math. Soc., Providence, RI, 2008. MR 2009h:53185 Zbl 1158.53064

[Brahic and Fernandes 2014] O. Brahic and R. L. Fernandes, "Integrability and reduction of Hamiltonian actions on Dirac manifolds", *Indag. Math.* (*N.S.*) **25**:5 (2014), 901–925. MR 3264780 Zbl 1298.53082

[Brahic and Zhu 2011] O. Brahic and C. Zhu, "Lie algebroid fibrations", *Adv. Math.* **226**:4 (2011), 3105–3135. MR 2012a:53156 Zbl 1210.53076

[Bursztyn et al. 2004] H. Bursztyn, M. Crainic, A. Weinstein, and C. Zhu, "Integration of twisted Dirac brackets", *Duke Math. J.* **123**:3 (2004), 549–607. MR 2005e:53128 Zbl 1067.58016

[Courant 1990] T. J. Courant, "Dirac manifolds", *Trans. Amer. Math. Soc.* **319**:2 (1990), 631–661. MR 90m:58065 Zbl 0850.70212

[Crainic 2004] M. Crainic, "Prequantization and Lie brackets", *J. Symplectic Geom.* **2**:4 (2004), 579–602. MR 2006j:53129 Zbl 1095.53060

[Crainic and Fernandes 2003] M. Crainic and R. L. Fernandes, "Integrability of Lie brackets", *Ann. of Math.* (2) **157**:2 (2003), 575–620. MR 2004h:58027 Zbl 1037.22003

[Crainic and Fernandes 2004] M. Crainic and R. L. Fernandes, "Integrability of Poisson brackets", *J. Differential Geom.* **66**:1 (2004), 71–137. MR 2006f:53119 Zbl 1066.53131

[Crainic and Fernandes 2011] M. Crainic and R. L. Fernandes, "Lectures on integrability of Lie brackets", pp. 1–107 in *Lectures on Poisson geometry* (Trieste, 2005), edited by T. Ratiu et al., Geom. Topol. Monogr. **17**, 2011. MR 2795150 Zbl 1227.22005

[Dufour and Wade 2008] J.-P. Dufour and A. Wade, "On the local structure of Dirac manifolds", *Compos. Math.* **144**:3 (2008), 774–786. MR 2010c:53121 Zbl 1142.53063

[Fernandes et al. 2009] R. L. Fernandes, J.-P. Ortega, and T. S. Ratiu, "The momentum map in Poisson geometry", *Amer. J. Math.* **131**:5 (2009), 1261–1310. MR 2011f:53199 Zbl 1180.53083

[Gualtieri and Li 2014] M. Gualtieri and S. Li, "Symplectic groupoids of log symplectic manifolds", *Int. Math. Res. Not.* **2014**:11 (2014), 3022–3074. MR 3214314 Zbl 1305.22004

[Guillemin et al. 1996] V. Guillemin, E. Lerman, and S. Sternberg, *Symplectic fibrations and multiplicity diagrams*, Cambridge University Press, Cambridge, 1996. MR 98d:58074 Zbl 0870.58023

[Mărcuţ and Frejlich 2013] I. Mărcuţ and P. Frejlich, "The normal form theorem around Poisson transversals", preprint, 2013. arXiv 1306.6055v2

[McDuff and Salamon 1998] D. McDuff and D. Salamon, *Introduction to symplectic topology*, 2nd ed., Clarendon Press, New York, 1998. MR 2000g:53098

[Vaisman 2006] I. Vaisman, "Foliation-coupling Dirac structures", *J. Geom. Phys.* **56**:6 (2006), 917–938. MR 2007f:53109 Zbl 1109.53076

[Vorobjev 2001] Y. Vorobjev, "Coupling tensors and Poisson geometry near a single symplectic leaf", pp. 249–274 in *Lie algebroids and related topics in differential geometry (Warsaw, 2000)* (Warsaw, 2000), edited by J. Kubarski and R. Urbański, Pawełand Wolak, Banach Center Publ. **54**, Polish Acad. Sci. Inst. Math., Warsaw, 2001. MR 2003d:53145 Zbl 1007.53062

[Wade 2008] A. Wade, "Poisson fiber bundles and coupling Dirac structures", *Ann. Global Anal. Geom.* **33**:3 (2008), 207–217. MR 2009d:53122 Zbl 1151.53070

[Weinstein 1980] A. Weinstein, "Fat bundles and symplectic manifolds", *Adv. in Math.* **37**:3 (1980), 239–250. MR 82a:53038 Zbl 0449.53035

Received October 6, 2014. Revised March 26, 2015.

OLIVIER BRAHIC
DEPARTMENT OF MATHEMATICS
FEDERAL UNIVERSITY OF PARANÁ
CP 19081
81531-980 CURITIBA-PR
BRAZIL
brahicolivier@gmail.com

RUI LOJA FERNANDES
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
1409 W. GREEN STREET
URBANA, IL 61801
UNITED STATES
ruiloja@illinois.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.278.369

ASYMPTOTIC BEHAVIOR OF PALAIS–SMALE SEQUENCES ASSOCIATED WITH FRACTIONAL YAMABE-TYPE EQUATIONS

YI FANG AND MARÍA DEL MAR GONZÁLEZ

In this paper, we analyze the asymptotic behavior of Palais–Smale sequences associated to fractional Yamabe-type equations on an asymptotically hyperbolic Riemannian manifold. We prove that Palais–Smale sequences can be decomposed into the solution of the limit equation plus a finite number of bubbles, which are the rescaling of the fundamental solution for the fractional Yamabe equation on Euclidean space. We also verify the non-interfering fact for multibubbles.

1. Introduction and statement of results

Let Ω be a smooth bounded domain in $\mathbb{R}^n$, $n \geq 3$. Fix a constant λ, and consider the Dirichlet boundary value problem of the elliptic PDE

$$\begin{cases} -\Delta u - \lambda u = u\,|u|^{\frac{4}{n-2}} & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{1-1}$$

The associated variational functional of (1-1) in the Sobolev space $W_0^{1,2}(\Omega)$ is

$$E(u) = \frac{1}{2}\int_\Omega (|\nabla u|^2 - \lambda u^2)\,dx - \frac{n-2}{2n}\int_\Omega |u|^{\frac{2n}{n-2}}\,dx.$$

Suppose that the sequence $\{u_\alpha\}_{\alpha\in\mathbb{N}} \subset W_0^{1,2}(\Omega)$ satisfies the Palais–Smale condition,

$$\{E(u_\alpha)\}_{\alpha\in\mathbb{N}} \text{ is uniformly bounded and } DE(u_\alpha) \to 0, \text{ strongly in } (W_0^{1,2}(\Omega))',$$

as $\alpha \to +\infty$, where $(W_0^{1,2}(\Omega))'$ is the dual space of $W_0^{1,2}(\Omega)$. In an elegant paper, M. Struwe [1984] considered the asymptotic behavior of $\{u_\alpha\}_{\alpha\in\mathbb{N}}$. In fact, in

Fang is fully supported by China Scholarship Council (CSC) for visiting the University of California, Santa Cruz. González is supported by Spain Government projects MTM2011-27739-C04-01 and MTM2014-52402-C3-1-P, and GenCat 2009SGR345.
MSC2010: primary 35J70; secondary 58J32.
Keywords: fractional Yamabe problem, Palais–Smale sequences.

the $W_0^{1,2}(\Omega)$ norm, u_α can be approximated by the solution to (1-1) plus a finite number of bubbles, which are the rescaling of the nontrivial entire solution of

$$-\Delta u = u|u|^{\frac{4}{n-2}} \text{ in } \mathbb{R}^n \quad \text{and} \quad u(x) \to 0 \text{ as } |x| \to +\infty.$$

One may pose the analogous problem on a manifold. Let (M^n, g) be a smooth compact Riemannian manifold without boundary. Consider a sequence of elliptic PDEs like

$$(E_\alpha) \qquad -\Delta_g u + h_\alpha u = u^{\frac{n+2}{n-2}},$$

where $\alpha \in \mathbb{N}$ and Δ_g denotes the Laplace–Beltrami operator of the metric g. Assume that h_α satisfies the condition that there exists $C > 0$ with $|h_\alpha(x)| \leq C$ for any α and any $x \in M$; also $h_\alpha \to h_\infty$ in $L^2(M)$ as $\alpha \to +\infty$. The limit equation is denoted by

$$(E_\infty) \qquad -\Delta_g u + h_\infty u = u^{\frac{n+2}{n-2}}.$$

The related variational functional for (E_α) is

$$E_g^\alpha(u) = \frac{1}{2}\int_M |\nabla u|_g^2 \, dv_g + \frac{1}{2}\int_M h_\alpha u^2 \, dv_g - \frac{n-2}{2n}\int_M |u|^{\frac{2n}{n-2}} \, dv_g.$$

Suppose that $\{u_\alpha \geq 0\}_{\alpha\in\mathbb{N}} \subset W^{1,2}(M)$ also satisfies the Palais–Smale condition. O. Druet, E. Hebey, and F. Robert [Druet et al. 2004] proved that, in the $W^{1,2}(M)$ sense, u_α can be decomposed into the solution of (E_∞) plus a finite number of bubbles, which are the rescaling of the nontrivial solution of

$$-\Delta u = u^{\frac{n+2}{n-2}} \quad \text{in } \mathbb{R}^n.$$

Let (M^n, g) be a compact Riemannian manifold with boundary ∂M. Recently, S. Almaraz [2014] considered the following sequence of equations with nonlinear boundary value condition:

$$(1\text{-}2) \qquad \begin{cases} -\Delta_g u = 0 & \text{in } M, \\ -\frac{\partial}{\partial \eta_g} u + h_\alpha u = u^{\frac{n}{n-2}} & \text{on } \partial M, \end{cases}$$

where $\alpha \in \mathbb{N}$ and η_g is the inward unit normal vector to ∂M. The associated energy functional for (1-2) is

$$\bar{E}_g^\alpha(u) = \frac{1}{2}\int_M |\nabla u|_g^2 \, dv_g + \frac{1}{2}\int_{\partial M} h_\alpha u^2 \, d\sigma_g - \frac{n-2}{2(n-1)}\int_{\partial M} |u|^{\frac{2(n-1)}{n-2}} \, d\sigma_g,$$

for $u \in H^1(M) := \{u \mid \nabla u \in L^2(M),\ u \in L^2(\partial M)\}$. Here dv_g and $d\sigma_g$ are the volume forms of M and ∂M, respectively. He also showed that a nonnegative Palais–Smale sequence $\{u_\alpha\}_{\alpha\in\mathbb{N}}$ of $\{\bar{E}_g^\alpha\}_{\alpha\in\mathbb{N}}$ converges, in the $H^1(M)$ sense, to a solution of the limit equation (the equation replacing h_α by h_∞ in (1-2)) plus a finite number of bubbles.

Motivated by these facts and the original study of the fractional Yamabe problem by M.d.M. González and J. Qing [2013] (see also [González 2015]), in this paper we shall be interested in the asymptotic behavior of nonnegative Palais–Smale sequences associated with the fractional Yamabe equation on an asymptotically hyperbolic Riemannian manifold.

Let (X^{n+1}, g^+), $n \geq 3$, be a smooth Riemannian manifold with smooth boundary $\partial X^{n+1} = M^n$. A function ρ_* is called a defining function of the boundary M^n in X^{n+1} if it satisfies

$$\rho_* > 0 \text{ in } X^{n+1}, \quad \rho_* = 0 \text{ on } M^n, \quad d\rho_* \neq 0 \text{ on } M^n.$$

We say that a metric g^+ is *conformally compact* if there exists a defining function ρ_* such that $(\overline{X^{n+1}}, \bar{g}_*)$ is compact for $\bar{g}_* = \rho_*^2 g^+$. This induces a conformal class of metrics $\hat{h} = \bar{g}_*|_{M^n}$ when defining functions vary. The conformal manifold $(M^n, [\hat{h}])$ is called the *conformal infinity* of (X^{n+1}, g^+). A metric g^+ is said to be *asymptotically hyperbolic* if it is conformally compact and the sectional curvature approaches -1 at infinity. It is easy to check then that $|d\rho_*|^2_{\bar{g}_*} = 1$ on M^n.

Using the meromorphic family of scattering operators $S(s)$ introduced by C.R. Graham and M. Zworski [2003], we will define the so-called fractional order scalar curvature. Given an asymptotically hyperbolic Riemannian manifold (X^{n+1}, g^+) and a representative $\hat{h}$ of the conformal infinity $(M^n, [\hat{h}])$, there is a unique geodesic defining function ρ_* such that, in $M^n \times (0, \delta)$ in X^{n+1} for small δ, g^+ has the normal form

$$g^+ = \rho_*^{-2}(d\rho_*^2 + h_{\rho_*}),$$

where h_{ρ_*} is a one parameter family of metric on M^n such that

$$h_{\rho_*} = \hat{h} + h^{(1)}\rho_* + O(\rho_*^2).$$

It is well-known [Graham and Zworski 2003] that, given $f \in \mathbb{C}^\infty(M^n)$ and $s \in \mathbb{C}$, $\mathrm{Re}(s) > n/2$ and $s(n-s)$ is not an L^2 eigenvalue for $-\Delta_{g^+}$, then the generalized eigenvalue problem

$$-\Delta_{g^+}\tilde{u} - s(n-s)\tilde{u} = 0 \quad \text{in } X^{n+1} \tag{1-3}$$

has a solution of the form

$$\tilde{u} = F(\rho_*)^{n-s} + G(\rho_*)^s, \quad F, G \in \mathscr{C}^\infty(\overline{X^{n+1}}), \quad F|_{\rho_*=0} = f.$$

The scattering operator on M^n is then defined as

$$S(s)f = G|_{M^n}.$$

Now we consider the normalized scattering operators

$$P_\gamma[g^+,\hat{h}] = d_\gamma S\big(\tfrac{n}{2}+\gamma\big), \quad d_\gamma = 2^{2\gamma}\frac{\Gamma(\gamma)}{\Gamma(-\gamma)}.$$

Note that $P_\gamma[g^+,\hat{h}]$ is a pseudodifferential operator whose principal symbol is equal to the one of $(-\Delta_{\hat{h}})^\gamma$. Moreover, $P_\gamma[g^+,\hat{h}]$ is conformally covariant, i.e., for any $\varphi, w \in \mathscr{C}^\infty(\overline{X^{n+1}})$ and $w > 0$,

$$P_\gamma[g^+, w^{\frac{4}{n-2\gamma}}\hat{h}](\varphi) = w^{-\frac{n+2\gamma}{n-2\gamma}} P_\gamma[g^+,\hat{h}](w\varphi). \tag{1-4}$$

Thus we shall call $P_\gamma[g^+,\hat{h}]$ the conformal fractional Laplacian for any $\gamma \in (0, n/2)$ such that $n^2/4 - \gamma^2$ is not an L^2 eigenvalue for $-\Delta_{g^+}$.

The fractional scalar curvature associated to the operator $P_\gamma[g^+,\hat{h}]$ is defined as

$$Q_\gamma^{\hat{h}} = P_\gamma[g^+,\hat{h}](1).$$

The scattering operator has a pole at the integer values γ. However, in such cases the residue may be calculated and, in particular, when g^+ is Poincaré-Einstein metric, for $\gamma = 1$,

$$P_1[g^+,\hat{h}] = -\Delta_{\hat{h}} + \frac{n-2}{4(n-1)}R_{\hat{h}},$$

which is exactly the so-called conformal Laplacian, and

$$Q_1^{\hat{h}} = \frac{n-2}{4(n-1)}R_{\hat{h}}.$$

Here, $R_{\hat{h}}$ is the scalar curvature of the metric $\hat{h}$.

For $\gamma = 2$, $P_2[g^+,\hat{h}]$ is precisely the Paneitz operator and its associated curvature is known as Q-curvature [2008]. In general, $P_k[g^+,\hat{h}]$ for $k \in \mathbb{N}$ are precisely the conformal powers of the Laplacian studied in [Graham et al. 1992].

We consider the conformal change $\hat{h}_w = w^{4/(n-2\gamma)}\hat{h}$ for some $w > 0$; then by (1-4),

$$P_\gamma[g^+,\hat{h}](w) = Q_\gamma^{\hat{h}_w} w^{\frac{n+2\gamma}{n-2\gamma}} \quad \text{in } (M^n,\hat{h}).$$

If for this conformal change $Q_\gamma^{\hat{h}_w}$ is a constant C_γ on M^n, this problem reduces to

$$P_\gamma[g^+,\hat{h}](w) = C_\gamma w^{\frac{n+2\gamma}{n-2\gamma}} \quad \text{in } (M^n,\hat{h}), \tag{1-5}$$

which is the so-called fractional Yamabe equation or the γ-Yamabe equation, studied in [González and Qing 2013].

Throughout the paper, we always suppose that $\gamma \in (0, 1)$, and such that the first eigenvalue for $-\Delta_{g^+}$ satisfies $\lambda_1 > n^2/4 - \gamma^2$, as was pointed out in [Case and Chang 2015; Case 2015].

It is well known that the above fractional Yamabe equation may be rewritten as a degenerate elliptic Dirichlet-to-Neumann boundary problem. For that, we first recall some results obtained by Chang and González in [2011] (see also the paper by J. Case and S.A. Chang [2015]). Suppose that u^* solves

$$\begin{cases} -\Delta_{g^+} u^* - s(n-s)u^* = 0 & \text{in } X^{n+1}, \\ \lim_{\rho_*\to 0} \rho_*^{s-n} u^* = 1 & \text{on } M^n. \end{cases} \tag{1-6}$$

Proposition 1.1 [Chang and González 2011; González and Qing 2013]. *Suppose that $f \in \mathscr{C}^\infty(M)$. Assume that $\tilde{u}, u^*$ are solutions to (1-3) and (1-6), respectively. Then $\rho = (u^*)^{1/(n-s)}$ is a geodesic defining function. Moreover, $u = \tilde{u}/u^* = \rho^{s-n}\tilde{u}$ solves*

$$\begin{cases} -\operatorname{div}(\rho^{1-2\gamma}\nabla u) = 0 & \text{in } X^{n+1}, \\ u = f & \text{on } M^n, \end{cases} \tag{1-7}$$

with respect to the metric $g = \rho^2 g^+$, and u is the unique minimizer of the energy functional

$$I(v) = \int_{X^{n+1}} \rho^{1-2\gamma} |\nabla v|_g^2 \, dv_g$$

among all the extensions $v \in W^{1,2}(X^{n+1}, \rho^{1-2\gamma})$ (see Definition 2.1) satisfying $v|_{M^n} = f$. Moreover,

$$\rho = \rho_* \left(1 + \frac{Q_\gamma^{\hat{h}}}{(n-s)d_\gamma} \rho_*^{2\gamma} + O(\rho_*^2) \right)$$

near the conformal infinity and

$$P_\gamma[g^+, \hat{h}](f) = -d_\gamma^* \lim_{\rho\to 0} \rho^{1-2\gamma} \partial_\rho u + Q_\gamma^{\hat{h}} f, \quad d_\gamma^* = -\frac{d_\gamma}{2\gamma} > 0,$$

provided that $\operatorname{Tr}_{\hat{h}} h^{(1)} = 0$ *when $\gamma \in (\frac{1}{2}, 1)$. Here $g|_{M^n} = \hat{h}$, and has asymptotic expansion*

$$g = d\rho^2[1 + O(\rho^{2\gamma})] + \hat{h}[1 + O(\rho^{2\gamma})].$$

We fix $\gamma \in (0, 1)$. By Proposition 1.1, one can rewrite the fractional Yamabe equation (1-5) into the following problem:

$$\begin{cases} -\operatorname{div}(\rho^{1-2\gamma}\nabla u) = 0 & \text{in } (X^{n+1}, g), \\ u = w & \text{on } (M^n, \hat{h}), \\ -d_\gamma^* \lim_{\rho\to 0} \rho^{1-2\gamma} \partial_\rho u + Q_\gamma^{\hat{h}} w = C_\gamma w^{\frac{n+2\gamma}{n-2\gamma}} & \text{on } (M^n, \hat{h}). \end{cases} \tag{1-8}$$

In this paper we consider the positive curvature case $C_\gamma > 0$. Without loss of generality, we assume that $C_\gamma = d_\gamma^*$.

In the particular case $\gamma = \frac{1}{2}$, one may check that (1-8) reduces to (1-2), which was considered in [Almaraz 2014]. The main difficulty we encounter here is the presence of the weight that makes the extension equation only degenerate elliptic.

Next, we introduce the so-called γ-Yamabe constant [González and Qing 2013]. For the defining function ρ mentioned above, we set

$$I_\gamma[u,g] = \frac{d_\gamma^* \int_X \rho^{1-2\gamma} |\nabla u|_g^2 \, dv_g + \int_M Q_\gamma^{\hat{h}} u^2 \, d\sigma_{\hat{h}}}{\left(\int_M |u|^{2^*} \, d\sigma_{\hat{h}}\right)^{\frac{2}{2^*}}},$$

then the γ-Yamabe constant is defined as

$$\Lambda_\gamma(M,[\hat{h}]) = \inf\{I_\gamma[u,g] \mid u \in W^{1,2}(X, \rho^{1-2\gamma})\}. \tag{1-9}$$

It was shown in [loc. cit.] that in the positive curvature case $C_\gamma > 0$ we must have $\Lambda_\gamma(M,[\hat{h}]) > 0$.

Now we take a perturbation of the linear term $Q_\gamma^{\hat{h}} w$ to a general $-d_\gamma^* Q_\alpha^\gamma w$, where $Q_\alpha^\gamma \in \mathscr{C}^\infty(M^n)$, $\alpha \in \mathbb{N}$. Suppose that for any $\alpha \in \mathbb{N}$ and any $x \in M^n$, there exists a constant $C > 0$ such that $|Q_\alpha^\gamma(x)| \le C$. Also assume that $Q_\alpha^\gamma \to Q_\infty^\gamma$ in $L^2(M^n, \hat{h})$ as $\alpha \to +\infty$. We will consider a family of equations

$$\begin{cases} -\operatorname{div}(\rho^{1-2\gamma} \nabla u) = 0 & \text{in } (X^{n+1}, g), \\ u = w & \text{on } (M^n, \hat{h}), \\ -\lim\limits_{\rho \to 0} \rho^{1-2\gamma} \partial_\rho u + Q_\alpha^\gamma w = w^{\frac{n+2\gamma}{n-2\gamma}} & \text{on } (M^n, \hat{h}). \end{cases} \tag{1-10}$$

The associated variational functional to (1-10) is

$$I_g^{\gamma,\alpha}(u) = \frac{1}{2} \int_{X^{n+1}} \rho^{1-2\gamma} |\nabla u|_g^2 \, dv_g + \frac{1}{2} \int_{M^n} Q_\alpha^\gamma u^2 \, d\sigma_{\hat{h}} - \frac{n-2\gamma}{2n} \int_{M^n} |u|^{\frac{2n}{n-2\gamma}} \, d\sigma_{\hat{h}}. \tag{1-11}$$

Hyperbolic space $(\mathbb{H}^{n+1}, g_{\mathbb{H}})$ is the first example of a conformally compact Einstein manifold. As $(\mathbb{H}^{n+1}, g_{\mathbb{H}})$ can be characterized as the upper half-space $\mathbb{R}_+^{n+1}$ endowed with metric $g^+ = y^{-2}(|dx|^2 + dy^2)$, where $x \in \mathbb{R}^n$, $y \in \mathbb{R}_+$, then the Dirichlet-to-Neumann problem (1-8) reduces to

$$\begin{cases} -\operatorname{div}(y^{1-2\gamma} \nabla u) = 0 & \text{in } (\mathbb{R}_+^{n+1}, |dx|^2 + dy^2), \\ u = w & \text{on } (\mathbb{R}^n, |dx|^2), \\ -\lim\limits_{y \to 0} y^{1-2\gamma} \partial_y u = w^{\frac{n+2\gamma}{n-2\gamma}} & \text{on } (\mathbb{R}^n, |dx|^2). \end{cases} \tag{1-12}$$

And the variational functional to (1-12) is defined as

$$\widetilde{E}(u) = \frac{1}{2} \int_{\mathbb{R}_+^{n+1}} y^{1-2\gamma} |\nabla u(x,y)|^2 \, dx \, dy - \frac{n-2\gamma}{2n} \int_{\mathbb{R}^n} |u(x,0)|^{\frac{2n}{n-2\gamma}} \, dx.$$

Up to multiplicative constants, the only solution to problem (1-12) is given by the standard

$$w(x) = w_a^{\lambda}(x) = \left(\frac{\lambda}{|x-a|^2+\lambda^2}\right)^{\frac{n-2\gamma}{2}}$$

for some $a \in \mathbb{R}^n$ and $\lambda > 0$ [González and Qing 2013; Jin et al. 2014]. By the Poisson formula of L. Caffarelli and L. Silvestre [2007], the corresponding extension can be expressed as

$$U_a^{\lambda}(x,y) = \int_{\mathbb{R}^n} \frac{y^{2\gamma}}{(|x-\xi|^2+y^2)^{(n+2\gamma)/2}}\, w_a^{\lambda}(\xi)\, d\xi. \tag{1-13}$$

Here U_a^{λ} is called a "bubble". Note that all of them have constant energy.

Remark 1.2. For any $a \in \mathbb{R}^n$ and $\lambda > 0$, we have

$$\widetilde{E}(U_a^{\lambda}) = \widetilde{E}(U_0^1) = \frac{\gamma}{n} \int_{\mathbb{R}^n} |U_0^1(x,0)|^{\frac{2n}{n-\gamma}}\, dx.$$

Now we give some notations which will be used in the following. In the half space $\mathbb{R}_+^{n+1} = \{(x,y) = (x^1,\dots,x^n,y) \in \mathbb{R}^{n+1} \mid y > 0\}$ we define, for $r > 0$,

$$\begin{aligned} B_r^+(z_0) &= \{z \in \mathbb{R}_+^{n+1} \mid |z-z_0| < r,\ z_0 \in \mathbb{R}_+^{n+1}\},\\ D_r(x_0) &= \{x \in \mathbb{R}^n \mid |x-x_0| < r,\ x_0 \in \mathbb{R}^n\},\\ \partial' B_r^+(z_0) &= B_r^+(z_0) \cap \mathbb{R}^n,\\ \partial^+ B_r^+(z_0) &= \partial B_r^+(z_0) \cap \mathbb{R}_+^{n+1}. \end{aligned}$$

Fix $\gamma \in (0,1)$. Suppose that (X, g^+) is an asymptotically hyperbolic manifold with boundary M satisfying, in addition, $\mathrm{Tr}_{\hat{h}}\, h^{(1)} = 0$ when $\gamma \in (1/2, 1)$. Let ρ be the special defining function given in Proposition 1.1 and set $g = \rho^2 g^+$ and $\hat{h} = g|_M$. Also, define

$$\begin{aligned} \mathfrak{B}_r^+(z_0) &= \{z \in X \mid d_g(z,z_0) < r,\ z_0 \in \overline{X}\},\\ \mathfrak{D}_r(x_0) &= \{x \in M \mid d_{\hat{h}}(x,x_0) < r,\ x_0 \in M\}, \end{aligned}$$

Now, modulo the definitions of the weighted Sobolev space $W^{1,2}(X, \rho^{1-2\gamma})$ and of a Palais–Smale sequence (see Section 2), the main result of this paper is the following fractional type blow up analysis theorem:

Theorem 1.3. *Let $\{u_\alpha \geq 0\}_{\alpha\in\mathbb{N}} \subset W^{1,2}(X, \rho^{1-2\gamma})$ be a Palais–Smale sequence for $\{I_g^{\gamma,\alpha}\}_{\alpha\in\mathbb{N}}$. Then there exists an integer $m \geq 1$, sequences $\{\mu_\alpha^j > 0\}_{\alpha\in\mathbb{N}}$ and $\{x_\alpha^j\}_{\alpha\in\mathbb{N}} \subset M$ for $j = 1,\dots,m$, a nonnegative solution $u^0 \in W^{1,2}(X, \rho^{1-2\gamma})$ to (2-4) and nontrivial nonnegative functions $U_{a_j}^{\lambda_j} \in W^{1,2}(\mathbb{R}_+^{n+1}, y^{1-2\gamma})$ for some $\lambda_j > 0$ and $a_j \in \mathbb{R}^n$ as given in (1-13), satisfying, up to a subsequence,*

(1) $\mu_\alpha^j \to 0$ *as* $\alpha \to +\infty$, *for* $j = 1, \dots, m$;

(2) $\{x_\alpha^j\}_{\alpha\in\mathbb{N}}$ *converges on* M *as* $\alpha \to +\infty$, *for* $j = 1, \dots, m$;

(3) *As* $\alpha \to +\infty$,

$$\Big\| u_\alpha - u^0 - \sum_{j=1}^m \eta_\alpha^j u_\alpha^j \Big\|_{W^{1,2}(X,\rho^{1-2\gamma})} \to 0,$$

where

$$u_\alpha^j(z) = (\mu_\alpha^j)^{-\frac{n-2\gamma}{2}} U_{a_j}^{\lambda_j}\big((\mu_\alpha^j)^{-1}\varphi_{x_\alpha^j}^{-1}(z)\big),$$

for $z \in \varphi_{x_\alpha^j}(B_{r_0}^+(0))$, *and* $\varphi_{x_\alpha^j}$ *are Fermi coordinates centered at* $x_\alpha^j \in M$ *with* $r_0 > 0$ *small, and* η_α^j *are cutoff functions such that*

$$\eta_\alpha^j \equiv 1 \ \text{in}\ \varphi_{x_\alpha^j}(B_{r_0}^+(0)) \quad \text{and} \quad \eta_\alpha^j \equiv 0 \ \text{in}\ M \setminus \varphi_{x_\alpha^j}(B_{2r_0}^+(0));$$

(4) *The energies*

$$I_g^{\gamma,\alpha}(u_\alpha) - I_g^\infty(u^0) - m\widetilde{E}(U_{a_j}^{\lambda_j}) \to 0, \quad \text{as}\ \alpha \to +\infty;$$

(5) *For any* $1 \le i, j \le m$, $i \ne j$,

$$\frac{\mu_\alpha^i}{\mu_\alpha^j} + \frac{\mu_\alpha^j}{\mu_\alpha^i} + \frac{d_{\hat{h}}(x_\alpha^i, x_\alpha^j)^2}{\mu_\alpha^i \mu_\alpha^j} \to +\infty, \quad \text{as}\ \alpha \to +\infty.$$

Remark 1.4. (i) We call $\eta_\alpha^j u_\alpha^j$ a bubble for $j = 1, \dots, m$.

(ii) If $u_\alpha \to u^0$ strongly in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$, then $m = 0$.

Although the local case $\gamma = 1$ is well known [Druet et al. 2004; Struwe 1984], the most interesting point in the fractional case is the fact that one still has an energy decomposition into bubbles, and that these bubbles are noninterfering, which is surprising since our operator is nonlocal.

We finally recall that in the flat case, compactness problems for the fractional Laplacian were considered in the nice papers by Palatucci and Pisante [2014; 2015], and also the paper by Yan, Yang, and Yu [Yan et al. 2015].

This paper is organized as follows: In Section 2, we will first recall the definition of weighted Sobolev spaces and Palais–Smale sequences. Then we will derive a criterion for the strong convergence of a given Palais–Smale sequence. At last, ε-regularity estimates will be established. In Section 3, we will extract the first bubble from the Palais–Smale sequence which is not strongly convergent. In Section 4, we will give the proof of Theorem 1.3. Finally, some regularity estimates of the degenerate elliptic PDE are given in the Appendix.

2. Preliminary results

Most of the arguments in this section are analogous to the results in [Druet et al. 2004, Chapter 3]. For the convenience to the reader, we also prove these lemmas with the necessary modifications.

From now on we use $2^* = 2n/(n-2\gamma)$, $\gamma \in (0,1)$ for simplicity, and always assume that Palais–Smale sequences are all nonnegative. Moreover, the notation $o(1)$ will be taken with respect to the limit $\alpha \to +\infty$.

Definition 2.1. The weighted Sobolev space $W^{1,2}(X, \rho^{1-2\gamma})$ is defined as the closure of $\mathscr{C}^\infty(\overline{X})$ with norm

$$\|u\|_{W^{1,2}(X,\rho^{1-2\gamma})} = \left(\int_X \rho^{1-2\gamma}|\nabla u|_g^2\,dv_g + \int_M u^2\,d\sigma_{\hat{h}}\right)^{\frac{1}{2}} \tag{2-1}$$

where dv_g is the volume form of the asymptotically hyperbolic Riemannian manifold (X, g) and $d\sigma_{\hat{h}}$ is the volume form of the conformal infinity $(M, [\hat{h}])$.

Proposition 2.2. *The norm defined above is equivalent to the following traditional norm*

$$\|u\|^*_{W^{1,2}(X,\rho^{1-2\gamma})} = \left(\int_X \rho^{1-2\gamma}(|\nabla u|_g^2 + u^2)\,dv_g\right)^{\frac{1}{2}}. \tag{2-2}$$

On one hand, $\|\cdot\|$ can be controlled by $\|\cdot\|^*$. This is a easy consequence of the following two propositions. The first one is a trace Sobolev embedding on Euclidean space.

Proposition 2.3 [Jin and Xiong 2013]. *For any $u \in \mathscr{C}_0^\infty(\mathbb{R}^{n+1}_+)$,*

$$\left(\int_{\mathbb{R}^n} |u(x,0)|^{2^*}\,dx\right)^{\frac{2}{2^*}} \le S(n,\gamma)\int_{\mathbb{R}^{n+1}_+} y^{1-2\gamma}|\nabla u(x,y)|^2 dx dy$$

where

$$S(n,\gamma) = \frac{1}{2\pi^\gamma}\frac{\Gamma(\gamma)}{\Gamma(1-\gamma)}\frac{\Gamma(\frac{n-2\gamma}{2})}{\Gamma(\frac{n+2\gamma}{2})}\left(\frac{\Gamma(n)}{\Gamma(\frac{n}{2})}\right)^{\frac{2\gamma}{n}}.$$

Using a standard partition of unity argument, one obtains a weighted trace Sobolev inequality on an asymptotically hyperbolic manifold:

Proposition 2.4 [Jin and Xiong 2013]. *For any $\varepsilon > 0$, there exists a constant $C_\varepsilon > 0$ such that*

$$\left(\int_M |u|^{2^*}\,d\sigma_{\hat{h}}\right)^{\frac{2}{2^*}} \le (S(n,\gamma)+\varepsilon)\int_X \rho^{1-2\gamma}|\nabla u|_g^2\,dv_g + C_\varepsilon\int_X \rho^{1-2\gamma}u^2\,dv_g.$$

On the other hand, $\|\cdot\|^*$ can be controlled by $\|\cdot\|$, which is implied by the following proposition.

Proposition 2.5. *For any $u \in W^{1,2}(X, \rho^{1-2\gamma})$, there exists a constant $C > 0$ such that*
$$\int_X \rho^{1-2\gamma} u^2 \, dv_g \le C\left(\int_X \rho^{1-2\gamma} |\nabla u|_g^2 \, dv_g + \int_M u^2 \, d\sigma_{\hat{h}}\right).$$

Proof. We use a contradiction argument. Thus, assume that for any $\alpha \ge 1$ there exists u_α satisfying
$$\int_X \rho^{1-2\gamma} u_\alpha^2 \, dv_g \ge \alpha\left(\int_X \rho^{1-2\gamma} |\nabla u_\alpha|_g^2 \, dv_g + \int_M u_\alpha^2 \, d\sigma_{\hat{h}}\right).$$
Without loss of generality, we can assume that $\int_X \rho^{1-2\gamma} u_\alpha^2 \, dv_g = 1$. Then we have
$$\int_X \rho^{1-2\gamma} (|\nabla u_\alpha|_g^2 + u_\alpha^2) \, dv_g \le 1 + \frac{1}{\alpha}.$$
Then there exists a weakly convergent subsequence, also denoted by $\{u_\alpha\}$, such that $u_\alpha \rightharpoonup u_0$ in $W^{1,2}(X, \rho^{1-2\gamma}, \|\cdot\|^*)$.

Since
$$\lim_{\alpha\to\infty} \int_X \rho^{1-2\gamma} |\nabla u_\alpha|_g^2 \, dv_g = 0 \quad \text{and} \quad \lim_{\alpha\to\infty} \int_M u_\alpha^2 \, d\sigma_{\hat{h}} = 0,$$
we get that $u_0 \equiv 0$. On the other hand, via the following Proposition 2.6, the embedding $W^{1,2}(X, \rho^{1-2\gamma}, \|\cdot\|^*) \hookrightarrow L^2(X, \rho^{1-2\gamma})$ is compact. So we have
$$\int_X \rho^{1-2\gamma} u_0^2 \, dv_g = 1,$$
which contradicts the fact that $u_0 \equiv 0$. Then the proof is completed. $\square$

Proposition 2.6 [Jin and Xiong 2013; Kufner 1985; Di Nezza et al. 2012]. *Let $1 \le p \le q < \infty$ with $\frac{1}{n+1} > \frac{1}{p} - \frac{1}{q}$.*

(i) *Suppose $2-2\gamma \le p$. Then $W^{1,p}(X, \rho^{1-2\gamma}, \|\cdot\|^*)$ is compactly embedded in $L^q(X, \rho^{1-2\gamma})$ if*
$$\frac{2-2\gamma}{p(n+2-2\gamma)} > \frac{1}{p} - \frac{1}{q}.$$

(ii) *Suppose $2-2\gamma > p$. Then $W^{1,p}(X, \rho^{1-2\gamma}, \|\cdot\|^*)$ is compactly embedded in $L^q(X, \rho^{1-2\gamma})$ if and only if*
$$\frac{1}{(n+2-2\gamma)} > \frac{1}{p} - \frac{1}{q}.$$

We will always use the norm in $W^{1,2}(X, \rho^{1-2\gamma})$ in the following unless otherwise stated.

Definition 2.7. The weighted Sobolev space $\overline{W}^{1,2}(X, \rho^{1-2\gamma})$ is the closure of $\mathscr{C}_0^\infty(X)$ in $W^{1,2}(X, \rho^{1-2\gamma})$ with the norm

$$\|u\|_{\overline{W}^{1,2}(X,\rho^{1-2\gamma})} = \left(\int_X \rho^{1-2\gamma} |\nabla u|_g^2 \, dv_g\right)^{\frac{1}{2}}.$$

Now we define Palais–Smale sequences for the functional (1-11) precisely.

Definition 2.8. The sequence $\{u_\alpha\}_{\alpha\in\mathbb{N}} \subset W^{1,2}(X, \rho^{1-2\gamma})$ is called a Palais–Smale sequence for $\{I_g^{\gamma,\alpha}\}_{\alpha\in\mathbb{N}}$ if:

(i) $\{I_g^{\gamma,\alpha}(u_\alpha)\}_{\alpha\in\mathbb{N}}$ is uniformly bounded; and

(ii) as $\alpha \to +\infty$,

$$DI_g^{\gamma,\alpha}(u_\alpha) \to 0, \text{ strongly in } W^{1,2}(X, \rho^{1-2\gamma})',$$

where we have defined $W^{1,2}(X, \rho^{1-2\gamma})'$ as the dual space of $W^{1,2}(X, \rho^{2\gamma-1})$, i.e., for any $\theta \in W^{1,2}(X, \rho^{1-2\gamma})$,

$$\begin{aligned} DI_g^{\gamma,\alpha}(u_\alpha)\cdot\theta &= \int_X \rho^{1-2\gamma} \langle \nabla u_\alpha, \nabla\theta\rangle_g \, dv_g + \int_M Q_\alpha^\gamma u_\alpha \theta \, d\sigma_{\hat{h}} - \int_M u_\alpha^{2^*-1}\theta \, d\sigma_{\hat{h}} \\ &= o(\|\theta\|_{W^{1,2}(X,\rho^{1-2\gamma})}), \quad \text{as } \alpha \to +\infty. \end{aligned} \tag{2-3}$$

The main properties of Palais–Smale sequences are contained in the next several lemmas:

Lemma 2.9. *Let $\{u_\alpha\}_{\alpha\in\mathbb{N}} \subset W^{1,2}(X, \rho^{1-2\gamma})$ be a Palais–Smale sequence for the functionals $\{I_g^{\gamma,\alpha}\}_{\alpha\in\mathbb{N}}$, then $\{u_\alpha\}_{\alpha\in\mathbb{N}}$ is uniformly bounded in $W^{1,2}(X, \rho^{1-2\gamma})$.*

Proof. We can take $\theta = u_\alpha \in W^{1,2}(X, \rho^{1-2\gamma})$ as a test function in (ii) of Definition 2.8. Then, we get

$$\int_X \rho^{1-2\gamma} |\nabla u_\alpha|_g^2 \, dv_g + \int_M Q_\alpha^\gamma u_\alpha^2 \, d\sigma_{\hat{h}} = \int_M u_\alpha^{2^*} \, d\sigma_{\hat{h}} + o(\|u_\alpha\|_{W^{1,2}(X,\rho^{1-2\gamma})}),$$

which yields that

$$\begin{aligned} I_g^{\gamma,\alpha}(u_\alpha) &= \frac{1}{2}\int_X \rho^{1-2\gamma}|\nabla u_\alpha|_g^2 \, dv_g + \frac{1}{2}\int_M Q_\alpha^\gamma u_\alpha^2 \, d\sigma_{\hat{h}} - \frac{1}{2^*}\int_M u_\alpha^{2^*} \, d\sigma_{\hat{h}} \\ &= \frac{\gamma}{n}\int_M u_\alpha^{2^*} \, d\sigma_{\hat{h}} + o(\|u_\alpha\|_{W^{1,2}(X,\rho^{1-2\gamma})}). \end{aligned}$$

Since $\{I_g^{\gamma,\alpha}(u_\alpha)\}_{\alpha\in\mathbb{N}}$ is uniformly bounded by (i) of Definition 2.8, there exists a constant $C > 0$ such that

$$\int_M u_\alpha^{2^*} \, d\sigma_{\hat{h}} \le C + o(\|u_\alpha\|_{W^{1,2}(X,\rho^{1-2\gamma})}),$$

which by Hölder's inequality yields

$$\int_M u_\alpha^2\, d\sigma_{\hat h} \le C\left(\int_M u_\alpha^{2^*}\, d\sigma_{\hat h}\right)^{\frac{2}{2^*}} \le C + o\big(\|u_\alpha\|^{2/2^*}_{W^{1,2}(X,\,\rho^{1-2\gamma})}\big).$$

Note that since $|Q^\gamma_\alpha| \le C$ for some constant $C>0$, we can choose sufficiently large $C_1>0$ such that $C_1 + Q^\gamma_\alpha \ge 1$ on M. It follows that

$$\begin{aligned}
\|u_\alpha\|^2_{W^{1,2}(X,\,\rho^{1-2\gamma})}
&= \int_X \rho^{1-2\gamma}|\nabla u_\alpha|_g^2\, dv_g + \int_M u_\alpha^2\, d\sigma_{\hat h}\\
&\le \int_X \rho^{1-2\gamma}|\nabla u_\alpha|_g^2\, dv_g + \int_M Q^\gamma_\alpha u_\alpha^2\, d\sigma_{\hat h} + C_1\int_M u_\alpha^2\, d\sigma_{\hat h}\\
&\le \int_M u_\alpha^{2^*}\, d\sigma_{\hat h} + o(\|u_\alpha\|_{W^{1,2}(X,\,\rho^{1-2\gamma})}) + C + o\big(\|u_\alpha\|^{2/2^*}_{W^{1,2}(X,\,\rho^{1-2\gamma})}\big)\\
&\le C + o(\|u_\alpha\|_{W^{1,2}(X,\,\rho^{1-2\gamma})}) + o\big(\|u_\alpha\|^{2/2^*}_{W^{1,2}(X,\,\rho^{1-2\gamma})}\big),
\end{aligned}$$

from which we conclude that $\{u_\alpha\}_{\alpha\in\mathbb{N}}$ is uniformly bounded in $W^{1,2}(X,\,\rho^{1-2\gamma})$ since $2/2^* < 1$. $\square$

Remark 2.10. From Lemma 2.9, it is easy to see that there exists a function u^0 in $W^{1,2}(X,\,\rho^{1-2\gamma})$ such that $u_\alpha \rightharpoonup u^0$ weakly in $W^{1,2}(X,\,\rho^{1-2\gamma})$ as $\alpha\to+\infty$.

Proposition 2.11. *The function u^0 is nonnegative in $\overline{X}$.*

Proof. Using Proposition 2.4, we can easily get that $u_\alpha \to u^0$ in $L^2(M,\hat h)$ as $\alpha\to+\infty$, so we have $u_\alpha\to u^0$ almost everywhere on M. Noting that $u_\alpha \ge 0$ on M, we obtain that $u^0\ge 0$ on M. On the other hand, by Proposition 2.6 and by the equivalence of the norms $\|\cdot\|$ and $\|\cdot\|^*$, we have $u_\alpha\to u^0$ in $L^2(X,\rho^{1-2\gamma})$ as $\alpha\to+\infty$. For any $z\in X$, take $d_z<\operatorname{dist}(z,M)$; then we also have $u_\alpha\to u^0$ in $L^2(\mathfrak{B}^+_{d_z}(z),\rho^{1-2\gamma})$. Since $\rho^{1-2\gamma}$ is bounded below by a positive constant in $\mathfrak{B}^+_{d_z}(z)$, we get $u_\alpha\to u^0$ almost everywhere in $\mathfrak{B}^+_{d_z}(z)$, up to passing to a subsequence. Noting that $u_\alpha\ge 0$ in X, we obtain $u^0\ge 0$ in $\mathfrak{B}^+_{d_z}(z)$. Since z is arbitrary in X, we have $u^0\ge 0$ in X. Combining the above arguments, we conclude that $u\ge 0$ in $\overline{X}$. $\square$

Next we define the two limit functionals

$$I^\gamma_g(u) = \frac12\int_X \rho^{1-2\gamma}|\nabla u|_g^2\, dv_g - \frac{1}{2^*}\int_M |u|^{2^*}\, d\sigma_{\hat h}$$

and

$$I^{\gamma,\infty}_g(u) = \frac12\int_X \rho^{1-2\gamma}|\nabla u|_g^2\, dv_g + \frac12\int_M Q^\gamma_\infty u^2\, d\sigma_{\hat h} - \frac{1}{2^*}\int_M |u|^{2^*}\, d\sigma_{\hat h}.$$

Lemma 2.12. *Let $\{u_\alpha\}_{\alpha\in\mathbb{N}} \subset W^{1,2}(X, \rho^{1-2\gamma})$ be a Palais–Smale sequence for $\{I_g^{\gamma,\alpha}\}_{\alpha\in\mathbb{N}}$, and $u_\alpha \rightharpoonup u^0$ weakly in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$. We also set $\hat{u}_\alpha = u_\alpha - u^0 \in W^{1,2}(X, \rho^{1-2\gamma})$. Then,*

(i) *u^0 is a nonnegative weak solution to the limit equation*

$$\begin{cases} -\operatorname{div}(\rho^{1-2\gamma}\nabla u) = 0 & \text{in } X, \\ -\lim\limits_{\rho\to 0} \rho^{1-2\gamma}\partial_\rho u + Q_\infty^\gamma u = u^{2^*-1} & \text{on } M; \end{cases} \tag{2-4}$$

(ii) *$I_g^{\gamma,\alpha}(u_\alpha) = I_g^\gamma(\hat{u}_\alpha) + I_g^{\gamma,\infty}(u^0) + o(1)$ as $\alpha \to +\infty$;*

(iii) *$\{\hat{u}_\alpha\}_{\alpha\in\mathbb{N}}$ is a Palais–Smale sequence for I_g^γ.*

Proof. (i) As $\mathscr{C}^\infty(\overline{X})$ is dense in $W^{1,2}(X, \rho^{1-2\gamma})$, we only consider the proof in $\mathscr{C}^\infty(\overline{X})$. Let $\theta \in \mathscr{C}^\infty(\overline{X})$. Since $Q_\alpha^\gamma \to Q_\infty^\gamma$ in $L^2(M, \hat{h})$ as $\alpha \to +\infty$ and $u_\alpha \rightharpoonup u^0$ weakly in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$,

$$\int_M Q_\alpha^\gamma u_\alpha \theta \, d\sigma_{\hat{h}} = \int_M Q_\infty^\gamma u^0 \theta \, d\sigma_{\hat{h}} + o(1).$$

Passing to the limit in (2-3), we get easily that

$$\int_X \rho^{1-2\gamma}\langle\nabla u^0, \nabla\theta\rangle_g \, dv_g + \int_M Q_\infty^\gamma u^0\theta \, d\sigma_{\hat{h}} = \int_M (u^0)^{2^*-1}\theta \, d\sigma_{\hat{h}},$$

i.e., u^0 is a weak solution to the limit equation (2-4).

For the proof of (ii), recall that

$$\int_M Q_\alpha^\gamma u_\alpha^2 \, d\sigma_{\hat{h}} = \int_M Q_\infty^\gamma (u^0)^2 \, d\sigma_{\hat{h}} + o(1),$$

and

$$\begin{aligned} I_g^{\gamma,\alpha}(u_\alpha) &= \frac{1}{2}\int_X \rho^{1-2\gamma}|\nabla u_\alpha|_g^2 \, dv_g + \frac{1}{2}\int_M Q_\alpha^\gamma u_\alpha^2 \, d\sigma_{\hat{h}} - \frac{1}{2^*}\int_M u_\alpha^{2^*} \, d\sigma_{\hat{h}}, \\ I_g^{\gamma,\infty}(u^0) &= \frac{1}{2}\int_X \rho^{1-2\gamma}|\nabla u^0|_g^2 \, dv_g + \frac{1}{2}\int_M Q_\infty^\gamma (u^0)^2 d\sigma_{\hat{h}} - \frac{1}{2^*}\int_M (u^0)^{2^*} \, d\sigma_{\hat{h}}, \\ I_g^\gamma(\hat{u}_\alpha) &= \frac{1}{2}\int_X \rho^{1-2\gamma}|\nabla \hat{u}_\alpha|_g^2 \, dv_g - \frac{1}{2^*}\int_M |\hat{u}_\alpha|^{2^*} \, d\sigma_{\hat{h}}, \end{aligned}$$

where $\hat{u}_\alpha = u_\alpha - u^0$. Then,

$$\begin{aligned} &I_g^{\gamma,\alpha}(u_\alpha) - I_g^{\gamma,\infty}(u^0) - I_g^\gamma(\hat{u}_\alpha) \\ &\qquad = \int_X \rho^{1-2\gamma}\langle\nabla u^0, \nabla\hat{u}_\alpha\rangle_g \, dv_g - \frac{1}{2^*}\int_M \Phi_\alpha \, d\sigma_{\hat{h}} + o(1), \end{aligned}$$

where

$$\Phi_\alpha = |\hat{u}_\alpha + u^0|^{2^*} - |\hat{u}_\alpha|^{2^*} - |u^0|^{2^*}.$$

Note that $\hat{u}_\alpha \rightharpoonup 0$ weakly in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$, thus

$$\int_X \rho^{1-2\gamma} \langle \nabla u^0, \nabla \hat{u}_\alpha \rangle_g \, dv_g \to 0, \quad \text{as } \alpha \to \infty.$$

On the other hand, it is easy to check that there exists a constant $C > 0$, independent of α, such that

$$\big| |\hat{u}_\alpha + u^0|^{2^*} - |\hat{u}_\alpha|^{2^*} - |u^0|^{2^*} \big| \leq C \big(|\hat{u}_\alpha|^{2^*-1} |u^0| + |u^0|^{2^*-1} |\hat{u}_\alpha| \big).$$

As a consequence, since $\hat{u}_\alpha \rightharpoonup 0$ weakly in $L^{2^*}(M, \hat{h})$ by Proposition 2.4, we have

$$\int_M |\Phi_\alpha| \, d\sigma_{\hat{h}} \to 0, \quad \text{as } \alpha \to +\infty.$$

The proof of (ii) is completed.

(iii) For any $\theta \in \mathscr{C}^\infty(\overline{X})$, by (i) we have

$$DI_g^{\gamma,\infty}(u^0) \cdot \theta = 0.$$

Since, in addition,

$$\int_M Q_\alpha^\gamma u_\alpha \theta \, d\sigma_{\hat{h}} = \int_M Q_\infty^\gamma u^0 \theta \, d\sigma_{\hat{h}} + o(\|\theta\|_{W^{1,2}(X, \rho^{1-2\gamma})}),$$

then

$$DI_g^{\gamma,\alpha}(u_\alpha) \cdot \theta = DI_g^\gamma(\hat{u}_\alpha) \cdot \theta - \int_M \Psi_\alpha \theta \, d\sigma_{\hat{h}} + o(\|\theta\|_{W^{1,2}(X, \rho^{1-2\gamma})}), \tag{2-5}$$

where $\Psi_\alpha = |\hat{u}_\alpha + u^0|^{2^*-2}(\hat{u}_\alpha + u^0) - |\hat{u}_\alpha|^{2^*-2}\hat{u}_\alpha - |u^0|^{2^*-2}u^0$, and it is easy to check that there exists a constant $C > 0$ independent of α such that

$$|\Psi_\alpha| \leq C \big(|\hat{u}_\alpha|^{2^*-2} \, |u^0| + |\hat{u}_\alpha| \, |u^0|^{2^*-2} \big).$$

By Hölder's inequality and the fact $\hat{u}_\alpha \rightharpoonup 0$ weakly in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$,

$$\begin{aligned} &\int_M \Psi_\alpha \theta \, d\sigma_{\hat{h}} \\ &\quad \leq \Big(\big\| |\hat{u}_\alpha|^{2^*-2} \, |u^0| \big\|_{L^{2^*/(2^*-1)}(M)} + \big\| |\hat{u}_\alpha| \, |u^0|^{2^*-2} \big\|_{L^{2^*/(2^*-1)}(M)} \Big) \|\theta\|_{L^{2^*}(M)} \\ &\quad = o(1) \|\theta\|_{L^{2^*}(M)}. \end{aligned}$$

Thus from (2-5),

$$DI_g^{\gamma,\alpha}(u_\alpha) \cdot \theta = DI_g^\gamma(\hat{u}_\alpha) \cdot \theta + o(1) \|\theta\|_{L^{2^*}(M)},$$

which implies that $DI_g^\gamma(\hat{u}_\alpha) \to 0$ in $W^{1,2}(X, \rho^{1-2\gamma})'$ as $\alpha \to +\infty$, since $\{u_\alpha\}_{\alpha \in \mathbb{N}}$ is a Palais–Smale sequence for $\{I_g^{\gamma,\alpha}\}_{\alpha \in \mathbb{N}}$.

Finally, from (ii), we know that $\{\hat{u}_\alpha\}_{\alpha\in\mathbb{N}}$ is a Palais–Smale sequence for I_g^γ. This completes the proof of the lemma. □

Now we give a criterion for strong convergence of Palais–Smale sequences.

Lemma 2.13. *Let $\{\hat{u}_\alpha\}_{\alpha\in\mathbb{N}}$ be a Palais–Smale sequence for I_g^γ such that $\hat{u}_\alpha \rightharpoonup 0$ weakly in $W^{1,2}(X,\rho^{1-2\gamma})$ as $\alpha\to+\infty$. If $I_g^\gamma(\hat{u}_\alpha)\to\beta$ and*

$$\beta<\beta_0=\frac{\gamma}{n}(d_\gamma^*)^{-\frac{n}{2\gamma}}\Lambda_\gamma(M,[\hat{h}])^{\frac{n}{2\gamma}}, \tag{2-6}$$

then $\hat{u}_\alpha\to 0$ in $W^{1,2}(X,\rho^{1-2\gamma})$ as $\alpha\to+\infty$.

Proof. By Lemma 2.9 (here $Q_\alpha^\gamma\equiv 0$), there exists a constant $C>0$ such that $\|\hat{u}_\alpha\|_{W^{1,2}(X,\rho^{1-2\gamma})}\le C$ for all $\alpha\in\mathbb{N}$, so

$$\begin{aligned} DI_g^\gamma(\hat{u}_\alpha)\cdot\hat{u}_\alpha &= \int_X \rho^{1-2\gamma}|\nabla\hat{u}_\alpha|_g^2\,dv_g-\int_M|\hat{u}_\alpha|^{2^*}\,d\sigma_{\hat{h}}\\ &= o(\|\hat{u}_\alpha\|_{W^{1,2}(X,\rho^{1-2\gamma})})=o(1). \end{aligned}$$

Then note that $I_g^\gamma(\hat{u}_\alpha)\to\beta$ as $\alpha\to+\infty$, so

$$\begin{aligned} \beta+o(1) &= I_g^\gamma(\hat{u}_\alpha)\\ &= \frac{1}{2}\int_X\rho^{1-2\gamma}|\nabla\hat{u}_\alpha|_g^2\,dv_g-\frac{1}{2^*}\int_M|\hat{u}_\alpha|^{2^*}\,d\sigma_{\hat{h}}\\ &= \frac{\gamma}{n}\int_X\rho^{1-2\gamma}|\nabla\hat{u}_\alpha|_g^2\,dv_g+o(1)\\ &= \frac{\gamma}{n}\int_M|\hat{u}_\alpha|^{2^*}\,d\sigma_{\hat{h}}+o(1). \end{aligned} \tag{2-7}$$

On the other hand, in the positive curvature case, it was shown in [González and Qing 2013] that the γ-Yamabe constant (1-9) must be positive: $\Lambda_\gamma(M,[\hat{h}])>0$. Moreover, by definition,

$$\begin{aligned} \Lambda_\gamma(M,[\hat{h}])\Big(\int_M|\hat{u}_\alpha|^{2^*}\,d\sigma_{\hat{h}}\Big)^{\frac{2}{2^*}}&\\ \le d_\gamma^*\int_X\rho^{1-2\gamma}|\nabla\hat{u}_\alpha|_g^2\,dv_g&+\int_M Q_\gamma^{\hat{h}}\,\hat{u}_\alpha^2\,d\sigma_{\hat{h}}. \end{aligned} \tag{2-8}$$

where $d_\gamma^*>0$. We also know that $|Q_\gamma^{\hat{h}}|\le C$ on M^n. Note that, by Proposition 2.4, $\hat{u}_\alpha\rightharpoonup 0$ in $L^{2^*}(M,\hat{h})$ as $\alpha\to+\infty$, so

$$\int_M\hat{u}_\alpha^2\,d\sigma_{\hat{h}}\to 0,\quad\text{as }\alpha\to+\infty,$$

since the embedding $L^{2^*}(M,\hat{h}) \subset L^2(M,\hat{h})$ is compact. So we get from (2-7) and (2-8) that

$$\left(\frac{n}{\gamma}\beta + o(1)\right)^{\frac{2}{2^*}} \le d_\gamma^* \Lambda_\gamma(M,[\hat{h}])^{-1} \frac{n}{\gamma}\beta + o(1).$$

Taking $\alpha \to +\infty$, we must have $\beta = 0$ because of our initial condition (2-6). $\square$

Note that the Palais–Smale condition (ii) is the weak form of a Dirichlet-to-Neumann problem for a degenerate elliptic PDE. In fact, as $DI_g^\gamma(\hat{u}_\alpha) \to 0$ in $W^{1,2}(X,\rho^{1-2\gamma})'$, it follows that, for any $\psi \in W^{1,2}(X,\rho^{1-2\gamma})$,

$$\int_X \rho^{1-2\gamma} \langle \nabla \hat{u}_\alpha, \nabla \psi \rangle_g \, dv_g - \int_M |\hat{u}_\alpha|^{2^*-2} \hat{u}_\alpha \psi \, d\sigma_{\hat{h}} = o(1)\|\psi\|_{W^{1,2}(X,\rho^{1-2\gamma})}. \tag{2-9}$$

In particular, for any $\overline{\psi} \in \overline{W}^{1,2}(X,\rho^{1-2\gamma})$,

$$\int_X \rho^{1-2\gamma} \langle \nabla \hat{u}_\alpha, \nabla \overline{\psi} \rangle_g \, dv_g = o(1)\|\overline{\psi}\|_{\overline{W}^{1,2}(X,\rho^{1-2\gamma})},$$

which is precisely the weak formulation of the asymptotic equation

$$-\operatorname{div}(\rho^{1-2\gamma} \nabla \hat{u}_\alpha) = o(1) \quad \text{in } X. \tag{2-10}$$

Multiplying both sides of (2-10) by $\psi \in W^{1,2}(X,\rho^{1-2\gamma})$ and integrating by parts, we obtain

$$\int_M \lim_{\rho\to 0} \rho^{1-2\gamma} \partial_\rho \hat{u}_\alpha \psi \, d\sigma_{\hat{h}} + \int_X \rho^{1-2\gamma} \langle \nabla \hat{u}_\alpha, \nabla \psi \rangle_g \, dv_g = o(1)\|\psi\|_{W^{1,2}(X,\rho^{1-2\gamma})},$$

which, combined with (2-9), yields that

$$\int_M \lim_{\rho\to 0} \rho^{1-2\gamma} \partial_\rho \hat{u}_\alpha \psi \, d\sigma_{\hat{h}} + \int_M |\hat{u}_\alpha|^{2^*-2} \hat{u}_\alpha \psi \, d\sigma_{\hat{h}} = o(1)\|\psi\|_{W^{1,2}(X,\rho^{1-2\gamma})},$$

and this is precisely the boundary equation in the weak sense

$$-\lim_{\rho\to 0} \rho^{1-2\gamma} \partial_\rho \hat{u}_\alpha = |\hat{u}_\alpha|^{2^*-2} \hat{u}_\alpha + o(1) \quad \text{on } M. \tag{2-11}$$

For (2-10) and (2-11) with $\{\hat{u}_\alpha\}_{\alpha\in\mathbb{N}}$, we have the following energy estimate, which will play an important role in the proof of the strong convergence in Section 3. We use the notation $\mathfrak{B}_r^+$ instead of $\mathfrak{B}_r^+(0)$ for convenience.

Lemma 2.14. (ε*-regularity estimates*) *Suppose that* $\{v_\alpha\}_{\alpha\in\mathbb{N}}$ *satisfies the following asymptotic boundary value problem*

$$\begin{cases} -\operatorname{div}(\rho^{1-2\gamma} \nabla v_\alpha) = o(1) & \text{in } X, \\ -\lim_{\rho\to 0} \rho^{1-2\gamma} \partial_\rho v_\alpha = |v_\alpha|^{2^*-2} v_\alpha + o(1) & \text{on } M. \end{cases} \tag{2-12}$$

If there exists small $\varepsilon > 0$ *depending on* n *and* γ *such that* $\int_{\partial'\mathfrak{B}_{2r}^+} |v_\alpha|^{2^*} d\sigma_{\hat{h}} \le \varepsilon$ *uniformly in* α *for some small* $r > 0$, *then*

$$\int_{\mathfrak{B}_r^+} \rho^{1-2\gamma} |\nabla v_\alpha|_g^2 \, dv_g \le \frac{C}{r^2} \int_{\mathfrak{B}_{2r}^+} \rho^{1-2\gamma} v_\alpha^2 \, dv_g + C \int_{\partial'\mathfrak{B}_{2r}^+} v_\alpha^2 \, d\sigma_{\hat{h}} + o(1) \int_{\mathfrak{B}_{2r}^+} |v_\alpha| \, dv_g,$$

where $C = C(n, \varepsilon, \gamma)$ *independent of* α.

Proof. Let η be a smooth cutoff function in $\overline{X}$ such that $0 \le \eta \le 1$, $\eta \equiv 1$ in $\mathfrak{B}_r^+(0)$ and $\eta \equiv 0$ in $\overline{X} \setminus \mathfrak{B}_{2r}^+(0)$. Multiplying both sides of the first equation in (2-12) by $\eta^2 v_\alpha$, integrating by parts and substituting the second equation in (2-12), we get

$$\begin{aligned}\int_{\mathfrak{B}_{2r}^+} \rho^{1-2\gamma} \langle \nabla v_\alpha, \nabla(\eta^2 v_\alpha) \rangle_g \, dv_g &= -\int_{\partial'\mathfrak{B}_{2r}^+} \lim_{\rho \to 0} \rho^{1-2\gamma} (\partial_\rho v_\alpha) \eta^2 v_\alpha \, d\sigma_{\hat{h}} + o(1) \int_{\mathfrak{B}_{2r}^+} \eta^2 v_\alpha \, dv_g \\ &= \int_{\partial'\mathfrak{B}_{2r}^+} \eta^2 |v_\alpha|^{2^*} d\sigma_{\hat{h}} + o(1) \int_{\mathfrak{B}_{2r}^+} \eta^2 v_\alpha \, dv_g,\end{aligned}$$

so we have

$$\begin{aligned}\int_{\mathfrak{B}_{2r}^+} \rho^{1-2\gamma} \eta^2 |\nabla v_\alpha|_g^2 \, dv_g &= -\int_{\mathfrak{B}_{2r}^+} \rho^{1-2\gamma} 2\eta v_\alpha \langle \nabla v_\alpha, \nabla \eta \rangle_g \, dv_g \\ &\quad + \int_{\partial'\mathfrak{B}_{2r}^+} \eta^2 |v_\alpha|^{2^*} d\sigma_{\hat{h}} + o(1) \int_{\mathfrak{B}_{2r}^+} \eta^2 |v_\alpha| \, dv_g \\ &\le \frac{1}{2} \int_{\mathfrak{B}_{2r}^+} \eta^2 \rho^{1-2\gamma} |\nabla v_\alpha|_g^2 \, dv_g + 2 \int_{\mathfrak{B}_{2r}^+} \rho^{1-2\gamma} |\nabla \eta|_g^2 v_\alpha^2 \, dv_g \\ &\quad + \int_{\partial'\mathfrak{B}_{2r}^+} \eta^2 |v_\alpha|^{2^*} d\sigma_{\hat{h}} + o(1) \int_{\mathfrak{B}_{2r}^+} \eta^2 |v_\alpha| \, dv_g,\end{aligned}$$

which implies that

$$\begin{aligned}&\int_{\mathfrak{B}_{2r}^+} \rho^{1-2\gamma} \eta^2 |\nabla v_\alpha|_g^2 \, dv_g \\ &\quad \le 4 \int_{\mathfrak{B}_{2r}^+} \rho^{1-2\gamma} |\nabla \eta|_g^2 v_\alpha^2 \, dv_g + 2 \int_{\partial'\mathfrak{B}_{2r}^+} \eta^2 |v_\alpha|^{2^*} d\sigma_{\hat{h}} + o(1) \int_{\mathfrak{B}_{2r}^+} \eta^2 |v_\alpha| \, dv_g \\ &\quad \le \frac{C}{r^2} \int_{\mathfrak{B}_{2r}^+} \rho^{1-2\gamma} v_\alpha^2 \, dv_g + 2 \int_{\partial'\mathfrak{B}_{2r}^+} (\eta v_\alpha)^2 |v_\alpha|^{2^*-2} d\sigma_{\hat{h}} + o(1) \int_{\mathfrak{B}_{2r}^+} \eta^2 |v_\alpha| \, dv_g.\end{aligned}$$

By Hölder's inequality and our initial hypothesis,

$$\begin{aligned}\int_{\partial'\mathfrak{B}_{2r}^+}(\eta v_\alpha)^2|v_\alpha|^{2^*-2}\,d\sigma_{\hat{h}} &\le \left(\int_{\partial'\mathfrak{B}_{2r}^+}|\eta v_\alpha|^{2^*}\,d\sigma_{\hat{h}}\right)^{\frac{2}{2^*}}\left(\int_{\partial'\mathfrak{B}_{2r}^+}|v_\alpha|^{2^*}\,d\sigma_{\hat{h}}\right)^{\frac{2^*-2}{2^*}}\\ &\le \varepsilon^{\frac{2^*-2}{2^*}}\left(\int_{\partial'\mathfrak{B}_{2r}^+}|\eta v_\alpha|^{2^*}\,d\sigma_{\hat{h}}\right)^{\frac{2}{2^*}}.\end{aligned}$$

Then it follows from above that

$$\begin{aligned}&\int_{\mathfrak{B}_{2r}^+}\rho^{1-2\gamma}|\nabla(\eta v_\alpha)|_g^2\,dv_g\\ &\quad\le 2\int_{\mathfrak{B}_{2r}^+}\rho^{1-2\gamma}(|\nabla\eta|_g^2 v_\alpha^2+\eta^2|\nabla v_\alpha|_g^2)\,dv_g\\ &\quad\le \frac{C}{r^2}\int_{\mathfrak{B}_{2r}^+}\rho^{1-2\gamma}v_\alpha^2\,dv_g + C\varepsilon^{\frac{2^*-2}{2^*}}\left(\int_{\partial'\mathfrak{B}_{2r}^+}|\eta v_\alpha|^{2^*}\,d\sigma_{\hat{h}}\right)^{\frac{2}{2^*}} + o(1)\int_{\mathfrak{B}_{2r}^+}\eta^2 v_\alpha\,dv_g.\end{aligned}$$

The trace Sobolev inequality on our manifold setting (Proposition 2.4) gives that

$$\left(\int_{\partial'\mathfrak{B}_{2r}^+}|\eta v_\alpha|^{2^*}\,d\sigma_{\hat{h}}\right)^{\frac{2}{2^*}} \le C\int_{\mathfrak{B}_{2r}^+}\rho^{1-2\gamma}|\nabla(\eta v_\alpha)|_g^2\,dv_g + C\int_{\partial'\mathfrak{B}_{2r}^+}(\eta v_\alpha)^2\,d\sigma_{\hat{h}}.$$

Therefore,

$$\begin{aligned}&\int_{\mathfrak{B}_{2r}^+}\rho^{1-2\gamma}|\nabla(\eta v_\alpha)|_g^2\,dv_g\\ &\qquad\le \frac{C}{r^2}\int_{\mathfrak{B}_{2r}^+}\rho^{1-2\gamma}v_\alpha^2\,dv_g + C\varepsilon^{\frac{2^*-2}{2^*}}\int_{\mathfrak{B}_{2r}^+}\rho^{1-2\gamma}|\nabla(\eta v_\alpha)|_g^2\,dv_g\\ &\qquad\qquad + C\varepsilon^{\frac{2^*-2}{2^*}}\int_{\partial'\mathfrak{B}_{2r}^+}(\eta v_\alpha)^2\,d\sigma_{\hat{h}} + o(1)\int_{\mathfrak{B}_{2r}^+}\eta^2|v_\alpha|\,dv_g.\end{aligned}$$

Now, fix $r>0$ small such that ε is small enough to satisfy $C\varepsilon^{\frac{2^*-2}{2^*}}\le\frac{1}{2}$. Then,

$$\begin{aligned}&\int_{\mathfrak{B}_{r}^+}\rho^{1-2\gamma}|\nabla v_\alpha|_g^2\,dv_g\\ &\qquad\le \frac{C}{r^2}\int_{\mathfrak{B}_{2r}^+}\rho^{1-2\gamma}v_\alpha^2\,dv_g + C\int_{\partial'\mathfrak{B}_{2r}^+}v_\alpha^2\,d\sigma_{\hat{h}} + o(1)\int_{\mathfrak{B}_{2r}^+}|v_\alpha|\,dv_g. \quad\square\end{aligned}$$

3. The first bubble argument

In this section, we focus on the blow up analysis of a Palais–Smale sequence which are not strongly convergent. In particular, using the ε-regularity estimates

(Lemma 2.14), we can figure out the first bubble. We will also show that the Palais–Smale sequence obtained by subtracting a bubble is also Palais–Smale sequence and that the energy is splitting.

Lemma 3.1. *Let $\{\hat{u}_\alpha\}_{\alpha\in\mathbb{N}}$ be a Palais–Smale sequence for I_g^γ such that $\hat{u}_\alpha \rightharpoonup 0$ weakly in $W^{1,2}(X, \rho^{1-2\gamma})$, but not strongly as $\alpha \to +\infty$. Then, there exists a sequence of real numbers $\{\mu_\alpha > 0\}_{\alpha\in\mathbb{N}}$, $\mu_\alpha \to 0$ as $\alpha \to +\infty$, a converging sequence of points $\{x_\alpha\}_{\alpha\in\mathbb{N}} \subset M$, and a nontrivial solution u to the equation*

$$\begin{cases} -\operatorname{div}(y^{1-2\gamma}\nabla u) = 0 & \text{in } \mathbb{R}^{n+1}_+, \\ -\lim\limits_{y\to 0} y^{1-2\gamma}\partial_y u = |u|^{2^*-2}u & \text{on } \mathbb{R}^n, \end{cases} \tag{3-1}$$

such that, up to a subsequence, if we take

$$\hat{v}_\alpha(z) = \hat{u}_\alpha(z) - \eta_\alpha(z)\mu_\alpha^{-\frac{n-2\gamma}{2}} u(\mu_\alpha^{-1}\varphi_{x_\alpha}^{-1}(z)), \quad z \in \varphi_{x_\alpha}(B^+_{2r_0}(0)),$$

where r_0, $\eta_\alpha(z)$, and $\varphi_{x_\alpha}(z)$ are the same as in Theorem 1.3, then we have the following three conclusions:

(1) *$\hat{v}_\alpha \rightharpoonup 0$ weakly in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$;*

(2) *$\{\hat{v}_\alpha\}_{\alpha\in\mathbb{N}}$ is also a Palais–Smale sequence for I_g^γ;*

(3) *$I_g^\gamma(\hat{v}_\alpha) = I_g^\gamma(\hat{u}_\alpha) - \widetilde{E}(u) + o(1)$ as $\alpha \to +\infty$.*

Proof. Without loss of generality, we assume that $\hat{u}_\alpha \in \mathscr{C}^\infty(\overline{X})$. By the proof of Lemma 2.13,

$$I_g^\gamma(\hat{u}_\alpha) = \frac{\gamma}{n}\int_X \rho^{1-2\gamma}|\nabla\hat{u}_\alpha|_g^2\, dv_g + o(1) = \frac{\gamma}{n}\int_M |\hat{u}_\alpha|^{2^*}\, d\sigma_{\hat{h}} + o(1).$$

Note that $\{\hat{u}_\alpha\}_{\alpha\in\mathbb{N}}$ is uniformly bounded in $W^{1,2}(X, \rho^{1-2\gamma})$ by Lemma 2.9, so there exist a subsequence, also denoted by $\{\hat{u}_\alpha\}_{\alpha\in\mathbb{N}}$ and a nonnegative constant β, such that

$$I_g^\gamma(\hat{u}_\alpha) = \beta + o(1), \quad \text{as } \alpha \to +\infty.$$

Since $\hat{u}_\alpha \rightharpoonup 0$ weakly in $W^{1,2}(X, \rho^{1-2\gamma})$ but not strongly as $\alpha \to +\infty$, again by Lemma 2.13,

$$\lim_{\alpha\to+\infty}\int_M |\hat{u}_\alpha|^{2^*}\, d\sigma_{\hat{h}} = \frac{n}{\gamma}\beta \geq \frac{n}{\gamma}\beta_0.$$

We will decompose the rest of the proof into several steps:

Step 1. Pick up the likely blow up points.

Claim 1. *For any $t_0 > 0$ small, there exist $x_0 \in M$ and $\varepsilon_0 > 0$ such that, up to a subsequence,*

$$\int_{\mathfrak{D}_{t_0}(x_0)} |\hat{u}_\alpha|^{2^*}\, d\sigma_{\hat{h}} \geq \varepsilon_0.$$

Proof. If the claim is not true, then there exists $t > 0$ small, such that for any $x \in M$,

$$\int_{\mathfrak{D}_t(x)} |\hat{u}_\alpha|^{2^*} \, d\sigma_{\hat{h}} \to 0, \quad \alpha \to +\infty.$$

On the other hand, since $(M, \hat{h})$ is compact and $M \subset \bigcup_{x \in M} \mathfrak{D}_t(x)$, there exists an integer $N \geq 1$ such that $M \subset \bigcup_{i=1}^{N} \mathfrak{D}_t(x_i)$. Thus,

$$\int_M |\hat{u}_\alpha|^{2^*} \, d\sigma_{\hat{h}} \leq \sum_{i=1}^{N} \int_{\mathfrak{D}_t(x_i)} |\hat{u}_\alpha|^{2^*} \, d\sigma_{\hat{h}} \to 0, \quad \alpha \to +\infty,$$

which is a contradiction. □

For $t > 0$, we set

$$\omega_\alpha(t) = \max_{x \in M} \int_{\mathfrak{D}_t(x)} |\hat{u}_\alpha|^{2^*} \, d\sigma_{\hat{h}}.$$

Then, by Claim 1, there exists $x_\alpha \in M$ such that

$$\omega_\alpha(t_0) = \int_{\mathfrak{D}_{t_0}(x_\alpha)} |\hat{u}_\alpha|^{2^*} \, d\sigma_{\hat{h}} \geq \varepsilon_0.$$

Note that

$$\int_{\mathfrak{D}_t(x_\alpha)} |\hat{u}_\alpha|^{2^*} \, d\sigma_{\hat{h}} \to 0, \quad \text{as } t \to 0.$$

Hence, for any $\varepsilon \in (0, \varepsilon_0)$, there exists $t_\alpha \in (0, t_0)$ such that

$$\varepsilon = \int_{\mathfrak{D}_{t_\alpha}(x_\alpha)} |\hat{u}_\alpha|^{2^*} \, d\sigma_{\hat{h}}. \tag{3-2}$$

Step 2. At each likely blow up point, we will establish weak convergence of a Palais–Smale sequence after properly rescaling.

For $r_0 > 0$ small, consider the Fermi coordinates at the likely blow up point $x_\alpha \in M$, $\varphi_{x_\alpha} : B^+_{2r_0}(0) \to X$. Here we restrict r_0 to $r_0 \leq i_g(X)/2$, where $i_g(X)$ is the injectivity radius of X. Then, for any $0 < \mu_\alpha \leq 1$, we define

$$\begin{aligned}
\tilde{u}_\alpha(z) &= \mu_\alpha^{(n-2\gamma)/2} \hat{u}_\alpha(\varphi_{x_\alpha}(\mu_\alpha z)), \\
\tilde{g}_\alpha(z) &= (\varphi^*_{x_\alpha} g)(\mu_\alpha z), \\
\tilde{h}_\alpha(x) &= (\varphi^*_{x_\alpha} \hat{h})(\mu_\alpha x),
\end{aligned}$$

if $z \in B^+_{\mu_\alpha^{-1} r_0}(0)$ and $x \in \partial' B^+_{\mu_\alpha^{-1} r_0}(0)$.

Given $z_0 \in \mathbb{R}^{n+1}_+$ and $r > 0$ such that $|z_0| + r < \mu_\alpha^{-1} r_0$, we have

$$\int_{B^+_r(z_0)} \tilde{\rho}_\alpha^{1-2\gamma} |\nabla \tilde{u}_\alpha|^2_{\tilde{g}_\alpha} \, dv_{\tilde{g}_\alpha} = \int_{\varphi_{x_\alpha}(\mu_\alpha B^+_r(z_0))} \rho^{1-2\gamma} |\nabla \hat{u}_\alpha|^2_g \, dv_g,$$

where

$$\tilde{\rho}_\alpha(z) = \mu_\alpha^{-1}\rho(\varphi_{x_\alpha}(\mu_\alpha z))$$

and $|d\tilde{\rho}_\alpha|_{\tilde{g}_\alpha} = 1$ on $\partial' B_r^+(z_0)$ since $|d\rho|_g = 1$ on M.

On the other hand, if $z_0 \in \mathbb{R}^n$, and $|z_0| + r < \mu_\alpha^{-1} r_0$, then

$$\int_{D_r(z_0)} |\tilde{u}_\alpha|^{2^*}\, d\sigma_{\tilde{h}_\alpha} = \int_{\varphi_{x_\alpha}(\mu_\alpha D_r(z_0))} |\hat{u}_\alpha|^{2^*}\, d\sigma_{\hat{h}} \le \int_{\mathfrak{D}_{2\mu_\alpha r}(\varphi_{x_\alpha}(\mu_\alpha z_0))} |\hat{u}_\alpha|^{2^*}\, d\sigma_{\hat{h}}.$$

Here we have used that $\varphi_{x_\alpha}(\mu_\alpha D_r(z_0)) = \varphi_{x_\alpha}(D_{\mu_\alpha r}(\mu_\alpha z_0))$, and for $x, y \in \mathbb{R}^n$, with $|x| < r_0$, $|y| < r_0$, we have $\frac{1}{2}|x-y| \le d_g(\varphi_{x_\alpha}(x), \varphi_{x_\alpha}(y)) \le 2|x-y|$.

Next, take $r \in (0, r_0)$ and choose t_0 in Claim 1 such that $0 < t_0 \le 2r$. For any $\varepsilon \in (0, \varepsilon_0)$, with ε to be determined later, and $t_\alpha \in (0, t_0)$, let

$$0 < \mu_\alpha = \tfrac{1}{2} r^{-1} t_\alpha \le \tfrac{1}{2} r^{-1} t_0 \le 1.$$

Then, by the definition of ε from (3-2), if $|z_0| + r < \mu_\alpha^{-1} r_0$,

$$\int_{\partial' B_r^+(z_0)} |\tilde{u}_\alpha|^{2^*}\, d\sigma_{\tilde{h}_\alpha} \le \varepsilon. \tag{3-3}$$

Note that $\varphi_{x_\alpha}(\partial' B_{2r\mu_\alpha}^+(0)) = \mathfrak{D}_{t_\alpha}(x_\alpha)$, we have

$$\begin{aligned}\varepsilon = \int_{\mathfrak{D}_{t_\alpha}(x_\alpha)} |\hat{u}_\alpha|^{2^*}\, d\sigma_{\hat{h}} &= \int_{\varphi_{x_\alpha}(\partial' B_{2r\mu_\alpha}^+(0))} |\hat{u}_\alpha|^{2^*}\, d\sigma_{\hat{h}} \\ &= \int_{\varphi_{x_\alpha}(\mu_\alpha \partial' B_{2r}^+(0))} |\hat{u}_\alpha|^{2^*}\, d\sigma_{\hat{h}} = \int_{\partial' B_{2r}^+(0)} |\tilde{u}_\alpha|^{2^*}\, d\sigma_{\tilde{h}_\alpha}.\end{aligned}$$

This $r_0 > 0$ can be chosen smaller again, such that for any $0 < \mu \le 1$ and any $x_0 \in M$, we can assume that

$$\begin{aligned}\frac{1}{2}\int_{\mathbb{R}_+^{n+1}} y^{1-2\gamma}|\nabla u|^2\, dx\, dy &\le \int_{\mathbb{R}_+^{n+1}} \tilde{\rho}_{x_0,\mu}^{1-2\gamma} |\nabla u|^2_{\tilde{g}_{x_0,\mu}}\, dv_{\tilde{g}_{x_0,\mu}} \\ &\le 2\int_{\mathbb{R}_+^{n+1}} y^{1-2\gamma}|\nabla u|^2\, dx\, dy,\end{aligned} \tag{3-4}$$

where $u \in \overline{W}^{1,2}(\mathbb{R}_+^{n+1}, \tilde{\rho}_{x_0,\mu}^{1-2\gamma})$, $\operatorname{supp}(u) \subset B_{2\mu^{-1}r_0}^+(0)$, $\tilde{\rho}_{x_0,\mu}(z) = \mu^{-1}\rho(\varphi_{x_0}(\mu z))$, and $\tilde{g}_{x_0,\mu}(z) = (\varphi_{x_0}^* g)(\mu z)$. And for $u \in L^1(\mathbb{R}^n)$ such that $\operatorname{supp}(u) \subset \partial' B_{2\mu^{-1}r_0}^+(0)$, we can also assume that

$$\frac{1}{2}\int_{\mathbb{R}^n} |u|\, dx \le \int_{\mathbb{R}^n} |u|\, d\sigma_{\tilde{h}_{x_0,\mu}} \le 2\int_{\mathbb{R}^n} |u|\, dx,$$

where $\tilde{h}_{x_0,\mu}(x) = (\varphi_{x_0}^* \hat{h})(\mu x)$.

Let $\tilde\eta \in \mathscr{C}_0^\infty(\mathbb{R}^{n+1}_+)$ be a cutoff function satisfying $0 \le \tilde\eta \le 1$,

$$\tilde\eta \equiv \begin{cases} 1, & \text{in } B^+_{1/4}(0), \\ 0, & \text{in } \mathbb{R}^{n+1}_+ \setminus B^+_{3/4}(0), \end{cases}$$

and set $\tilde\eta_\alpha(z) = \tilde\eta(r_0^{-1}\mu_\alpha z)$.

Claim 2. *$\{\tilde\eta_\alpha \tilde u_\alpha\}_{\alpha\in\mathbb{N}}$ is uniformly bounded in $W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$.*

Proof. Note that

$$\begin{aligned}
\int_{\mathbb{R}^{n+1}_+} \tilde\rho_\alpha^{1-2\gamma} |\nabla(\tilde\eta_\alpha \tilde u_\alpha)|^2_{\tilde g_\alpha}\, dv_{\tilde g_\alpha} &+ \int_{\mathbb{R}^{n+1}_+} \tilde\rho_\alpha^{1-2\gamma} (\tilde\eta_\alpha \tilde u_\alpha)^2\, dv_{\tilde g_\alpha} \\
&\le \int_{\mathbb{R}^{n+1}_+} \tilde\rho_\alpha^{1-2\gamma} (2|\nabla \tilde\eta_\alpha|^2_{\tilde g_\alpha} + \tilde\eta_\alpha^2)\tilde u_\alpha^2\, dv_{\tilde g_\alpha} + 2\int_{\mathbb{R}^{n+1}_+} \tilde\rho_\alpha^{1-2\gamma} \tilde\eta_\alpha^2 |\nabla \tilde u_\alpha|^2_{\tilde g_\alpha}\, dv_{\tilde g_\alpha} \\
&\le C\int_X \rho^{1-2\gamma} \hat u_\alpha^2\, dv_g + C\int_X \rho^{1-2\gamma} |\nabla \hat u_\alpha|^2_g\, dv_g \le C,
\end{aligned}$$

since $\{\hat u_\alpha\}_{\alpha\in\mathbb{N}}$ is uniformly bounded in $W^{1,2}(X, \rho^{1-2\gamma})$. Combining this with (3-4), we get that $\{\tilde\eta_\alpha \tilde u_\alpha\}_{\alpha\in\mathbb{N}}$ is uniformly bounded in $W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$. □

Due to the weak compactness of $W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$, there exists some u in $W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$ such that $\tilde\eta_\alpha \tilde u_\alpha \rightharpoonup u$ in $W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$ as $\alpha \to +\infty$.

Step 3. The weak convergence is in fact strong via ε-regularity estimates.

Claim 3. *Let $r_1 = \frac{r_0}{8}$. Then, there exists $\varepsilon_1 = \varepsilon_1(\gamma, n)$ such that for any $0 < r < r_1$, $0 < \varepsilon < \min\{\varepsilon_0, \varepsilon_1\}$, we have $\tilde\eta_\alpha \tilde u_\alpha \to u$ in $W^{1,2}(B^+_{2r}(0), y^{1-2\gamma})$ as $\alpha \to +\infty$.*

Proof. Given r sufficiently small, to be determined later, for any $z_0 \in \mathbb{R}^{n+1}_+$, let $\psi \in \mathscr{C}_0^\infty(B_r^+(z_0)) \cap W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$. Let

$$\hat\psi_\alpha(z) = \mu_\alpha^{-\frac{n-2\gamma}{2}} \psi(\mu_\alpha^{-1}\varphi_{x_\alpha}^{-1}(z)) \quad \text{for } z \in \varphi_{x_\alpha}(B_r^+(z_0)).$$

Since $\{\hat u_\alpha\}$ satisfies the asymptotic equation (2-10),

$$\begin{aligned}
o(1)\,\|\psi\|_{\overline{W}^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})} &= o(1)\,\|\hat\psi_\alpha\|_{\overline{W}^{1,2}(X, \rho^{1-2\gamma})} \\
&= \int_{\varphi_{x_\alpha}(\mu_\alpha B_r^+(z_0))} \rho^{1-2\gamma} \langle \nabla \hat u_\alpha, \nabla \hat\psi_\alpha\rangle_g\, dv_g \\
&= \int_{B_r^+(z_0)} (\mu_\alpha^{-1}\rho)^{1-2\gamma} \langle \nabla(\tilde\eta_\alpha \tilde u_\alpha), \nabla\psi\rangle_{\tilde g_\alpha}\, dv_{\tilde g_\alpha},
\end{aligned}$$

since $\tilde\eta$ is supported in $B^+_{3/4}(0)$ and $\tilde\eta \equiv 1$ in $B^+_{1/4}(0)$. Also, note that since $\tilde\eta_\alpha(z) = \tilde\eta(\mu_\alpha r_0^{-1} z)$, we have $\tilde\eta_\alpha \equiv 1$ in $B^+_{1/4\mu_\alpha^{-1} r_0}$; thus, we need $|z_0| + r < \frac{1}{4}\mu_\alpha^{-1} r_0$.

It is easy to check that $\mu_\alpha^{-1}\rho \to y$ as $\alpha \to +\infty$ since $|d(\mu_\alpha^{-1}\rho)|_{\tilde{g}_\alpha} = 1$ on $\mathbb{R}^n$ and $\tilde{g}_\alpha \to (|dx|^2 + dy^2)$. Then we have the asymptotic equation

$$-\operatorname{div}(y^{1-2\gamma}\nabla(\tilde{\eta}_\alpha \tilde{u}_\alpha)) = o(1) \quad \text{in } B_r^+(z_0). \tag{3-5}$$

Since $\tilde{\eta}_\alpha \tilde{u}_\alpha \rightharpoonup u$ weakly in $W^{1,2}(\mathbb{R}_+^{n+1}, y^{1-2\gamma})$, we simultaneously get that

$$-\operatorname{div}(y^{1-2\gamma}\nabla u) = 0 \quad \text{in } B_r^+(z_0). \tag{3-6}$$

Now, let $\psi \in W^{1,2}(B_r^+(z_0), y^{1-2\gamma})$. Then, multiplying both sides of (3-5) by ψ and integrating by parts, we get

$$\begin{aligned} o(1)\,\|\psi\|_{W^{1,2}(B_r^+(z_0), y^{1-2\gamma})} = \int_{\partial' B_r^+(z_0)} \lim_{y\to 0} y^{1-2\gamma}\partial_y(\tilde{\eta}_\alpha \tilde{u}_\alpha)\psi \, d\sigma_{\tilde{h}_\alpha} \\ + \int_{B_r^+(z_0)} y^{1-2\gamma}\langle \nabla(\tilde{\eta}_\alpha \tilde{u}_\alpha), \nabla\psi\rangle_{\tilde{g}_\alpha}\, dv_{\tilde{g}_\alpha}. \end{aligned} \tag{3-7}$$

On the other hand, using (2-10), (2-11), and the definition of $\hat{\psi}_\alpha$,

$$\begin{aligned} &\int_{B_r^+(z_0)} y^{1-2\gamma}\langle \nabla(\tilde{\eta}_\alpha \tilde{u}_\alpha), \nabla\psi\rangle_{\tilde{g}_\alpha}\, dv_{\tilde{g}_\alpha} \\ &\qquad = \int_{\varphi_{x_\alpha}(\mu_\alpha B_r^+(z_0))} \rho^{1-2\gamma}\langle \nabla\hat{u}_\alpha, \nabla\hat{\psi}_\alpha\rangle_g \, dv_g \\ &\qquad = -\int_M \lim_{\rho\to 0} \rho^{1-2\gamma}(\partial_\rho \hat{u}_\alpha)\psi_\alpha \, d\sigma_{\hat{h}} + o(1)\,\|\hat{\psi}_\alpha\|_{W^{1,2}(X,\rho^{1-2\gamma})} \\ &\qquad = \int_M |\hat{u}_\alpha|^{2^*-2}\hat{u}_\alpha \hat{\psi}_\alpha \, d\sigma_{\hat{h}} + o(1)\|\hat{\psi}_\alpha\|_{W^{1,2}(X,\rho^{1-2\gamma})} \\ &\qquad = \int_{\partial' B_r^+(z_0)} |\tilde{\eta}_\alpha \tilde{u}_\alpha|^{2^*-2}(\tilde{\eta}_\alpha \tilde{u}_\alpha)\psi \, d\sigma_{\tilde{h}_\alpha} + o(1)\,\|\hat{\psi}_\alpha\|_{W^{1,2}(X,\rho^{1-2\gamma})}. \end{aligned} \tag{3-8}$$

Since $\|\psi\|_{W^{1,2}(B_r^+(z_0), y^{1-2\gamma})} = \|\hat{\psi}_\alpha\|_{W^{1,2}(X,\rho^{1-2\gamma})}$, combining expressions (3-7) and (3-8) yields

$$\begin{aligned} o(1)\,\|\psi\|_{W^{1,2}(B_r^+(z_0), y^{1-2\gamma})} = \int_{\partial' B_r^+(z_0)} \lim_{y\to 0} y^{1-2\gamma}\partial_y(\tilde{\eta}_\alpha \tilde{u}_\alpha)\psi \, d\sigma_{\tilde{h}_\alpha} \\ + \int_{\partial' B_r^+(z_0)} |\tilde{\eta}_\alpha \tilde{u}_\alpha|^{2^*-2}(\tilde{\eta}_\alpha \tilde{u}_\alpha)\psi \, d\sigma_{\tilde{h}_\alpha}, \end{aligned}$$

i.e.,

$$-\lim_{y\to 0} y^{1-2\gamma}\partial_y(\tilde{\eta}_\alpha \tilde{u}_\alpha) = |\tilde{\eta}_\alpha \tilde{u}_\alpha|^{2^*-2}(\tilde{\eta}_\alpha \tilde{u}_\alpha) + o(1) \quad \text{on } \partial' B_r^+(z_0).$$

Meanwhile, since $\tilde\eta_\alpha\tilde u_\alpha \rightharpoonup u$ weakly in $W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$, the same argument as above gives that

$$-\lim_{y\to 0} y^{1-2\gamma}\partial_y u = |u|^{2^*-2}u \quad \text{on } \partial' B_r^+(z_0).$$

If we denote by

$$\Gamma_\alpha := |\tilde\eta_\alpha\tilde u_\alpha|^{2^*-2}(\tilde\eta_\alpha\tilde u_\alpha) - |u|^{2^*-2}u - |\tilde\eta_\alpha\tilde u_\alpha - u|^{2^*-2}(\tilde\eta_\alpha\tilde u_\alpha - u),$$

then

$$\text{(3-9)}\qquad \begin{cases} -\operatorname{div}(y^{1-2\gamma}\nabla(\tilde\eta_\alpha\tilde u_\alpha - u)) = o(1) & \text{in } B_r^+(z_0),\\ -\lim\limits_{y\to 0} y^{1-2\gamma}\partial_y(\tilde\eta_\alpha\tilde u_\alpha - u) & \\ \qquad\qquad = |\tilde\eta_\alpha\tilde u_\alpha - u|^{2^*-2}(\tilde\eta_\alpha\tilde u_\alpha - u) + \Gamma_\alpha + o(1) & \text{on } \partial' B_r^+(z_0). \end{cases}$$

We have proved in (3-3) that for any $r>0$ and $\varepsilon_1>0$, there exists a sequence $\{\mu_\alpha\}_{\alpha\in\mathbb{N}}$ such that, if $|z_0|+r\le r_0\le \mu_\alpha^{-1}r_0$, then

$$\int_{\partial' B_r^+(z_0)} |\tilde u_\alpha|^{2^*}\,dx \le \varepsilon_1.$$

Therefore, we can also choose r small enough such that, if $|z_0|+3r<r_0$, then

$$\int_{\partial' B_r^+(z_0)} |\tilde\eta_\alpha\tilde u_\alpha - u|^{2^*}\,dx \le \varepsilon_1.$$

We claim that $\Gamma_\alpha = o(1)$ in the sense that for any $\theta\in W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})'$,

$$\int_{\partial' B_r^+(z_0)} |\Gamma_\alpha\theta|\,d\sigma_{\hat h} = o(1)\,\|\theta\|_{L^{2^*}(\partial' B_r^+(z_0))}, \quad \text{as } \alpha\to+\infty.$$

We can use the same arguments as in the proof of Lemma 2.12 to show this claim.

Then by the ε-regularity estimates and the compact embedding of the weighted Sobolev space, we can prove that $\tilde\eta_\alpha\tilde u_\alpha\to u$ in $W^{1,2}(B_r^+(z_0), y^{1-2\gamma})$. Then, by the finite covering we can prove that $\tilde\eta_\alpha\tilde u_\alpha\to u$ in $W^{1,2}(B_{2r}^+(0), y^{1-2\gamma})$. □

Applying Claim 3, noting that $\tilde\eta_\alpha\tilde u_\alpha\to u$ in $W^{1,2}(B_{2r}^+(0), y^{1-2\gamma})$ and that $\tilde\eta_\alpha\equiv 1$ in $\partial' B^+_{1/4\mu_\alpha^{-1}r_0}$ since $0<\mu_\alpha\le 1$ and $2r<\frac{r_0}{4}$,

$$\varepsilon = \int_{\partial' B_{2r}^+(0)} |\tilde u_\alpha|^{2^*}\,d\sigma_{\tilde h_\alpha} = \int_{\partial' B_{2r}^+(0)} |\tilde\eta_\alpha\tilde u_\alpha|^{2^*}\,d\sigma_{\tilde h_\alpha} \le 2\int_{\partial' B_{2r}^+(0)} |u|^{2^*}\,dx + o(1),$$

where we used $\tilde\eta_\alpha\tilde u_\alpha\to u$ in $L^{2^*}(\partial' B_{2r}^+(0), |dx|^2)$ as $\alpha\to+\infty$ by Proposition 2.4. So, $u\neq 0$.

Claim 4. $\lim\limits_{\alpha\to+\infty}\mu_\alpha = 0$.

In fact, if $\mu_\alpha\to\mu_0>0$, then $\tilde\eta_\alpha\tilde u_\alpha\rightharpoonup 0$ in $W^{1,2}(B_{2r}^+(0), y^{1-2\gamma})$ since $\hat u_\alpha\rightharpoonup 0$ in $W^{1,2}(X, \rho^{1-2\gamma})$. But, $u\neq 0$, which is a contradiction.

Claim 5. *For any $0 < \mu_0 \le 1$, $\tilde{u}_\alpha \to u$ strongly in $W^{1,2}(B^+_{\mu_0^{-1}}(0), y^{1-2\gamma})$ as $\alpha \to +\infty$, and u is a weak solution of (3-1).*

Proof. Let $0 < \mu_0 \le 1$. By Claim 4, we know that $0 < \mu_\alpha \le \mu_0$ for α large. Then, (3-3) holds for $|z_0| + r < \mu_0^{-1} r_0$. By the same arguments, it is easy to check that

$$\tilde{\eta}_\alpha \tilde{u}_\alpha \to u \quad \text{in } W^{1,2}(B^+_{2r\mu_0^{-1}}(0), y^{1-2\gamma}).$$

For α large, $\tilde{\eta}_\alpha \equiv 1$ in $B^+_{2r\mu_0^{-1}}(0)$, so we have

$$\tilde{u}_\alpha \to u \;\text{ in }\; W^{1,2}(B^+_{2r\mu_0^{-1}}(0),\, y^{1-2\gamma})$$

strongly as $\alpha \to +\infty$.

Finally, we claim that u solves the boundary problem

$$\begin{cases} -\operatorname{div}(y^{1-2\gamma}\nabla u) = 0 & \text{in } \mathbb{R}^{n+1}_+, \\ -\lim\limits_{y\to 0} y^{1-2\gamma}\partial_y u = |u|^{2^*-2}u & \text{on } \mathbb{R}^n. \end{cases} \tag{3-10}$$

Since $0 < \mu_0 \le 1$ is arbitrary, $\tilde{u}_\alpha \to u$ strongly in $W^{1,2}(B^+_R(0), y^{1-2\gamma})$ for any large $R > 0$. Without loss of generality, let $\psi \in \mathscr{C}_0^\infty(\mathbb{R}^{n+1}_+)$ and $\operatorname{supp}\psi \subset B^+_{R_0}(0)$ for some $R_0 > 0$. Set

$$\psi_\alpha(z) = \mu_\alpha^{-\frac{n-2\gamma}{2}} \psi(\mu_\alpha^{-1}\varphi_{x_\alpha}^{-1}(z)).$$

For α large enough,

$$\int_X \rho^{1-2\gamma}\langle \nabla\hat{u}_\alpha, \nabla\psi_\alpha\rangle_g \, dv_g = \int_{\mathbb{R}^{n+1}_+} \tilde{\rho}_\alpha^{1-2\gamma}\langle \nabla(\tilde{\eta}_\alpha\tilde{u}_\alpha), \nabla\psi\rangle_{\tilde{g}_\alpha}\, dv_{\tilde{g}_\alpha},$$

and

$$\int_M |\hat{u}_\alpha|^{2^*-2}\hat{u}_\alpha\psi_\alpha \, dv_g = \int_{\mathbb{R}^n} |\tilde{\eta}_\alpha\tilde{u}_\alpha|^{2^*-2}(\tilde{\eta}_\alpha\tilde{u}_\alpha)\psi \, dv_{\tilde{g}_\alpha}.$$

Note that $\tilde{g}_\alpha \to |dx|^2 + dy^2$ in $\mathscr{C}^1(B^+_R(0))$ as $\alpha \to +\infty$, $\{\hat{u}_\alpha\}$ is a Palais–Smale sequence for I_g^γ and $\tilde{\eta}_\alpha\tilde{u}_\alpha \to u$ in $W^{1,2}(B^+_R(0))$ for any $R > 0$. Then, we have

$$\int_{\mathbb{R}^{n+1}_+} y^{1-2\gamma}\langle \nabla u, \nabla\psi\rangle \, dx\, dy - \int_{\mathbb{R}^n} |u|^{2^*-2}u\psi \, dx\, dy = 0,$$

which yields our desired result. □

Step 4. The Palais–Smale sequence minus a bubble is still a Palais–Smale sequence. Define

$$\begin{cases} \hat{w}_\alpha(z) = \hat{\eta}_\alpha(z)\mu_\alpha^{-\frac{n-2\gamma}{2}} u(\mu_\alpha^{-1}\varphi_{x_\alpha}^{-1}(z)), & z \in \varphi_{x_\alpha}(B^+_{2r_0}(0)), \\ \hat{w}_\alpha(z) = 0, & \text{otherwise}, \end{cases} \tag{3-11}$$

where $\hat{\eta}_\alpha$ is a cut-off function satisfying $\hat{\eta}_\alpha = 1$ in $\varphi_{x_\alpha}(B^+_{r_0}(0))$ and $\hat{\eta}_\alpha = 0$ in $M \setminus \varphi_{x_\alpha}(B^+_{2r_0}(0))$. Here we have $\mathfrak{B}^+_{2r_0}(x_\alpha) = \varphi_{x_\alpha}(B^+_{2r_0}(0))$. Let $\hat{v}_\alpha = \hat{u}_\alpha - \hat{w}_\alpha$. We claim:

(i) $\hat{v}_\alpha \rightharpoonup 0$ in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$;

(ii) $DI_g^\gamma(\hat{v}_\alpha) \to 0$ in $W^{1,2}(X, \rho^{1-2\gamma})'$ as $\alpha \to +\infty$;

(iii) $I_g^\gamma(\hat{v}_\alpha) = I_g^\gamma(\hat{u}_\alpha) - \widetilde{E}(u) + o(1)$ as $\alpha \to +\infty$;

(iv) $\{\hat{v}_\alpha\}_{\alpha\in\mathbb{N}}$ is also a Palais–Smale sequence for I_g^γ.

The remainder of the proof of Lemma 3.1 consists of proving these claims.

(i) Since $\hat{u}_\alpha \rightharpoonup 0$ in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$, it suffices to prove $\hat{w}_\alpha \rightharpoonup 0$ in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$. First, we prove that $\int_M \hat{w}_\alpha \psi \, d\sigma_{\hat{h}} = o(1)$ as $\alpha \to +\infty$ for any $\psi \in \mathscr{C}^\infty(\overline{X})$. Given $R > 0$,

$$\int_M \hat{w}_\alpha \psi \, d\sigma_{\hat{h}} = \int_{\mathfrak{D}_{\mu_\alpha R}(x_\alpha)} \hat{w}_\alpha \psi \, d\sigma_{\hat{h}} + \int_{M\setminus\mathfrak{D}_{\mu_\alpha R}(x_\alpha)} \hat{w}_\alpha \psi \, d\sigma_{\hat{h}}. \tag{3-12}$$

Note that $\tilde{h}_\alpha(x) = (\varphi_{x_\alpha}^* \hat{h})(\mu_\alpha x)$. Using (3-11),

$$\begin{aligned}\int_{\mathfrak{D}_{\mu_\alpha R}(x_\alpha)} \hat{w}_\alpha \psi \, d\sigma_{\hat{h}} &= \int_{\mathfrak{D}_{\mu_\alpha R}(x_\alpha)} \hat{\eta}_\alpha(x) \mu_\alpha^{-\frac{n-2\gamma}{2}} u(\mu_\alpha^{-1}\varphi_{x_\alpha}^{-1}(x))\psi(x) \, d\sigma_{\hat{h}} \\ &= \mu_\alpha^{\frac{n+2\gamma}{2}} \int_{D_R(0)} \hat{\eta}_\alpha(\varphi_{x_\alpha}(\mu_\alpha x)) u(x) \psi(\varphi_{x_\alpha}(\mu_\alpha x)) \, d\sigma_{\tilde{h}_\alpha} \\ &\le C \|\psi\|_{L^\infty(M)} \mu_\alpha^{\frac{n+2\gamma}{2}} \int_{D_R(0)} |u(x)| \, dx.\end{aligned}$$

Similarly, we can deal with the second term in the right hand side of (3-12):

$$\begin{aligned}\int_{M\setminus\mathfrak{D}_{\mu_\alpha R}(x_\alpha)} \hat{w}_\alpha \psi \, d\sigma_{\hat{h}} &= \int_{\mathfrak{D}_{2r_0}(x_\alpha)\setminus\mathfrak{D}_{\mu_\alpha R}(x_\alpha)} \hat{w}_\alpha \psi \, d\sigma_{\hat{h}} \\ &\le C \|\psi\|_{L^\infty(M)} \mu_\alpha^{\frac{n+2\gamma}{2}} \int_{D_{2r_0\mu_\alpha^{-1}}(0)\setminus D_R(0)} |u(x)| \, dx \\ &\le C \|\psi\|_{L^\infty(M)} \mu_\alpha^{\frac{n+2\gamma}{2}} \left(\int_{D_{2r_0\mu_\alpha^{-1}}(0)\setminus D_R(0)} |u(x)|^{2^*} dx \right)^{\frac{1}{2^*}} \\ &\qquad \times \left(\int_{D_{2r_0\mu_\alpha^{-1}}(0)\setminus D_R(0)} dx \right)^{\frac{n+2\gamma}{2n}} \\ &\le C \|\psi\|_{L^\infty(M)} \left(\int_{D_{2r_0\mu_\alpha^{-1}}(0)\setminus D_R(0)} |u(x)|^{2^*} dx \right)^{\frac{1}{2^*}}.\end{aligned}$$

Since $u \in L^{2^*}(\mathbb{R}^n, |dx|^2)$ and $\mu_\alpha \to 0$ as $\alpha \to +\infty$, taking R large enough we get

$$\int_M \hat{w}_\alpha \psi \, d\sigma_{\hat{h}} = o(1) \quad \text{as } \alpha \to +\infty.$$

Next, we will show that

$$\int_X \rho^{1-2\gamma} \langle \nabla \hat{w}_\alpha, \nabla \psi \rangle_g \, dv_g = o(1) \quad \text{as } \alpha \to +\infty$$

for any $\psi \in \mathscr{C}^\infty(\overline{X})$. Let $\tilde{\eta}_\alpha(z) = \hat{\eta}_\alpha(\varphi_{x_\alpha}(\mu_\alpha z))$, $\tilde{\rho}_\alpha(z) = \mu_\alpha^{-1} \rho(\varphi_{x_\alpha}(\mu_\alpha z))$. Noting that $\hat{w}_\alpha \equiv 0$ in $X \setminus \mathfrak{B}^+_{2r_0}(x_\alpha)$, then for any $R > 0$ and α large,

$$\begin{aligned} \int_X \rho^{1-2\gamma} \langle \nabla \hat{w}_\alpha, \nabla \psi \rangle_g \, dv_g &= \int_{\mathfrak{B}^+_{2r_0}(x_\alpha)} \rho^{1-2\gamma} \langle \nabla \hat{w}_\alpha, \nabla \psi \rangle_g \, dv_g \\ &= \int_{\mathfrak{B}^+_{2r_0}(x_\alpha) \setminus \mathfrak{B}^+_{R\mu_\alpha}(x_\alpha)} \rho^{1-2\gamma} \langle \nabla \hat{w}_\alpha, \nabla \psi \rangle_g \, dv_g \\ &\quad + \int_{\mathfrak{B}^+_{R\mu_\alpha}(x_\alpha)} \rho^{1-2\gamma} \langle \nabla \hat{w}_\alpha, \nabla \psi \rangle_g \, dv_g \\ &=: I_1 + I_2. \end{aligned} \tag{3-13}$$

By Hölder's inequality and the fact that $u \in W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$,

$$\begin{aligned} I_1 &\le \left(\int_{\mathfrak{B}^+_{2r_0}(x_\alpha) \setminus \mathfrak{B}^+_{R\mu_\alpha}(x_\alpha)} \rho^{1-2\gamma} |\nabla \hat{w}_\alpha|_g^2 \, dv_g \right)^{\frac{1}{2}} \\ &\qquad \times \left(\int_{\mathfrak{B}^+_{2r_0}(x_\alpha) \setminus \mathfrak{B}^+_{R\mu_\alpha}(x_\alpha)} \rho^{1-2\gamma} |\nabla \psi|_g^2 \, dv_g \right)^{\frac{1}{2}} \\ &= \left(\int_{B^+_{2r_0\mu_\alpha^{-1}}(0) \setminus B^+_R(0)} \tilde{\rho}_\alpha^{1-2\gamma} |\nabla (\tilde{\eta}_\alpha u)|^2_{\tilde{g}_\alpha} \, dv_{\tilde{g}_\alpha} \right)^{\frac{1}{2}} \\ &\qquad \times \left(\int_{\mathfrak{B}^+_{2r_0}(x_\alpha) \setminus \mathfrak{B}^+_{R\mu_\alpha}(x_\alpha)} \rho^{1-2\gamma} |\nabla \psi|_g^2 \, dv_g \right)^{\frac{1}{2}} =: \beta(R), \end{aligned}$$

where

$$\lim_{R \to +\infty} \lim_{\alpha \to +\infty} \sup \beta(R) = 0. \tag{3-14}$$

The previous limit is estimated because $u \in W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$, so for any α, R,

$$\left(\int_{B^+_{2r_0\mu_\alpha^{-1}}(0) \setminus B^+_R(0)} \tilde{\rho}_\alpha^{1-2\gamma} |\nabla (\tilde{\eta}_\alpha u)|^2_{\tilde{g}_\alpha} \, dv_{\tilde{g}_\alpha} \right)^{\frac{1}{2}} \le C \|u\|_{W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})},$$

and for any $\varepsilon > 0$ and any α large, there exists $R_0 > 0$ such that for $R > R_0$,

$$\left(\int_{\mathfrak{B}^+_{2r_0}(x_\alpha) \setminus \mathfrak{B}^+_{R\mu_\alpha}(x_\alpha)} \rho^{1-2\gamma} |\nabla \psi|_g^2 \, dv_g \right)^{\frac{1}{2}} \le \varepsilon.$$

Meanwhile,

$$\begin{aligned} I_2 &\le \left(\int_{\mathfrak{B}^+_{R\mu_\alpha}(x_\alpha)} \rho^{1-2\gamma} |\nabla \hat{w}_\alpha|^2_g \, dv_g\right)^{\frac{1}{2}} \left(\int_{\mathfrak{B}^+_{R\mu_\alpha}(x_\alpha)} \rho^{1-2\gamma} |\nabla \psi|^2_g \, dv_g\right)^{\frac{1}{2}} \\ &= \left(\int_{B^+_R(0)} \tilde{\rho}_\alpha^{1-2\gamma} |\nabla (\tilde{\eta}_\alpha u)|^2_{\tilde{g}_\alpha} \, dv_{\tilde{g}_\alpha}\right)^{\frac{1}{2}} \left(\int_{\mathfrak{B}^+_{R\mu_\alpha}(x_\alpha)} \rho^{1-2\gamma} |\nabla \psi|^2_g \, dv_g\right)^{\frac{1}{2}} = o(1), \end{aligned}$$

uniformly in R as $\alpha \to +\infty$. To see this, for any $R > 0$,

$$\left(\int_{B^+_R(0)} \tilde{\rho}_\alpha^{1-2\gamma} |\nabla (\tilde{\eta}_\alpha u)|^2_{\tilde{g}_\alpha} \, dv_{\tilde{g}_\alpha}\right)^{\frac{1}{2}} \le C \|u\|_{W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})},$$

also in Claim 4 we have proved that

$$\lim_{\alpha \to +\infty} \mu_\alpha = 0$$

and note that $\psi \in W^{1,2}(X, \rho^{1-2\gamma})$. Since $R > 0$ is arbitrary, (3-13) implies that

$$\int_X \rho^{1-2\gamma} \langle \nabla \hat{w}_\alpha, \nabla \psi \rangle_g \, dv_g = o(1)$$

as $\alpha \to +\infty$.

(ii) For any $\psi \in W^{1,2}(X, \rho^{1-2\gamma})$, the proof of (i) and Propositions 2.4 and 2.6 imply that

$$d\sigma_{\hat{h}} \to 0, \quad \text{as } \alpha \to +\infty.$$

On the other hand, we have

$$\begin{aligned} DI^\gamma_g(\hat{v}_\alpha) \cdot \psi &= \int_X \rho^{1-2\gamma} \langle \nabla \hat{v}_\alpha, \nabla \psi \rangle_g \, dv_g - \int_M |\hat{v}_\alpha|^{2^*-2} \hat{v}_\alpha \psi \, d\sigma_{\hat{h}} \\ &= DI^\gamma_g(\hat{u}_\alpha) \cdot \psi - DI^\gamma_g(\hat{w}_\alpha) \cdot \psi - \int_M \Phi_\alpha \psi \, d\sigma_{\hat{h}}, \end{aligned}$$

where

$$\Phi_\alpha = |\hat{u}_\alpha - \hat{w}_\alpha|^{2^*-2} (\hat{u}_\alpha - \hat{w}_\alpha) + |\hat{w}_\alpha|^{2^*-2} \hat{w}_\alpha - |\hat{u}_\alpha|^{2^*-2} \hat{u}_\alpha.$$

Following the same argument of [Druet et al. 2004, pp. 39–40], we can prove that

$$\int_M \Phi_\alpha \psi \, d\sigma_{\hat{h}} \to 0, \quad \text{as } \alpha \to +\infty.$$

Then, we get that $DI^\gamma_g(\hat{v}_\alpha) \to 0$ in $W^{1,2}(X, \rho^{1-2\gamma})'$ as $\alpha \to +\infty$, since $\{\hat{u}_\alpha\}_{\alpha \in \mathbb{N}}$ is a Palais–Smale sequence for I^γ_g.

(iii) Note that $\hat{v}_\alpha = \hat{u}_\alpha - \hat{w}_\alpha$ and $\hat{w}_\alpha \equiv 0$ in $X \setminus \mathfrak{B}^+_{2r_0}(x_\alpha)$. Given $R > 0$, for α large,

$$
\begin{aligned}
\text{(3-15)}\quad & \int_X \rho^{1-2\gamma} |\nabla \hat{v}_\alpha|_g^2 \, dv_g \\
&= \int_{\mathfrak{B}^+_{2r_0}(x_\alpha)} \rho^{1-2\gamma} |\nabla \hat{v}_\alpha|_g^2 \, dv_g + \int_{X \setminus \mathfrak{B}^+_{2r_0}(x_\alpha)} \rho^{1-2\gamma} |\nabla \hat{u}_\alpha|_g^2 \, dv_g \\
&= \int_{\mathfrak{B}^+_{\mu_\alpha R}(x_\alpha)} \rho^{1-2\gamma} |\nabla \hat{v}_\alpha|_g^2 \, dv_g + \int_{\mathfrak{B}^+_{2r_0}(x_\alpha) \setminus \mathfrak{B}^+_{\mu_\alpha R}(x_\alpha)} \rho^{1-2\gamma} |\nabla \hat{v}_\alpha|_g^2 \, dv_g \\
&\qquad + \int_{X \setminus \mathfrak{B}^+_{2r_0}(x_\alpha)} \rho^{1-2\gamma} |\nabla \hat{u}_\alpha|_g^2 \, dv_g \\
&=: I_1 + I_2 + \int_{X \setminus \mathfrak{B}^+_{2r_0}(x_\alpha)} \rho^{1-2\gamma} |\nabla \hat{u}_\alpha|_g^2 \, dv_g .
\end{aligned}
$$

Since $\tilde{\eta}_\alpha \tilde{u}_\alpha \to u$ in $W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$ as $\alpha \to +\infty$ because of Claim 5,

$$
\begin{aligned}
I_1 &= \int_{\mathfrak{B}^+_{\mu_\alpha R}(x_\alpha)} \rho^{1-2\gamma} |\nabla (\hat{u}_\alpha - \hat{w}_\alpha)|_g^2 \, dv_g \\
&= \int_{B^+_R(0)} \tilde{\rho}_\alpha^{1-2\gamma} |\nabla (\tilde{u}_\alpha - u)|_{\tilde{g}_\alpha}^2 \, dv_{\tilde{g}_\alpha} \\
&\le 2 \int_{B^+_R(0)} y^{1-2\gamma} |\nabla (\tilde{u}_\alpha - u)|^2 \, dx \, dy = o(1), \quad \text{as } \alpha \to +\infty,
\end{aligned}
$$

where we have used that $\tilde{\eta}_\alpha \equiv 1$ in $B^+_R(0)$ for α large.

On the other hand, direct computations give that

$$
\begin{aligned}
\int_{\mathfrak{B}^+_{2r_0}(x_\alpha) \setminus \mathfrak{B}^+_{\mu_\alpha R}(x_\alpha)} \rho^{1-2\gamma} |\nabla \hat{w}_\alpha|_g^2 \, dv_g &= \int_{B^+_{2r_0 \mu_\alpha^{-1}}(0) \setminus B^+_R(0)} \tilde{\rho}_\alpha^{1-2\gamma} |\nabla u|_{\tilde{g}_\alpha}^2 \, dv_{\tilde{g}_\alpha} \\
&\le 2 \int_{B^+_{2r_0 \mu_\alpha^{-1}}(0) \setminus B^+_R(0)} y^{1-2\gamma} |\nabla u|^2 \, dx \, dy \\
&= \beta(R),
\end{aligned}
$$

since $u \in W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$ and $\mu_\alpha \to 0$ as $\alpha \to +\infty$, where $\beta(R)$ is defined as in (3-14). Hence, we get

$$
\begin{aligned}
I_2 &= \int_{\mathfrak{B}^+_{2r_0}(x_\alpha) \setminus \mathfrak{B}^+_{\mu_\alpha R}(x_\alpha)} \rho^{1-2\gamma} \big(|\nabla \hat{u}_\alpha|_g^2 + |\nabla \hat{w}_\alpha|_g^2 - 2 \langle \nabla \hat{u}_\alpha, \nabla \hat{w}_\alpha \rangle_g \big) \, dv_g \\
&= \int_{\mathfrak{B}^+_{2r_0}(x_\alpha) \setminus \mathfrak{B}^+_{\mu_\alpha R}(x_\alpha)} \rho^{1-2\gamma} |\nabla \hat{u}_\alpha|_g^2 \, dv_g + \beta(R).
\end{aligned}
$$

Here we have used Hölder's inequality and the fact that $\{\hat{u}_\alpha\}$ is uniformly bounded in $W^{1,2}(X, \rho^{1-2\gamma})$ to get

$$\int_{\mathfrak{B}^+_{2r_0}(x_\alpha)\setminus\mathfrak{B}^+_{\mu_\alpha R}(x_\alpha)} \rho^{1-2\gamma}\langle\nabla\hat{u}_\alpha, \nabla\hat{w}_\alpha\rangle_g \, dv_g = \beta(R).$$

Therefore, noting that $\tilde{u}_\alpha \to u$ in $W^{1,2}(\mathbb{R}^{n+1}_+, y^{1-2\gamma})$ as $\alpha \to +\infty$, by (3-15),

$$\begin{aligned}
\int_X \rho^{1-2\gamma}|\nabla\hat{v}_\alpha|_g^2 \, dv_g
&= \int_X \rho^{1-2\gamma}|\nabla\hat{u}_\alpha|_g^2 \, dv_g - \int_{\mathfrak{B}^+_{\mu_\alpha R}(x_\alpha)} \rho^{1-2\gamma}|\nabla\hat{u}_\alpha|_g^2 \, dv_g + \beta(R) + o(1)\\
&= \int_X \rho^{1-2\gamma}|\nabla\hat{u}_\alpha|_g^2 \, dv_g - \int_{B^+_R(0)} \tilde{\rho}_\alpha^{1-2\gamma}|\nabla\tilde{u}_\alpha|_{\tilde{g}_\alpha}^2 \, dv_{\tilde{g}_\alpha} + \beta(R) + o(1)\\
&= \int_X \rho^{1-2\gamma}|\nabla\hat{u}_\alpha|_g^2 \, dv_g - \int_{B^+_R(0)} y^{1-2\gamma}|\nabla u|^2 \, dx\, dy + \beta(R) + o(1)\\
&= \int_X \rho^{1-2\gamma}|\nabla\hat{u}_\alpha|_g^2 \, dv_g - \int_{\mathbb{R}^{n+1}_+} y^{1-2\gamma}|\nabla u|^2 \, dx\, dy + \beta(R) + o(1).
\end{aligned}$$

In a similar way,

$$\int_M |\hat{v}_\alpha|^{2^*} d\sigma_{\hat{h}} = \int_M |\hat{u}_\alpha|^{2^*} d\sigma_{\hat{h}} - \int_{\mathbb{R}^n} |u|^{2^*} dx + \beta(R) + o(1).$$

These imply that

$$I_g^\gamma(\hat{v}_\alpha) = I_g^\gamma(\hat{u}_\alpha) - \widetilde{E}(u) + \beta(R) + o(1).$$

Since $R > 0$ is arbitrary, we get conclusion (iii).

(iv) It is a direct consequence of (ii) and (iii). □

4. Proof of the main results

Proof of Theorem 1.3. From Remark 2.10, we have $u_\alpha \rightharpoonup u^0$ in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$. And $u_\alpha \to u^0$ a.e. on M as $\alpha \to +\infty$. Then, $u^0 \geq 0$ on M since $u_\alpha \geq 0$. Also, $\hat{u}_\alpha = u_\alpha - u^0$ satisfies the Palais–Smale condition and

$$I_g^\gamma(\hat{u}_\alpha) = I_g^{\gamma,\alpha}(u_\alpha) - I_g^{\gamma,\infty}(u^0) + o(1).$$

If $\hat{u}_\alpha \to 0$ in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$, then the theorem is proved. If $\hat{u}_\alpha \rightharpoonup 0$ but not strongly in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$, then, using Lemma 3.1, we can obtain a new Palais–Smale sequence $\{\hat{u}^1_\alpha\}_{\alpha\in\mathbb{N}}$ satisfying

$$I_g^\gamma(\hat{u}^1_\alpha) = I_g^\gamma(\hat{u}_\alpha) - \widetilde{E}(u) + o(1).$$

Now, either $\hat{u}^1_\alpha \to 0$ in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$, in which case the theorem holds, or $\hat{u}^1_\alpha \rightharpoonup 0$ but not strongly in $W^{1,2}(X, \rho^{1-2\gamma})$ as $\alpha \to +\infty$, in which case we again use Lemma 3.1.

Since $\{I_g^{\gamma,\alpha}(u_\alpha)\}_{\alpha\in\mathbb{N}}$ is uniformly bounded, after a finite number of induction steps, we get the last Palais–Smale sequence (for $m > 1$)

$$\{\hat{u}^m_\alpha\}_{\alpha\in\mathbb{N}} \quad \text{with } I_g^\gamma(\hat{u}^m_\alpha) \to \beta < \beta_0.$$

Then, by Lemma 2.13, we can get that

$$\hat{u}^m_\alpha \to 0 \quad \text{in } W^{1,2}(X, \rho^{2\gamma-1}) \text{ as } \alpha \to +\infty.$$

Applying Lemma 3.1 in the process, we can get that $\{u^j\}_{j=1}^m$ are solutions to (3-1). We will prove the positivity of u^j, $j = 1, \dots, m$, in Lemma 4.2, and the relation (5) of Theorem 1.3 in Lemma 4.1.

For the regularity of u^j, we can use Lemmas A.1 and A.2. □

Lemma 4.1. *For any integer k in $[1, m]$, and any integer l in $[0, k-1]$, there exist an integer s and sequences $\{y^j_\alpha\}_{\alpha\in\mathbb{N}} \subset M$ and $\{\lambda^j_\alpha > 0\}_{\alpha\in\mathbb{N}}$, $j = 1, \dots, s$, such that $d_{\hat{h}}(x^k_\alpha, y^j_\alpha)/\mu^k_\alpha$ is bounded, $\lambda^j_\alpha/\mu^k_\alpha \to 0$ as $\alpha \to +\infty$, and for any $R, R' > 0$,*

$$\int_{\mathfrak{D}_{R\mu^k_\alpha}(x^k_\alpha)\setminus\bigcup_{j=1}^s \mathfrak{D}_{R'\lambda^j_\alpha}(y^j_\alpha)} \left|\hat{u}_\alpha - \sum_{i=1}^{l} u^i_\alpha - u^k_\alpha\right|^{2^*} d\sigma_{\hat{h}} = o(1) + \varepsilon(R'), \tag{4-1}$$

where

$$\lim_{R'\to+\infty} \lim_{\alpha\to+\infty} \sup \varepsilon(R') = 0,$$

and $\{u^i_\alpha\}$ is derived from the rescaling of u^i we obtained in the above proof of Theorem 1.3, and $\{x^i_\alpha\}$ is the i-th likely blow up points sequence.

Proof. We prove this lemma by iteration on l. For any integer k $(1 \le k \le m)$, if $l = k-1$, then combining the above proof of Theorem 1.3 with Lemma 3.1 and Proposition 2.4,

$$\int_{\mathfrak{D}_{R\mu^k_\alpha}(x^k_\alpha)} \left|\hat{u}_\alpha - \sum_{i=1}^{k-1} u^i_\alpha - u^k_\alpha\right|^{2^*} d\sigma_{\hat{h}} = o(1),$$

so (4-1) holds for $s = 0$.

Suppose that (4-1) holds for some l, $1 \le l \le k-1$, we need to show that (4-1) holds for $l - 1$.

Case 1: $d_{\hat h}(x_\alpha^l, x_\alpha^k) \nrightarrow 0$ as $\alpha \to +\infty$. Then, for any $\bar R > 0$, up to a subsequence, $\mathfrak{D}_{\bar R \mu_\alpha^l}(x_\alpha^l) \cap \mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k) = \varnothing$, so we have

$$\begin{aligned}\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)\setminus\bigcup_{j=1}^s \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)} |u_\alpha^l|^{2^*}\, d\sigma_{\hat h} &\le \int_{M\setminus \mathfrak{D}_{\bar R\mu_\alpha^l}(x_\alpha^l)} |u_\alpha^l|^{2^*}\, d\sigma_{\hat h}\\ &\le C\int_{\mathbb{R}^n\setminus D_{\bar R(0)}} |u^l|^{2^*}\, d\sigma_{\tilde h_\alpha}\\ &\le C\int_{\mathbb{R}^n\setminus D_{\bar R(0)}} |u^l|^{2^*}\, dx.\end{aligned}$$

Since $\bar R > 0$ is arbitrary and $u^l \in L^{2^*}(\mathbb{R}^n)$,

$$\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)\setminus\bigcup_{j=1}^s \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)} |u_\alpha^l|^{2^*}\, d\sigma_{\hat h} = o(1), \quad \text{as } \alpha \to +\infty. \tag{4-2}$$

So by the induction hypothesis for l and (4-2), we obtain

$$\begin{aligned}\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)\setminus\bigcup_{j=1}^s \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)} \Big|\hat u_\alpha - \sum_{i=1}^{l-1} u_\alpha^i - u_\alpha^k\Big|^{2^*} d\sigma_{\hat h}\\ \le 2^{2^*-1}\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)\setminus\bigcup_{j=1}^s \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)} \Big|\hat u_\alpha - \sum_{i=1}^{l} u_\alpha^i - u_\alpha^k\Big|^{2^*} d\sigma_{\hat h}\\ + 2^{2^*-1}\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)\setminus\bigcup_{j=1}^s \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)} |u_\alpha^l|^{2^*}\, d\sigma_{\hat h}\\ = o(1) + \varepsilon(R').\end{aligned}$$

Thus we have proven that (4-1) holds for $l-1$.

Case 2: $d_{\hat h}(x_\alpha^l, x_\alpha^k) \to 0$ as $\alpha \to +\infty$. Let r_0 be sufficiently small such that for any $P \in M$, $x, y \in \mathbb{R}^n$, and $|x|, |y| \le r_0$,

$$\tfrac12 |x-y| \le d_{\hat h}(\varphi_P(x), \varphi_P(y)) \le 2|x-y|.$$

Let $\tilde x_\alpha^l = (\mu_\alpha^k)^{-1}\varphi_{x_\alpha^k}^{-1}(x_\alpha^l)$ and $\tilde y_\alpha^j = (\mu_\alpha^k)^{-1}\varphi_{x_\alpha^k}^{-1}(y_\alpha^j)$. Then,

$$\begin{cases} D_{\frac R2 \mu_\alpha^l/\mu_\alpha^k}(\tilde x_\alpha^l) \subset (\mu_\alpha^k)^{-1}\varphi_{x_\alpha^k}^{-1}(\mathfrak{D}_{R\mu_\alpha^l}(x_\alpha^l)) \subset D_{2R\mu_\alpha^l/\mu_\alpha^k}(\tilde x_\alpha^l),\\ D_{\frac R2 \lambda_\alpha^j/\mu_\alpha^k}(\tilde y_\alpha^j) \subset (\mu_\alpha^k)^{-1}\varphi_{x_\alpha^k}^{-1}(\mathfrak{D}_{R\lambda_\alpha^j}(y_\alpha^j)) \subset D_{2R\lambda_\alpha^j/\mu_\alpha^k}(\tilde y_\alpha^j).\end{cases} \tag{4-3}$$

Given $\tilde R > 0$, from Lemma 3.1, Proposition 2.4, and the proof of Theorem 1.3,

$$\int_{\mathfrak{D}_{\tilde R\mu_\alpha^l}(x_\alpha^l)} \Big|\hat u_\alpha - \sum_{i=1}^{l} u_\alpha^i\Big|^{2^*} d\sigma_{\hat h} = o(1). \tag{4-4}$$

By the assumption for $1 \le l \le k-1$, i.e.,

$$\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)\setminus\bigcup_{j=1}^s \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)} \Big|\hat{u}_\alpha - \sum_{i=1}^{l} u_\alpha^i - u_\alpha^k\Big|^{2^*} d\sigma_{\hat{h}} = o(1) + \varepsilon(R'),$$

combined with (4-4),

$$\int_{[\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)\setminus\bigcup_{j=1}^s \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)]\cap \mathfrak{D}_{\tilde{R}\mu_\alpha^l}(x_\alpha^l)} |u_\alpha^k|^{2^*} d\sigma_{\hat{h}} = o(1) + \varepsilon(R'),$$

so using (4-3) we arrive at

$$\int_{[D_R(0)\setminus\bigcup_{j=1}^s D_{2R'\lambda_\alpha^j/\mu_\alpha^k}(\tilde{y}_\alpha^j)]\cap D_{1/2\tilde{R}\mu_\alpha^l/\mu_\alpha^k}(\tilde{x}_\alpha^l)} |u^k|^{2^*} d\sigma_{\tilde{h}_\alpha} = o(1) + \varepsilon(R'). \tag{4-5}$$

Next, we consider two scenarios: first, assume that $d_{\hat{h}}(x_\alpha^l, x_\alpha^k)/\mu_\alpha^k \to +\infty$ as $\alpha \to +\infty$. We claim that $d_{\hat{h}}(x_\alpha^l, x_\alpha^k)/\mu_\alpha^l \to +\infty$ as $\alpha \to +\infty$. If not, then (4-5) with $\tilde{R}$ large enough yields that $\mu_\alpha^l/\mu_\alpha^k \to 0$ as $\alpha \to +\infty$. Moreover,

$$\frac{d_{\hat{h}}(x_\alpha^l, x_\alpha^k)}{\mu_\alpha^l} = \frac{d_{\hat{h}}(x_\alpha^l, x_\alpha^k)}{\mu_\alpha^k}\,\frac{\mu_\alpha^k}{\mu_\alpha^l},$$

so we can choose $\tilde{R} > 0$ such that $\mathfrak{D}_{\tilde{R}\mu_\alpha^k}(x_\alpha^k) \cap \mathfrak{D}_{\tilde{R}\mu_\alpha^l}(x_\alpha^l) = \varnothing$, which reduces to the previous Case 1; as a consequence, (4-1) holds for $l-1$.

Second, if $d_{\hat{h}}(x_\alpha^l, x_\alpha^k)/\mu_\alpha^k \nrightarrow +\infty$ as $\alpha \to +\infty$, then, up to a subsequence, $d_{\hat{h}}(x_\alpha^l, x_\alpha^k)/\mu_\alpha^k$ converges. So, (4-5) implies that $\mu_\alpha^l/\mu_\alpha^k \to +\infty$. Set $y_\alpha^{s+1} = x_\alpha^l$ and $\lambda_\alpha^{s+1} = \mu_\alpha^l$. Then,

$$\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)\setminus\bigcup_{j=1}^{s+1} \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)} \Big|\hat{u}_\alpha - \sum_{i=1}^{l} u_\alpha^i - u_\alpha^k\Big|^{2^*} d\sigma_{\hat{h}} = o(1) + \varepsilon(R')$$

and

$$\begin{aligned}\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)\setminus\bigcup_{j=1}^{s+1} \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)} |u_\alpha^l|^{2^*} d\sigma_{\hat{h}} &\le \int_{M\setminus\mathfrak{D}_{R'\mu_\alpha^l}(x_\alpha^l)} |u_\alpha^l|^{2^*} d\sigma_{\hat{h}} \\ &\le C\int_{\mathbb{R}^n\setminus D_{R'}(0)} |u^l|^{2^*} dx \le \varepsilon(R'),\end{aligned}$$

which yield that

$$\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)\setminus\bigcup_{j=1}^{s+1} \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)} \Big|\hat{u}_\alpha - \sum_{i=1}^{l-1} u_\alpha^i - u_\alpha^k\Big|^{2^*} d\sigma_{\hat{h}} = o(1) + \varepsilon(R').$$

In particular, (4-1) holds for $l-1$, as desired. The iteration process is thus completed.

Moreover, we have also shown that for any $i \neq j$

$$\frac{\mu_\alpha^i}{\mu_\alpha^j} + \frac{\mu_\alpha^j}{\mu_\alpha^i} + \frac{d_{\hat{h}}(x_\alpha^i, x_\alpha^j)^2}{\mu_\alpha^i \mu_\alpha^j} \to +\infty \quad \text{as } \alpha \to +\infty;$$

compare [Almaraz 2014; Druet et al. 2004; Struwe 1984]. Note that this convergence contains two kinds of bubbles: one case is that $\mu_\alpha^i = O(\mu_\alpha^j)$ when $\alpha \to +\infty$; then the two blow up points are far away from each other. The other case is that $\mu_\alpha^i = o(\mu_\alpha^j)$ or $\mu_\alpha^j = o(\mu_\alpha^i)$ when $\alpha \to +\infty$; then the distance of the two blow up point cannot be determined. Also we get that $\lambda_\alpha^j / \mu_\alpha^k \to 0$ as $\alpha \to +\infty$. □

Lemma 4.2. *The u^i (for $i = 0, 1, \dots, m$) that we get in the Theorem 1.3 are all nonnegative. In particular, for $i \geq 1$, u^i is of the form $U_{a_i}^{\lambda_i}$ for some $\lambda_i > 0$ and $a_i \in \mathbb{R}^n$, where $U_{a_i}^{\lambda_i}$ is as in (1-13).*

Proof. First of all, note that $u^0 \geq 0$ in $\overline{X}$ by Proposition 2.11. So, we just need to prove the positivity of u^i for $i \geq 1$. For any $k \in [1, m]$, taking $l = 0$ in Lemma 4.1,

$$\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k) \setminus \bigcup_{j=1}^s \mathfrak{D}_{R'\lambda_\alpha^j}(y_\alpha^j)} |\hat{u}_\alpha - U_\alpha^k|^{2^*} \, d\sigma_{\hat{h}} = o(1) + \varepsilon(R') \tag{4-6}$$

where

$$U_\alpha^k(x) = (\mu_\alpha^k)^{-\frac{n-2\gamma}{2}} u^k((\mu_\alpha^k)^{-1} \varphi_{x_\alpha^k}^{-1}(x)) \quad \text{for } x \in \mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)$$

is called a bubble. Since $u_\alpha = \hat{u}_\alpha + u^0$, for $x \in D_{r_0/\mu_\alpha^k}(0) \subset \mathbb{R}^n$, where r_0 is the same as the one mentioned in Theorem 1.3,

$$u_\alpha^k(x) = \tilde{u}_\alpha^k(x) + \tilde{u}_\alpha^{0,k}(x),$$

where

$$\begin{aligned} u_\alpha^k(x) &= (\mu_\alpha^k)^{\frac{n-2\gamma}{2}} u_\alpha(\varphi_{x_\alpha^k}(\mu_\alpha^k x)), \\ \tilde{u}_\alpha^k(x) &= (\mu_\alpha^k)^{\frac{n-2\gamma}{2}} \hat{u}_\alpha(\varphi_{x_\alpha^k}(\mu_\alpha^k x)), \\ \tilde{u}_\alpha^{0,k}(x) &= (\mu_\alpha^k)^{\frac{n-2\gamma}{2}} u^0(\varphi_{x_\alpha^k}(\mu_\alpha^k x)). \end{aligned}$$

Then, (4-6) implies that

$$\int_{D_R(0) \setminus \bigcup_{j=1}^s D_{2R'\lambda_\alpha^j/\mu_\alpha^k}(\tilde{y}_\alpha^j)} |\tilde{u}_\alpha^k - u^k|^{2^*} \, dx = o(1) + \varepsilon(R'), \tag{4-7}$$

where $\tilde{y}_\alpha^j = (\mu_\alpha^k)^{-1} \varphi_{x_\alpha^k}^{-1}(y_\alpha^j)$. Since $\{d_{\hat{h}}(x_\alpha^k, y_\alpha^j)/\mu_\alpha^k\}_{\alpha \in \mathbb{N}}$ is uniformly bounded by Lemma 4.1, $\{\tilde{y}_\alpha^j\}_{\alpha \in \mathbb{N}}$ is bounded and there exists a subsequence, also denoted by $\{\tilde{y}_\alpha^j\}$, such that $\tilde{y}_\alpha^j \to \tilde{y}^j$ as $\alpha \to +\infty$ for $j = 1, \dots, s$. Combining (4-7) with $\lambda_\alpha^j / \mu_\alpha^k \to 0$ as $\alpha \to +\infty$, we get

$$\tilde{u}_\alpha^k \to u^k \quad \text{in } L^{2^*}_{\text{loc}}(D_R(0) \setminus Y),$$

as $\alpha \to +\infty$ for $Y = \{\tilde{y}^j\}_{j=1}^s$, so

$$\tilde{u}_\alpha^k \to u^k \quad \text{a.e. in } \mathbb{R}^n,$$

since $R > 0$ is arbitrary.

Also note that

$$\int_{\mathfrak{D}_{R\mu_\alpha^k}(x_\alpha^k)} |u^0|^{2^*} \, d\sigma_{\hat{h}} = \int_{D_R(0)} |\tilde{u}_\alpha^{0,k}|^{2^*} \, d\sigma_{\tilde{h}_\alpha^k},$$

where $\tilde{h}_\alpha^k(x) = (\varphi_{x_\alpha^k}^* \hat{h})(\mu_\alpha^k x)$. Then, $\mu_\alpha^k \to 0$ as $\alpha \to +\infty$ and $u^0 \in L^{2^*}(M, \hat{h})$ yield that

$$\tilde{u}_\alpha^{0,k} \to 0, \quad \text{in } L^{2^*}(D_R(0), |dx|^2)$$

as $\alpha \to +\infty$, so

$$\tilde{u}_\alpha^{0,k} \to 0 \quad \text{a.e. in } \mathbb{R}^n$$

since $R > 0$ is arbitrary.

In particular, we have shown that $u_\alpha^k \to u^k$ almost everywhere on $\mathbb{R}^n$ as $\alpha \to +\infty$. Note that u_α is nonnegative by definition, so $u_\alpha^k \geq 0$ on $\mathbb{R}^n$. We conclude that $u^k \geq 0$ on $\mathbb{R}^n$. Then by the maximum principle, it follows that $u^k \geq 0$ in $\mathbb{R}_+^{n+1}$. Due to the previous arguments, u^k is of the form $U_{a_k}^{\lambda_k}$ for some $\lambda_k > 0$ and $a_k \in \mathbb{R}^n$, where $U_{a_k}^{\lambda_k}$ is as in (1-13). □

Appendix

We will prove the $\mathscr{C}^\infty$ estimates from the L^∞ estimates by the Harnack inequality. The two important lemmas are given here.

Lemma A.1 [González and Qing 2013]. *Let $R > 0$ and u be a weak solution of*

$$\text{(A-8)} \qquad \begin{cases} -\operatorname{div}(y^{1-2\gamma} \nabla u) = 0 & \text{in } B_{2R}^+(0), \\ -\lim_{y \to 0} y^{1-2\gamma} \partial_y u = f(x)u + g(x)|u|^{2^*-2}u & \text{on } D_{2R}(0). \end{cases}$$

Here, f and g are smooth functions on $D_{2R}(0)$. Assume that

$$\lambda = \int_{D_{2R}(0)} |u|^{2^*} \, dx < \infty.$$

Then, for any $p > 1$, there exists a constant $C_p = C(p, \lambda)$ such that

$$\sup_{B_R^+(0)} |u| + \sup_{D_R(0)} |u| \leq C_p \big(R^{-\frac{n+2-2\gamma}{p}} \|u\|_{L^p(B_{2R}^+(0))} + R^{-\frac{n}{p}} \|u\|_{L^p(D_{2R}(0))} \big).$$

Lemma A.2 [Jin et al. 2014]. *Let $a(x), b(x) \in \mathscr{C}^{\alpha}(D_2(0))$ for some $0 < \alpha \notin \mathbb{N}$ and let $u \in W^{1,2}(B_2^+(0), y^{1-2\gamma})$ be a weak solution of*

$$\begin{cases} -\operatorname{div}(y^{1-2\gamma}\nabla u) = 0 & \text{in } B_2^+(0), \\ -\lim_{y\to 0} y^{1-2\gamma}\,\partial_y u = a(x)u + b(x) & \text{on } D_2(0). \end{cases} \tag{A-9}$$

If $2\gamma + \alpha \notin \mathbb{N}$, then $u(\cdot, 0)$ is of $\mathscr{C}^{2\gamma+\alpha}(D_1(0))$, and

$$\|u(\cdot,0)\|_{C^{2\gamma+\alpha}(D_1(0))} \le C(\|u\|_{L^\infty(B_2^+(0))} + \|b\|_{C^\alpha(D_2(0))})$$

where $C > 0$ depends only on n, γ, α, and $\|a\|_{\mathscr{C}^\alpha(D_2(0))}$.

Acknowledgments

The authors thank Professor Jie Qing for useful discussions and comments on this work.

References

[Almaraz 2014] S. Almaraz, "The asymptotic behavior of Palais–Smale sequences on manifolds with boundary", *Pacific J. Math.* **269**:1 (2014), 1–17. MR 3233907 Zbl 1307.35116

[Caffarelli and Silvestre 2007] L. Caffarelli and L. Silvestre, "An extension problem related to the fractional Laplacian", *Comm. Partial Differential Equations* **32** (2007), 1245–1260. MR 2009k:35096 Zbl 1143.26002

[Case 2015] J. S. Case, "Some energy inequalities involving fractional GJMS operators", preprint, 2015. arXiv 1509.08347

[Case and Chang 2015] J. S. Case and S.-Y. A. Chang, "On fractional GJMS operators", *Comm. Pure Appl. Math.* (online publication February 2015).

[Chang and González 2011] S.-Y. A. Chang and M. d. M. González, "Fractional Laplacian in conformal geometry", *Adv. Math.* **226**:2 (2011), 1410–1432. MR 2012k:58057 Zbl 1214.26005

[Di Nezza et al. 2012] E. Di Nezza, G. Palatucci, and E. Valdinoci, "Hitchhiker's guide to the fractional Sobolev spaces", *Bull. Sci. Math.* **136**:5 (2012), 521–573. MR 2944369 Zbl 1252.46023

[Druet et al. 2004] O. Druet, E. Hebey, and F. Robert, *Blow-up theory for elliptic PDEs in Riemannian geometry*, Mathematical Notes **45**, Princeton University Press, 2004. MR 2005g:53058 Zbl 1059.58017

[González 2015] M. d. M. González, "Further results on the fractional Yamabe problem: the umbilic case", preprint, 2015. arXiv 1503.02862v1

[González and Qing 2013] M. d. M. González and J. Qing, "Fractional conformal Laplacians and fractional Yamabe problems", *Anal. PDE* **6**:7 (2013), 1535–1576. MR 3148060 Zbl 1287.35039

[Graham and Zworski 2003] C. R. Graham and M. Zworski, "Scattering matrix in conformal geometry", *Invent. Math.* **152**:1 (2003), 89–118. MR 2004c:58064 Zbl 1030.58022

[Graham et al. 1992] C. R. Graham, R. Jenne, L. J. Mason, and G. A. J. Sparling, "Conformally invariant powers of the Laplacian, I: Existence", *J. London Math. Soc.* (2) **46**:3 (1992), 557–565. MR 94c:58226 Zbl 0726.53010

[Jin and Xiong 2013] T. Jin and J. Xiong, "Sharp constants in weighted trace inequalities on Riemannian manifolds", *Calc. Var. Partial Differential Equations* **48**:3-4 (2013), 555–585. MR 3116023 Zbl 1288.46026

[Jin et al. 2014] T. Jin, Y. Li, and J. Xiong, "On a fractional Nirenberg problem, I: Blow up analysis and compactness of solutions", *J. Eur. Math. Soc.* (*JEMS*) **16**:6 (2014), 1111–1171. MR 3226738 Zbl 1300.53041

[Kufner 1985] A. Kufner, *Weighted Sobolev spaces*, Wiley, New York, 1985. MR 86m:46033 Zbl 0567.46009

[Palatucci and Pisante 2014] G. Palatucci and A. Pisante, "Improved Sobolev embeddings, profile decomposition, and concentration-compactness for fractional Sobolev spaces", *Calc. Var. Partial Differential Equations* **50**:3-4 (2014), 799–829. MR 3216834 Zbl 1296.35064

[Palatucci and Pisante 2015] G. Palatucci and A. Pisante, "A global compactness type result for Palais–Smale sequences in fractional Sobolev spaces", *Nonlinear Anal.* **117** (2015), 1–7. MR 3316602 Zbl 1312.35092

[Paneitz 2008] S. M. Paneitz, "A quartic conformally covariant differential operator for arbitrary pseudo-Riemannian manifolds (summary)", *SIGMA Symmetry Integrability Geom. Methods Appl.* **4** (2008), Paper 036, 3. MR 2009b:58080 Zbl 1145.53053

[Struwe 1984] M. Struwe, "A global compactness result for elliptic boundary value problems involving limiting nonlinearities", *Math. Z.* **187**:4 (1984), 511–517. MR 86k:35046 Zbl 0535.35025

[Yan et al. 2015] S. Yan, J. Yang, and X. Yu, "Equations involving fractional Laplacian operator: compactness and application", *J. Funct. Anal.* **269**:1 (2015), 47–79. MR 3345604 Zbl 1317.35287

Received October 20, 2014.

YI FANG
SCHOOL OF MATHEMATICAL SCIENCES
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA
GUANLI-KEYAN BUILDING
96 JINZHAI ROAD, HEFEI
230026 ANHUI
CHINA
fangyi1915@gmail.com

MARÍA DEL MAR GONZÁLEZ
ETSEIB – DEPARTAMENT DE MATEMATICA APLICADA I
UNIVERSITAT POLITECNICA DE CATALUNYA
AVENIDA DIAGONAL, 647
08028 BARCELONA
SPAIN
mar.gonzalez@upc.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.278.407

K-THEORY AND HOMOTOPIES OF 2-COCYCLES ON HIGHER-RANK GRAPHS

ELIZABETH GILLASPY

This paper continues our investigation into the question of when a homotopy of 2-cocycles on a locally compact Hausdorff groupoid gives rise to an isomorphism of the *K*-theory groups of the twisted groupoid C^*-algebras. Our main result, which builds on work by Kumjian, Pask, and Sims, shows that a homotopy of 2-cocycles on a row-finite higher-rank graph Λ gives rise to twisted groupoid C^*-algebras with isomorphic *K*-theory groups. (The groupoid in question is the path groupoid of Λ.) We also establish a technical result: any homotopy of 2-cocycles on a locally compact Hausdorff groupoid $\mathcal{G}$ gives rise to an upper semicontinuous bundle of C^*-algebras.

1. Introduction

Higher-rank graphs, or k-graphs, provide a k-dimensional analogue of directed graphs. They were introduced by Kumjian and Pask [2000] to provide a combinatorial model for the higher-rank Cuntz–Krieger algebras studied by Robertson and Steger [1999]. Much of the interest in the C^*-algebras $C^*(\Lambda)$ associated to k-graphs Λ stems from the multiple ways one can model $C^*(\Lambda)$ — the k-graph Λ reflects many of the properties of $C^*(\Lambda)$, but we can also describe $C^*(\Lambda)$ as a universal C^*-algebra for certain generators and relations, or as a groupoid C^*-algebra $C^*(\Lambda) \cong C^*(\mathcal{G}_\Lambda)$.

The class of groupoids includes groups, group actions, equivalence relations, and group bundles. Renault [1980] initiated the study of groupoid C^*-algebras, and the theory and applications of groupoid C^*-algebras have since been developed by many researchers. Given a 2-cocycle $\omega \in Z^2(\mathcal{G}, \mathbb{T})$ on a groupoid $\mathcal{G}$, Renault also explained how to construct the twisted groupoid C^*-algebra $C^*(\mathcal{G}, \omega)$. These objects have received relatively little attention until quite recently, but it has now become clear that twisted groupoid C^*-algebras can help answer many questions about the structure of untwisted groupoid C^*-algebras (see [Muhly and Williams 1992; Muhly et al. 1996; Clark and an Huef 2012; an Huef et al. 2011; Brown

MSC2010: 46L05, 46L80.

Keywords: higher-rank graph, twisted groupoid C^*-algebra, *K*-theory, twisted k-graph C^*-algebra, upper semicontinuous C^*-bundle, $C_0(X)$-algebra, groupoid, 2-cocycle.

and an Huef 2014]), as well as classifying those C^*-algebras which admit diagonal subalgebras (also known as Cartan subalgebras) — see [Kumjian 1986]. In another direction, [Tu et al. 2004] establishes a connection between the K-theory of twisted groupoid C^*-algebras and the classification of D-brane charges in string theory.

Two recent papers have explored the effect of a homotopy $\{\omega_t\}_{t\in[0,1]}$ of 2-cocycles on the K-theory of the twisted groupoid C^*-algebras. Echterhoff, Lück, Phillips, and Walters showed in [Echterhoff et al. 2010, Theorem 1.9] that if G is a group that satisfies the Baum–Connes conjecture with respect to the coefficient algebras $\mathcal{K}$ and $C([0,1],\mathcal{K})$, and if $\{\omega_t\}_{t\in[0,1]}$ is a homotopy of 2-cocycles on G, then the K-theory groups of the reduced twisted group C^*-algebras are unperturbed by the homotopy

$$K_*(C_r^*(G,\omega_0)) \cong K_*(C_r^*(G,\omega_1)). \tag{1}$$

In particular, taking $G=\mathbb{Z}^2$, we obtain another proof of the fact, established by Pimsner and Voiculescu [1980], that all of the rotation algebras $\{A_\theta\}_{\theta\in[0,1]}$ have isomorphic K-theory groups.

Kumjian, Pask, and Sims also studied the effect of a homotopy of 2-cocycles on K-theory in [Kumjian et al. 2013]. Theorem 5.4 of [Kumjian et al. 2013] establishes that if Λ is a row-finite, source-free k-graph and c is a 2-cocycle on Λ such that $c(\lambda,\mu)=e^{2\pi i\sigma(\lambda,\mu)}$ for some $\mathbb{R}$-valued 2-cocycle σ, then we have $K_*(C^*(\Lambda))\cong K_*(C^*(\Lambda,c))$. Defining $c_t(\lambda,\mu)=e^{2\pi i t\sigma(\lambda,\mu)}$ for $t\in[0,1]$ gives us a homotopy of 2-cocycles linking c and the trivial 2-cocycle. Moreover, Corollary 7.8 of [Kumjian et al. 2015] tells us that $C^*(\Lambda,c)$ is isomorphic to a twisted groupoid C^*-algebra $C^*(\mathcal{G}_\Lambda,\omega_c)$. Thus, we can view [Kumjian et al. 2013, Theorem 5.4] as a result about homotopic 2-cocycles on groupoids.

Inspired by the above-mentioned results, we have begun exploring the question of when a homotopy of 2-cocycles on a locally compact Hausdorff groupoid $\mathcal{G}$ induces an isomorphism of the K-theory groups of the (full or reduced) twisted groupoid C^*-algebras. In a previous article [Gillaspy 2015], we extended the above-mentioned Theorem 1.9 of [Echterhoff et al. 2010] to the case when $\mathcal{G}=G\ltimes X$ is a transformation group, where X is locally compact Hausdorff and G satisfies the Baum–Connes conjecture with coefficients.

We prove the following generalization of [Kumjian et al. 2013, Theorem 5.4].

Theorem 4.1. *Let Λ be a row-finite k-graph with no sources and let $\{c_t\}_{t\in[0,1]}$ be a homotopy of 2-cocycles in $\underline{Z}^2(\Lambda,\mathbb{T})$. Then $\{c_t\}_{t\in[0,1]}$ gives rise to a homotopy $\{\sigma_{c_t}\}_{t\in[0,1]}$ of 2-cocycles on $\mathcal{G}_\Lambda$ such that*

$$K_*(C^*(\mathcal{G}_\Lambda,\sigma_{c_0}))\cong K_*(C^*(\mathcal{G}_\Lambda,\sigma_{c_1})).$$

As of this writing, we are unaware of any examples of groupoids $\mathcal{G}$ and homotopies $\omega=\{\omega_t\}_{t\in[0,1]}$ of 2-cocycles on $\mathcal{G}$ where the homotopy does not induce an isomorphism of the K-theory groups of the twisted groupoid C^*-algebras.

Outline. This paper begins by recalling the definitions of a higher-rank graph and of a groupoid in Section 2, as well as the definition of a 2-cocycle in each category, and sketching the procedure by which we can construct a C^*-algebra from these objects. In Section 3 we define a homotopy of 2-cocycles on a k-graph and on a groupoid, and show that the definitions are compatible. We also prove a technical result (Theorem 3.3), namely, that a homotopy $\{\omega_t\}_{t\in[0,1]}$ of 2-cocycles on a groupoid $\mathcal{G}$ gives rise to a $C([0,1])$-algebra with fiber algebra $C^*(\mathcal{G},\omega_t)$ at $t\in[0,1]$. We expect that this result will prove useful in future work, as we search for more classes of groupoids where a homotopy of 2-cocycles induces an isomorphism of the K-theory groups of the twisted groupoid C^*-algebras.

In Section 4 we begin the proof of Theorem 4.1. Our proof technique consists of proving a stronger version of Theorem 4.1 in a simple case, and then showing how to use this simple case to obtain our desired result for general k-graphs. To be precise, Proposition 4.2 shows that when the degree map d on Λ satisfies $d(\lambda) = b(s(\lambda)) - b(r(\lambda))$ for all $\lambda \in \Lambda$, the $C([0,1])$-algebra associated to a homotopy of 2-cocycles on Λ is actually a trivial continuous field. We then show how to exploit the triviality of the continuous field in this special case to see that a homotopy of 2-cocycles on any row-finite, source-free k-graph Λ induces an isomorphism $K_*(C^*(\Lambda,c_0)) \cong K_*(C^*(\Lambda,c_1))$. The argument closely parallels Section 5 of [Kumjian et al. 2013].

2. Groupoids and k-graphs

Definition 2.1 [Kumjian and Pask 2000, Definition 1.1]. A *higher-rank graph* of degree k, or a k-graph, is a nonempty countable small category Λ equipped with a functor $d:\Lambda\to\mathbb{N}^k$ (the *degree map*) satisfying the following *factorization property*: Given a morphism $\lambda\in\Lambda$ with $d(\lambda)=m+n$, there exist unique $\mu,\nu\in\Lambda$ such that $\lambda=\mu\nu$, $d(\mu)=m$, and $d(\nu)=n$.

The simplest example of a k-graph is $\mathbb{N}^k$, equipped with the identity morphism $\mathrm{id}:\mathbb{N}^k\to\mathbb{N}^k$.

In this article we use the arrows-only picture of category theory, so that we think of the objects of a category Λ as identity morphisms. Hence, $\lambda\in\Lambda$ means that λ is a morphism in Λ. Given an element λ in a category Λ, write $s(\lambda)$ for the domain, or *source*, of the morphism λ, and write $r(\lambda)$ for its target, or *range*. We say a k-graph Λ is *row-finite* if, for any $v\in\mathrm{Obj}(\Lambda)$ and any $n\in\mathbb{N}^k$, the set

$$v\Lambda^n := \{\lambda\in\Lambda : r(\lambda)=v,\ d(\lambda)=n\}$$

is finite. We say Λ has *no sources* if $v\Lambda^n\neq\varnothing$ for every $v\in\mathrm{Obj}(\Lambda)$ and every $n\in\mathbb{N}^k$. We only consider k-graphs which are row-finite and have no sources, since these are the k-graphs which we can study via the groupoid method that was introduced in [Kumjian and Pask 2000] and which we will explain in Section 2.

Definition 2.2 [Renault 1980, Definition I.1.12; Kumjian et al. 2015, Section 3]. For a category Λ, let $\Lambda^{*2} = \{(\lambda_1, \lambda_2) \in \Lambda \times \Lambda : s(\lambda_1) = r(\lambda_2)\}$. A function $c : \Lambda^{*2} \to \mathbb{T}$ is called a *2-cocycle* on Λ if

$$c(\lambda, \mu\nu)c(\mu, \nu) = c(\lambda\mu, \nu)c(\lambda, \mu) \tag{2}$$

whenever $(\lambda, \mu), (\mu, \nu) \in \Lambda^{*2}$ and $c(\lambda, s(\lambda)) = c(r(\lambda), \lambda) = 1$ for all $\lambda \in \Lambda$. We write $\underline{Z}^2(\Lambda, \mathbb{T})$ for the set of 2-cocycles on Λ.

If c, $\tilde{c}$ are two 2-cocycles on Λ, we say that c, $\tilde{c}$ are *cohomologous* if there exists a function $b : \Lambda \to \mathbb{T}$ such that

$$\tilde{c}(\mu, \nu) := b(\mu)b(\nu)b(\mu\nu)^{-1}c(\mu, \nu) = \delta b(\mu, \nu)c(\mu, \nu) \quad \text{for all } (\mu, \nu) \in \Lambda^{*2}.$$

We note that cohomologous 2-cocycles give rise to isomorphic twisted C^*-algebras (see [Kumjian et al. 2015, Proposition 5.6; Renault 1980, Proposition II.1.2]).

The only cocycles we consider in this paper are 2-cocycles, so we will occasionally drop the "2" and refer to them simply as *cocycles*.

Definition 2.3 [Kumjian et al. 2015, Definition 5.2]. The *twisted higher-rank-graph algebra* $C^*(\Lambda, c)$ associated to a k-graph Λ and a 2-cocycle c on Λ is the universal C^*-algebra generated by a collection $\{s_\lambda\}_{\lambda\in\Lambda}$ of partial isometries satisfying the following *twisted Cuntz–Krieger relations*:

(CK1) $\{s_v\}_{v\in \mathrm{Obj}(\Lambda)}$ is a collection of mutually orthogonal projections;

(CK2) $s_\mu s_\nu = c(\mu, \nu)s_{\mu\nu}$ whenever $s(\mu) = r(\nu)$;

(CK3) $s_\mu^* s_\mu = s_{s(\mu)}$ for all $\mu \in \Lambda$;

(CK4) $s_v = \sum_{\mu\in v\Lambda^n} s_\mu s_\mu^*$ for all $v \in \mathrm{Obj}(\Lambda)$ and all $n \in \mathbb{N}^k$.

Note that every k-graph Λ admits at least one 2-cocycle: the trivial cocycle, obtained by setting $c(\lambda, \mu) = 1$ for all $(\lambda, \mu) \in \Lambda^{*2}$. In this case, the definition above of $C^*(\Lambda, c)$ agrees with that of $C^*(\Lambda)$ given in [Kumjian and Pask 2000, Definition 1.5]. For example, if Λ is $\mathbb{N}^k$ and c is the trivial cocycle, then we have $C^*(\Lambda, c) \cong C(\mathbb{T}^k)$. More generally, if Λ is $\mathbb{N}^2$, let $c_\theta : \Lambda^{*2} \to \mathbb{T}$ be given by $c_\theta((m, n), (j, k)) = e^{2\pi i\theta nj}$. Then c_θ is a 2-cocycle on Λ and $C^*(\Lambda, c_\theta)$ is isomorphic to the rotation algebra A_θ.

Groupoids. In this section, we review the construction of a twisted groupoid C^*-algebra set forth in [Renault 1980], as well as the procedure given in the seminal article [Kumjian and Pask 2000] for associating a groupoid to a k-graph. Theorem 3.3 applies to arbitrary locally compact Hausdorff groupoids, so we present in full generality all the definitions necessary for the construction of a twisted groupoid C^*-algebra.

A *groupoid* $\mathcal{G}$ is a small category with inverses. We use the notation of [Renault 1980] to denote groupoid elements and operations; for example, $\mathcal{G}^{(2)} \subseteq \mathcal{G} \times \mathcal{G}$ denotes the set of composable pairs and $\mathcal{G}^{(0)}$ denotes the unit space. If $u \in \mathcal{G}^{(0)}$, we write

$$\mathcal{G}_u = \{x \in \mathcal{G} : s(x) = u\}, \quad \mathcal{G}^u = \{x \in \mathcal{G} : r(x) = u\}.$$

In this article, we restrict our attention to groupoids which admit a locally compact Hausdorff topology in which the operations of composition (or multiplication) and inversion are continuous.

In addition to the groupoids associated to k-graphs, examples of groupoids include groups, vector bundles, and transformation groups. For details and examples, see [Goehle 2009].

Given a row-finite, source-free k-graph Λ, Section 2 of [Kumjian and Pask 2000] describes how to form the associated path groupoid $\mathcal{G}_\Lambda$:

Definition 2.4 [Kumjian and Pask 2000, Example 1.7(ii)]. Define the k-graph Ω_k to be the category with $\mathrm{Obj}(\Omega_k) = \mathbb{N}^k$ and morphisms $\Omega_k = \{(m, n) \in \mathbb{N}^k \times \mathbb{N}^k : n \geq m\}$. We have $r(m, n) = m$, $s(m, n) = n$, $d(m, n) = n - m$. Composition in Ω_k is given by $(m, n)(n, \ell) = (m, \ell)$.

For a k-graph Λ, let Λ^∞ denote the set of degree-preserving functors $x : \Omega_k \to \Lambda$. When $k = 1$, the elements $x \in \Lambda^\infty$ are the infinite paths in Λ.

Given $p \in \mathbb{N}^k$, define $\sigma^p : \Lambda^\infty \to \Lambda^\infty$ by $\sigma^p(x)(m, n) = x(m + p, n + p)$. When Λ is row-finite and source-free, Proposition 2.3 in [Kumjian and Pask 2000] shows that if $\lambda \in \Lambda$, $x \in \Lambda^\infty$ satisfy $s(\lambda) = x(0)$, there is a unique $y \in \Lambda^\infty$ such that $\sigma^{d(\lambda)}(y) = x$; we often write $y = \lambda x$.

Definition 2.5 [Kumjian and Pask 2000, Definition 2.1]. Given a row-finite, source-free k-graph Λ, the associated path groupoid $\mathcal{G}_\Lambda$ is the groupoid associated to the equivalence relation on Λ^∞ of "shift equivalence with lag". In other words,

$$\mathcal{G}_\Lambda := \{(x, n - m, y) \in \Lambda^\infty \times \mathbb{Z}^k \times \Lambda^\infty : n, m \in \mathbb{N}^k,\ \sigma^n(x) = \sigma^m(y)\}$$

and $\mathcal{G}_\Lambda^{(0)} = \Lambda^\infty$, with $r(x, \ell, y) = x$, $s(x, \ell, y) = y$ and multiplication and inversion in $\mathcal{G}_\Lambda$ given by $(x, \ell, y)(y, m, z) = (x, \ell + m, z)$, $(x, \ell, y)^{-1} = (y, -\ell, x)$.

When Λ is a row-finite, source-free k-graph, Proposition 2.8 in [Kumjian and Pask 2000] tells us that the sets

$$Z(\mu, \nu) := \{(\mu x, d(\mu) - d(\nu), \nu x) : x(0) = s(\mu) = s(\nu)\}$$

form a basis of compact open sets for a locally compact Hausdorff topology on $\mathcal{G}_\Lambda$ (in fact, with this topology $\mathcal{G}_\Lambda$ is an ample étale groupoid).

To build a C^*-algebra out of a groupoid $\mathcal{G}$ we start by putting a $*$-algebra structure on $C_c(\mathcal{G})$, and to do this we need to integrate over the groupoid $\mathcal{G}$. A Haar system $\{\lambda^u\}_{u\in\mathcal{G}^{(0)}}$ (the groupoid analogue of Haar measure for groups; see [Renault 1980, Definition I.2.2]) will allow us to do this. Unlike in the group case, one cannot make existence or uniqueness statements about Haar systems for groupoids, so one usually starts by hypothesizing the existence of a fixed Haar system. For example, we obtain a Haar system $\{\lambda^x\}_{x\in\Lambda^\infty}$ on $\mathcal{G}_\Lambda$ by setting

$$\lambda^x(E) = \#\{e \in E : e = (x, n, y) \text{ for some } n \in \mathbb{Z}^k,\ y \in \Lambda^\infty\}.$$

We will always use this Haar system on $\mathcal{G}_\Lambda$ in this paper.

Definition 2.6. Let $\mathcal{G}$ be a locally compact Hausdorff groupoid equipped with a Haar system $\{\lambda^u\}_{u\in\mathcal{G}^{(0)}}$ and a continuous 2-cocycle ω. We define a $*$-algebra structure on $C_c(\mathcal{G})$ as follows: for $f, g \in C_c(\mathcal{G})$ let

$$f *_\omega g(a) = \int_{\mathcal{G}_{s(a)}} f(ab)g(b^{-1})\omega(ab, b^{-1})\, d\lambda^{s(a)}(b),$$
$$f^*(a) = \overline{f(a^{-1})\omega(a, a^{-1})}.$$

From [Renault 1980, Proposition II.1.1] we know that the multiplication is well-defined (that is, that $f *_\omega g \in C_c(\mathcal{G})$ as claimed) and associative, and that $(f^*)^* = f$ so that the involution is involutive. The proof of associativity relies on the cocycle condition (2).

Given the fundamental role that the cocycle ω plays in the multiplication and involution on $C_c(\mathcal{G})$, we will often write $C_c(\mathcal{G}, \omega)$ to denote the set $C_c(\mathcal{G})$ equipped with the $*$-algebra structure of Definition 2.6. We define $C^*(\mathcal{G}, \omega)$ to be the completion of $C_c(\mathcal{G}, \omega)$ in the maximal C^*-norm, as described in Chapter II of [Renault 1980].

Definition 2.7. When $\mathcal{G} = \mathcal{G}_\Lambda$ is the groupoid associated to a row-finite k-graph Λ with no sources, Lemma 6.3 of [Kumjian et al. 2015] explains how, given a cocycle $c \in \underline{Z}^2(\Lambda, \mathbb{T})$, we can construct a cocycle $\sigma_c \in Z^2(\mathcal{G}_\Lambda, \mathbb{T})$. Then Corollary 7.8 of [Kumjian et al. 2015] shows that $C^*(\Lambda, c) \cong C^*(\mathcal{G}_\Lambda, \sigma_c)$. The construction of σ_c is rather technical, but since we will need the details later, we present it here.

Lemma 6.6 of [Kumjian et al. 2015] establishes the existence of a subset

$$\mathcal{P} \subseteq \{Z(\mu, \nu) : s(\mu) = s(\nu)\}$$

that partitions $\mathcal{G}_\Lambda$. In other words, every $a \in \mathcal{G}_\Lambda$ has exactly one representation of the form $a = (\mu_a x, d(\mu_a) - d(\nu_a), \nu_a x)$ with $Z(\mu_a, \nu_a) \in \mathcal{P}$. Note that if $(a, b) \in \mathcal{G}_\Lambda^{(2)}$, we need not have $\mu_a = \mu_{ab}$ or $\nu_b = \nu_{ab}$. However, given $(a, b) \in \mathcal{G}_\Lambda^{(2)}$, Lemma 6.3(i) of [Kumjian et al. 2015] shows that we can always find $y \in \Lambda^\infty$ and $\alpha, \beta, \gamma \in \Lambda$

such that

$$\begin{aligned} a &= (\mu_a \alpha y, d(\mu_a) - d(\nu_a), \nu_a \alpha y), \\ b &= (\mu_b \beta y, d(\mu_b) - d(\nu_b), \nu_b \beta y), \\ ab &= (\mu_{ab} \gamma y, d(\mu_{ab}) - d(\nu_{ab}), \nu_{ab} \gamma y). \end{aligned}$$

Then, given a 2-cocycle c on Λ, we define a 2-cocycle σ_c on $\mathcal{G}_\Lambda$ by

$$\sigma_c(a, b) = c(\mu_a, \alpha) c(\mu_b, \beta) c(\nu_{ab}, \gamma) \overline{c(\nu_a, \alpha) c(\nu_b, \beta) c(\mu_{ab}, \gamma)}.$$

Since c satisfies the cocycle condition (2), it's straightforward to check that σ_c does also. Lemma 6.3 of [Kumjian et al. 2015] checks that σ_c is well-defined and continuous, so we can construct the groupoid C^*-algebra $C^*(\mathcal{G}_\Lambda, \sigma_c)$ as outlined above. Corollary 7.8 of [Kumjian et al. 2015] tells us that $C^*(\mathcal{G}_\Lambda, \sigma_c) \cong C^*(\Lambda, c)$.

Theorem 6.5 of [Kumjian et al. 2015] establishes that different choices of partitions $\mathcal{P}$ give rise to cohomologous groupoid cocycles, and hence to isomorphic twisted groupoid C^*-algebras.

3. Homotopies of cocycles

In order to define a homotopy of groupoid 2-cocycles, we begin by observing that, given any locally compact Hausdorff groupoid $\mathcal{G}$, we can make $\mathcal{G} \times [0, 1]$ into a locally compact Hausdorff groupoid by equipping it with the product topology and setting $(\mathcal{G} \times [0, 1])^{(2)} := \mathcal{G}^{(2)} \times [0, 1]$. In other words, $(\mathcal{G} \times [0, 1])^{(0)} = \mathcal{G}^{(0)} \times [0, 1]$ and

$$r(\gamma, t) = (r(\gamma), t), \quad s(\gamma, t) = (s(\gamma), t).$$

Moreover, if $\mathcal{G}$ has a Haar system $\{\lambda^u\}_{u \in \mathcal{G}^{(0)}}$, then setting $\lambda^{u,t} := \lambda^u$ for every $t \in [0, 1]$ gives rise to a Haar system on $\mathcal{G} \times [0, 1]$. We will always use this Haar system on $\mathcal{G} \times [0, 1]$ in this paper.

Definition 3.1 [Gillaspy 2015, Definition 2.12]. A *homotopy of* (2-)*cocycles* on a locally compact Hausdorff groupoid $\mathcal{G}$ is a 2-cocycle $\omega \in Z^2(\mathcal{G} \times [0, 1], \mathbb{T})$. We say that two cocycles $\omega_0, \omega_1 \in Z^2(\mathcal{G}, \mathbb{T})$ are *homotopic* if there exists a homotopy $\omega \in Z^2(\mathcal{G} \times [0, 1], \mathbb{T})$ such that $\omega_i = \omega|_{\mathcal{G} \times \{i\}}$ for $i = 0, 1$.

If ω is a homotopy of cocycles on $\mathcal{G}$ linking ω_0, ω_1, Theorem 3.3 below tells us that $C^*(\mathcal{G}, \omega_0)$ and $C^*(\mathcal{G}, \omega_1)$ are quotients of $C^*(\mathcal{G} \times [0, 1], \omega)$. This is fundamental to the proof of our main result, Theorem 4.1.

Definition 3.2 [Williams 2007, Definition C.1]. Let X be a locally compact Hausdorff space. A C^*-algebra A is a $C_0(X)$-*algebra* if we have a $*$-homomorphism $\Phi : C_0(X) \to ZM(A)$ such that

$$A = \overline{\operatorname{span}}\{\Phi(f)a : f \in C_0(X), \ a \in A\}.$$

We usually write $f \cdot a$ for $\Phi(f)a$.

If A is a $C_0(X)$-algebra, then, for any $x \in X$, $\overline{\text{span}} C_0(X \backslash x) \cdot A$ is an ideal I_x. We call $A_x := A/I_x$ the *fiber* of A at $x \in X$.

Theorem 3.3. *Let ω be a homotopy of cocycles on a locally compact Hausdorff groupoid $\mathcal{G}$ with Haar system $\{\lambda^u\}_{u \in \mathcal{G}^{(0)}}$. Then $C^*(\mathcal{G} \times [0,1], \omega)$ is a $C([0,1])$-algebra, with fiber $C^*(\mathcal{G}, \omega_t)$ at $t \in [0,1]$.*

Proof. We begin by checking that $C^*(\mathcal{G} \times [0,1], \omega)$ is a $C([0,1])$-algebra. For $f \in C([0,1])$, $\phi \in C_c(\mathcal{G} \times [0,1], \omega)$, define

$$f \cdot \phi(a, t) = f(t)\phi(a, t).$$

It's not difficult to check that this action extends to a $*$-homomorphism

$$\Phi : C([0,1]) \to ZM(C^*(\mathcal{G} \times [0,1], \omega))$$

such that $\|\Phi(f)\phi\| \le \|f\|_\infty \|\phi\|$, or to check that

$$\Phi(C([0,1])) \cdot C_c(\mathcal{G} \times [0,1], \omega) = C_c(\mathcal{G} \times [0,1], \omega)$$

is dense in $C^*(\mathcal{G} \times [0,1], \omega)$. In other words, Φ makes $C^*(\mathcal{G} \times [0,1], \omega)$ into a $C([0,1])$-algebra as claimed.

Fix $t \in [0,1]$ and denote by $q_t : C_c(G \times [0,1], \omega) \to C_c(G, \omega_t)$ the evaluation map. Then q_t is bounded by the I-norm (see [Renault 1980, Section II.1]), and hence extends to a surjective $*$-homomorphism $q_t : C^*(\mathcal{G} \times [0,1], \omega) \to C^*(\mathcal{G}, \omega_t)$. In other words, $C^*(\mathcal{G}, \omega_t)$ is a quotient of $C^*(\mathcal{G} \times [0,1], \omega)$. To see that $C^*(\mathcal{G}, \omega_t) \cong C^*(\mathcal{G} \times [0,1], \omega)_t$, we need to check that $\ker q_t = I_t$. A standard approximation argument will show that $\ker q_t \supseteq I_t$; thus, we will only detail the proof that $\ker q_t \subseteq I_t$.

Note that the fiber algebra $C^*(\mathcal{G} \times [0,1], \omega)_t \cong C^*(\mathcal{G} \times [0,1], \omega)/I_t$ can be calculated as a completion $\overline{C_c(\mathcal{G} \times [0,1], \omega)}$ with respect to the norm given by

$$\|f\|_t := \sup\{\|L(f)\| : L(I_t) = 0,\ L \text{ is an } I\text{-norm-bounded representation}\}.$$

Thus, to show that $\ker q_t \subseteq I_t$, we will show that each such representation L factors through q_t.

Given such a representation $L : C_c(\mathcal{G} \times [0,1], \omega) \to B(\mathcal{H})$, define $L' : C_c(\mathcal{G}, \omega_t) \to B(\mathcal{H})$ by $L'(q_t(f)) := L(f)$. We claim that L' is an I-norm-bounded representation of $C_c(\mathcal{G}, \omega_t)$. To see this, it suffices to check that L' is well-defined and bounded.

Lemma 3.4. *If $f, g \in C_c(\mathcal{G} \times [0,1], \omega)$ satisfy $q_t(f) = q_t(g)$, then the function $h = f - g \in C_c(\mathcal{G} \times [0,1], \omega)$ lies in I_t. Consequently, $L(f) = L(g)$ and L' is well-defined on $C_c(\mathcal{G}, \omega_t)$.*

Proof. Let $\{f_i\}_{i \in I}$ be an approximate unit for $C_0([0,1] \backslash t)$ such that $f_i(s) \nearrow 1$ for every $s \neq t$; moreover, suppose that for each i there exists a $\delta_i > 0$ such that $f_i(s) = 1$ if $|s - t| \ge \delta_i$. We will show that the I-norm $\|h - f_i h\|_I$ tends to 0. Consequently, $h = \lim_i f_i h$ in $C^*(\mathcal{G} \times [0,1], \omega)$, so $h \in I_t$.

For any $k \in C_c(\mathcal{G} \times [0,1], \omega)$, the axioms of a Haar system tell us that the function $(u,t) \mapsto \int |k(a,t)| d\lambda^{u,t}(a)$ is in $C_0(\mathcal{G}^{(0)} \times [0,1])$. In particular, if we take k to be a function that equals 1 where h is nonzero and vanishes rapidly off $\operatorname{supp} h$, this shows us that $\phi(u,t) := \lambda^{u,t}(\operatorname{supp} h)$ is a pointwise limit of functions in $C_0(\mathcal{G}^{(0)} \times [0,1])$, and hence is bounded. Let $K = \max \phi$.

Let $\epsilon > 0$ be given. Since h is compactly supported and since $h(a,t) = 0$ for all $a \in \mathcal{G}$, we can choose a $\delta > 0$ such that $|h(a,s)| < \epsilon/K$ for all $a \in \mathcal{G}$ whenever $|s-t| < \delta$, and we can choose a j such that $i \geq j$ means $\delta_i < \delta$. Then, if $|s-t| < \delta$,

$$\int_{\mathcal{G}^u} |h(a,s) - f_i(s)h(a,s)| \, d\lambda^u(a) = (1 - f_i(s)) \int_{G^u} |h(a,s)| \, d\lambda^u(a) < 1 \cdot \epsilon.$$

On the other hand, if $|s-t| \geq \delta > \delta_i$, then $f_i(s) = 1$ and

$$\int_{\mathcal{G}^u} |h(a,s) - f_i(s)h(a,s)| \, d\lambda^u(a) = 0$$

for any $u \in \mathcal{G}^{(0)}$. In either case, given any $\epsilon > 0$ we can always choose j such that $i \geq j$ implies

$$\|h - f_i h\|_I = \max\Big\{ \sup_{s\in[0,1]} \sup_{u\in\mathcal{G}^{(0)}} \int_{\mathcal{G}^u} |h(a,s) - f_i(s)h(a,s)| \, d\lambda^u(a), \\ \sup_{s\in[0,1]} \sup_{u\in\mathcal{G}^{(0)}} \int_{\mathcal{G}^u} |h(a^{-1},s) - f_i(s)h(a^{-1},s)| \, d\lambda^u(a) \Big\} < \epsilon.$$

Since $\|h - f_i h\| \leq \|h - f_i h\|_I$ and I_t is closed, it follows that $h \in I_t$ as desired, and so $L(h) = 0$. □

Having seen that L' is well-defined, we now proceed to show that it is bounded.

Lemma 3.5. *For any fixed $f \in C_c(\mathcal{G} \times [0,1], \omega)$, the map $s \mapsto \|q_s(f)\|_I$ is continuous.*

Proof. Fix $f \in C_c(\mathcal{G} \times [0,1], \omega)$ and fix $t \in [0,1]$. As in the proof of Lemma 3.4, let K denote the supremum of the function $(u,s) \mapsto \lambda^{u,s}(\operatorname{supp} f)$. Since f has compact support, given $\epsilon > 0$ we can choose a δ such that

$$|s-t| < \delta \Rightarrow |f(a,t) - f(a,s)| < \frac{\epsilon}{2K} \quad \text{for all } a \in \mathcal{G}.$$

Now, by definition of the I-norm, there exists a $u \in \mathcal{G}^{(0)}$ such that either

$$\|q_s(f)\|_I < \int_{\mathcal{G}^u} |f(a,s)| \, d\lambda^u(a) + \frac{\epsilon}{2}$$

or

$$\|q_s(f)\|_I < \int_{\mathcal{G}^u} |f(a^{-1},s)| \, d\lambda^u(a) + \frac{\epsilon}{2}.$$

It follows that either

$$\|q_s(f)\|_I < \int_{\mathcal{G}^u} |f(a,t)| + \frac{\epsilon}{2K}\, d\lambda^u(a) + \frac{\epsilon}{2} \le \int_{\mathcal{G}^u} |f(a,t)|\, d\lambda^u(a) + \epsilon$$

or

$$\|q_s(f)\|_I < \int_{\mathcal{G}^u} |f(a^{-1},t)| + \frac{\epsilon}{2K}\, d\lambda^u(a) + \frac{\epsilon}{2} \le \int_{\mathcal{G}^u} |f(a^{-1},t)|\, d\lambda^u(a) + \epsilon.$$

Thus,

$$\begin{aligned}
\|q_s(f)\|_I &< \max\left\{ \int_{\mathcal{G}^u} |f(a,t)|\, d\lambda^u(a),\ \int_{\mathcal{G}^u} |f(a^{-1},t)|\, d\lambda^u(a) \right\} + \epsilon \\
&\le \max\left\{ \sup_{u \in \mathcal{G}^{(0)}} \int_{\mathcal{G}^u} |f(a,t)|\, d\lambda^u(a),\ \sup_{u \in \mathcal{G}^{(0)}} \int_{\mathcal{G}^u} |f(a^{-1},t)|\, d\lambda^u(a) \right\} + \epsilon \\
&= \|q_t(f)\|_I + \epsilon
\end{aligned}$$

if $|s-t| < \delta$. Reversing the roles of s and t in the above argument tells us that

$$|s-t| < \delta \Rightarrow \left| \|q_s(f)\|_I - \|q_t(f)\|_I \right| < \epsilon. \qquad \square$$

Now we can finish the proof of Theorem 3.3. Set $S_t = \{\psi \in C([0,1]) : \psi(t) = 1\}$. For any $\psi \in S_t$ and any $f \in C_c(\mathcal{G} \times [0,1], \omega)$, we have

$$\|L(\psi \cdot f)\| = \|L'(q_t(\psi \cdot f))\| = \|L'(q_t(f))\|.$$

Consequently,

$$\begin{aligned}
\|L'(q_t(f))\| &= \inf_{\psi \in S} \|L(\psi \cdot f)\| \le \inf_{\psi} \|\psi \cdot f\|_I \\
&= \inf_{\psi} \max\left\{ \sup_{s \in [0,1]} \sup_{u \in \mathcal{G}^{(0)}} \int |\psi(s) f(a,s)|\, d\lambda^u(a), \right. \\
&\qquad\qquad \left. \sup_{s \in [0,1]} \sup_{u \in \mathcal{G}^{(0)}} \int |\psi(s) f(a^{-1},s)|\, d\lambda^u(a) \right\} \\
&= \inf_{\psi} \sup_{s \in [0,1]} \|q_s(\psi \cdot f)\|_I.
\end{aligned}$$

Let $\epsilon > 0$ be given. Choose a δ such that $|s-t| < \delta \Rightarrow \left| \|q_s(f)\|_I - \|q_t(f)\|_I \right| < \epsilon$; choose $\psi_\epsilon \in C([0,1])$ such that $\psi_\epsilon(t) = 1$ and $|s-t| \ge \delta \Rightarrow \psi_\epsilon(s) = 0$. Then, since $\psi_\epsilon \in S_t$, we have

$$\|q_s(\psi_\epsilon \cdot f)\|_I = \psi_\epsilon(s) \|q_s(f)\|_I < \psi_\epsilon(s)(\|q_t(f)\|_I + \epsilon) \le \|q_t(f)\|_I + \epsilon \tag{3}$$

if $|s-t| < \delta$; otherwise we have $\|q_s(\psi_\epsilon \cdot f)\|_I = 0$, and (3) still holds.

Since we can find such a ψ_ϵ for any $\epsilon > 0$, it follows that

$$\begin{aligned}\|L'(q_t(f))\| &\le \inf_{\psi\in S_t}\sup_{s\in[0,1]} \|q_s(\psi\cdot f)\|_I \le \inf_\epsilon \sup_s \|q_s(\psi_\epsilon\cdot f)\|_I \\ &\le \inf_\epsilon \|q_t(f)\|_I + \epsilon = \|q_t(f)\|_I.\end{aligned}$$

The fact that q_t is onto now tells us that L' is a bounded representation of $C_c(\mathcal{G}, \omega_t)$ as claimed. In other words, every representation L of $C_c(\mathcal{G}\times[0,1], \omega)$ that kills I_t also factors through q_t, so $\ker q_t \subseteq I_t$. This completes the proof that the fiber algebra $C^*(\mathcal{G}\times[0,1], \omega)/I_t$ of the $C([0,1])$-algebra $C^*(\mathcal{G}\times[0,1], \omega)$ is simply $C^*(\mathcal{G}, \omega_t)$. □

In order to apply Theorem 3.3 to a homotopy of cocycles on a k-graph, we first need to define such a homotopy. Unlike for groupoids, there is no obvious way to make $\Lambda\times[0,1]$ into a higher-rank graph, so our definition of a homotopy of k-graph cocycles will look rather different than Definition 3.1 above. However, Proposition 3.8 below shows that the two definitions are compatible.

Definition 3.6. Let Λ be a k-graph. A family $\{c_t\}_{t\in[0,1]}$ of 2-cocycles in $\underline{Z}^2(\Lambda, \mathbb{T})$ is a *homotopy of* (2-)*cocycles* on Λ if for each pair $(\lambda, \mu) \in \Lambda^{*2}$ the function $t \mapsto c_t(\lambda, \mu) \in \mathbb{T}$ is continuous.

Definition 3.7. Let $\{c_t\}_{t\in[0,1]}$ be a homotopy of cocycles on a k-graph Λ. Define $\omega \in Z^2(\mathcal{G}_\Lambda\times[0,1], \mathbb{T})$ by

$$\omega((a,t),(b,t)) = \sigma_{c_t}(a,b),$$

where σ_{c_t} is the cocycle on $\mathcal{G}_\Lambda$ associated to c_t as in Definition 2.7.

A moment's thought will reveal that ω satisfies the cocycle condition (2), since each σ_{c_t} is a cocycle. Thus, in order to see that ω is a homotopy of cocycles on $\mathcal{G}_\Lambda$, we merely need to check that $\omega : (\mathcal{G}_\Lambda\times[0,1])^{(2)} \to \mathbb{T}$ is continuous.

Proposition 3.8. *The cocycle ω described in Definition 3.7 is continuous, and hence is a homotopy of groupoid cocycles on $\mathcal{G}_\Lambda$.*

Proof. We will show that if $\{(a_i, b_i, t_i)\}_{i\in I} \subseteq \mathcal{G}_\Lambda^{(2)}\times[0,1]$ is a net which converges to (a,b,t), then

$$\omega((a_i,t_i),(b_i,t_i)) := \sigma_{c_{t_i}}(a_i,b_i) = \sigma_{c_{t_i}}(a,b) \tag{4}$$

for large enough i. Recall from Definition 2.7 that $\sigma_{c_{t_i}}(a,b)$ is a finite product of terms of the form $c_{t_i}(\mu,\nu)$ and their inverses, where the elements μ, ν depend only on the elements a, b and on the choice of partition $\mathcal{P}$ of $\mathcal{G}_\Lambda$ — but not on the 2-cocycle c_{t_i}. Thus, (4) and the continuity of the maps $t \mapsto c_t(\mu,\nu)$ will imply that $\omega((a_i,t_i),(b_i,t_i)) \to \sigma_{c_t}(a,b) = \omega((a,t),(b,t))$.

In what follows, we use the notation of Definition 2.7. If $(a_i, b_i, t_i) \to (a, b, t)$, then, for large enough i, we have $a_i \in Z(\mu_a, \nu_a)$, $b_i \in Z(\mu_b, \nu_b)$, and in addition $a_i b_i \in Z(\mu_{ab}, \nu_{ab})$. In other words, we can write

$$\begin{aligned} a &= (\mu_a \alpha y, d(\mu_a) - d(\nu_a), \nu_a \alpha y), \quad & a_i &= (\mu_a \alpha_i y_i, d(\mu_a) - d(\nu_a), \nu_a \alpha_i y_i), \\ b &= (\mu_b \beta y, d(\mu_b) - d(\nu_b), \nu_b \beta y), \quad & b_i &= (\mu_b \beta_i y_i, d(\mu_b) - d(\nu_b), \nu_b \beta_i y_i), \end{aligned}$$
$$ab = (\mu_{ab} \gamma y, d(\mu_{ab}) - d(\nu_{ab}), \nu_{ab} \gamma y),$$
$$a_i b_i = (\mu_{ab} \gamma_i y_i, d(\mu_{ab}) - d(\nu_{ab}), \nu_{ab} \gamma_i y_i)$$

for some $\alpha, \beta, \gamma, \alpha_i, \beta_i, \gamma_i \in \Lambda$ and $y, y_i \in \Lambda^\infty$.

Since $a_i \to a$ we must also have $\alpha_i y_i \to \alpha y$ in Λ^∞. Thus, for large enough i, $\alpha_i y_i \in Z(\alpha) := \{\alpha y : y \in \Lambda^\infty,\ y(0) = s(\alpha)\}$ (see Proposition 2.8 of [Kumjian and Pask 2000]). It follows that

$$a_i = (\mu_a \alpha y_i', d(\mu_a) - d(\nu_a), \nu_a \alpha y_i'), \quad b_i = (\mu_b \beta z_i', d(\mu_b) - d(\nu_b), \nu_b \beta z_i')$$
$$a_i b_i = (\mu_{ab} \gamma w_i', d(\mu_{ab}) - d(\nu_{ab}), \nu_{ab} \gamma w_i'),$$

where (since each pair (a_i, b_i) is in $\mathcal{G}_\Lambda^{(2)}$ by hypothesis)

$$\nu_a \alpha y_i' = \mu_b \beta z_i', \quad \mu_a \alpha y_i' = \mu_{ab} \gamma w_i', \quad \nu_b \beta z_i' = \nu_{ab} \gamma w_i'.$$

Now, $\nu_a \alpha = \mu_b \beta$ by [Kumjian et al. 2015, Lemma 6.3], and thus $y_i' = z_i'$. A similar argument gives $z_i' = w_i'$ as well, so $y_i' = z_i' = w_i'$. In other words, for large enough i,

$$\begin{aligned} \sigma_{c_{t_i}}(a_i, b_i) &= c_{t_i}(\mu_a, \alpha) c_{t_i}(\mu_b, \beta) c_{t_i}(\nu_{ab}, \gamma) \overline{c_{t_i}(\nu_a, \alpha) c_{t_i}(\nu_b, \beta) c_{t_i}(\mu_{ab}, \gamma)} \\ &= \sigma_{c_{t_i}}(a, b), \end{aligned}$$

as claimed. As observed in the first paragraph of the proof, it now follows that ω is a homotopy of cocycles on $\mathcal{G}_\Lambda$ as desired. □

Corollary 3.9. *Let $\{c_t\}$ be a homotopy of cocycles on a k-graph Λ, and define a cocycle ω on $\mathcal{G}_\Lambda \times [0, 1]$ as in Definition 3.7. Then $C^*(\mathcal{G}_\Lambda \times [0, 1], \omega)$ is a $C([0, 1])$-algebra with fiber algebra $C^*(\mathcal{G}_\Lambda, \sigma_{c_t}) \cong C^*(\Lambda, c_t)$ at $t \in [0, 1]$.*

Proof. Proposition 3.8 tells us that ω is a homotopy of cocycles on $\mathcal{G}_\Lambda$. Theorem 3.3 tells us that the fiber over $t \in [0, 1]$ of the $C([0, 1])$-algebra $C^*(\mathcal{G}_\Lambda \times [0, 1], \omega)$ is $C^*(\mathcal{G}_\Lambda, \sigma_{c_t})$. The final isomorphism is provided by Corollary 7.8 of [Kumjian et al. 2015]. □

4. The main theorem

Theorem 4.1. *Let Λ be a row-finite k-graph with no sources and let $\{c_t\}_{t \in [0,1]}$ be a homotopy of cocycles on Λ. Then*

$$K_*(C^*(\Lambda, c_0)) \cong K_*(C^*(\Lambda, c_1)).$$

Moreover, this isomorphism preserves the K-theory class of the vertex projection s_v *for each* $v \in \mathrm{Obj}(\Lambda)$.

We begin by proving a stronger version of Theorem 4.1 in the simpler case when the degree functor d satisfies $d(\lambda) = \delta b(\lambda) := b(s(\lambda)) - b(r(\lambda))$ for some function $b : \mathrm{Obj}(\Lambda) \to \mathbb{Z}^k$; this is Proposition 4.2 below. We then combine Proposition 4.2 with techniques from [Kumjian et al. 2013] to prove Theorem 4.1 in full generality.

The AF case. If (Λ, d) is a k-graph such that $d = \delta b$, then Lemma 8.4 of [Kumjian et al. 2015] tells us that $C^*(\Lambda, c)$ and $C^*(\Lambda)$ are both AF-algebras, with the same approximating subalgebras and multiplicities of partial inclusions. Consequently, $C^*(\Lambda, c) \cong C^*(\Lambda)$. In order to fix notation for what follows, we describe this isomorphism in some detail.

Lemma 3.1 of [Kumjian and Pask 2000] shows that if Λ is a row-finite, source-free k-graph, then $\{s_\lambda s_\mu^* : s(\lambda) = s(\mu)\}$ spans a dense $*$-subalgebra of $C^*(\Lambda)$. Moreover, when $d = \delta b$, Lemma 5.4 of [Kumjian and Pask 2000] tells us that $\{s_\lambda s_\mu^* : b(s(\lambda)) = b(s(\mu)) = n\}$ forms a collection of matrix units for the subalgebra

$$A_n = \overline{\mathrm{span}}\{s_\lambda s_\mu^* : b(s(\lambda)) = b(s(\mu)) = n\} \cong \bigoplus_{b(v)=n} \mathcal{K}\big(\ell^2(s^{-1}(v))\big).$$

Observe that we can think of A_n as a subalgebra of $C^*(\Lambda)$ or of $C^*(\Lambda, c)$. In fact, these subalgebras allow us to exhibit $C^*(\Lambda, c)$ and $C^*(\Lambda)$ as AF-algebras:

$$C^*(\Lambda, c) = \varinjlim(A_n, \phi^c_{m,n}) \quad \text{and} \quad C^*(\Lambda) = \varinjlim(A_n, \phi_{m,n}),$$

where the connecting maps $\phi_{m,n}, \phi^c_{m,n} : A_n \to A_m$ are given by

$$\phi^c_{m,n}(s_\lambda s_\mu^*) = \sum_{\substack{r(\alpha)=s(\lambda)\\ b(s(\alpha))=m}} c(\lambda, \alpha)\overline{c(\mu, \alpha)} s_{\lambda\alpha} s_{\mu\alpha}^*,$$

$$\phi_{m,n}(s_\lambda s_\mu^*) = \sum_{\substack{r(\alpha)=s(\lambda)\\ b(s(\alpha))=m}} s_{\lambda\alpha} s_{\mu\alpha}^*.$$

We can now describe explicitly the isomorphism $C^*(\Lambda, c) \cong C^*(\Lambda)$. As in Theorem 4.2 of [Kumjian et al. 2013], write $\mathbb{1}$ for $(1, \dots, 1) \in \mathbb{N}^k$, and define $\kappa : \Lambda \to \mathbb{T}$ by

$$\kappa(\lambda) = \begin{cases} 1 & d(\lambda) \not\geq \mathbb{1}, \\ \kappa(\mu)c(\mu, \alpha) & d(\alpha) = \mathbb{1} \text{ and } \lambda = \mu\alpha. \end{cases}$$

For $n \in \mathbb{Z}^k$, let $U_n = \sum_{b(s(\lambda))=n} \kappa(\lambda) s_\lambda s_\lambda^* \in U(M(A_n))$. A quick computation will show that, for any λ, μ with $s_\lambda s_\mu^* \in A_n$,

$$\mathrm{Ad}\, U_n(s_\lambda s_\mu^*) = \kappa(\lambda)\overline{\kappa(\mu)} s_\lambda s_\mu^*. \tag{5}$$

Moreover, the factorization property tells us that, for any $h \in \mathbb{Z}$,

$$\phi^c_{(h+1)\mathbb{1},h\mathbb{1}} \circ \operatorname{Ad} U_{h\mathbb{1}} = \operatorname{Ad} U_{(h+1)\mathbb{1}} \circ \phi_{(h+1)\mathbb{1},h\mathbb{1}}.$$

In other words, $\operatorname{Ad} U_*$ intertwines the connecting maps $\phi^c_{m,n}$, $\phi_{m,n}$, and hence implements the isomorphism $C^*(\Lambda) \to C^*(\Lambda, c)$.

We can now use this isomorphism to prove that a homotopy of cocycles on Λ gives rise to a trivial continuous field when $d = \delta b$.

Proposition 4.2. *Let (Λ, d) be a row-finite, source-free k-graph such that $d = \delta b$ for some function $b : \mathrm{Obj}(\Lambda) \to \mathbb{Z}^k$, let $\{c_t\}_{t\in[0,1]}$ be a homotopy of cocycles on Λ, and let ω be the cocycle on $\mathcal{G}_\Lambda \times [0, 1]$ associated to $\{c_t\}_{t\in[0,1]}$ as in Definition 3.7. We have an isomorphism of $C([0, 1])$-algebras*

$$C^*(\mathcal{G}_\Lambda \times [0, 1], \omega) \cong C^*(\mathcal{G}_\Lambda \times [0, 1]) \cong C([0, 1]) \otimes C^*(\Lambda).$$

Proof. Recall that

$$C^*(\mathcal{G}_\Lambda \times [0, 1])_t \cong C^*(\mathcal{G}_\Lambda, \sigma_{c_t}) \cong C^*(\Lambda, c_t) \cong C^*(\Lambda)$$

if $d = \delta b$. Thus, the $C([0, 1])$-algebras $C^*(\mathcal{G}_\Lambda \times [0, 1], \omega)$ and $C([0, 1]) \otimes C^*(\Lambda)$ have isomorphic fibers over each point $t \in [0, 1]$.

In order to prove the proposition, we need to show that these isomorphisms $C^*(\mathcal{G}_\Lambda, \sigma_{c_t}) \cong C^*(\Lambda)$ vary continuously in t, so that they patch together to give us an isomorphism of $C([0, 1])$-algebras $C^*(\mathcal{G}_\Lambda \times [0, 1], \omega) \cong C([0, 1]) \otimes C^*(\Lambda)$.

For each $t \in [0, 1]$, let $\pi^t : C^*(\Lambda, c_t) \to C^*(\mathcal{G}_\Lambda, \sigma_{c_t})$ denote the isomorphism described in Theorem 6.7 of [Kumjian et al. 2015]. Let $\pi : C^*(\Lambda) \to C^*(\mathcal{G}_\Lambda)$ denote the equivalent isomorphism for the case of a trivial cocycle c. For each $n \in \mathbb{Z}^k$, write U^t_n for the unitary $U^t_n : A_n \to A_n$ associated to the cocycle c_t as above. Setting

$$\Psi_t := \pi^t \circ \operatorname{Ad} U^t_* \circ \pi^{-1}$$

consequently gives an isomorphism of C^*-algebras $\Psi_t : C^*(\mathcal{G}_\Lambda) \to C^*(\mathcal{G}_\Lambda, \sigma_{c_t})$.

We claim that $\Psi := \{\Psi_t\}_{t\in[0,1]}$ defines an isomorphism of $C([0, 1])$-algebras

$$\Psi : C^*(\mathcal{G}_\Lambda \times [0, 1]) \to C^*(\mathcal{G}_\Lambda \times [0, 1], \omega).$$

In order to prove this assertion, we begin by writing down an explicit formula for Ψ_t on the characteristic functions $1_{Z(\mu,\nu)} \in C_c(\mathcal{G}_\Lambda)$ where $Z(\mu, \nu) \in \mathcal{P}$ and where $\mathcal{P}$ is the partition of $\mathcal{G}_\Lambda$ described in Lemma 6.6 of [Kumjian et al. 2015].

Recall that the value of $\sigma_{c_t}(a, b)$ depends only on the sets $Z(\mu, \nu) \in \mathcal{P}$ containing the points a, b, and ab in $\mathcal{G}_\Lambda$. Moreover, the proof of [Kumjian et al. 2015, Theorem 6.7] establishes that, if $1_{Z(\mu,\nu)}$ denotes the characteristic function on $Z(\mu,\nu)\subseteq\mathcal{G}_\Lambda$, and if we write $a\in Z(\mu,\nu)$ as $a=bd$ where $b\in Z(\mu,s(\mu))$, $d\in Z(s(\nu),\nu)$, then

$$\pi^t(s_\mu s_\nu^*)(a)=1_{Z(\mu,\nu)}(a)\sigma_{c_t}(b,d)\overline{\sigma_{c_t}(d^{-1},d)}=1_{Z(\mu,\nu)}(a)\overline{\sigma_{c_t}(bd,d^{-1})}.$$

In addition, we have $Z(\mu,s(\mu))\in\mathcal{P}$ for all $\mu\in\Lambda$ by Lemma 6.6 of [Kumjian et al. 2015]. If we also have $Z(\mu,\nu)\in\mathcal{P}$, then the elements α, β, γ in the formula for $\sigma_{c_t}(bd,d^{-1})$ given in Definition 2.7 are all units, so, for any t, we have $\sigma_{c_t}(bd,d^{-1})=1$ by our hypothesis that any cocycle c satisfies the equality $c(\lambda,s(\lambda))=c(r(\lambda),\lambda)=1$. Thus,

$$Z(\mu,\nu)\in\mathcal{P}\Rightarrow\pi^t(s_\mu s_\nu^*)=1_{Z(\mu,\nu)}\Rightarrow\Psi_t(1_{Z(\mu,\nu)})=\kappa_t(\mu)\overline{\kappa_t(\nu)}1_{Z(\mu,\nu)}.$$

Now observe that each $f\in C_c(\mathcal{G}_\Lambda\times[0,1])$ can be written as a finite sum $f(a,t)=\sum_{i\in N}f_i(a,t)$, where, for all i, we have $f_i\in C(Z(\mu_i,\nu_i)\times[0,1])$ and $Z(\mu_i,\nu_i)\in\mathcal{P}$. Consequently, on $C_c(\mathcal{G}_\Lambda\times[0,1])$, our map Ψ becomes

$$\Psi\Bigg(\sum_{i\in N}f_i\Bigg)(a,t)=\sum_{i\in N}\Psi_t(f_i(\,\cdot\,,t))(a)=\sum_{i\in N}\kappa_t(\mu_i)\overline{\kappa_t(\nu_i)}f_i(a,t);\tag{6}$$

the fact that all the sums are finite implies that Ψ takes $C_c(\mathcal{G}_\Lambda\times[0,1])$ onto $C_c(\mathcal{G}_\Lambda\times[0,1])$.

Since Ψ is evidently $C([0,1])$-linear and is a $*$-isomorphism in each fiber, Proposition C.10 of [Williams 2007] tells us that Ψ is norm-preserving. Moreover, Ψ is a $*$-homomorphism since the operations in $C_c(\mathcal{G}_\Lambda\times[0,1])$ preserve the fiber over $t\in[0,1]$, and each Ψ_t is a $*$-homomorphism.

In other words, Ψ extends to an isomorphism of $C([0,1])$-algebras

$$\Psi:C^*(\mathcal{G}_\Lambda\times[0,1])\cong C^*(\mathcal{G}_\Lambda\times[0,1],\omega).$$

A straightforward check will establish that the identity map on $C_c(\mathcal{G}_\Lambda\times[0,1])$ induces an isomorphism $\mathrm{id}:C^*(\mathcal{G}_\Lambda\times[0,1])\to C([0,1],C^*(\mathcal{G}_\Lambda))$ of $C([0,1])$-algebras; the isomorphism $C^*(\mathcal{G}_\Lambda)\cong C^*(\Lambda)$ of [Kumjian and Pask 2000, Corollary 3.5(i)] now finishes the proof. □

Remark 4.3. Note that Ψ induces an isomorphism $\Phi:C([0,1])\otimes C^*(\Lambda)\to C^*(\mathcal{G}_\Lambda\times[0,1],\omega)$ as follows. If $Z(\mu,\nu)\in\mathcal{P}$ and $f\in C([0,1])$, then

$$\Phi(f\otimes s_\mu s_\nu^*)(x,t)=f(t)1_{Z(\mu,\nu)}(x)\kappa_t(\mu)\overline{\kappa_t(\nu)}.\tag{7}$$

Remark 4.4. Since evaluation at $t\in[0,1]$ induces a homotopy equivalence between $C([0,1],C^*(\Lambda))$ and $C^*(\Lambda)$, the isomorphism established in the previous proposition implies that evaluation at t also induces a homotopy equivalence between $C^*(\mathcal{G}_\Lambda\times[0,1],\omega)$ and its fiber algebra $C^*(\mathcal{G}_\Lambda,\sigma_{c_t})$ when $d=\delta b$.

To leverage Proposition 4.2 into the proof of Theorem 4.1, we will use the skew-product k-graphs $\Lambda \times_d \mathbb{Z}^k$.

Definition 4.5 [Kumjian and Pask 2000, Definition 5.1]. Given a k-graph (Λ, d), the *skew-product k-graph* $\Lambda \times_d \mathbb{Z}^k$ is the set $\Lambda \times \mathbb{Z}^k$, with the structure maps

$$r(\lambda, n) = (r(\lambda), n), \quad s(\lambda, n) = (s(\lambda), n + d(\lambda)), \quad d(\lambda, n) = d(\lambda),$$

and with multiplication given by $(\lambda, n)(\mu, n + d(\lambda)) = (\lambda\mu, n)$ for $(\lambda, \mu) \in \Lambda^{*2}$.

Observe that the function $b : (\Lambda \times_d \mathbb{Z}^k)^{(0)} = \Lambda^{(0)} \times \mathbb{Z}^k \to \mathbb{Z}^k$ given by $b(v, n) = n$ satisfies $\delta b = d$ on $\Lambda \times_d \mathbb{Z}^k$. Moreover, if Λ is row-finite and source-free, then so is $\Lambda \times_d \mathbb{Z}^k$.

We can now complete the proof of Theorem 4.1.

Proof of Theorem 4.1. Let $\phi : \Lambda \times_d \mathbb{Z}^k \to \Lambda$ be the projection onto the first coordinate: $\phi(\lambda, n) = \lambda$. A cocycle c on Λ induces a cocycle $c \circ \phi$ on the skew-product k-graph $\Lambda \times_d \mathbb{Z}^k$:

$$c \circ \phi\big((\lambda, n), (\mu, n + d(\lambda))\big) := c(\lambda, \mu)$$

whenever $(\lambda, \mu) \in \Lambda^{*2}$. Note that if $\{c_t\}_{t \in [0,1]}$ is a homotopy of cocycles on Λ then $\{c_t \circ \phi\}_t$ is also a homotopy of cocycles on $\Lambda \times_d \mathbb{Z}^k$.

If ω is the homotopy of cocycles on $\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k}$ associated to the homotopy $\{c_t\}_{t \in [0,1]}$ of cocycles on Λ, then Proposition 4.2 tells us that

$$C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k} \times [0, 1], \omega) \cong C([0, 1]) \otimes C^*(\Lambda \times_d \mathbb{Z}^k).$$

Now we define an action of $\mathbb{Z}^k$ on $C([0, 1]) \otimes C^*(\Lambda \times_d \mathbb{Z}^k)$ by setting

$$f \otimes s_{\lambda,n} \cdot m := f \otimes s_{\lambda,n+m}. \tag{8}$$

To see that this formula gives a well-defined action of $\mathbb{Z}^k$ on $C([0, 1]) \otimes C^*(\Lambda \times_d \mathbb{Z}^k)$, one checks first that, for each $m \in \mathbb{Z}^k$, the set $\{s_{\lambda,m+n} : \lambda \in \Lambda,\ n \in \mathbb{Z}^k\}$ is a collection of partial isometries satisfying the defining axioms (CK1)–(CK4) for $C^*(\Lambda \times_d \mathbb{Z}^k)$. Consequently, the universal property of $C^*(\Lambda \times_d \mathbb{Z}^k)$ implies that, for each fixed $m \in \mathbb{Z}^k$, the map $s_{\lambda,n} \mapsto s_{\lambda,n+m}$ determines a $*$-homomorphism

$$\alpha_m : C^*(\Lambda \times_d \mathbb{Z}^k) \to C^*(\Lambda \times_d \mathbb{Z}^k).$$

Each α_m is invertible with inverse α_{-m}; it follows that $m \mapsto \alpha_m$ defines a group action of $\mathbb{Z}^k$ on $C^*(\Lambda \times_d \mathbb{Z}^k)$. Thus, (8) describes a well-defined action $\mathrm{id} \otimes \alpha$ of $\mathbb{Z}^k$ on $C([0, 1]) \otimes C^*(\Lambda \times_d \mathbb{Z}^k)$, given by $m \mapsto \mathrm{id} \otimes \alpha_m$. The fact that the degree map on $\Lambda \times_d \mathbb{Z}^k$ is a coboundary now allows us to combine the action $\mathrm{id} \otimes \alpha$ with the isomorphism $\Phi : C([0, 1]) \otimes C^*(\Lambda \times_d \mathbb{Z}^k) \to C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k} \times [0, 1], \omega)$ of Remark 4.3 to obtain an action β of $\mathbb{Z}^k$ on $C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k} \times [0, 1], \omega)$:

$$\beta_n(\Phi(f \otimes s_{\mu,m} s^*_{\nu,m+d(\mu)-d(\nu)})) := \Phi(\mathrm{id} \otimes \alpha_n(f \otimes s_{\mu,m} s^*_{\nu,m+d(\mu)-d(\nu)})).$$

Moreover, since both $\mathrm{id} \otimes \alpha$ and Φ (and hence β) fix $C([0,1])$ by construction, Lemma 5.3 of [Kumjian et al. 2013] tells us that the crossed product

$$C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k} \times [0,1], \omega) \rtimes_\beta \mathbb{Z}^k \cong (C([0,1]) \otimes C^*(\Lambda \times_d \mathbb{Z}^k)) \rtimes_{\mathrm{id}\otimes\alpha} \mathbb{Z}^k$$

is a $C([0,1])$-algebra with fiber $C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k}, \sigma_{c_t \circ \phi}) \rtimes_{\beta_t} \mathbb{Z}^k$, where

$$\begin{aligned}(\beta_t)_n(\Phi_t(s_{\mu,m} s^*_{\nu,m+d(\mu)-d(\nu)})) &= \Phi_t(\alpha_n(s_{(\mu,m)} s^*_{(\nu,m+d(\mu)-d(\nu))}))\\ &= \kappa_t(\mu)\overline{\kappa_t(\nu)} 1_{Z((\mu,m+n),(\nu,m+n+d(\mu)-d(\nu)))}\end{aligned}$$

whenever $Z\big((\mu, m+n), (\nu, m+n+d(\mu)-d(\nu))\big)$ is in the partition $\mathcal{P}$ of $\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k}$ that we used in the proof of Proposition 4.2.

Recall that we have a homotopy equivalence $q_t : C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k} \times [0,1], \omega) \to C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k}, \sigma_{c_t})$. A computation will show that q_t is equivariant with respect to the actions β, β_t of $\mathbb{Z}^k$; thus, Theorem 5.1 of [Kumjian et al. 2013] tells us that

$$K_*(C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k} \times [0,1], \omega) \rtimes_\beta \mathbb{Z}^k) \cong K_*(C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k}, \sigma_{c_t \circ \phi}) \rtimes_{\beta_t} \mathbb{Z}^k). \tag{9}$$

Thanks to Lemma 5.2 of [Kumjian et al. 2013], we know that

$$C^*(\Lambda \times_d \mathbb{Z}^k, c_t \circ \phi) \rtimes_{lt} \mathbb{Z}^k \sim_{ME} C^*(\Lambda, c_t),$$

where $lt_m(s_{\lambda,n}) = s_{\lambda,n+m}$. To make use of this result, we need to show that β_t induces the action lt on $C^*(\Lambda \times_d \mathbb{Z}^k, c_t \circ \phi)$.

Recall from the proof of Proposition 4.2 that $\pi^t(s_{\lambda,m}) = 1_{Z((\lambda,m),(s(\lambda),m+s(\lambda)))}$, since $Z\big((\lambda,m),(s(\lambda), m+s(\lambda))\big) \in \mathcal{P}$ always. Observe that

$$C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k}, \sigma_{c_t \circ \phi}) \rtimes_{\beta_t} \mathbb{Z}^k \cong C^*(\Lambda \times_d \mathbb{Z}^k, c_t \circ \phi) \rtimes_{\gamma_t} \mathbb{Z}^k, \tag{10}$$

where

$$\begin{aligned}(\gamma_t)_n(s_{\lambda,m}) &:= (\pi^t)^{-1}\big((\beta_t)_n(\pi^t(s_{\lambda,m}))\big) = (\pi^t)^{-1}(\beta_t)_n(1_{Z((\lambda,m),(s(\lambda),m))})\\ &= (\pi^t)^{-1}(\beta_t)_n\big(\Phi_t(\overline{\kappa_t(\lambda)} s_{(\lambda,m)})\big) = (\pi^t)^{-1}\big(\Phi_t(\alpha_n(\overline{\kappa_t(\lambda)} s_{\lambda,m}))\big)\\ &= (\pi^t)^{-1}\big(\Phi_t(\overline{\kappa_t(\lambda)} s_{\lambda,m+n})\big) = (\pi^t)^{-1}(1_{Z((\lambda,m+n),(s(\lambda),m+n))})\\ &= s_{\lambda,m+n}.\end{aligned}$$

It follows that the action (γ_t) induced by β_t agrees with lt, as desired. Now, the Morita equivalence of Lemma 5.2 of [Kumjian et al. 2013] and (10) tell us that

$$C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k}, \sigma_{c_t \circ \phi}) \rtimes_{\beta_t} \mathbb{Z}^k \sim_{ME} C^*(\Lambda, c_t). \tag{11}$$

Combining (9) and (11) now yields

$$K_*(C^*(\Lambda, c_t)) \cong K_*(C^*(\mathcal{G}_{\Lambda \times_d \mathbb{Z}^k} \times [0,1], \omega) \rtimes_\beta \mathbb{Z}^k)$$

for any $t \in [0,1]$. It follows that, if $\{c_t\}_{t\in[0,1]}$ is a homotopy of cocycles on a row-finite k-graph Λ with no sources, then, for any $s, t \in [0,1]$,

$$K_*(C^*(\Lambda, c_t)) \cong K_*(C^*(\Lambda, c_s)).$$

It remains to show that this isomorphism preserves the K-theory class of each vertex projection s_v. Essentially, this follows because the cocycles c_t, and thus the functions κ_t, are all trivial on any $v \in \mathrm{Obj}(\Lambda)$.

To be precise, let $v \in \mathrm{Obj}(\Lambda)$ and define $f_v \in C_c(\mathbb{Z}^k, C_c(\mathcal{G}_{\Lambda\times_d \mathbb{Z}^k} \times [0,1])) \subseteq C^*(\mathcal{G}_{\Lambda\times_d \mathbb{Z}^k} \times [0,1], \omega) \rtimes_\beta \mathbb{Z}^k$ by

$$f_v(n)(a,t) = \begin{cases} 1 & a \in Z_{(v,0),(v,0)} \text{ and } n = 0,\\ 0 & \text{else.}\end{cases}$$

Then the projection $q_t \rtimes \mathrm{id}(f_v)$ of f_v onto the fiber algebra $C^*(\mathcal{G}_{\Lambda\times_d \mathbb{Z}^k}, \omega_t) \rtimes_{\beta_t} \mathbb{Z}^k$ is independent of the choice of $t \in [0,1]$:

$$q_t \rtimes \mathrm{id}(f_v)(n)(a) = \begin{cases} 1 & a \in Z_{(v,0),(v,0)} \text{ and } n = 0,\\ 0 & \text{else,}\end{cases}$$

for any $t \in [0,1]$. Moreover, the isomorphism $\Phi_t : C^*(\Lambda \times_d \mathbb{Z}^k, c_t \circ \phi) \to C^*(\mathcal{G}_{\Lambda\times_d \mathbb{Z}^k}, \sigma_{c_t\circ\phi})$ of Remark 4.3 satisfies

$$\Phi_t \rtimes \mathrm{id}(j(s_{(v,0)})) = q_t \rtimes \mathrm{id}(f_v), \tag{12}$$

where $j : C^*(\Lambda \times_d \mathbb{Z}^k, c_t \circ \phi) \to C^*(\Lambda \times_d \mathbb{Z}^k, c_t \circ \phi) \rtimes_{lt} \mathbb{Z}^k$ is the canonical embedding of $C^*(\Lambda \times_d \mathbb{Z}^k, c_t \circ \phi)$ into the crossed product.

The fact that the Morita equivalence $C^*(\Lambda, c_t) \sim_{ME} C^*(\Lambda \times_d \mathbb{Z}^k, c_t \circ \phi) \rtimes_{lt} \mathbb{Z}^k$ takes $s_v \in C^*(\Lambda, c_t)$ to $j(s_{(v,0)})$ (see Lemma 5.2 in [Kumjian et al. 2013]) thus implies that our K-theoretic isomorphism $K_*(C^*(\mathcal{G}_{\Lambda\times_d \mathbb{Z}^k} \times [0,1], \omega) \rtimes_\beta \mathbb{Z}^k) \to K_*(C^*(\Lambda, c_t))$, which is given by the composition of the Morita equivalence (11) with the $*$-homomorphism

$$\begin{aligned} q_t \rtimes \mathrm{id} : C^*(\mathcal{G}_{\Lambda\times_d \mathbb{Z}^k} \times [0,1], \omega) \rtimes_\beta \mathbb{Z}^k &\to C^*(\mathcal{G}_{\Lambda\times_d \mathbb{Z}^k}, \omega_t) \rtimes_{\beta_t} \mathbb{Z}^k \\ &\cong C^*(\Lambda \times_d \mathbb{Z}^k, c_t \circ \phi) \rtimes_{\gamma_t} \mathbb{Z}^k, \end{aligned}$$

takes $[f_v]$ to $[s_v]$ for any $v \in \mathrm{Obj}(\Lambda)$ and any $t \in [0,1]$. Consequently, the isomorphism $K_*(C^*(\Lambda, c_t)) \cong K_*(C^*(\Lambda, c_s))$ preserves the class of s_v, as claimed. □

Remark 4.6. It's tempting to think that, since $C^*(\Lambda \times_d \mathbb{Z}^k, c \circ \phi) \cong C^*(\Lambda \times_d \mathbb{Z}^k)$ and $C^*(\Lambda, c) \sim_{ME} C^*(\Lambda \times_d \mathbb{Z}^k, c \circ \phi) \rtimes_{lt} \mathbb{Z}^k$ for any cocycle c on Λ, any two twisted k-graph C^*-algebras should be Morita equivalent. This statement is false, however (the rotation algebras provide a counterexample). The flaw lies in the fact that the isomorphism $\mathrm{Ad}\, U_* : C^*(\Lambda \times_d \mathbb{Z}^k, c \circ \phi) \to C^*(\Lambda \times_d \mathbb{Z}^k)$ is not equivariant with respect to the left-translation action of $\mathbb{Z}^k$, so the isomorphism

$$C^*(\Lambda \times_d \mathbb{Z}^k, c \circ \phi) \cong C^*(\Lambda \times_d \mathbb{Z}^k)$$

does not pass to an isomorphism $C^*(\Lambda \times_d \mathbb{Z}^k, c \circ \phi) \rtimes_{lt} \mathbb{Z}^k \to C^*(\Lambda \times_d \mathbb{Z}^k) \rtimes_{lt} \mathbb{Z}^k$. In other words, a K-theoretic equivalence of twisted k-graph C^*-algebras is the best result we can hope for in general.

5. Future work

The standing hypotheses of this paper, that our k-graphs be row-finite and source-free, are slightly more restrictive than the current standard for k-graphs. Thus, we would like to extend Theorem 4.1 to apply to all finitely aligned k-graphs. Finitely aligned k-graphs were introduced in [Raeburn and Sims 2005; Raeburn et al. 2004], and it seems that they constitute the largest class of k-graphs to which one can profitably associate a C^*-algebra. However, the Kumjian–Pask construction of a groupoid $\mathcal{G}_\Lambda$ associated to a k-graph Λ, which we described in Section 2 and which we use throughout the proof of Theorem 4.1, only works when Λ is row-finite and source-free. Farthing, Muhly, and Yeend [Farthing et al. 2005] provide an alternate construction of a groupoid $\mathcal{G}$ which can be associated to an arbitrary finitely aligned k-graph, and we hope that this approach will allow us to apply groupoid results such as Theorem 3.3 to study the effect on K-theory of homotopies of cocycles for finitely aligned k-graphs.

Acknowledgements

This research constitutes part of my PhD thesis, and I would like to thank my advisor, Erik van Erp, for his encouragement and assistance. I would also especially like to thank Prof. Jody Trout for his help and support throughout my PhD work.

References

[Brown and an Huef 2014] J. H. Brown and A. an Huef, "Decomposing the C^*-algebras of groupoid extensions", *Proc. Amer. Math. Soc.* **142**:4 (2014), 1261–1274. MR 3162248 Zbl 1302.46053

[Clark and an Huef 2012] L. O. Clark and A. an Huef, "The representation theory of C^*-algebras associated to groupoids", *Math. Proc. Cambridge Philos. Soc.* **153**:1 (2012), 167–191. MR 3043775 Zbl 1270.46047

[Echterhoff et al. 2010] S. Echterhoff, W. Lück, N. C. Phillips, and S. Walters, "The structure of crossed products of irrational rotation algebras by finite subgroups of $\mathrm{SL}_2(\mathbb{Z})$", *J. Reine Angew. Math.* **639** (2010), 173–221. MR 2011c:46127 Zbl 1202.46081

[Farthing et al. 2005] C. Farthing, P. S. Muhly, and T. Yeend, "Higher-rank graph C^*-algebras: An inverse semigroup and groupoid approach", *Semigroup Forum* **71**:2 (2005), 159–187. MR 2006h:46052 Zbl 1099.46036

[Gillaspy 2015] E. Gillaspy, "K-theory and homotopies of 2-cocycles on transformation groups", *J. Operator Theory* **73**:2 (2015), 465–490. MR 3346133 Zbl 06465653

[Goehle 2009] G. Goehle, *Groupoid crossed products*, Ph.D. thesis, Dartmouth College, 2009, available at http://search.proquest.com/docview/230889835.

[an Huef et al. 2011] A. an Huef, A. Kumjian, and A. Sims, "A Dixmier–Douady theorem for Fell algebras", *J. Funct. Anal.* **260**:5 (2011), 1543–1581. MR 2012i:46066 Zbl 1251.46028

[Kumjian 1986] A. Kumjian, "On C^*-diagonals", *Canad. J. Math.* **38**:4 (1986), 969–1008. MR 88a:46060 Zbl 0627.46071

[Kumjian and Pask 2000] A. Kumjian and D. Pask, "Higher rank graph C^*-algebras", *New York J. Math.* **6** (2000), 1–20. MR 2001b:46102 Zbl 0946.46044

[Kumjian et al. 2013] A. Kumjian, D. Pask, and A. Sims, "On the K-theory of twisted higher-rank-graph C^*-algebras", *J. Math. Anal. Appl.* **401**:1 (2013), 104–113. MR 3011251 Zbl 1279.46053

[Kumjian et al. 2015] A. Kumjian, D. Pask, and A. Sims, "On twisted higher-rank graph C^*-algebras", *Trans. Amer. Math. Soc.* **367**:7 (2015), 5177–5216. MR 3335414 Zbl 06429159

[Muhly and Williams 1992] P. S. Muhly and D. P. Williams, "Continuous trace groupoid C^*-algebras, II", *Math. Scand.* **70**:1 (1992), 127–145. MR 93i:46117 Zbl 0783.46028

[Muhly et al. 1996] P. S. Muhly, J. N. Renault, and D. P. Williams, "Continuous-trace groupoid C^*-algebras, III", *Trans. Amer. Math. Soc.* **348**:9 (1996), 3621–3641. MR 96m:46125 Zbl 0859.46039

[Pimsner and Voiculescu 1980] M. Pimsner and D. Voiculescu, "Exact sequences for K-groups and Ext-groups of certain cross-product C^*-algebras", *J. Operator Theory* **4**:1 (1980), 93–118. MR 82c:46074 Zbl 0474.46059

[Raeburn and Sims 2005] I. Raeburn and A. Sims, "Product systems of graphs and the Toeplitz algebras of higher-rank graphs", *J. Operator Theory* **53**:2 (2005), 399–429. MR 2006d:46073 Zbl 1093.46032

[Raeburn et al. 2004] I. Raeburn, A. Sims, and T. Yeend, "The C^*-algebras of finitely aligned higher-rank graphs", *J. Funct. Anal.* **213**:1 (2004), 206–240. MR 2005e:46103 Zbl 1063.46041

[Renault 1980] J. Renault, *A groupoid approach to C^*-algebras*, Lecture Notes in Mathematics **793**, Springer, Berlin, 1980. MR 82h:46075 Zbl 0433.46049

[Robertson and Steger 1999] G. Robertson and T. Steger, "Affine buildings, tiling systems and higher rank Cuntz–Krieger algebras", *J. Reine Angew. Math.* **513** (1999), 115–144. MR 2000j:46109 Zbl 1064.46504

[Tu et al. 2004] J.-L. Tu, P. Xu, and C. Laurent-Gengoux, "Twisted K-theory of differentiable stacks", *Ann. Sci. École Norm. Sup.* (4) **37**:6 (2004), 841–910. MR 2005k:58037 Zbl 1069.19006

[Williams 2007] D. P. Williams, *Crossed products of C^*-algebras*, Mathematical Surveys and Monographs **134**, Amer. Math. Soc., Providence, RI, 2007. MR 2007m:46003 Zbl 1119.46002

Received April 2, 2014. Revised March 25, 2015.

ELIZABETH GILLASPY
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF COLORADO - BOULDER
CAMPUS BOX 395
BOULDER, CO 80309-0395
UNITED STATES
elizabeth.gillaspy@colorado.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.278.427

FUSION PRODUCTS AND TOROIDAL ALGEBRAS

DENIZ KUS AND PETER LITTELMANN

We study the category of finite-dimensional bigraded representations of toroidal current algebras associated to finite-dimensional complex simple Lie algebras. Using the theory of graded representations for current algebras, we construct in different ways objects in that category and prove them to be isomorphic. As a consequence we obtain generators and relations for certain types of fusion products, including the N-fold fusion product of $V(\lambda)$. This result shows that the fusion product of these types is independent of the chosen parameters, proving a special case of a conjecture by Feigin and Loktev. Moreover, we prove a conjecture by Chari, Fourier and Sagaki on truncated Weyl modules for certain classes of dominant integral weights and show that they are realizable as fusion products. In the last section we consider the case $\mathfrak{g} = \mathfrak{sl}_2$ and compute a PBW type basis for truncated Weyl modules of the associated current algebra.

1. Introduction

Let $\mathfrak{g}$ be a finite-dimensional complex simple Lie algebra with highest root θ. The current algebra $\mathfrak{g}[t]$ associated to $\mathfrak{g}$ is the algebra of polynomial maps $\mathbb{C} \to \mathfrak{g}$; equivalently, it is the complex vector space $\mathfrak{g} \otimes \mathbb{C}[t]$ with Lie bracket the $\mathbb{C}[t]$-bilinear extension of the Lie bracket on $\mathfrak{g}$. The toroidal current algebra $\mathfrak{g}[t, u]$ associated to $\mathfrak{g}$ is the algebra of polynomial maps $\mathbb{C}^2 \to \mathfrak{g}$ and can be identified with the complex vector space $\mathfrak{g} \otimes \mathbb{C}[t, u]$ with similar Lie bracket. The Lie algebra $\mathfrak{g}[t]$ is graded by the nonnegative integers, where the r-th graded component is $\mathfrak{g} \otimes t^r$ and $\mathfrak{g}[t, u]$ is bigraded by pairs of nonnegative integers, where the (r, s)-th graded component is $\mathfrak{g} \otimes t^r u^s$. We are interested in the category of finite-dimensional graded representations of $\mathfrak{g}[t]$ and finite-dimensional bigraded representations of $\mathfrak{g}[t, u]$. The former category contains a large number of interesting objects, for example local Weyl modules (see for instance [Chari et al. 2014b; Chari and Pressley 2001; Fourier and Littelmann 2007; Fourier et al. 2012]), $\mathfrak{g}$-stable Demazure modules (see [Chari et al. 2014c; Fourier and Littelmann 2006; 2007]) and fusion products.

This work has been partially supported by the DFG Priority Program SPP 1388 (Representation Theory) and SFB/TR 12 (Symmetries and Universality in Mesoscopic Systems).
MSC2010: primary 17B67; secondary 17B10.
Keywords: fusion products, toroidal algebras, Demazure modules.

The latter class of representations was introduced in a paper by Feigin and Loktev [1999]: given finite-dimensional cyclic $\mathfrak{g}[t]$-modules $V_1, \dots, V_N$ with cyclic vectors $v_1, \dots, v_N$ and a tuple of pairwise distinct complex numbers $z = (z_1, \dots, z_N)$ one can define a filtration on the tensor product $V_1^{z_1} \otimes \cdots \otimes V_N^{z_N}$ and build the associated graded space with respect to this filtration. This space is called the fusion product and is denoted by $V_1^{z_1} * \cdots * V_N^{z_N}$, where V^z is a nongraded $\mathfrak{g}[t]$-module (see Section 3 for more details). The Feigin–Loktev conjecture states that under suitable conditions on V_s and v_s the fusion product is independent of the chosen fusion parameters z. This conjecture has been proved for several classes of representations. For example it has been proved in [Chari and Loktev 2006; Chari and Pressley 2001; Fourier and Littelmann 2007; Naoi 2012] that the fusion product of local Weyl modules is again a local Weyl module and hence independent of the chosen parameters. Other examples are fusion products of Kirillov–Reshetikhin modules (see [Ardonne and Kedem 2007; Kedem 2011]) and fusion products of $\mathfrak{g}$-stable Demazure modules (see [Chari et al. 2014c; Fourier and Littelmann 2007; Kus and Venkatesh 2014]).

Another interesting class of $\mathfrak{g}[t]$-modules comprises those which are obtained as fusion products of finite-dimensional simple $\mathfrak{g}$-modules, where a $\mathfrak{g}$-module V is made into a $\mathfrak{g}[t]$-module by requiring $(\mathfrak{g} \otimes t\mathbb{C}[t])\, V = 0$. Hence for any tuple $(\lambda_1, \dots, \lambda_N)$ of dominant integral weights the fusion product $V^{z_1}(\lambda_1) * \cdots * V^{z_N}(\lambda_N)$ can be defined and studied. For these types of representations the Feigin–Loktev conjecture has been proved in the case of $\mathfrak{sl}_2$ and in some other special cases (see for instance [Chari and Venkatesh 2015; Feigin and Feigin 2002; Feigin et al. 2004; Ravinder 2014]). Moreover, in the case of $\mathfrak{sl}_2$ a presentation for the fusion product $V(k_1) * \cdots * V(k_N)$ has been established in terms of generators and relations of the enveloping algebra (see [Chari and Venkatesh 2015; Feigin and Feigin 2002]) and a PBW type basis has been computed [Chari and Venkatesh 2015]. An easy calculation shows that the aforementioned presentation can be greatly simplified if the highest weights are equal. In particular, $V(k) * \cdots * V(k)$ is a cyclic $\boldsymbol{U}(\mathfrak{sl}_2[t])$-module generated by a vector v subject to the same relations as the highest weight vector of the local Weyl module $W_{\mathrm{loc}}(kN)$ with the only additional relation $(\mathfrak{sl}_2 \otimes t^N)v = 0$.

This paper is motivated by the idea of generalizing the above observation for arbitrary $\mathfrak{g}$: Is the fusion product $V^{z_1}(\lambda) * \cdots * V^{z_N}(\lambda)$ independent of the fusion parameters for arbitrary $\mathfrak{g}$? Is there a simple presentation considered as a $\boldsymbol{U}(\mathfrak{g}[t])$-module? Is the truncated Weyl module $W(N\lambda, N)$ realizable as a fusion product? For the definition of truncated Weyl modules see Section 4A. In this paper we give a positive answer to these questions. Our approach is based on the theory of finite-dimensional bigraded modules for the toroidal current algebra $\mathfrak{g}[t, u]$. In particular we construct an associated graded version of a $\mathfrak{g}$-stable Demazure module $\mathrm{gr}_{t^N} T(\ell, N)$ and a bigraded version of a fusion product $\overline{D^u(\ell, \ell\lambda)} * \cdots * \overline{D^u(\ell, \ell\lambda)} *$

$\overline{\mathrm{D}^u(\ell, \ell\lambda+\lambda^0)}$ (for the precise definitions see Sections 3C and 3D) such that the zeroth graded space (with respect to the u-grading) of the second construction is isomorphic to the fusion product of finite-dimensional simple $\mathfrak{g}$-modules. Our first result is the following; for the precise definition of the ingredients see Section 3. *We remark that if* $\lambda^0 \neq 0$, *then the Lie algebra* $\mathfrak{g}$ *is assumed to be classical or* $\mathtt{G}_2$; *for* $\lambda^0 = 0$ *there is no restriction on* $\mathfrak{g}$.

Theorem. *Let* $\ell \in \mathbb{N}$, λ *be a dominant integral coweight and* λ^0 *be a dominant integral weight such that* $\lambda^0(\theta^\vee) \leq \ell$. *We have an isomorphism of* $\boldsymbol{U}(\mathfrak{g}[t, u])$-*modules*

$$\operatorname{gr}_{t^N} \mathrm{T}(\ell, N) \cong \overline{\mathrm{D}^u(\ell, \ell\lambda)} * \cdots * \overline{\mathrm{D}^u(\ell, \ell\lambda)} * \overline{\mathrm{D}^u(\ell, \ell\lambda+\lambda^0)}.$$

Our second result gives a connection to truncated Weyl modules, where the first part is a direct consequence of the previous theorem and the second part proves a special case of a conjecture by Chari, Fourier and Sagaki. Again for the precise definition of the ingredients see Section 4A.

Theorem. *Let* $\ell \in \mathbb{N}$, λ *be a dominant integral coweight and* λ^0 *be a dominant integral weight such that* $\lambda^0(\theta^\vee) \leq \ell$.

(1) *The fusion product* $\mathrm{V}(\ell\lambda)^{*(N-1)} * \mathrm{V}(\ell\lambda+\lambda^0)$ *is independent of the fusion parameters.*

(2) *If* $\lambda^0(\theta^\vee) \leq 1$ *and* $|N\lambda+\lambda^0| \geq N$, *then*

$$\mathrm{W}(N\lambda+\lambda^0, N) \cong \mathrm{V}(\lambda)^{*(N-1)} * \mathrm{V}(\lambda+\lambda^0).$$

As a special case of the previous theorem we can choose $\ell = 1$ and $\lambda^0 = 0$. This yields that the N-fold fusion product of $V(\lambda)$ is independent of the fusion parameters for any dominant integral coweight λ. The second part of the theorem shows that the N-fold fusion product of $V(\lambda)$ has a remarkably simple presentation.

In Sections 4C and 4D we deal with the case of $\mathfrak{sl}_2$ and prove that the truncated Weyl module is realizable as a fusion product. Moreover, we compute a PBW type basis for truncated Weyl modules which differs from the basis described in [Chari and Venkatesh 2015, Section 6]. For the precise definition of the set $S(k^{N-j}, (k+1)^j)$ see Section 4B.

Theorem. *Let* $m \in \mathbb{Z}_+$ *and write* $m = kN + j$, $0 \leq j < N$.

(1) *We have an isomorphism of* $\boldsymbol{U}(\mathfrak{sl}_2 \otimes \mathbb{C}[t]/t^N)$-*modules*

$$\mathrm{W}(m, N) \cong V(k)^{*(N-j)} * V(k+1)^{*j}.$$

(2) *A* PBW *type basis of* $\mathrm{W}(m, N)$ *is given by*

$$\{(x_{-\alpha} \otimes 1)^{i_0} \cdots (x_{-\alpha} \otimes t^{N-1})^{i_{N-1}} w_{m,N} \mid (i_0, \ldots, i_{N-1}) \in S(k^{N-j}, (k+1)^j)\}.$$

Our paper is organized as follows. Section 2 establishes the basic notation needed in the rest of the paper. In Section 3, we construct in different ways two bigraded modules and prove them to be isomorphic. As a consequence we obtain that the fusion product is independent of the chosen parameters. In Section 4, we give some applications regarding the conjecture on truncated Weyl modules and compute a PBW type basis.

2. Preliminaries

2A. Throughout the paper $\mathbb{C}$ denotes the field of complex numbers, $\mathbb{Z}$ the ring of integers, $\mathbb{Z}_+$ the set of nonnegative integers and $\mathbb{N}$ the set of positive integers. Given any complex Lie algebra $\mathfrak{a}$ we let $\boldsymbol{U}(\mathfrak{a})$ be the universal enveloping algebra of $\mathfrak{a}$. Further, let $\mathfrak{a}[t]$ be the Lie algebra of polynomial maps from $\mathbb{C}$ to $\mathfrak{a}$ with the obvious pointwise Lie bracket:

$$[x\otimes f, y\otimes g]=[x, y]\otimes fg, \quad x, y\in\mathfrak{a},\ f, g\in\mathbb{C}[t].$$

The Lie algebra $\mathfrak{a}[t]$ and its universal enveloping algebra inherit a grading from the degree grading of $\mathbb{C}[t]$; thus an element $a_1\otimes t^{r_1}\cdots a_s\otimes t^{r_s}$, $a_j\in\mathfrak{a}$, $r_j\in\mathbb{Z}_+$ for $1\le j\le s$, will have grade $r_1+\cdots+r_s$. We shall be interested in $\mathbb{Z}$-graded modules $V=\bigoplus_{s\in\mathbb{Z}} V[s]$ for $\mathfrak{a}[t]$.

2B. We refer to [Kac 1990] for the general theory of affine Lie algebras. Let $\mathfrak{g}$ be a finite-dimensional complex simple Lie algebra and $\hat{\mathfrak{g}}$ be the corresponding untwisted affine algebra. We fix $\mathfrak{h}\subset\hat{\mathfrak{h}}$ Cartan subalgebras of $\mathfrak{g}$ and $\hat{\mathfrak{g}}$, respectively, and denote by R the set of roots of $\mathfrak{g}$ with respect to $\mathfrak{h}$ and by $\widehat{R}$ the set of roots of $\hat{\mathfrak{g}}$ with respect to $\hat{\mathfrak{h}}$. The corresponding sets of positive and negative roots are denoted as usual by $R^\pm$ and $\widehat{R}^\pm$, respectively. We fix a basis $\Delta=\{\alpha_1,\dots,\alpha_n\}$ for R such that $\widehat{\Delta}=\Delta\cup\{\alpha_0\}$ is a basis for $\widehat{R}$. For $\alpha\in\widehat{R}$, let $\alpha^\vee$ be the corresponding coroot. We fix $d\in\hat{\mathfrak{h}}$ such that $\alpha_0(d)=1$ and $\alpha_i(d)=0$ for $i\neq 0$; d is called the scaling element and it is unique modulo the center of $\hat{\mathfrak{g}}$. For $1\le i\le n$, define $\omega_i\in\mathfrak{h}^*$ by $\omega_i(\alpha_j^\vee)=\delta_{i,j}$ for $1\le j\le n$, where $\delta_{i,j}$ is Kronecker's delta symbol. The element ω_i is the fundamental weight of $\mathfrak{g}$ corresponding to $\alpha_i^\vee$. Let $(\cdot,\cdot)$ be the nondegenerate symmetric bilinear form on $\mathfrak{h}^*$ normalized so that the square length of a long root is two. For $\alpha\in R^+$ we set

$$d_\alpha=\frac{2}{(\alpha,\alpha)},\quad d_i:=d_{\alpha_i}\text{ for }1\le i\le n.$$

The weight lattices P and P^+ are the $\mathbb{Z}$-span and $\mathbb{Z}_+$-span, respectively, of the fundamental weights. The coweight lattice L is the sublattice of P spanned by the elements $d_i\omega_i$, $1 \le i \le n$, and the subset L^+ is defined in the obvious way. For $\lambda \in P^+$ we define

$$|\lambda| = \sum_{i=1}^{n} \lambda(\alpha_i^\vee) \in \mathbb{Z}_+.$$

2C. Given $\alpha \in \widehat{R}^+$ let $\hat{\mathfrak{g}}_\alpha \subset \hat{\mathfrak{g}}$ be the corresponding root space; note that $\hat{\mathfrak{g}}_\alpha \subset \mathfrak{g}$ if $\alpha \in R$. For a real root α we denote by x_α a generator of $\hat{\mathfrak{g}}_\alpha$. The element d defines a $\mathbb{Z}_+$-graded Lie algebra structure on $\mathfrak{g}[t]$: for $\alpha \in \widehat{R}$ we say that $\mathfrak{g}_\alpha$ has grade k if

$$[d, x_\alpha] = kx_\alpha$$

or, equivalently, if $\alpha(d) = k$. With respect to this grading, the zero homogeneous component of the current algebra is $\mathfrak{g}[t][0] \cong \mathfrak{g}$ and the subspace spanned by the positive homogeneous components is an ideal denoted by $\mathfrak{g}[t]_+$. We have a short exact sequence of Lie algebras

$$0 \to \mathfrak{g}[t]_+ \to \mathfrak{g}[t] \xrightarrow{\mathrm{ev}_0} \mathfrak{g} \to 0.$$

Clearly the pull-back of any $\mathfrak{g}$-module V by ev_0 defines the structure of a graded $\mathfrak{g}[t]$-module on V, and we denote this module by $\mathrm{ev}_0^* \mathrm{V}$.

2D. For $\lambda \in P^+$, denote by $\mathrm{V}(\lambda)$ the simple finite-dimensional $\mathfrak{g}$-module generated by an element v_λ with defining relations

$$\mathfrak{n}^+ v_\lambda = 0, \quad \alpha_i^\vee v_\lambda = \lambda(\alpha_i^\vee) v_\lambda, \quad (x_{-\alpha_i})^{\lambda(\alpha_i^\vee)+1} v_\lambda = 0, \quad 1 \le i \le n.$$

It is well known that $\mathrm{V}(\lambda) \cong \mathrm{V}(\mu)$ if and only if $\lambda = \mu$ and that any finite-dimensional $\mathfrak{g}$-module is isomorphic to a direct sum of modules $\mathrm{V}(\lambda)$, $\lambda \in P^+$. If V is an $\mathfrak{h}$-semisimple $\mathfrak{g}$-module (in particular if $\dim \mathrm{V} < \infty$), we have

$$\mathrm{V} = \bigoplus_{\mu \in \mathfrak{h}^*} \mathrm{V}_\mu, \quad \mathrm{V}_\mu = \{v \in \mathrm{V} \mid hv = \mu(h)v, \ h \in \mathfrak{h}\},$$

and we set $\mathrm{wt}\, \mathrm{V} = \{\mu \in \mathfrak{h}^* \mid V_\mu \neq 0\}$. By our previous comments, for any $\lambda \in P^+$ we obtain a graded $\mathfrak{g}[t]$-module $\mathrm{ev}_0^* \mathrm{V}(\lambda)$.

We also define the local Weyl module $\mathrm{W}_{\mathrm{loc}}(\lambda)$, which is a finite-dimensional $\mathfrak{g}[t]$-module generated by an element w_λ with defining relations

$$\mathfrak{n}^+[t] w_\lambda = 0, \quad (\alpha_i^\vee \otimes t^s) w_\lambda = \delta_{s,0} \lambda(\alpha_i^\vee) w_\lambda, \quad (x_{-\alpha_i} \otimes 1)^{\lambda(\alpha_i^\vee)+1} w_\lambda = 0 \qquad \forall s \ge 0, \ 1 \le i \le n.$$

For more details regarding the theory of local Weyl modules see [Chari et al. 2014b; Chari and Loktev 2006; Chari and Pressley 2001; Fourier et al. 2012; Fourier and Kus 2013; Fourier and Littelmann 2007; Naoi 2012].

2E. We recall a general construction from [Feigin and Loktev 1999]. Let $\boldsymbol{U}(\mathfrak{g}[t])[k]$ be the homogeneous component of degree k (with respect to the grading induced by d) and recall that it is a $\mathfrak{g}$-module for all $k \in \mathbb{Z}_+$. Suppose now that we are given a $\mathfrak{g}[t]$-module V which is generated by v. Define an increasing filtration $0 \subset \mathrm{V}^0 \subset \mathrm{V}^1 \subset \cdots$ of $\mathfrak{g}$-submodules of V by

$$\mathrm{V}^k = \bigoplus_{s=0}^{k} \boldsymbol{U}(\mathfrak{g}[t])[s]v.$$

The associated graded vector space $\operatorname{gr}\mathrm{V}$ admits an action of $\mathfrak{g}[t]$ given by

$$x(v+\mathrm{V}^k) = xv + \mathrm{V}^{k+s}, \quad x \in \mathfrak{g}[t][s],\ v \in \mathrm{V}^{k+1}.$$

Furthermore, $\operatorname{gr}\mathrm{V}$ is a cyclic $\mathfrak{g}[t]$-module with cyclic generator $\bar{v}$, the image of v in $\operatorname{gr}\mathrm{V}$. Given a $\mathfrak{g}[t]$-module V and $z \in \mathbb{C}$, let V^z be the $\mathfrak{g}[t]$-module with action

$$(x \otimes t^r)w = (x \otimes (t+z)^r)w, \quad x \in \mathfrak{g},\ w \in \mathrm{V},\ r \in \mathbb{Z}_+.$$

Starting with finite-dimensional cyclic $\mathfrak{g}[t]$-modules $\mathrm{V}_1, \dots, \mathrm{V}_N$ with cyclic vectors $v_1, \dots, v_N$ and a tuple of pairwise distinct complex numbers $\boldsymbol{z} = (z_1, \dots, z_N)$, the fusion product is defined to be $\mathrm{V}^{z_1} * \cdots * \mathrm{V}^{z_N} := \operatorname{gr}(\mathrm{V}^{z_1} \otimes \cdots \otimes \mathrm{V}^{z_N})$. It was proved in [Feigin and Loktev 1999] that the tensor product $\mathrm{V}^{z_1} \otimes \cdots \otimes \mathrm{V}^{z_N}$ is cyclic and generated by $v_1 \otimes \cdots \otimes v_N$. Clearly the definition of the fusion product depends on the parameters z_s, $1 \le s \le N$. However, it was conjectured in that paper (and proved in special cases; see [Chari and Loktev 2006; Feigin and Feigin 2002; Feigin and Loktev 1999; Fourier and Littelmann 2007; Kus and Venkatesh 2014], for example) that under suitable conditions on V_s and v_s, the fusion product is independent of the choice of the complex numbers. In this paper we cover another class of representations, where the construction of the fusion product is independent of the parameters. To keep the notation as simple as possible we almost always omit the parameters in the notation for the fusion product and write $\mathrm{V}_1 * \cdots * \mathrm{V}_N$ for $\mathrm{V}_1^{z_1} * \cdots * \mathrm{V}_N^{z_N}$.

3. Filtrations and bigraded modules

The aim of this section is to construct two finite-dimensional bigraded modules in different ways and prove them to be isomorphic. The advantage of this construction is that a comparison of the zeroth graded components leads to a realization of the fusion product associated to rectangular partitions.

3A. Let us start with our first construction. Let $\lambda \in P^+$ and $\ell \in \mathbb{N}$. The $\mathfrak{g}$-stable Demazure module $\mathrm{D}(\ell, \lambda)$ is a finite-dimensional submodule of a level ℓ highest weight representation for the affine algebra $\hat{\mathfrak{g}}$. For these representations, generators and relations are known if we consider them as $\boldsymbol{U}(\mathfrak{g}[t])$-modules; see [Fourier and Kus 2013; Fourier and Littelmann 2007; Mathieu 1988] for more details. We remark that these relations are greatly simplified for Demazure modules for untwisted affine algebras in [Chari and Venkatesh 2015] and for twisted affine algebras in [Kus and Venkatesh 2014]. For instance, one can use these simplified relations to see directly that level one Demazure modules are isomorphic to local Weyl modules for simply laced affine algebras and twisted affine algebras, initially proved in [Fourier and Littelmann 2007] and [Chari et al. 2014b; Fourier and Kus 2013], respectively. We recall the simplified presentation of $\mathfrak{g}$-stable Demazure modules. Write

$$\lambda(\beta^\vee) = (p_\beta - 1)d_\beta\ell + m_\beta, \quad 0 < m_\beta \leq d_\beta\ell, \text{ for } \beta \in R^+. \tag{3-1}$$

Proposition 3.1. *The Demazure module* $\mathrm{D}(\ell, \lambda)$ *is isomorphic to the cyclic* $\boldsymbol{U}(\mathfrak{g}[t])$-*module generated by a vector* $v \neq 0$ *subject to the following relations*:

$$\mathfrak{n}^+[t]v = 0, \quad (h \otimes t^s)v = \delta_{s,0}\lambda(h)v \quad \forall h \in \mathfrak{h},\ s \geq 0, \tag{3-2}$$

$$(x_{-\beta} \otimes 1)^{\lambda(\beta^\vee)+1}v = 0, \quad (x_{-\beta} \otimes t^{p_\beta})v = 0 \quad \forall \beta \in R^+, \tag{3-3}$$

$$(x_{-\beta} \otimes t^{p_\beta - 1})^{m_\beta+1}v = 0 \quad \forall \beta \in R^+ \textit{ such that } m_\beta < d_\beta\ell. \tag{3-4}$$

We can decompose $\mathrm{D}(\ell, \lambda)$ into simple finite-dimensional $\mathfrak{g}$-modules. We remark that the vector v in Proposition 3.1 corresponds to the highest weight vector of $\mathrm{ev}_0^* \mathrm{V}(\lambda)$ in the $\mathfrak{g}$-module decomposition of $\mathrm{D}(\ell, \lambda)$. We call it a highest weight vector of the module.

3B. We are concerned with Demazure modules of the form $\mathrm{D}(\ell, \ell N\lambda^1 + \lambda^0)$, where $\lambda^1 \in L^+$ and $\lambda^0 \in P^+$ such that $\lambda^0(\theta^\vee) \leq \ell$. *For the rest of this paper we assume that either* $\lambda^0 = 0$ *and* $\mathfrak{g}$ *is arbitrary or* $\lambda^0 \neq 0$ *and* $\mathfrak{g}$ *is of classical type or* $\mathtt{G}_2$. By the results of [Chari et al. 2014c; Fourier and Littelmann 2007], the Demazure module $\mathrm{D}(\ell, \ell N\lambda^1 + \lambda^0)$ is isomorphic to the fusion product of $N - 1$ copies of the Demazure module $\mathrm{D}(\ell, \ell\lambda^1)$ with $\mathrm{D}(\ell, \ell\lambda^1 + \lambda^0)$:

$$\mathrm{D}(\ell, \ell N\lambda^1 + \lambda^0) \cong \mathrm{D}(\ell, \ell\lambda^1) * \cdots * \mathrm{D}(\ell, \ell\lambda^1) * \mathrm{D}(\ell, \ell\lambda^1 + \lambda^0). \tag{3-5}$$

This decomposition holds for all fusion parameters $\boldsymbol{z} = (z_1, \ldots, z_N)$ with $z_i \neq z_j$ for all $i \neq j$. We emphasize that the restriction on $\mathfrak{g}$ is only made because (3-5) is not proved for the remaining exceptional Lie algebras if λ^0 is nonzero. In other words, our results are applicable whenever we have such a fusion product decomposition. We will need the following lemma.

Lemma 3.2. *Let* β *be a positive root and* $\lambda \in P^+$. *We write* $\theta - \beta = \sum_j \gamma_j$ *as a sum of positive roots. Then we have*

$$\lambda(\beta^\vee)(\beta, \beta) \leq \lambda(\theta^\vee)(\theta, \theta) \quad \textit{with equality if and only if } \lambda(\gamma_j^\vee) = 0\ \forall j.$$

Proof. Since λ is a dominant integral weight we have $\lambda(\beta^\vee) \geq 0$ for a positive root β. We obtain

$$\theta^\vee = \Big(\beta + \sum_j \gamma_j\Big)^\vee = \frac{(\beta, \beta)}{2}\beta^\vee + \sum_j \frac{(\gamma_j, \gamma_j)}{2}\gamma_j^\vee,$$

which gives

$$\lambda(\theta^\vee)(\theta, \theta) = (\beta, \beta)\lambda(\beta^\vee) + \sum_j (\gamma_j, \gamma_j)\lambda(\gamma_j^\vee) \geq (\beta, \beta)\lambda(\beta^\vee).$$

Note that equality is only possible if $\lambda(\gamma_j^\vee) = 0$ for all j, since $(\gamma_j, \gamma_j) > 0$. □

By Lemma 3.2 and Equation (3-3) of Proposition 3.1 we get the following result.

Corollary 3.3. $(x_{-\beta} \otimes t^{(\lambda^1(\theta^\vee)+1)N})v = 0$ *for all roots* $\beta \in R^+$.

Proof. Write $(\ell N\lambda^1 + \lambda^0)(\beta^\vee)$ as in (3-1). Since $\lambda^1 \in L^+$ we have $m_\beta = d_\beta \ell$ if $\lambda^0(\beta^\vee) = 0$ and $m_\beta = \lambda^0(\beta^\vee)$ else. Then $(x_{-\beta} \otimes t^{p_\beta})v = 0$ and

$$p_\beta = N\frac{\lambda^1(\beta^\vee)}{d_\beta} + \frac{\lambda^0(\beta^\vee) - m_\beta}{d_\beta \ell} + 1 \leq N(\lambda^1(\theta^\vee) + 1). \qquad \square$$

Hence $\mathrm{D}(\ell, \ell N\lambda^1 + \lambda^0)$ is a $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t]/t^{(\lambda^1(\theta^\vee)+1)N})$-module.

3C. We define a decreasing filtration on $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t]/t^{(\lambda^1(\theta^\vee)+1)N})$

$$\mathrm{T}_0(N) \supseteq \mathrm{T}_1(N) \supseteq \mathrm{T}_2(N) \supseteq \cdots,$$

with

$$\begin{aligned} \mathrm{T}_0(N) &= \boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t]/t^{(\lambda^1(\theta^\vee)+1)N}), \\ \mathrm{T}_j(N) &= (\mathfrak{g} \otimes t^N\mathbb{C}[t])\,\mathrm{T}_{j-1}(N) \quad \text{for } j \geq 1, \end{aligned}$$

and study the induced decreasing filtration on our Demazure module given by

$$\mathrm{D}(\ell, \ell N\lambda^1 + \lambda^0) = \mathrm{T}_0(N)v =: \mathrm{T}_0(\ell, N) \supseteq \mathrm{T}_1(N)v := \mathrm{T}_1(\ell, N) \supseteq \mathrm{T}_2(\ell, N) \supseteq \cdots.$$

To be consistent with the notation in [Feigin 2008], we refer to it as the t^N-filtration. Let $\mathrm{gr}_{t^N} \mathrm{T}(N)$ and $\mathrm{gr}_{t^N} \mathrm{T}(\ell, N)$, respectively, be the associated graded spaces

$$\mathrm{gr}_{t^N} T(N) = \mathrm{T}_0(N)/\mathrm{T}_1(N) \oplus \mathrm{T}_1(N)/\mathrm{T}_2(N) \oplus \cdots$$

and

$$\mathrm{gr}_{t^N} \mathrm{T}(\ell, N) = \mathrm{T}_0(\ell, N)/\mathrm{T}_1(\ell, N) \oplus \mathrm{T}_1(\ell, N)/\mathrm{T}_2(\ell, N) \oplus \cdots.$$

Since $\mathrm{D}(\ell, \ell N\lambda^1 + \lambda^0)$ is a module for $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t]/t^{(\lambda^1(\theta^\vee)+1)N})$ we obtain that $\mathrm{gr}_{t^N} \mathrm{T}(\ell, N)$ is a module for $\mathrm{gr}_{t^N} \mathrm{T}(N)$. By the following lemma $\mathrm{gr}_{t^N} \mathrm{T}(\ell, N)$ is also a module for the toroidal current algebra $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t, u]/\langle t^N, u^{\lambda^1(\theta^\vee)+1}\rangle)$.

Lemma 3.4. *We have an isomorphism of algebras*

$$\Psi : \mathrm{gr}_{t^N} \mathrm{T}(N) \xrightarrow{\sim} \boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t, u]/\langle t^N, u^{\lambda^1(\theta^\vee)+1}\rangle),$$

where $\Psi(x \otimes t^{jN+s}) = x \otimes u^j t^s$ *for* $x \in \mathfrak{g}$ *and* $0 \leq s < N$.

Proof. The map Ψ is clearly an isomorphism of vector spaces. In order to show that this map is an algebra homomorphism, we have to check that the naive way of defining Ψ on a product of elements is well defined. Hence we will verify that

$$(x \otimes u^j t^s)(y \otimes u^i t^q) - (y \otimes u^i t^q)(x \otimes u^j t^s) = [x, y] \otimes u^{i+j} t^{s+q}$$

holds on the right-hand side whenever we have

$$(x \otimes t^{jN+s})(y \otimes t^{iN+q}) - (y \otimes t^{iN+q})(x \otimes t^{jN+s}) = [x, y] \otimes t^{(i+j)N+(s+q)}$$

on the left-hand side. This is obvious for $s + q < N$. Otherwise the variables $x \otimes u^j t^s$ and $y \otimes u^i t^q$ commute in $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t, u]/\langle t^N, u^{\lambda^1(\theta^\vee)+1}\rangle)$. By the definition of the associated graded space we also obtain that the variables $x \otimes t^{jN+s}$ and $y \otimes t^{iN+q}$ commute in $\mathrm{gr}_{t^N} \mathrm{T}(N)$ since on the one hand

$$(x \otimes t^{jN+s})(y \otimes t^{iN+q}) - (y \otimes t^{iN+q})(x \otimes t^{jN+s}) \in \mathrm{T}_{i+j}(N)$$

and on the other hand

$$[x, y] \otimes t^{(i+j)N+(s+q)} \in \mathrm{T}_{i+j+1}(N). \qquad \square$$

3D. Now we present a quite different construction of the module $\mathrm{gr}_{t^N} \mathrm{T}(\ell, N)$. In fact, it is one of the main results of this paper that the two constructions give isomorphic modules. We start with the $(N-1)$-fold tensor product of Demazure modules $\mathrm{D}(\ell, \ell\lambda^1)$ with $\mathrm{D}(\ell, \ell\lambda^1 + \lambda^0)$. The gambit: we switch the variables and consider now the current algebra $\mathfrak{g}[u]$ which operates on the Demazure modules $\mathrm{D}^u(\ell, \ell\lambda^1)$. We add the index u to emphasize that here the algebra $\mathfrak{g}[u]$ is acting. We extend the action trivially to $\mathfrak{g}[t, u]$ and denote the corresponding module by $\overline{\mathrm{D}^u(\ell, \ell\lambda^1)}$; i.e., $\mathfrak{g} \otimes t\mathbb{C}[t, u]$ acts trivially. Recall that we get a highly nontrivial action of $\mathfrak{g}[t, u]$ when we consider fusion products of these modules with respect to the variable t. The bigraded fusion product

$$\overline{\mathrm{D}^u(\ell, \ell\lambda^1)} * \cdots * \overline{\mathrm{D}^u(\ell, \ell\lambda^1)} * \overline{\mathrm{D}^u(\ell, \ell\lambda^1 + \lambda^0)} \tag{3-6}$$

is a cyclic module for the Lie algebra $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t, u]/\langle t^N, u^{\lambda^1(\theta^\vee)+1}\rangle)$. Note the similarity but also the difference between (3-5) and (3-6). In (3-5) we consider the fusion product (with respect to the variable t) of $\mathfrak{g}[t]$-modules. The $\mathfrak{g}[t, u]$-module structure comes into the picture only by the filtration defined in Section 3C. We would like to remind the reader that if $\lambda^0 \neq 0$, then $\mathfrak{g}$ is of classical type or $\mathtt{G}_2$.

Theorem 3.5. *Let $\lambda^1 \in L^+$ and $\lambda^0 \in P^+$ such that $\lambda^0(\theta^\vee) \le \ell$. We have an isomorphism of $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t,u]/\langle t^N, u^{\lambda^1(\theta^\vee)+1}\rangle)$-modules*

$$\operatorname{gr}_{t^N} \mathrm{T}(\ell, N) \cong \overline{\mathrm{D}^u(\ell, \ell\lambda^1)} * \cdots * \overline{\mathrm{D}^u(\ell, \ell\lambda^1)} * \overline{\mathrm{D}^u(\ell, \ell\lambda^1 + \lambda^0)}.$$

Proof. Let $v_\ell^{*(N-1)} * v_0$ be the highest weight vector of the right-hand side. The isomorphism between $\operatorname{gr}_{t^N} \mathrm{T}(N)$ and $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t,u]/\langle t^N, u^{\lambda^1(\theta^\vee)+1}\rangle)$ (see Lemma 3.4) induces a natural surjective map

$$\operatorname{gr}_{t^N} \mathrm{T}(N) \twoheadrightarrow \boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t,u]/\langle t^N, u^{\lambda^1(\theta^\vee)+1}\rangle) \circ (v_\ell^{*(N-1)} * v_0).$$

It remains to prove that this map induces an isomorphism between the cyclic module $\operatorname{gr}_{t^N} \mathrm{T}(\ell, N)$ and the fusion product. Since the dimensions of the modules coincide it is enough to show that all relations which hold in $\operatorname{gr}_{t^N} \mathrm{T}(\ell, N)$ also hold on the right-hand side.

Recall that a presentation of $\operatorname{gr}_{t^N} \mathrm{T}(\ell, N)$ is given by two types of relations, the ones coming from the presentation of the Demazure module and the ones coming from going to the associated graded space with respect to the t^N-filtration. We start by proving that the defining relations of $\mathrm{D}(\ell, \ell N\lambda^1 + \lambda^0)$ given for v in Proposition 3.1 are satisfied by $v_\ell * \cdots * v_\ell * v_0$. Since the relations (3-2) and the first part of (3-3) are obviously satisfied it remains to verify the second part of (3-3) and (3-4). Write $(\ell N\lambda^1 + \lambda^0)(\beta^\vee)$ as in (3-1). We start by proving that

$$(x_{-\beta} \otimes u^{j_\beta} t^{r_\beta})(v_\ell^{*(N-1)} * v_0) = 0, \quad \text{where } p_\beta = j_\beta N + r_\beta,\ 0 \le r_\beta < N. \tag{3-7}$$

Since $\lambda^0(\beta^\vee) \le d_\beta \ell$, we have

$$p_\beta = \begin{cases} N\lambda^1(\beta^\vee)d_\beta^{-1} & \text{if } \lambda^0(\beta^\vee) = 0, \\ N\lambda^1(\beta^\vee)d_\beta^{-1} + 1 & \text{else.} \end{cases}$$

In either case $j_\beta \ge \lambda^1(\beta^\vee)d_\beta^{-1}$ and thus $(x_{-\beta} \otimes u^{j_\beta} t^{r_\beta})v_\ell = 0$ follows immediately from the defining relations of $\mathrm{D}(\ell, \ell\lambda^1)$. If $r_\beta \ne 0$ we can replace t^{r_β} by $(t - z_N)^{r_\beta}$ in the associated graded space and obtain that the element in (3-7) acts trivially on v_0. If $r_\beta = 0$, then p_β is divisible by N, which forces $\lambda^0(\beta^\vee) = 0$. Therefore, in this case we obtain $j_\beta = \lambda^1(\beta^\vee)d_\beta^{-1}$, and $(x_{-\beta} \otimes u^{j_\beta})v_0 = 0$ follows immediately from the defining relations of $\mathrm{D}(\ell, \ell\lambda^1 + \lambda^0)$. It remains to consider the relations (3-4). So suppose we have

$$p_\beta - 1 = N\frac{\lambda^1(\beta^\vee)}{d_\beta} + \frac{\lambda^0(\beta^\vee) - m_\beta}{d_\beta \ell} = j_\beta N + r_\beta, \quad 0 \le r_\beta < N.$$

Since $m_\beta < d_\beta \ell$, we must have $m_\beta = \lambda^0(\beta^\vee) \ne 0$ and hence $p_\beta - 1 = N\lambda^1(\beta^\vee)d_\beta^{-1}$. It follows that $j_\beta = \lambda^1(\beta^\vee)d_\beta^{-1}$ and therefore $(x_{-\beta} \otimes u^{j_\beta})v_\ell = 0$. So we have

$$(x_{-\beta} \otimes u^{j_\beta})^{m_\beta+1}(v_\ell^{*(N-1)} * v_0) = v_\ell^{*(N-1)} * (x_{-\beta} \otimes u^{j_\beta})^{m_\beta+1} v_0 = 0.$$

We now consider the relations coming from the t^N-filtration. Suppose

$$M=\sum_m \sum_{\substack{i_1,\dots,i_m\\ j_1,\dots,j_m}} k(m)_{\substack{i_1,\dots,i_m\\ j_1\dots,j_m}}(x_{i_1}\otimes t^{i_1N+j_1})\cdots(x_{i_m}\otimes t^{i_mN+j_m}) \in \boldsymbol{U}(\mathfrak{g}\otimes\mathbb{C}[t]/t^{(\lambda(\theta^\vee)+1)N})$$

is a linear combination of monomials with fixed t^N-degree such that $w=Mv\neq 0$ in $\mathrm{D}(\ell,\ell N\lambda^1+\lambda^0)$ but the image $\bar{w}=0$ in $\mathrm{gr}_{t^N}\,\mathrm{T}(\ell,N)$. This is only possible if there exists a linear combination of monomials of greater t^N-degree

$$M'=\sum_{m'} \sum_{\substack{p_1,\dots,p_{m'}\\ q_1\dots,q_{m'}}} k(m')_{\substack{p_1,\dots,p_{m'}\\ q_1,\dots,q_{m'}}}(x_{p_1}\otimes t^{p_1N+q_1})\cdots(x_{p_{m'}}\otimes t^{p_{m'}N+q_{m'}}) \in \boldsymbol{U}(\mathfrak{g}\otimes\mathbb{C}[t]/t^{(\lambda(\theta^\vee)+1)N})$$

such that $w=Mv=M'v$ in $\mathrm{D}(\ell,\ell N\lambda^1+\lambda^0)$. We assume in what follows that M' is of maximal t^N-degree. We have $(M-M')v=0$, so the difference $M-M'$ is an element in the left ideal generated by the elements in (3-2)–(3-4). Since M' is of higher t^N-degree we get $\overline{M-M'}=\overline{M}$ in $\mathrm{gr}_{t^N}\,\mathrm{T}(N)$, and since all defining relations of $\mathrm{D}(\ell,\ell N\lambda^1+\lambda^0)$ are satisfied by $v_\ell^{*(N-1)}*v_0$ we get $\Psi(\overline{M})\circ(v_\ell^{*(N-1)}*v_0)=0$, which shows that the natural surjective map

$$\mathrm{gr}_{t^N}\,\mathrm{T}(N)\twoheadrightarrow \boldsymbol{U}(\mathfrak{g}\otimes\mathbb{C}[t,u]/\langle t^N,u^{\lambda^1(\theta^\vee)+1}\rangle)\circ(v_\ell^{*(N-1)}*v_0)$$

induces an isomorphism of cyclic modules $\mathrm{gr}_{t^N}\,\mathrm{T}(\ell,N)\cong\overline{\mathrm{D}^u(\ell,\ell\lambda^1)}*\cdots*\overline{\mathrm{D}^u(\ell,\ell\lambda^1)}*\overline{\mathrm{D}^u(\ell,\ell\lambda^1+\lambda^0)}$. □

For the rest of this section we discuss a crucial consequence of our result.

Corollary 3.6. *Let $\ell\in\mathbb{N}$, $\lambda^1\in L^+$ and $\lambda^0\in P^+$ such that $\lambda^0(\theta^\vee)\le\ell$.*

(1) *The fusion product $\overline{\mathrm{D}^u(\ell,\ell\lambda^1)}*\cdots*\overline{\mathrm{D}^u(\ell,\ell\lambda^1)}*\overline{\mathrm{D}^u(\ell,\ell\lambda^1+\lambda^0)}$ is independent of the fusion parameters.*

(2) *The fusion product $\mathrm{V}(\ell\lambda^1)^{*(N-1)}*\mathrm{V}(\ell\lambda^1+\lambda^0)$ is independent of the fusion parameters.*

(3) *We have an isomorphism of $\boldsymbol{U}(\mathfrak{g}\otimes\mathbb{C}[t]/t^N)$-modules*

$$\mathrm{V}(\ell\lambda^1)^{*(N-1)}*\mathrm{V}(\ell\lambda^1+\lambda^0)\cong\mathrm{D}(\ell,\ell N\lambda^1+\lambda^0)/(\mathfrak{g}\otimes t^N\mathbb{C}[t])\,\mathrm{D}(\ell,\ell N\lambda^1+\lambda^0).$$

(4) *If $\lambda^0(\theta^\vee)\le 1$, the truncated level one Demazure module is isomorphic to the truncated level ℓ Demazure module*

$$\mathrm{D}(1,\ell N\lambda^1+\lambda^0)/(\mathfrak{g}\otimes t^N\mathbb{C}[t])\,\mathrm{D}(1,\ell N\lambda^1+\lambda^0) \cong\mathrm{D}(\ell,\ell N\lambda^1+\lambda^0)/(\mathfrak{g}\otimes t^N\mathbb{C}[t])\,\mathrm{D}(\ell,\ell N\lambda^1+\lambda^0).$$

Proof. Since the fusion product $\mathrm{V}(\ell\lambda^1)^{*(N-1)}*V(\ell\lambda^1+\lambda^0)$ is isomorphic to the zeroth graded component of $\overline{\mathrm{D}^u(\ell,\ell\lambda^1)}*\cdots*\overline{\mathrm{D}^u(\ell,\ell\lambda^1)}*\overline{\mathrm{D}^u(\ell,\ell\lambda^1+\lambda^0)}$ (with respect to the u-grading) the statement follows from Theorem 3.5. □

Remark. Theorem 3.5 generalizes a result of [Feigin 2008], where the theorem was proved for $\ell = 1$, $\lambda^0 = 0$ and $\lambda^1 = \theta$. Unfortunately, the proof in that paper has a gap (personal communication by the author), which is now fixed by the proof above. Ravinder [2014] used the result of [Feigin 2008] to prove a presentation for the fusion product $\mathrm{V}(\theta)^{*N} * \mathrm{D}(1,\theta)^{*M}$.

4. Truncated Weyl modules and PBW type basis

In this section we give some evidence for the conjecture made by Chari, Fourier and Sagaki on truncated Weyl modules (see [Chari et al. 2014a; Fourier 2015]). For the reader's convenience we state the precise conjecture in this paper (Conjecture 4.1). Finally, we consider the case $\mathfrak{g} = \mathfrak{sl}_2$ and compute a PBW type basis.

4A. Let $P^+(\lambda, N)$ be the set of N-tuples of dominant integral weights $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$ such that $\sum_i \lambda_i = \lambda$. Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N) \in P^+(\lambda, N)$. For a positive root β define

$$r_{\beta,k}(\boldsymbol{\lambda}) = \min\{(\lambda_{i_1} + \cdots + \lambda_{i_k})(\beta^\vee) \mid 1 \le i_1 < \cdots < i_k \le N\}.$$

We say $\boldsymbol{\lambda} \preceq \boldsymbol{\mu}$ if

$$r_{\beta,k}(\boldsymbol{\lambda}) \le r_{\beta,k}(\boldsymbol{\mu}) \quad \text{for all } \beta \in R^+ \text{ and } 1 \le k \le N.$$

The above partial order was considered by Chari et al. [2014a], who observed that for a tuple $\boldsymbol{\lambda}$ the dimension of the tensor product of the corresponding finite-dimensional simple $\mathfrak{g}$-modules increases along $\preceq$. Moreover, they proved in certain cases (for instance when λ is a multiple of a fundamental minuscule weight) that there exists an inclusion of tensor products along with the partial order and conjectured that this remains true for $N = 2$ and arbitrary λ (see [Chari et al. 2014a, Conjecture 2.3]). Using the unique maximal element in the partially ordered set $P^+(\lambda, N)$ one can formulate a conjecture on truncated Weyl modules, which we will explain now.

Definition. Let $\lambda \in P^+$. The truncated Weyl module $\mathrm{W}(\lambda, N)$ is a cyclic module for $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t]/t^N)$ generated by $w_{\lambda,N}$ with relations

$$(\mathfrak{n}^+ \otimes \mathbb{C}[t]/t^N) w_{\lambda,N} = 0, \quad (h \otimes t^s) w_{\lambda,N} = \delta_{s,0} \lambda(h) w_{\lambda,N} \ \forall h \in \mathfrak{h},\ s \ge 0, \tag{4-1}$$

$$(x_{-\beta} \otimes 1)^{\lambda(\beta^\vee)+1} w_{\lambda,N} = 0 \quad \forall \beta \in R^+. \tag{4-2}$$

The following conjecture gives a connection between truncated Weyl modules and fusion products of irreducible finite-dimensional $\mathfrak{g}$-modules.

Conjecture 4.1. Let $\lambda \in P^+$ such that $|\lambda| \ge N$, and let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$ be the unique maximal element in $P^+(\lambda, N)$. Then we have an isomorphism of $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t]/t^N)$-modules

$$\mathrm{W}(\lambda, N) \cong V(\lambda_1) * \cdots * V(\lambda_N).$$

The following result proves the above conjecture for certain classes of dominant integral weights.

Theorem 4.2. *Let $\lambda \in L^+$ and $\lambda^0 \in P^+$ such that $\lambda^0(\theta^\vee) \le 1$ and $|N\lambda + \lambda^0| \ge N$. Then we have an isomorphism of $\boldsymbol{U}(\mathfrak{g} \otimes \mathbb{C}[t]/t^N)$-modules*

$$\mathrm{W}(N\lambda + \lambda^0, N) \cong V(\lambda) * \cdots * V(\lambda) * V(\lambda + \lambda^0).$$

Proof. If $\lambda = 0$, there is nothing to prove. By Corollary 3.6 we obtain that

$$V(\lambda) * \cdots * V(\lambda) * V(\lambda + \lambda^0) \cong \mathrm{D}(1, N\lambda + \lambda^0)/(\mathfrak{g} \otimes t^N \mathbb{C}[t])\, \mathrm{D}(1, N\lambda + \lambda^0).$$

We show that the defining relations of $\mathrm{D}(1, N\lambda + \lambda^0)/(\mathfrak{g} \otimes t^N \mathbb{C}[t])\, \mathrm{D}(1, N\lambda + \lambda^0)$ hold in the truncated Weyl module. We shall prove only the nonobvious relations. Let $\beta \in R^+$ and write $(N\lambda + \lambda^0)(\beta^\vee)$ as in (3-1). Then, as before,

$$p_\beta - 1 = \begin{cases} N\lambda(\beta^\vee) d_\beta^{-1} & \text{if } \lambda^0(\beta^\vee) \neq 0, \\ N\lambda(\beta^\vee) d_\beta^{-1} - 1 & \text{else.} \end{cases}$$

We consider four cases. If $\lambda(\beta^\vee) \neq 0$ and $\lambda^0(\beta^\vee) \neq 0$, then $p_\beta \ge p_\beta - 1 \ge N$ and hence

$$(x_{-\beta} \otimes t^{p_\beta}) w_{N\lambda+\lambda^0,N} = (x_{-\beta} \otimes t^{p_\beta - 1}) w_{N\lambda+\lambda^0,N} = 0.$$

If $\lambda(\beta^\vee) \neq 0$ and $\lambda^0(\beta^\vee) = 0$, then $p_\beta \ge N$ and $m_\beta = d_\beta$ (recall that (3-4) was only considered when $m_\beta < d_\beta$). If $\lambda(\beta^\vee) = 0$ and $\lambda^0(\beta^\vee) = 0$, there is nothing to show; so consider the last case, $\lambda(\beta^\vee) = 0$ and $\lambda^0(\beta^\vee) \neq 0$. In this case $p_\beta = 1$ and $m_\beta = \lambda^0(\beta^\vee)$. Thus we have to prove

$$(x_{-\beta} \otimes t) w_{N\lambda+\lambda^0,N} = (x_{-\beta} \otimes 1)^{m_\beta + 1} w_{N\lambda+\lambda^0,N} = 0,$$

where the last equality is clear. Note that it is enough to prove that $(x_{-\beta} \otimes t)$ acts by zero on the highest weight vector of the local Weyl module $\mathrm{W}_{\mathrm{loc}}(N\lambda + \lambda^0)$. Since $\mathrm{W}_{\mathrm{loc}}(N\lambda + \lambda^0) \cong \mathrm{W}_{\mathrm{loc}}^{z_1}(N\lambda) * \mathrm{W}_{\mathrm{loc}}^{z_2}(\lambda^0)$ we get

$$(x_{-\beta} \otimes t)(w_{N\lambda} * w_{\lambda^0}) = (x_{-\beta} \otimes (t - z_2))(w_{N\lambda} * w_{\lambda^0}) = w_{N\lambda} * (x_{-\beta} \otimes t) w_{\lambda^0}.$$

If $\mathfrak{g}$ is not of type $\mathtt{G}_2$, then $\mathrm{W}_{\mathrm{loc}}(\lambda^0)$ is irreducible and the statement follows. If $\mathfrak{g}$ is $\mathtt{G}_2$ it is easy to see that the only positive root β with $(x_{-\beta} \otimes t) w_{\lambda^0} \neq 0$ is the longest short root $\beta = \alpha_1 + 2\alpha_2$. But then $\lambda(\beta^\vee) \neq 0$. □

We shall show that $(\lambda, \dots, \lambda, \lambda + \lambda^0)$ is in fact the unique maximal element in $P^+(N\lambda + \lambda^0, N)$. Since $\lambda^0(\theta^\vee) \le 1$, there exists at most one simple root α such that $\lambda^0(\alpha^\vee) > 0$. Without loss of generality we suppose $\lambda^0(\alpha_j^\vee) = 0$ for all $j > 1$. Assume that $(\mu_1, \dots, \mu_N) \in P^+(N\lambda + \lambda^0, N)$ such that $(\lambda, \dots, \lambda, \lambda + \lambda^0) \preceq (\mu_1, \dots, \mu_N)$. We fix a simple root α_j and a permutation σ_j such that

$$\mu_{\sigma_j(1)}(\alpha_j^\vee) \le \cdots \le \mu_{\sigma_j(N)}(\alpha_j^\vee).$$

We write $\mu_{\sigma_j(i)}(\alpha_j^\vee) = \lambda(\alpha_j^\vee) + \epsilon_i(j) + \delta_{i,N}\lambda^0(\alpha_j^\vee)$ for integers $\epsilon_i(j)$. By our assumptions we obtain

$$0 \le \epsilon_1(j) \le \cdots \le \epsilon_{N-1}(j) \le \epsilon_N(j) + \lambda^0(\alpha_j^\vee) \quad \text{and} \quad \sum_{p=1}^{N} \epsilon_p(j) = 0.$$

Hence, up to a permutation we have $\mu_i = \lambda$ for $1 \le i \le N-1$ and $\mu_N = \lambda + \lambda^0$.

4B. For the rest of this section we prove the conjecture for $\mathfrak{sl}_2$ and compute a PBW type basis. For $0 \le j < N$, let $S(k^{N-j}, (k+1)^j)$ be the set of tuples $(i_0, \dots, i_{N-1})$ satisfying

$$\sum_{p=0}^{N-1} \frac{N!}{N-p} i_p \le N!\,k - \sum_{\ell=0}^{N-4} \frac{N!}{(N-\ell)!}(N-\ell-2)!\,b_\ell + j(N-1)! \tag{4-3}$$

for integers b_ℓ defined as follows: $0 \le b_\ell < N - \ell$ and

$$\begin{aligned} &i_0 - j = b_0 \bmod N, \\ &i_\ell + (b_{\ell-1} \bmod N-\ell) = b_\ell \bmod N-\ell \quad \text{for } \ell = 1, \dots, N-4. \end{aligned}$$

The theorem we shall prove is the following.

Theorem 4.3. *Let $m \in \mathbb{Z}_+$ and write $m = kN + j$ for $0 \le j < N$.*

(1) *We have an isomorphism of $\boldsymbol{U}(\mathfrak{sl}_2 \otimes \mathbb{C}[t]/t^N)$-modules*

$$\mathrm{W}(m, N) \cong V(k)^{*(N-j)} * V(k+1)^{*j}.$$

(2) *A* PBW *type basis of* $\mathrm{W}(m, N)$ *is given by*

$$\{(x_{-\alpha} \otimes 1)^{i_0} \cdots (x_{-\alpha} \otimes t^{N-1})^{i_{N-1}} w_{m,N} \mid (i_0, \dots, i_{N-1}) \in S(k^{N-j}, (k+1)^j)\}.$$

A simple calculation similar to the one above shows that $(k, \dots, k, k+1, \dots, k+1) \in P^+(m, N)$ is in fact the unique maximal element.

The rest of this section is dedicated to the proof of Theorem 4.3.

4C. We start by proving the first part of the theorem. A presentation of the fusion product as a $\boldsymbol{U}(\mathfrak{sl}_2 \otimes \mathbb{C}[t])$ was given in [Chari and Venkatesh 2015]. So by their results it is enough to show that the highest weight vector of $\mathrm{W}(m, N)$ satisfies the defining relations of $V(k)^{*(N-j)} * V(k+1)^{*j}$ given in [Chari and Venkatesh 2015, Proposition 2.7], which are

$$\boldsymbol{x}_{-\alpha}(r, s) = \sum_{(b_p)_{p \ge 0} \in \boldsymbol{S}(r,s)} (x_{-\alpha} \otimes 1)^{b_0} (x_{-\alpha} \otimes t)^{b_1} \cdots (x_{-\alpha} \otimes t^s)^{b_s},$$
$$s, r, \ell \in \mathbb{N},\ r + s \ge 1 + r\ell + q + p,$$

where $q = \max\{0, (N-\ell)k\}$, $p = \max\{0, j-\ell\}$ and $\boldsymbol{S}(r, s)$ is the set of tuples $(b_p)_{p\geq 0}$ satisfying $b_0 + \cdots + b_s = r$ and $b_1 + 2b_2 + \cdots + sb_s = s$. We assume that $r+s \leq m$, because otherwise the claim follows from the following result of Garland [1978]:

$$(x_\alpha \otimes t)^{(s)}(x_{-\alpha} \otimes 1)^{(s+r)} - (-1)^s \boldsymbol{x}_{-\alpha}(r, s) \in \boldsymbol{U}(\mathfrak{g}[t])\mathfrak{n}^+[t].$$

Our aim is to prove that for any tuple $(b_p)_{p\geq 0} \in \boldsymbol{S}(r, s)$ there exists $p \geq N$ such that $b_p \neq 0$. Assume this is not the case. If $\ell \geq N$ we obtain

$$rN \geq r+s \geq 1+r\ell \geq 1+rN,$$

which is obviously a contradiction. So assume $l \leq N-1$. It follows that

$$m \geq r+s \geq 1+r\ell+(N-\ell)k+p = 1+\ell(r-k)+m-j+p$$

and thus $r \leq k$. Therefore we obtain the contradiction

$$1+\ell(r-k)+m-j+p \leq r+s \leq rN \Rightarrow 1 \leq (N-\ell)(r-k)-p.$$

Hence

$$\mathrm{W}(m, N) \cong V(k)^{*(N-j)} * V(k+1)^{*j}.$$

4D. Now we will prove the second part of the theorem. For simplicity we write f_i for $x_{-\alpha} \otimes t^i$, $1 \leq i \leq N-1$, and consider the map $\mathrm{sh} : \boldsymbol{U}(\mathfrak{n}^-[t]) \to \boldsymbol{U}(\mathfrak{n}^-[t])$ given by $\mathrm{sh}(f_i) = f_{i+1}$. We will need the following result from [Feigin and Feigin 2002].

Proposition 4.4. *Let $k_1 \leq k_2 \leq \cdots \leq k_N$. We have a short exact sequence of $\boldsymbol{U}(\mathfrak{n}^-[t])$-modules*

$$0 \to V(k_1)*\cdots*V(k_{N-1}) \xrightarrow{\mathrm{sh}} V(k_1)*\cdots*V(k_N) \xrightarrow{f_0^{-1}} V(k_1)*\cdots*V(k_N-1) \to 0.$$

Using this proposition one can construct inductively a PBW type basis of the fusion product. To be more precise, we have

$$B(k_1, \ldots, k_N) = B(k_1, \ldots, k_{N-1})_{\mathrm{sh}} \cup f_0 B(k_1, \ldots, k_N - 1), \tag{4-4}$$

where $B(\cdot)$ denotes a basis of the appropriate fusion product.

Example. We have $B(1, 2) = B(1)_{\mathrm{sh}} \cup f_0 B(1, 1)$ and hence

$$B(1, 2) = \{1, f_0\}_{\mathrm{sh}} \cup f_0\{1, f_0, f_0^2, f_1\} = \{1, f_1, f_0 f_1, f_0, f_0^2, f_0^3\}.$$

Lemma 4.5. *We have the recursion formula*

$$B(k^N) = \bigcup_{r=0}^{k} f_0^{Nr} B((k-r)^{N-1})_{\mathrm{sh}} \cup \bigcup_{j=1}^{N-1} \bigcup_{r=1}^{k} f_0^{Nr-j} B((k-r)^{N-j}, (k-r+1)^{j-1})_{\mathrm{sh}}.$$

Proof. The proof follows by repeated applications of (4-4); for the convenience of the reader we present the first step:

$$\begin{aligned}B(k^N) &= B(k^{N-1})_{\mathrm{sh}} \cup f_0 B((k-1)^1, k^{N-1})\\ &= B(k^{N-1})_{\mathrm{sh}} \cup f_0 B((k-1)^1, k^{n-2})_{\mathrm{sh}} \cup f_0^2 B((k-1)^2, k^{N-2})\\ &= \cdots = \bigcup_{r=0}^{N-1} f_0^r B((k-1)^r, k^{N-1-r})_{\mathrm{sh}} \cup f_0^N B((k-1)^N).\end{aligned}$$

The formula now follows by proceeding in the same way with $B((k-1)^N)$. □

Theorem 4.6. *A* PBW *type basis of the truncated Weyl module* $\mathrm{W}(kN, N)$ *is given by*

$$B(k^N) = \{f_0^{i_0} f_1^{i_1} \cdots f_{N-1}^{i_{N-1}} \mid (i_0, \dots, i_{N-1}) \in S(k^N)\}.$$

Example. (1) For $N = 1$ we get that $S(k)$ is the set of 1-tuples (i_0) satisfying

$$i_0 = \sum_{j=0}^{0} \frac{1!}{1-j} i_j \leq 1!\,k - \sum_{\ell=0}^{-3} \frac{1!}{(1-\ell)!}(1-\ell-2)!\,b_\ell = k,$$

so $S(k) = \{0, 1, \dots, k\}$ and $B(k) = \{f_0^j \mid j = 0, \dots, k\}$.

(2) For $N = 4$ and $k = 2$ we get that $S(2^4)$ is the set of quadruples (i_0, i_1, i_2, i_3) satisfying

$$6i_0 + 8i_1 + 12i_2 + 24i_3 \leq 48 - 2b_0,$$

where $i_0 = b_0 \bmod 4$ and

$$B(2^4) = \{f_0^{i_0} f_1^{i_1} f_2^{i_2} f_3^{i_3} \mid (i_0, i_1, i_2, i_3) \in S(2^4)\}.$$

Proof. The proof of Theorem 4.6 proceeds by upward induction on N. The initial step is obvious (see also the previous example) and the induction begins. So suppose that the theorem holds for all integers less than N.

Claim. For all $M < N$ we have

$$B(k^{M-j}, (k+1)^j) = \{f_0^{i_0} f_1^{i_1} \cdots f_{M-1}^{i_{M-1}} \mid (i_0, \dots, i_{M-1}) \in S(k^{M-j}, (k+1)^j)\}.$$

Proof of the claim. We use induction. There is nothing to prove if $j = 0$. Assuming $j > 0$, we obtain

$$\begin{aligned}&B(k^{M-j}, (k+1)^j)\\ &\quad= B(k^{M-j}, (k+1)^{j-1})_{\mathrm{sh}} \cup f_0 B(k^{M-j+1}, (k+1)^{j-1})\\ &\quad= \{f_0^{i_0} f_1^{i_1} \cdots f_{M-2}^{i_{M-2}} \mid (i_0, \dots, i_{M-2}) \in S(k^{M-j}, (k+1)^{j-1})\}_{\mathrm{sh}}\\ &\qquad\qquad \cup f_0\{f_0^{i_0} f_1^{i_1} \cdots f_{M-1}^{i_{M-1}} \mid (i_0, \dots, i_{M-1}) \in S(k^{M-j+1}, (k+1)^{j-1})\}.\end{aligned}$$

The shift by the map sh leads to the following description:

$$\{f_0^{i_0} f_1^{i_1} \cdots f_{M-2}^{i_{M-2}} \mid (i_0, \ldots, i_{M-2}) \in S(k^{M-j}, (k+1)^{j-1})\}_{\mathrm{sh}}$$
$$= \Bigg\{ f_1^{i_1} f_2^{i_2} \cdots f_{M-1}^{i_{M-1}} \mid \sum_{p=1}^{M-1} \frac{M!}{M-p} i_p \le M!\,k - \sum_{\ell=1}^{M-4} \frac{M!}{(M-\ell)!}(M-\ell-2)!\,b_\ell + M(j-1)(M-2)! \Bigg\}$$

with

$$i_1 - j + 1 = b_1 \mod M-1,$$
$$i_\ell + (b_{\ell-1} \mod M-\ell) = b_\ell \mod M-\ell \quad \text{for } \ell = 2, \ldots, M-4,$$

and

$$f_0\{f_0^{i_0} f_1^{i_1} \cdots f_{M-1}^{i_{M-1}} \mid (i_0, \ldots, i_{M-1}) \in S(k^{M-j+1}, (k+1)^{j-1})\}$$
$$= \Bigg\{ f_0^{i_0+1} f_1^{i_1} \cdots f_{M-1}^{i_{M-1}} \mid \sum_{p=0}^{M-1} \frac{M!}{M-p} i_p \le M!\,k - \sum_{\ell=0}^{M-4} \frac{M!}{(M-\ell)!}(M-\ell-2)!\,b_\ell + (j-1)(M-1)! \Bigg\}$$
$$= \Bigg\{ f_0^{i_0} f_1^{i_1} \cdots f_{M-1}^{i_{M-1}} \mid \sum_{p=0}^{M-1} \frac{M!}{M-p} i_p \le M!\,k - \sum_{\ell=0}^{M-4} \frac{M!}{(M-\ell)!}(M-\ell-2)!\,b_\ell + j(M-1)!\,,\ i_0 \ge 1 \Bigg\}$$

with

$$i_0 - j = b_0 \mod M,$$
$$i_\ell + (b_{\ell-1} \mod M-\ell) = b_\ell \mod M-\ell \quad \text{for } \ell = 1, \ldots, M-4.$$

Therefore, the claim follows with

$$\{f_0^{i_0} f_1^{i_1} \cdots f_{M-2}^{i_{M-2}} \mid (i_0, \ldots, i_{M-2}) \in S(k^{M-j}, (k+1)^{j-1})\}_{\mathrm{sh}}$$
$$= \Bigg\{ f_0^{0} f_1^{i_1} \cdots f_{M-1}^{i_{M-1}} \mid \sum_{p=0}^{M-1} \frac{M!}{M-p} i_p \le M!\,k - \sum_{\ell=0}^{M-4} \frac{M!}{(M-\ell)!}(M-\ell-2)!\,b_\ell + j(M-1)! \Bigg\}.$$

Now it is easy to verify with Lemma 4.5 that the theorem holds. □

The proof of Theorem 4.6 gives the following.

Corollary 4.7. *A* PBW *type basis of the truncated Weyl module* $\mathrm{W}(kN+j, N)$ *is given by*

$$\{f_0^{i_0} f_1^{i_1} \cdots f_{N-1}^{i_{N-1}} \mid (i_0, \ldots, i_{N-1}) \in S(k^{N-j}, (k+1)^j)\}.$$

Remark. The fusion product $V(1)^{*N}$ is isomorphic to the truncated Weyl module $W_{\mathrm{loc}}(N, N)$ and also to the local Weyl module $W_{\mathrm{loc}}(N)$. The inductively obtained basis $B(1^N)$ coincides with the basis of the Weyl module $W_{\mathrm{loc}}(N)$ constructed in [Chari and Pressley 2001]. However, we would like to emphasize that the PBW type basis of the truncated Weyl module $\mathrm{W}(m, N)$ described in Theorem 4.3 is different from the basis described in [Chari and Venkatesh 2015, Section 6]. For example, we have $f_1^3 \in B(1^4)$ but f_1^3 is not contained in their basis.

References

[Ardonne and Kedem 2007] E. Ardonne and R. Kedem, "Fusion products of Kirillov–Reshetikhin modules and fermionic multiplicity formulas", *J. Algebra* **308**:1 (2007), 270–294. MR 2008a:17028 Zbl 1122.17014

[Chari and Loktev 2006] V. Chari and S. Loktev, "Weyl, Demazure and fusion modules for the current algebra of $\mathfrak{sl}_{r+1}$", *Adv. Math.* **207**:2 (2006), 928–960. MR 2008a:17029 Zbl 1161.17318

[Chari and Pressley 2001] V. Chari and A. Pressley, "Weyl modules for classical and quantum affine algebras", *Represent. Theory* **5** (2001), 191–223. MR 2002g:17027 Zbl 0989.17019

[Chari and Venkatesh 2015] V. Chari and R. Venkatesh, "Demazure modules, fusion products and Q-systems", *Comm. Math. Phys.* **333**:2 (2015), 799–830. MR 3296163 Zbl 06394855

[Chari et al. 2014a] V. Chari, G. Fourier, and D. Sagaki, "Posets, tensor products and Schur positivity", *Algebra Number Theory* **8**:4 (2014), 933–961. MR 3248990 Zbl 06331627

[Chari et al. 2014b] V. Chari, B. Ion, and D. Kus, "Weyl modules for the hyperspecial current algebra", *Int. Math. Res. Not.* (online publication August 2014).

[Chari et al. 2014c] V. Chari, P. Shereen, R. Venkatesh, and J. Wand, "A Steinberg type decomposition theorem for higher level Demazure modules", preprint, 2014. arXiv 1408.4090

[Feigin 2008] E. Feigin, "The PBW filtration, Demazure modules and toroidal current algebras", *SIGMA Symmetry Integrability Geom. Methods Appl.* **4** (2008), Paper 070. MR 2010b:17029 Zbl 1215.17015

[Feigin and Feigin 2002] B. Feigin and E. Feigin, "Q-characters of the tensor products in $\mathfrak{sl}_2$-case", *Mosc. Math. J.* **2**:3 (2002), 567–588. MR 2004e:17002 Zbl 1027.05007

[Feigin and Loktev 1999] B. Feigin and S. Loktev, "On generalized Kostka polynomials and the quantum Verlinde rule", pp. 61–79 in *Differential topology, infinite-dimensional Lie algebras, and applications*, edited by A. Astashkevich and S. Tabachnikov, Amer. Math. Soc. Transl. Ser. 2 **194**, Amer. Math. Soc., Providence, RI, 1999. MR 2002b:17007 Zbl 0974.17008

[Feigin et al. 2004] B. Feigin, M. Jimbo, R. Kedem, S. Loktev, and T. Miwa, "Spaces of coinvariants and fusion product, II: $\widehat{\mathfrak{sl}}_2$ character formulas in terms of Kostka polynomials", *J. Algebra* **279**:1 (2004), 147–179. MR 2005m:17033 Zbl 1160.17312

[Fourier 2015] G. Fourier, "New homogeneous ideals for current algebras: filtrations, fusion products and Pieri rules", *Mosc. Math. J.* **15**:1 (2015), 49–72.

[Fourier and Kus 2013] G. Fourier and D. Kus, "Demazure modules and Weyl modules: the twisted current case", *Trans. Amer. Math. Soc.* **365**:11 (2013), 6037–6064. MR 3091275 Zbl 1287.17016

[Fourier and Littelmann 2006] G. Fourier and P. Littelmann, "Tensor product structure of affine Demazure modules and limit constructions", *Nagoya Math. J.* **182** (2006), 171–198. MR 2007e:17021 Zbl 1143.22010

[Fourier and Littelmann 2007] G. Fourier and P. Littelmann, "Weyl modules, Demazure modules, KR-modules, crystals, fusion products and limit constructions", *Adv. Math.* **211**:2 (2007), 566–593. MR 2008k:17005 Zbl 1114.22010

[Fourier et al. 2012] G. Fourier, T. Khandai, D. Kus, and A. Savage, "Local Weyl modules for equivariant map algebras with free abelian group actions", *J. Algebra* **350** (2012), 386–404. MR 2859894 Zbl 1268.17022

[Garland 1978] H. Garland, "The arithmetic theory of loop algebras", *J. Algebra* **53**:2 (1978), 480–551. MR 80a:17012 Zbl 0383.17012

[Kac 1990] V. G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge University Press, Cambridge, 1990. MR 92k:17038 Zbl 0716.17022

[Kedem 2011] R. Kedem, "A pentagon of identities, graded tensor products, and the Kirillov–Reshetikhin conjecture", pp. 173–193 in *New trends in quantum integrable systems*, edited by B. Feigin et al., World Sci. Publ., Hackensack, NJ, 2011. MR 2012g:17011 Zbl 1221.82037

[Kus and Venkatesh 2014] D. Kus and R. Venkatesh, "Twisted Demazure modules, fusion product decomposition and twisted Q-systems", preprint, 2014. arXiv 1409.1201

[Mathieu 1988] O. Mathieu, "Construction du groupe de Kac–Moody et applications", *C. R. Acad. Sci. Paris Sér. I Math.* **306**:5 (1988), 227–230. MR 89e:17013 Zbl 0666.17006

[Naoi 2012] K. Naoi, "Weyl modules, Demazure modules and finite crystals for non-simply laced type", *Adv. Math.* **229**:2 (2012), 875–934. MR 2855081 Zbl 1305.17009

[Ravinder 2014] B. Ravinder, "Demazure modules, Chari–Venkatesh modules and fusion products", *SIGMA Symmetry Integrability Geom. Methods Appl.* **10** (2014), Paper 110. MR 3298990 Zbl 06421406

Received November 26, 2014. Revised March 18, 2015.

DENIZ KUS
MATHEMATICAL INSTITUTE OF THE UNIVERSITY OF COLOGNE
D-50931 COLOGNE
GERMANY
dkus@math.uni-koeln.de

PETER LITTELMANN
MATHEMATICAL INSTITUTE OF THE UNIVERSITY OF COLOGNE
D-50931 COLOGNE
GERMANY
peter.littelmann@math.uni-koeln.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.278.447

DIFFERENTIAL HARNACK ESTIMATES FOR POSITIVE SOLUTIONS TO HEAT EQUATION UNDER FINSLER–RICCI FLOW

SAJJAD LAKZIAN

We prove first order differential Harnack estimates for positive solutions of the heat equation (in the sense of distributions) under closed Finsler–Ricci flows. We assume suitable Ricci curvature bounds throughout the flow and also assume that the *S*-curvature vanishes along the flow. One of the key tools we use is the Bochner identity for Finsler structures proved by Ohta and Sturm (*Adv. Math.* 252 (2014), 429–448).

1. Introduction

In the past few decades, geometric flows and, more notably among them, the Ricci flow have proved very useful in attacking long standing geometry and topology questions. One important application is finding the so-called round (of constant curvature, Einstein, soliton, etc.) metrics on manifolds by homogenizing a given initial metric.

There is also a hope that similar methods can be applied in the Finsler setting. One might hope to find an answer for, for instance, Professor Chern's question about the existence of Finsler–Einstein metrics on every smooth manifold by using a suitable geometric flow resembling the Ricci flow.

In the Finsler setting, there are notions of Ricci and sectional curvatures, and Bao [2007] has proposed an evolution of Finsler structures that in essence shares a great resemblance with the Ricci flow of Riemannian metrics. The flow Bao suggests is $\partial F^2/\partial t = -2F^2 R$ where $R = (1/F^2)\,\mathrm{Ric}$. In terms of the symmetric metric tensor associated with F and Akbarzadeh's Ricci tensor, this flow takes the form of $\partial g_{ij}/\partial t = -2\,\mathrm{Ric}_{ij}$ which is the familiar Ricci flow.

The notion of Finsler–Ricci flow is very recent and very little has been done about it. Some partial results regarding the existence and uniqueness of such flows

Lakzian is supported by the Hausdorff postdoctoral fellowship at the Hausdorff Center for Mathematics, University of Bonn, Germany.

MSC2010: primary 35K55; secondary 53C21.

Keywords: Finsler–Ricci flow, differential Harnack, gradient estimate, weighted Ricci curvature, heat equation, curvature-dimension.

are obtained in [Azami and Razavi 2013]. Also, the solitons of this flow have been studied in [Bidabad and Yarahmadi 2014]. Our focus in these notes will be to consider a positive solution of the heat equation (in the sense of distributions) under Finsler–Ricci flow and prove first order differential Harnack estimates that are similar to those in the Riemannian case (see [Liu 2009; Sun 2011]). The key tools we use are the Bochner identity for Finsler metrics (pointwise and in the sense of distributions) proven by Ohta and Sturm [2014] and, as is customary in such estimates, the maximum principle.

We should mention that, in this paper, we are not dealing with the existence and Sobolev regularity of such solutions (which is very important and extremely delicate — for example, in the static case, solutions will be C^2 if and only if the structure is Riemannian). For existence and regularity in the static case see [Ohta and Sturm 2009]. Our main theorem is the following.

Theorem 1.1. *Let $(M^n, F(t))$, $t \in [0, T]$ be a closed Finsler–Ricci flow. Suppose there is a real number $K \in \mathbb{R}$ and positive real numbers K_1 and K_2 such that, for all $t \in [0, T]$,*

(i) $-K_1 \le (\mathrm{Ric}_{ij}(\boldsymbol{v}))_{i,j=1}^n \le K_2$ *as quadratic forms on T_xM for all $\boldsymbol{v} \in T_xM \setminus \{0\}$, in any coordinate system, $\{\partial/\partial x_i\}$, that is orthonormal with respect to $g_{\boldsymbol{v}}$, and*

(ii) ***S**-curvature vanishes (see Section 2.2.7).*

Let $u(x,t) \in L^2([0,T], H^1(M)) \cap H^1([0,T], H^{-1}(M))$ be a positive global solution (*in the sense of distributions*) *of the heat equation under Finsler–Ricci flow*; *i.e., for any test function $\phi \in C^\infty(M)$ and for all $t \in [0, T]$,*

$$\int_M \phi \partial_t u(t,\cdot)\, dm = -\int_M D\phi(\nabla u(t,\cdot))\, dm\, dt. \tag{1}$$

Then, u satisfies

$$F^2(\nabla(\log u)(t,x)) - \theta \partial_t (\log u)(t,x) \le \frac{n\theta^2}{t} + \frac{n\theta^3 C_1}{\theta - 1} + n^{3/2}\theta^2\sqrt{C_2}, \tag{2}$$

for any $\theta > 1$ and where

$$C_1 = K_1 \quad \textit{and} \quad C_2 = \max\{K_1^2, K_2^2\}. \tag{3}$$

Remark 1.2. Our results can be applied to any Finsler–Ricci flow of Berwald metrics on closed manifolds, since the ***S***-curvature vanishes for Berwald metrics (for example, see [Ohta 2011]).

We will note that it might be possible to obtain stronger results with fewer curvature bound conditions by using different methods such as Nash–Moser iteration (as is done by Xia [2014] for harmonic functions in the static case).

Integrating the differential Harnack inequalities, in a standard manner, leads to Harnack-type inequalities.

Corollary 1.3. *Let $(M, F(t))$, $t \in [0, T]$ be as in Theorem 1.1. Then for any two points $(x, t_1), (y, t_2) \in M \times (0, T]$ with $t_1 < t_2$, we get*

(4)
$$u(x, t_1) \le u(y, t_2)\left(\frac{t_2}{t_1}\right)^{2n\epsilon} \exp\left\{\int_0^1 \frac{\epsilon F^2(\gamma'(s))|_\tau}{2(t_2-t_1)}\, ds + C(n, \epsilon)(t_2-t_1)(C_1+\sqrt{C_2})\right\},$$

whenever $\epsilon > 1/2$, and for C depending on n and ϵ only, and where the dependencies of C_1 and C_2 on our parameters are as in Theorem 1.1. Here γ is a curve joining x and y, with $\gamma(1) = x$ and $\gamma(0) = y$, and $F(\gamma'(s))|_\tau$ is the speed of γ at time $\tau = (1-s)t_2 + st_1$.

The organization of this paper is as follows: in Section 2, we first briefly review some facts and results about differential Harnack estimates in the Riemannian setting and about Finsler geometry; in Section 3, we present lemmas and computations that we need in order to obtain a useful parabolic partial differential inequality; and in Section 4, we will complete the proof of our main theorem.

2. Background

2.1. ***Differential Harnack estimates for heat equations in Riemannian Ricci flow.*** The Ricci flow equation, $\partial g/\partial t = -2\,\mathrm{Ric}$, was first proposed by Richard Hamilton in his seminal paper [1982]. Ricci flow is a heat-type quasilinear partial differential equation but, as is well-known, it enjoys a short-time existence and uniqueness theorem (see [Hamilton 1982]) and has been the key tool in proving the Poincaré and geometrization conjectures.

The gradient estimates for solutions of parabolic equations under Ricci flow are a very important part of Ricci flow theory. Perelman in his groundbreaking work [2002] proves such estimates for the conjugate heat equation; he then benefited from these estimates in the analysis of his $\mathcal{W}$-entropy functional. Since then there have been many important results in this direction (for both heat equation and conjugate heat equation) in, for example, [Kuang and Zhang 2008; Bailesteanu et al. 2010; Cao et al. 2013; Cao and Hamilton 2009; Cao 2008], to name a few.

Since our proof, in spirit, is closer to ones in Liu [2009] and Sun [2011], we will only mention their result without commenting on the other literature in this direction. Their estimates for positive solutions of the heat equation under a closed Ricci flow can be stated as follows.

Theorem [Liu 2009; Sun 2011]. *Let $(M, g(t))$; $t \in [0, T]$ be a closed Ricci flow solution with $-K_1 \le \mathrm{Ric} \le K_2$ $(K_1, K_2 > 0)$ along the flow. For $u(x, t)$, a positive solution of the heat equation $(\Delta_{g(t)} - \partial_t)u(x, t) = 0$, one has the first order* *gradient estimate*

$$\frac{|\nabla u(x,t)|^2}{u^2(x,t)} - \theta \frac{\partial_t u(x,t)}{u(x,t)} \leq \frac{n\theta^2}{t} + \frac{n\theta^3 K_1}{\theta - 1} + n^{\frac{3}{2}}\theta^2 (K_1 + K_2), \tag{5}$$

where $\theta > 1$.

Their method of proof is to take $f = \log u$ and

$$\alpha := t\left(\frac{|\nabla u(x,t)|^2}{u^2(x,t)} - \theta \frac{\partial_t u(x,t)}{u(x,t)}\right) = t(|\nabla f|^2 - \theta \partial_t f) \tag{6}$$

and apply the maximum principle to the parabolic partial differential inequality

$$(\Delta_{g(t)} - \partial_t)\alpha + 2Df(\nabla\alpha) \geq -\frac{\alpha}{t} + \frac{t}{n}(|\nabla f|^2 - \partial_t f)^2 - 2\theta K_1 t |\nabla f|^2 - t\theta^2 n^2 (K_1 + K_2)^2.$$

This is the method that we will adopt throughout the paper.

2.2. *Finsler structures.*

2.2.1. *Finsler metric.* Let M be a C^∞-connected manifold. A Finsler structure on M consists of a C^∞ Finsler norm $F : TM \to \mathbb{R}$ satisfying the following conditions:

(F1) F is C^∞ on $TM \setminus 0$.

(F2) F restricted to the fibers is positively 1-homogeneous.

(F3) For any nonzero tangent vector $\boldsymbol{y} \in TM$, the approximated symmetric metric tensor defined by

$$g_{\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v}) := \frac{1}{2} \frac{\partial^2}{\partial s \partial t} F^2(\boldsymbol{y} + s\boldsymbol{u} + t\boldsymbol{v})|_{s=t=0} \tag{7}$$

is positive definite.

2.2.2. *Cartan tensor.* One way to measure the nonlinearity of a Finsler structure is to introduce the so-called *Cartan tensor* defined by

$$C_{\boldsymbol{y}} : \otimes^3 TM \to \mathbb{R}, \quad C_{\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) := \frac{1}{2} \frac{d}{dt}[g_{\boldsymbol{y}+t\boldsymbol{w}}(\boldsymbol{u}, \boldsymbol{v})]. \tag{8}$$

2.2.3. *Legendre transform.* In order to define the *gradient* of a function, we need the Legendre transform, $\mathcal{L}^* : T^*M \to TM$. For $\omega \in T^*M$, let $\mathcal{L}^*(\omega)$ be the unique vector $\boldsymbol{y} \in TM$ such that

$$\omega(\boldsymbol{y}) = F^*(\omega)^2 \quad \text{and} \quad F(\boldsymbol{y}) = F^*(\omega), \tag{9}$$

where F^* is the dual norm to F.

For a smooth function $u : M \to \mathbb{R}$, the gradient of u is $\nabla u(x) := \mathcal{L}^*(Du(x))$.

2.2.4. *Geodesic spray, Chern connection and curvature tensor.* It is easy to see that the geodesic spray in the Finsler setting is of the form $\boldsymbol{G} = y^i \partial/\partial x_i - 2G^i(x, \boldsymbol{y})\partial/\partial y^i$, where

$$G^i(x, \boldsymbol{y}) = \frac{1}{4} g_{\boldsymbol{y}}^{ik} \left\{ 2 \frac{\partial (g_{\boldsymbol{y}})_{jk}}{\partial x_l} - \frac{\partial (g_{\boldsymbol{y}})_{jl}}{\partial x_k} \right\} y^j y^l. \tag{10}$$

The nonlinear connection that we will be using in this work is the Chern connection, the connection coefficients of which are given by

$$\Gamma^i_{jk} = \Gamma^i_{kj} := \frac{1}{2} g^{il} \left\{ \frac{\partial g_{lj}}{\partial x_k} - \frac{\partial g_{jk}}{\partial x_l} + \frac{\partial g_{kl}}{\partial x_j} - \frac{\partial g_{lj}}{\partial y^r} G^r_k + \frac{\partial g_{jk}}{\partial y^r} G^r_l - \frac{\partial g_{kl}}{\partial y^r} G^r_j \right\}, \tag{11}$$

where $G^i_j := \partial G^i / \partial y^j$ and g is in fact $g_{\boldsymbol{y}}$.

For Berwald metrics, the geodesic coefficients G^i are quadratic in terms of y (by definition) which immensely simplifies the formula for connection coefficients. In fact for Berwald metrics we have $\Gamma^i_{jk} = \partial^2 G^i / \partial y^j \partial y^k$.

Similar to the Riemannian setting, one uses the Chern connection (and the associated covariant differentiation) to define the curvature tensor

$$R^V(X, Y)Z := [\nabla^V_X, \nabla^V_Y]Z - \nabla^V_{[X,Y]} Z, \tag{12}$$

which, of course, depends on a nonzero vector field V.

2.2.5. *Flag and Ricci curvatures.* *Flag curvature* is defined similar to the sectional curvature in the Riemannian setting. For a fixed flag pole $\boldsymbol{v} \in T_x M$ and for $\boldsymbol{w} \in T_x M$, the flag curvature is defined by

$$\mathcal{K}^{\boldsymbol{v}}(\boldsymbol{v}, \boldsymbol{w}) := \frac{g_{\boldsymbol{v}}(R^{\boldsymbol{v}}(\boldsymbol{v}, \boldsymbol{w})\boldsymbol{w}, \boldsymbol{v})}{g_{\boldsymbol{v}}(\boldsymbol{v}, \boldsymbol{v}) g_{\boldsymbol{v}}(\boldsymbol{w}, \boldsymbol{w}) - g_{\boldsymbol{v}}(\boldsymbol{v}, \boldsymbol{w})^2}. \tag{13}$$

The Ricci curvature is then the trace of the flag curvature, i.e.,

$$\mathrm{Ric}(\boldsymbol{v}) := F^2(\boldsymbol{v}) \sum_{i=1}^{n-1} \mathcal{K}^{\boldsymbol{v}}(\boldsymbol{v}, \boldsymbol{e}_i), \tag{14}$$

where $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}, \frac{\boldsymbol{v}}{F(\boldsymbol{v})}\}$ constitutes a $g_{\boldsymbol{v}}$-orthonormal basis of $T_x M$.

2.2.6. *Akbarzadeh's Ricci tensor.* Akbarzadeh's Ricci tensor is defined by

$$\mathrm{Ric}_{ij} := \frac{\partial^2}{\partial y^i \partial y^j} \left(\frac{\mathrm{Ric}}{2} \right). \tag{15}$$

It can be shown that the scalar Ricci curvature, Ric, and Akbarzadeh's Ricci tensor, Ric_{ij}, have the same geometrical implications. For further details regarding this tensor, see [Bao and Robles 2004].

2.2.7. ***S**-curvature.* Associated with any Finsler structure, there is one canonical measure, called the Busemann–Hausdorff measure, which is given by

$$dV_F := \sigma_F(x)\,dx_1 \wedge \cdots \wedge dx_n, \tag{16}$$

where $\sigma_F(x)$ is the volume ratio

$$\sigma_F(x) := \frac{\mathrm{vol}(B_{\mathbb{R}^n}(1))}{\mathrm{vol}(\boldsymbol{y} \in T_xM : F(\boldsymbol{y}) < 1)}. \tag{17}$$

The set whose volume appears in the denominator of (17) is called the indicatrix, and there is often no known way to express its volume in terms of F.

The $\boldsymbol{S}$-curvature, which is another measure of nonlinearity, is then defined by

$$\boldsymbol{S}(\boldsymbol{y}) := \frac{\partial G^i}{\partial y^i}(x, \boldsymbol{y}) - y^i \frac{\partial}{\partial x_i}(\ln \sigma_F(x)). \tag{18}$$

For more details, see [Shen 2004], for example.

2.2.8. *Hessian, divergence and Laplacian.* The Hessian in a Finsler structure is defined by

$$\mathrm{Hess}(u)(X, Y) := XY(u) - \nabla_X^{\nabla u} Y(u) = g_{\nabla u}(\nabla_X^{\nabla u} \nabla u, Y). \tag{19}$$

As usual, for a twice differentiable function u,

$$\mathrm{Hess}(u)\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right) = \frac{\partial^2 u}{\partial x_i \partial x_j} - \Gamma_{ij}^k \frac{\partial u}{\partial x_k} \tag{20}$$

For a smooth measure $\mu = e^{-\Psi} dx_1 \wedge \cdots \wedge dx_n$ and a vector field V, the divergence is defined by

$$\mathrm{div}_\mu V := \sum_{i=1}^n \left(\frac{\partial V_i}{\partial x_i} - V_i \frac{\partial \Psi}{\partial x_i}\right). \tag{21}$$

Now, using this divergence, one can define the distributional Laplacian of a function $u \in H^1(M)$ by $\Delta u := \mathrm{div}_\mu(\nabla u)$, i.e.,

$$\int_M \phi \Delta u \, d\mu := -\int_M D\phi(\nabla u)\, d\mu, \tag{22}$$

for $\phi \in C^\infty(M)$.

The Finsler distributional Laplacian is nonlinear but fortunately there is a way to relate it to the trace of the Hessian by adding an $\boldsymbol{S}$-curvature term. Indeed, one has

$$\Delta u = \mathrm{tr}_{\nabla u} \mathrm{Hess}(u) - \boldsymbol{S}(\nabla u). \tag{23}$$

For a proof of (23), see for instance [Wu and Xin 2007].

2.3. ***Weighted Ricci curvature and Bochner–Weitzenböck formula.*** The notion of the *weighted Ricci curvature*, Ric_N, of a Finsler structure equipped with a measure μ was introduced by Ohta [2009]. Take a unit vector $\boldsymbol{v} \in T_xM$ and let $\gamma : [-\epsilon, +\epsilon] \to M$ be a short geodesic whose velocity at time $t = 0$ is $\dot\gamma(0) = \boldsymbol{v}$. Decompose the measure μ along γ with respect to the Riemannian volume form; i.e., let $\mu = e^{-\Psi} d\mathrm{vol}_{\dot\gamma}$. Then

$$\mathrm{Ric}_n(\boldsymbol{v}) := \begin{cases} \mathrm{Ric}(\boldsymbol{v}) + (\Psi \circ \gamma)''(0) & \text{if } (\Psi \circ \gamma)'(0) = 0, \\ -\infty & \text{otherwise,} \end{cases} \tag{24}$$

$$\mathrm{Ric}_N(\boldsymbol{v}) := \mathrm{Ric}(\boldsymbol{v}) + (\Psi \circ \gamma)''(0) - \frac{(\Psi \circ \gamma)'(0)^2}{N - n} \quad \text{when } n < N < \infty, \tag{25}$$

$$\mathrm{Ric}(\boldsymbol{v}) := \mathrm{Ric}(\boldsymbol{v}) + (\Psi \circ \gamma)''(0). \tag{26}$$

Also $\mathrm{Ric}_N(\lambda \boldsymbol{v}) := \lambda^2 \,\mathrm{Ric}_N(\boldsymbol{v})$ for $\lambda \geq 0$.

It is proven in [Ohta 2009] that the curvature bound $\mathrm{Ric}_N \geq KF^2$ is equivalent to the Lott–Villani–Sturm $CD(K, N)$ condition.

Using the weighted Ricci curvature bounds, Ohta and Sturm [2014] proved the Bochner–Weitzenböck formulae (both pointwise and integrated versions) for Finsler structures. For $u \in C^\infty(M)$, the pointwise version of the identity and inequality are

$$\Delta^{\nabla u}\left(\frac{F^2(\nabla u)}{2}\right) - D(\Delta u)(\nabla u) = \mathrm{Ric}_\infty(\nabla u) + \|\nabla^2 u\|^2_{HS(\nabla u)} \quad \text{(identity)}, \tag{27}$$

$$\Delta^{\nabla u}\left(\frac{F^2(\nabla u)}{2}\right) - D(\Delta u)(\nabla u) \geq \mathrm{Ric}_N(\nabla u) + \frac{(\Delta u)^2}{N} \quad \text{(inequality)}. \tag{28}$$

3. Estimates

In this section we will gather all the required lemmas and estimates that will be needed to apply the maximum principle.

Evolution of the Legendre transform. Since in the Finsler setting the gradient is nonlinear and depends on the Legendre transform, we will need to know the evolution of the Legendre transform under Finsler–Ricci flow.

Let (M, F) be a Finsler structure evolving under Finsler–Ricci flow. Then the inverse of the Legendre transform is defined by

$$(\mathcal{L}^*)^{-1} : TM \to T^*M, \quad (\mathcal{L}^*)^{-1}(x, y) = (x, p), \quad \text{where } p_i = g_{ij}(x, y) y^j. \tag{29}$$

To explicitly formulate the Legendre transform, we have, for any given $\omega \in T_x^*M$, that $\mathcal{L}^*(\omega) = y \in T_xM$, where y is the unique solution to the *nonlinear* system

$$g(x, y)_{k1} \cdot y^1 + \cdots + g(x, y)_{kn} \cdot y^n = \omega_k, \quad \text{for } k = 1, \ldots, n, \tag{30}$$

or, in the matrix form,

$$g(\boldsymbol{y})\boldsymbol{y} = \omega. \tag{31}$$

Lemma 3.1. *Let $(M, F(t))$ be a Finsler structure evolving by Finsler–Ricci flow. Then the Legendre transform $\mathcal{L}^* : T^*M \to TM$ satisfies*

$$\partial_t \mathcal{L}^* = 2\,\mathrm{Ric}^i_j\,\mathcal{L}^*; \tag{32}$$

i.e., for any fixed 1-form ω with $\mathcal{L}^*(\omega) = \boldsymbol{y} = y^i\,\partial/\partial x_i \in TM$, we have

$$\partial_t y^i = 2\,\mathrm{Ric}^i_r\,y^r, \tag{33}$$

where $\mathrm{Ric}^i_r := g^{ij}\,\mathrm{Ric}_{jr}$.

Proof. Fix ω and differentiate both sides of (31) with respect to t to get

$$[\partial_t g(\boldsymbol{y})]\boldsymbol{y} + g(\boldsymbol{y})\,\partial_t \boldsymbol{y} = 0. \tag{34}$$

Therefore,

$$\partial_t \boldsymbol{y} = -g(\boldsymbol{y})^{-1}\,\partial_t g(\boldsymbol{y})\boldsymbol{y}. \tag{35}$$

Expanding the right-hand side of (35), we have, for every i,

$$\begin{aligned} \partial_t y^i &= -g(\boldsymbol{y})^{ij}(\partial_t g(\boldsymbol{y}))_{jr}y^r \\ &= 2g(\boldsymbol{y})^{ij}\,\mathrm{Ric}_{jr}(\boldsymbol{y})y^r - g(\boldsymbol{y})^{ij}\left(\frac{\partial g_{jr}}{\partial y^k}\partial_t y^k\right)y^r \\ &= 2\,\mathrm{Ric}^i_r(\boldsymbol{y})y^r. \end{aligned} \tag{36}$$

Notice that the second term in the second line of (36) vanishes by Euler's theorem. □

Evolution of $F^2(\nabla f)$. One crucial step in the proof of the gradient estimates is to be able to estimate the evolution of the term $F^2(\nabla f)$.

Lemma 3.2. *Let $(M, F(t))$ be a time-dependent Finsler structure. Then*

$$\partial_t[F^2(\nabla f)] = 2g^{ij}(Df)[\partial_t f]_i f_j + [\partial_t g^{ij}](Df) f_i f_j. \tag{37}$$

Proof. Simple differentiation gives

$$\begin{aligned} \partial_t[F^2(\nabla f)] &= \partial_t[F^*(Df)^2] \\ &= \partial_t[g^{ij}(Df) f_i f_j] \\ &= 2g^{ij}(Df)[\partial_t f]_i f_j + \partial_t[g^{ij}(Df)] f_i f_j. \end{aligned} \tag{38}$$

Expanding the second term of the last line in (38), we have

$$\partial_t[g^{ij}(Df)] f_i f_j = [\partial_t g^{ij}](Df) f_i f_j + \frac{\partial g^{ij}}{\partial y^k}\,\partial_t y^k(Df) f_i f_j. \tag{39}$$

Using Euler's theorem, the second term of the right-hand side of (39) vanishes. □

Lemma 3.3. *Suppose F is evolving by the Finsler–Ricci flow equation. Then*

$$\partial_t[F^2(\nabla f)] = 2D(\partial_t f)(\nabla f) + 2\,\mathrm{Ric}^{ij}(Df) f_i f_j. \tag{40}$$

Proof. It is standard to see that under Finsler–Ricci flow, we have

$$\partial_t g^{ij} = 2\,\mathrm{Ric}^{ij}, \tag{41}$$

where, as before, $\mathrm{Ric}^{ij} := g^{ir} g^{js}\,\mathrm{Ric}_{rs}$. □

4. Proof of main theorem

In this section we will complete the proof of our main theorem. Throughout the rest of these notes, we consider a solution u of the heat equation. The Laplacian, gradient and Legendre transform are all with respect to $V := \nabla u$ and are valid on $M_u := \{x \in M : \nabla u(x) \neq 0\}$.

Let $\sigma(t,x) = t\partial_t f(t,x)$ where $f = \log u$. Then we have $g_{\nabla f} = g_V$. Let

$$\alpha(t,x) := t\{F^2(\nabla f(t,x)) - \theta\,\partial_t f(t,x)\} = tF^2(\nabla f(t,x)) - \theta\sigma. \tag{42}$$

Lemma 4.1. *In the sense of distributions, $\sigma(t,x)$ satisfies the parabolic differential equality*

$$\Delta\sigma - \partial_t\sigma + \frac{\sigma}{t} + 2D\sigma(\nabla f) = t\{-2\,\mathrm{Ric}^{ij}(\nabla f) f_i f_j - 2(\mathrm{Ric})^{kl}(\nabla f) f_{kl}\}. \tag{43}$$

Proof. We first note that, for any nonnegative test function $\phi \in H^1([0,T]\times M)$ whose support is included in the domain of the local coordinate,

$$\partial_t(D(t\phi)(\nabla f)) = D(\partial_t(t\phi))(\nabla f) + D(t\phi)(\nabla(\partial_t f)) + 2(\mathrm{Ric})^{ij}(\nabla f)\frac{\partial(t\phi)}{\partial x_i}\frac{\partial f}{\partial x_j}. \tag{44}$$

Indeed,

$$\begin{aligned}
&\partial_t(D(t\phi)(\nabla f)) \\
&\quad= D(\partial_t(t\phi))(\nabla f) + D(t\phi)(\partial_t(\mathcal{L}^*(Df)) \\
&\quad= D(\partial_t(t\phi))(\nabla f) + D(t\phi)(\partial_t(\mathcal{L}^*)(Df) + \mathcal{L}^*(D\partial_t f)) \\
&\quad= D(\partial_t(t\phi))(\nabla f) + D(t\phi)(\partial_t(\mathcal{L}^*)(Df)) + D(t\phi)(\mathcal{L}^*(D\partial_t f)) \\
&\quad= D(\partial_t(t\phi))(\nabla f) + D(t\phi)(\nabla(\partial_t f)) + 2g^{sj}(\mathrm{Ric})^i_s(\nabla f)\frac{\partial(t\phi)}{\partial x_i}\frac{\partial f}{\partial x_j} \\
&\quad= D(\partial_t(t\phi))(\nabla f) + D(t\phi)(\nabla(\partial_t f)) + 2(\mathrm{Ric})^{ij}(\nabla f)\frac{\partial(t\phi)}{\partial x_i}\frac{\partial f}{\partial x_j}.
\end{aligned} \tag{45}$$

That is,

(46)
$$-D(t\phi)(\nabla(\partial_t f)) = -\partial_t(D(t\phi)(\nabla f)) + D(\partial_t(t\phi))(\nabla f) + 2(\mathrm{Ric})^{ij}(\nabla f)\frac{\partial(t\phi)}{\partial x_i}\frac{\partial f}{\partial x_j}.$$

Multiplying the left-hand side of (43) by ϕ, integrating and then substituting (46), we get

$$\begin{aligned}
\text{(47)}\quad \boldsymbol{A} &= \int_0^T\int_M \{-D\phi(\nabla\sigma) + \partial_t\phi\cdot\sigma + \frac{\phi\sigma}{t} + 2\phi D\sigma(\nabla f)\}\,dm\,dt \\
&= \int_0^T\int_M \{-D(t\phi)(\nabla(\partial_t f)) + \partial_t(t\phi)\partial_t f + 2t\phi D(\partial_t f)(\nabla f)\}\,dm\,dt \\
&= \int_0^T\int_M \Big\{D(\partial_t(t\phi))(\nabla f) + \partial_t(t\phi)(\Delta f + F^2(\nabla f)) \\
&\qquad + 2(\mathrm{Ric})^{ij}(\nabla f)\frac{\partial(t\phi)}{\partial x_i}\frac{\partial f}{\partial x_j} + 2t\phi D(\partial_t f)(\nabla f)\Big\}\,dm\,dt.
\end{aligned}$$

Using the estimates we have obtained for $\partial_t[F(\nabla f)^2]$ in Lemmas 3.2 and 3.3, we arrive at

$$\begin{aligned}
\text{(48)}\quad \boldsymbol{A} &= \int_0^T\int_M \Big\{D(\partial_t(t\phi))(\nabla f) + \partial_t(t\phi)(\Delta f) + \partial_t(t\phi)(F^2(\nabla f)) \\
&\qquad + 2(\mathrm{Ric})^{ij}(\nabla f)\frac{\partial(t\phi)}{\partial x_i}\frac{\partial f}{\partial x_j} + 2t\phi D(\partial_t f)(\nabla f)\Big\}\,dm\,dt \\
&= \int_0^T\int_M \Big\{\partial_t(t\phi)(F^2(\nabla f)) + 2(\mathrm{Ric})^{ij}(\nabla f)\frac{\partial(t\phi)}{\partial x_i}\frac{\partial f}{\partial x_j} + t\phi\partial_t[F(\nabla f)^2] \\
&\qquad - 2t\phi\,\mathrm{Ric}^{ij}(\nabla f) f_i f_j\Big\}\,dm\,dt \\
&= \int_0^T\int_M t\phi\{-2\,\mathrm{Ric}^{ij}(\nabla f) f_i f_j - 2\,\mathrm{Ric}^{ij}(\nabla f) f_{ij}\}\,dm\,dt.
\end{aligned}$$

Notice that Euler's theorem has been used in the last line of (48). □

Now we can compute a parabolic partial differential inequality for $\alpha(t,x)$ with a similar left-hand side.

Lemma 4.2. *In the sense of distributions, $\alpha(t,x)$ satisfies*

$$\Delta^V\alpha + 2D\alpha(\nabla f) - \partial_t\alpha + \frac{\alpha}{t} = \boldsymbol{B}, \tag{49}$$

where

$$\begin{aligned}
\boldsymbol{B} &= \theta(2t\,\mathrm{Ric}^{ij}(\nabla f) f_i f_j + 2t\,\mathrm{Ric}^{kl}(\nabla f) f_{kl}) \\
&\qquad + 2t\,\mathrm{Ric}(\nabla f) + 2t\|\nabla^2 f\|^2_{HS(\nabla f)} - 2t\,\mathrm{Ric}^{ij}(\nabla f) f_i f_j.
\end{aligned}$$

Proof. For a nonnegative test function ϕ, one computes

$$\begin{aligned}
(50)\quad &\int_0^T\int_M \Big\{-D\phi(\nabla\alpha)+\partial_t\phi\alpha+\frac{\phi\alpha}{t}+2\phi D\alpha(\nabla f)\Big\}\,dm\,dt\\
&= -\theta\boldsymbol{A}+\int_0^T\int_M \big\{-tD\phi\big(\nabla(F^2(\nabla f))\big)+\partial_t\phi(tF^2(\nabla f))\\
&\qquad\qquad +\phi(F^2(\nabla f))+2t\phi D(F^2(\nabla f))(\nabla f)\big\}\,dm\,dt\\
&= -\theta\boldsymbol{A}+\int_0^T\int_M \big\{-tD\phi\big(\nabla(F^2(\nabla f))\big)-\phi\cdot\partial_t\big(t(F^2(\nabla f))\big)\\
&\qquad\qquad +\phi(F^2(\nabla f))+2t\phi D(F^2(\nabla f))(\nabla f)\big\}\,dm\,dt\\
&= -\theta\boldsymbol{A}+\int_0^T\int_M \big\{-tD\phi\big(\nabla(F^2(\nabla f))\big)-\phi\cdot\big(F^2(\nabla f)+t\partial_t(F^2(\nabla f))\big)\\
&\qquad\qquad +\phi(F^2(\nabla f))+2t\phi D(F^2(\nabla f))(\nabla f)\big\}\,dm\,dt\\
&= -\theta\boldsymbol{A}+\int_0^T\int_M \big\{-tD\phi\big(\nabla(F^2(\nabla f))\big)-\phi\cdot t\partial_t(F^2(\nabla f))\\
&\qquad\qquad +2t\phi D(F^2(\nabla f))(\nabla f)\big\}\,dm\,dt,
\end{aligned}$$

where $\boldsymbol{A}$ is as in (48).

Again using the estimates for $\partial_t[F(\nabla f)^2]$ (as in Lemmas 3.2 and 3.3), we arrive at

$$\begin{aligned}
(51)\quad &\int_0^T\int_M \Big\{-D\phi(\nabla\alpha)+\partial_t\phi\cdot\alpha+\frac{\phi\alpha}{t}+2\phi D\alpha(\nabla f)\Big\}\,dm\,dt\\
&= -\theta\boldsymbol{A}+\int_0^T\int_M \big\{-tD\phi\big(\nabla(F^2(\nabla f))\big)-\phi\cdot t\partial_t(F^2(\nabla f))\\
&\qquad\qquad +2t\phi D(F^2(\nabla f))(\nabla f)\big\}\,dm\,dt\\
&= -\theta\boldsymbol{A}+\int_0^T\int_M \big\{-tD\phi\big(\nabla(F^2(\nabla f))\big)-2t\phi D(\partial_t f)(\nabla f)\\
&\qquad\qquad -2t\phi\,\mathrm{Ric}^{ij}(\nabla f)f_i f_j+2t\phi D(F^2(\nabla f))(\nabla f)\big\}\,dm\,dt\\
&= -\theta\boldsymbol{A}+\int_0^T\int_M \big\{-tD\phi\big(\nabla(F^2(\nabla f))\big)-2t\phi D(\Delta f)(\nabla f)\\
&\qquad\qquad -2t\phi D(F^2(\nabla f))(\nabla f)-2t\phi\,\mathrm{Ric}^{ij}f_i f_j\\
&\qquad\qquad +2t\phi D(F^2(\nabla f))(\nabla f)\big\}\,dm\,dt\\
&= -\theta\boldsymbol{A}+\int_0^T\int_M \big\{-tD\phi\big(\nabla(F^2(\nabla f))\big)\\
&\qquad\qquad -2t\phi D(\Delta f)(\nabla f)-2t\phi\,\mathrm{Ric}^{ij}(\nabla f)f_i f_j\big\}\,dm\,dt.
\end{aligned}$$

By applying the Bochner–Weitzenböck formula (proven in [Ohta and Sturm 2014]; see also Section 2.3) and noticing that $\boldsymbol{S}=0$ implies $\mathrm{Ric}_\infty(\boldsymbol{v})=\mathrm{Ric}(\boldsymbol{v})$, we can continue as follows:

$$\begin{aligned}-\theta\boldsymbol{A}+\int_0^T\int_M\big\{-tD\phi\big(\nabla(F^2(\nabla f))\big)-2t\phi D(\Delta f)(\nabla f)-2t\phi\,\mathrm{Ric}^{ij}f_if_j\big\}\,dm\,dt\\ =-\theta\boldsymbol{A}+\int_0^T\int_M\phi\big\{2t\,\mathrm{Ric}(\nabla f)+2t\|\nabla^2 f\|^2_{HS(\nabla f)}-2t\,\mathrm{Ric}^{ij}(\nabla f)f_if_j\big\}\,dm\,dt.\end{aligned}$$

Now, substituting $\boldsymbol{A}$ from (47), we have

$$\begin{aligned}\boldsymbol{B}=\theta(2t\,\mathrm{Ric}^{ij}(\nabla f)f_if_j+2t\,\mathrm{Ric}^{kl}(\nabla f)f_{kl})+2t\,\mathrm{Ric}(\nabla f)\\ +2t\|\nabla^2 f\|^2_{HS(\nabla f)}-2t\,\mathrm{Ric}^{ij}(\nabla f)f_if_j.\quad\square\end{aligned}$$

Proof of Theorem 1.1. Assume the curvature bounds given in the statement of Theorem 1.1, and assume that the $\boldsymbol{S}$-curvature vanishes. The constants obtained below all depend on our curvature bounds and the ellipticity of the flow.

Let's start with $B(t,x)$:

$$\begin{aligned}B(t,x)=\theta(2t\,\mathrm{Ric}^{ij}(\nabla f)f_if_j+2t\,\mathrm{Ric}^{kl}(\nabla f)f_{kl})+2t\,\mathrm{Ric}(\nabla f)\\ +2t\|\nabla^2 f\|^2_{HS(\nabla f)}-2t\,\mathrm{Ric}^{ij}(\nabla f)f_if_j.\end{aligned}$$

Young's inequality tells us that

$$|\,\mathrm{Ric}^{kl}f_{kl}|\le\frac{\theta}{2}(\mathrm{Ric}^{kl})^2+\frac{1}{2\theta}f_{kl}^2,\tag{52}$$

and therefore

$$2\theta t|\,\mathrm{Ric}^{kl}f_{kl}|\le t\theta^2(\mathrm{Ric}^{kl})^2+tf_{kl}^2.\tag{53}$$

Pick a normal coordinate system with respect to $g_{\nabla f}$, with $\nabla f(x)=\partial/\partial x_1$ as well as $\Gamma^1_{ij}(\nabla f(x))=0$ for all i, j. Then

$$\mathrm{Ric}^{ij}(\nabla f)=\mathrm{Ric}_{ij}(\nabla f),\qquad\|\nabla^2 f\|^2_{HS(\nabla f)}=\sum f_{ij}^2,\qquad\sum_{i=1}^n f_{ii}=\Delta f(x),\tag{54}$$

and consequently

$$\begin{aligned}B(t,x)\ge 2t\theta\,\mathrm{Ric}_{ij}(\nabla f)f_if_j-t\sum\theta^2(\mathrm{Ric}_{kl})^2-t\sum f_{kl}^2\\ +2t\,\mathrm{Ric}(\nabla f)+2t\|\nabla^2 f\|^2_{HS(\nabla f)}-2t\,\mathrm{Ric}_{ij}(\nabla f)f_if_j\\ \ge-2t\theta K_1F^2(\nabla f)-2tK_1F^2(\nabla f)+t\sum f_{ij}^2\\ -t\theta^2n^2C_2+2tK_1F^2(\nabla f).\end{aligned}\tag{55}$$

On the other hand, one computes

$$\sum f_{ij}^2 \geq \sum f_{ii}^2 \geq \frac{1}{n}\Big(\sum f_{ii}\Big)^2 = \frac{1}{n}(\Delta f)^2. \tag{56}$$

Hence,

$$t \sum f_{ij}^2 \geq \frac{t}{n}(\Delta f)^2. \tag{57}$$

Putting all the above estimates together and noting that $\theta > 1$, we get

$$\begin{aligned} B(t,x) &\geq \frac{t}{n}(\Delta f)^2 - 2t\theta K_1 F^2(\nabla f) - 2t K_1 F^2(\nabla f) - t\theta^2 n^2 C_2 + 2t K_1 F^2(\nabla f) \\ &\geq \frac{t}{n}(\Delta f)^2 - 2t\theta K_1 F^2(\nabla f) - t\theta^2 n^2 C_2. \end{aligned}$$

Replacing the term Δf with $(F(\nabla f)^2 - \partial_t f)$, we get the inequality

$$B(t,x) \geq \frac{t}{n}(F(\nabla f)^2 - \partial_t f)^2 - 2t\theta C_1 F^2(\nabla f) - t\theta^2 n^2 C_2, \tag{58}$$

where

$$C_1 = K_1, \tag{59}$$

$$C_2 = \max\{K_1^2, K_2^2\}. \tag{60}$$

This means that

$$\begin{aligned} &\Delta^V \alpha + 2D\alpha(\nabla f) - \partial_t \alpha \\ &\qquad \geq -\frac{\alpha}{t} + \frac{t}{n}(F(\nabla f)^2 - \partial_t f)^2 - 2t\theta C_1 F^2(\nabla f) - t\theta^2 n^2 C_2. \end{aligned} \tag{61}$$

This inequality is exactly of the form that appears in [Liu 2009], and a computation similar to the one at the end of the proof of [Liu 2009, Theorem 2] (using the quadratic formula and maximum principle) gives the desired result. For the sake of clarity, we will repeat the computation here.

Let

$$\bar{\alpha} := \alpha - t\frac{n\theta^3 C_1}{(\theta - 1)} - t n^{3/2}\theta^2\sqrt{C_2}. \tag{62}$$

Suppose the maximum of $\bar{\alpha}$ is attained at (x_0, t_0) and suppose $\bar{\alpha}(x_0, t_0) > n\theta^2$ (which implicitly implies $t_0 > 0$). Therefore, at (x_0, t_0), we have

$$0 \geq (\Delta - \partial_t)\bar{\alpha} \geq (\Delta - \partial_t)\alpha. \tag{63}$$

Let $w := F^2(\nabla f)$ and $z := \partial_t f$. Then in terms of w and z we have

$$0 \geq -\frac{\alpha}{t_0} + \frac{t_0}{n}(w - z)^2 - 2t_0\theta C_1 w - t_0\theta^2 n^2 C_2. \tag{64}$$

By the quadratic formula, we get

$$
\begin{aligned}
(65)\quad & \frac{t_0}{n}(w-z)^2 - 2t_0\theta C_1 w \\
&= \frac{t_0}{n}\left(\frac{1}{\theta^2}(w-\theta z)^2 + \left(\frac{\theta-1}{\theta}\right)^2 w^2 - 2\theta n C_1 w + 2\left(\frac{\theta-1}{\theta^2}w\right)(w-\theta z)\right) \\
&\geq \frac{t_0}{n}\left(\frac{1}{\theta^2}(w-\theta z)^2 - \frac{\theta^4 n^2 C_1^2}{(\theta-1)^2} + 2\left(\frac{\theta-1}{\theta^2}w\right)(w-\theta z)\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
(66)\quad 0 &\geq \frac{t_0}{n\theta^2}\left(\frac{\alpha}{t_0}\right)^2 - \frac{\alpha}{t_0} - \frac{n\theta^4 C_1^2}{(\theta-1)^2}t_0 - t_0\theta^2 n^2 C_2 + \frac{2t_0}{n}\frac{\theta-1}{\theta^2}F^2(\nabla f)\left(\frac{\alpha}{t_0}\right) \\
&\geq \frac{t_0}{n\theta^2}\left(\frac{\alpha}{t_0}\right)^2 - \frac{\alpha}{t_0} - \frac{n\theta^4 C_1^2}{(\theta-1)^2}t_0 - t_0\theta^2 n^2 C_2.
\end{aligned}
$$

Using the quadratic formula one more time, (66) implies that

$$
(67)\qquad \frac{\alpha}{t_0} \leq \frac{n\theta^2}{t_0} + \frac{n\theta^3 C_1}{\theta-1} + n^{\frac{3}{2}}\theta^2\sqrt{C_2},
$$

which in turn implies

$$
(68)\qquad \bar{\alpha}(x_0, t_0) \leq n\theta^2,
$$

and this is a contradiction. Therefore,

$$
(69)\qquad F^2(\nabla(\log u)(t,x)) - \theta\partial_t(\log u)(t,x) \leq \frac{n\theta^2}{t} + \frac{n\theta^3 C_1}{(\theta-1)} + n^{3/2}\theta^2\sqrt{C_2},
$$

with C_1 and C_2 as in (59) and (60). □

Proof of Corollary 1.3. From Theorem 1.1, we know that

$$
(70)\qquad F^2(\nabla(\log u)(t,x)) - \theta\partial_t(\log u)(t,x) \leq \frac{n\theta^2}{t} + C(n,\theta)(C_1 + \sqrt{C_2}).
$$

Let $l(s) := \ln u(\gamma(s), \tau(s)) = f(\gamma(s), \tau(s))$. Then

$$
\begin{aligned}
(71)\quad \frac{\partial l(s)}{\partial s} &= (t_2 - t_1)\left(\frac{Df(\dot\gamma(s))}{t_2-t_1} - \partial_t f\right) \\
&\leq (t_2 - t_1)\left(\frac{F(\nabla f)F(\dot\gamma)}{t_2-t_1} - \partial_t f\right) \\
&\leq (t_2 - t_1)\left(\frac{\epsilon F^2(\dot\gamma)|_\tau}{2(t_2-t_1)^2} + \frac{1}{2\epsilon}F^2(\nabla f) - \partial_t f\right) \\
&\leq \frac{\epsilon F^2(\dot\gamma)|_\tau}{2(t_2-t_1)} + (t_2-t_1)\left(\frac{2n\epsilon}{\tau} + C(n,\epsilon)(C_1 + \sqrt{C_2})\right).
\end{aligned}
$$

Integrating this inequality gives

$$\begin{aligned}\ln\frac{u(x,t_1)}{u(y,t_2)} &= \int_0^1 \frac{\partial l(s)}{\partial s}\,ds\\ &\le \int_0^1 \frac{\epsilon F^2(\dot{\gamma})|_\tau}{2(t_2-t_1)}\,ds + C(n,\epsilon)(t_2-t_1)(C_1+\sqrt{C_2}) + 2\epsilon n\ln\frac{t_2}{t_1}. \qquad \square\end{aligned}$$

Acknowledgements

The author would like to thank Professor K.-T. Sturm for the opportunity of working under his supervision as a postdoctoral fellow and for his great insights. The author also would like to thank Professor Christina Sormani for her constant support and encouragement. Many thanks go to the stochastic analysis group in Bonn and also my colleagues Mike Munn and Lashi Bandara.

The author is deeply grateful for valuable comments made and numerous corrections suggested by the anonymous reviewer of this paper.

References

[Azami and Razavi 2013] S. Azami and A. Razavi, "Existence and uniqueness for solution of Ricci flow on Finsler manifolds", *Int. J. Geom. Methods Mod. Phys.* **10**:3 (2013), Article ID #1250091. MR 3011654 Zbl 06160396

[Bailesteanu et al. 2010] M. Bailesteanu, X. Cao, and A. Pulemotov, "Gradient estimates for the heat equation under the Ricci flow", *J. Funct. Anal.* **258**:10 (2010), 3517–3542. MR 2011b:53153 Zbl 1193.53139

[Bao 2007] D. Bao, "On two curvature-driven problems in Riemann–Finsler geometry", pp. 19–71 in *Finsler geometry: in memory of Makoto Matsumoto* (Sapporo, 2005), edited by S. V. Sabau and H. Shimada, Adv. Stud. Pure Math. **48**, Math. Soc. Japan, Tokyo, 2007. MR 2009d:53102 Zbl 1147.53018

[Bao and Robles 2004] D. Bao and C. Robles, "Ricci and flag curvatures in Finsler geometry", pp. 197–259 in *A sampler of Riemann–Finsler geometry*, edited by D. Bao et al., Math. Sci. Res. Inst. Publ. **50**, Cambridge University Press, 2004. MR 2005k:53124 Zbl 1076.53093

[Bidabad and Yarahmadi 2014] B. Bidabad and M. Yarahmadi, "On quasi-Einstein Finsler spaces", *Bull. Iranian Math. Soc.* **40**:4 (2014), 921–930. MR 3255407

[Cao 2008] X. Cao, "Differential Harnack estimates for backward heat equations with potentials under the Ricci flow", *J. Funct. Anal.* **255**:4 (2008), 1024–1038. MR 2009e:35121 Zbl 1146.58014

[Cao and Hamilton 2009] X. Cao and R. S. Hamilton, "Differential Harnack estimates for time-dependent heat equations with potentials", *Geom. Func. Anal.* **19** (2009), 989–1000. MR 2010j:53124 Zbl 1183.53059

[Cao et al. 2013] X. Cao, B. Fayyazuddin Ljungberg, and B. Liu, "Differential Harnack estimates for a nonlinear heat equation", *J. Funct. Anal.* **265**:10 (2013), 2312–2330. MR 3091816 Zbl 1285.58009

[Hamilton 1982] R. S. Hamilton, "Three-manifolds with positive Ricci curvature", *J. Differential Geom.* **17**:2 (1982), 255–306. MR 84a:53050 Zbl 0504.53034

[Kuang and Zhang 2008] S. Kuang and Q. S. Zhang, "A gradient estimate for all positive solutions of the conjugate heat equation under Ricci flow", *J. Funct. Anal.* **255**:4 (2008), 1008–1023. MR 2009h:53150 Zbl 1146.58017

[Liu 2009] S. Liu, "Gradient estimates for solutions of the heat equation under Ricci flow", *Pacific J. Math.* **243**:1 (2009), 165–180. MR 2010g:53122 Zbl 1180.58017

[Ohta 2009] S.-I. Ohta, "Finsler interpolation inequalities", *Calc. Var. Partial Differential Equations* **36**:2 (2009), 211–249. MR 2011m:58027 Zbl 1175.49044

[Ohta 2011] S.-I. Ohta, "Vanishing $\mathbb{S}$-curvature of Randers spaces", *Differential Geom. Appl.* **29**:2 (2011), 174–178. MR 2012b:53156 Zbl 1215.53067

[Ohta and Sturm 2009] S.-I. Ohta and K.-T. Sturm, "Heat flow on Finsler manifolds", *Comm. Pure Appl. Math.* **62**:10 (2009), 1386–1433. MR 2010j:58058 Zbl 1176.58012

[Ohta and Sturm 2014] S.-I. Ohta and K.-T. Sturm, "Bochner–Weitzenböck formula and Li–Yau estimates on Finsler manifolds", *Adv. Math.* **252** (2014), 429–448. MR 3144236 Zbl 06284328

[Perelman 2002] G. Perelman, "The entropy formula for the Ricci flow and its geometric applications", preprint, 2002. Zbl 1130.53001 arXiv math/0211159

[Shen 2004] Z. Shen, "Landsberg curvature, S-curvature and Riemann curvature", pp. 303–355 in *A sampler of Riemann–Finsler geometry*, edited by D. Bao et al., Math. Sci. Res. Inst. Publ. **50**, Cambridge University Press, 2004. MR 2005k:53132 Zbl 1074.53063

[Sun 2011] J. Sun, "Gradient estimates for positive solutions of the heat equation under geometric flow", *Pacific J. Math.* **253**:2 (2011), 489–510. MR 2878821 Zbl 1235.53070

[Wu and Xin 2007] B. Y. Wu and Y. L. Xin, "Comparison theorems in Finsler geometry and their applications", *Math. Ann.* **337**:1 (2007), 177–196. MR 2007h:53121 Zbl 1111.53060

[Xia 2014] C. Xia, "Local gradient estimate for harmonic functions on Finsler manifolds", *Calc. Var. Partial Differential Equations* **51**:3-4 (2014), 849–865. MR 3268873 Zbl 1316.53083

Received May 8, 2014. Revised March 10, 2015.

SAJJAD LAKZIAN
HAUSDORFF CENTER FOR MATHEMATICS
UNIVERSITÄT BONN
VILLA MARIA
ENDENICHER ALLEE 62
D-53115 BONN
GERMANY

lakzians@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.278.463

ON THE ONE-ENDEDNESS OF GRAPHS OF GROUPS

NICHOLAS TOUIKAN

We give a technical result that implies a straightforward necessary and sufficient condition for a graph of groups with virtually cyclic edge groups to be one-ended. For arbitrary graphs of groups, we show that if their fundamental group is not one-ended, then we can blow up vertex groups to graphs of groups with simpler vertex and edge groups. As an application, we generalize a theorem of Swarup to decompositions of virtually free groups.

1. Introduction

A finitely generated group $G = \langle S \rangle$ is said to be *one-ended* if the corresponding Cayley graph $\mathrm{Cay}(G, S)$ cannot be separated into two or more infinite components by removing a finite subset. Otherwise G is said to be *many-ended*. It is a classical result due to Stallings [1971] that a many-ended group decomposes as either an amalgamated free product or an HNN extension over a finite group.

Given the Bass–Serre correspondence between group actions on simplicial trees and their decompositions, or splittings, as (fundamental groups of) graphs of groups (see [Serre 1980]), a finitely generated group G is many-ended if and only if it acts minimally, without inversions, and cocompactly on a simplicial tree T in which for some edge e the stabilizer G_e is finite.

It is often the case that a graph of groups with many-ended vertex groups is itself one-ended. For example, the fundamental group of a closed surface is one-ended but it is an amalgamated free product of free groups, which are many-ended. Theorem 3.1, stated and proved in Section 3, essentially characterizes one-ended graphs of groups. This result is rather technical, but has many "nontechnical" corollaries which we now present.

We say that G is *one-ended relative to a collection $\mathcal{H}$ of subgroups* if for any minimal nontrivial G-tree T with finite edge stabilizers, there exists a subgroup $H \in \mathcal{H}$ that acts without a global fixed point. Otherwise G is said to be *many-ended relative to $\mathcal{H}$*. In this case, G admits a nontrivial splitting as a graph of groups relative to $\mathcal{H}$ (i.e., groups in $\mathcal{H}$ are conjugate into vertex groups) with finite edge groups.

MSC2010: primary 20E06, 20E08; secondary 57M07, 57M60, 57M20.

Keywords: group theory, groups acting on trees, Bass–Serre theory, ends of groups, square complexes.

Corollary 1.1. *If G_1 is one-ended relative to a collection $\mathcal{H}_1 \cup \{C_1\}$, and G_2 is one-ended relative to $\mathcal{H}_2 \cup \{C_2\}$ with $C_1 \approx C_2$ virtually cyclic groups, then any free product with amalgamation of the form*

$$G_1 *_{C_1=C_2} G_2$$

is one-ended relative to $\mathcal{H}_1 \cup \mathcal{H}_2$.

In the case of graphs of free groups with cyclic edge groups, this corollary (actually its natural generalization, see Corollary 1.5) is proved in [Wilton 2012, Theorem 18] and implied by results in [Diao and Feighn 2005]. Corollary 1.1 is false if we do not require the amalgamating subgroups to be virtually cyclic or, synonymously, two-ended. Nonetheless, we can still understand the failure of one-endedness of general graphs of groups.

Definition 1.2. A G-equivariant map $S \to T$ of simplicial G-trees is called a *collapse* if T is obtained by identifying some edge orbits of S to points. In this case we also say that S is obtained from T by a *blow up*. We call the preimage $\check{T}_v \subset S$ of a vertex $v \in T$ its *blowup*.

Definition 1.3. We write $H \preccurlyeq G$ to signify that G splits essentially as a graph of groups with finite edge groups and H is a vertex group. A group G is *accessible* if it admits no infinite proper chains

$$G \succ G_1 \succ G_2 \succ \dots.$$

For example, if F is a free group and $H \preccurlyeq F$, then H is a free factor of F. This next theorem, a formal consequence of Theorem 3.1, states that if a graph of groups with finitely generated infinite edge group is not one-ended, then we can blow up some of its vertex groups.

Theorem 1.4. *Suppose that T is a G-tree (in which a collection of subgroups $\mathcal{H}$ act elliptically) with infinite edge groups, and that G is not one-ended relative to $\mathcal{H}$. Then there is a vertex $v \in$ Vertices(T) and an edge $e \in$ Edges(T) with $v \in e$ such that the orbit of v can be blown up with G_v acting minimally on the nontrivial blowups $\check{T}_v$ satisfying the following properties:*

- *$G_e \leq G_v$ is the stabilizer of a vertex in $\check{T}_v$.*
- *The edge groups of $\check{T}_v$ are conjugate in G_v to the vertex groups of an essential amalgamated free product or HNN decomposition of G_e with a finite edge group.*

In particular, in the tree S obtained by blowing up the orbit of v in T to $\check{T}_v$, each vertex or edge stabilizer of S is $\preccurlyeq$ a vertex or edge stabilizer of T, and at least one of these inclusions is strict. Furthermore the groups in $\mathcal{H}$ act elliptically on S.

We note that blowing up a G-tree is equivalent to *refining* a graph of groups. If G acts on a tree with accessible vertex and edge stabilizers then the order $\prec$ actually tells us that the vertex groups of the blowup given by Theorem 1.4 have lower complexity, in the sense that the process of successively blowing up vertex groups in this manner must terminate in finitely many steps.

Accessible groups, in turn, are abundant. Linnell [1983] showed that if there is a global bound on the order of finite order elements in a finitely generated group, then the group is accessible. Dunwoody [1985] showed that finitely presented groups are accessible. We now use Theorem 1.4 to give a proof of Corollary 1.1.

Proof of Corollary 1.1. We show the contrapositive. Let T be the Bass–Serre tree dual to the splitting $G = G_1 *_C G_2$, and suppose that G is not one-ended relative to $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$. Note that any decomposition of a virtually cyclic group as an HNN extension or an essential amalgamated free product must have finite edge groups. It follows that in all cases, by Theorem 1.4, some orbit of vertices Gv can be blown up to minimal gG_vg^{-1}-trees with finite edge groups. This implies that one of the vertex groups G_i fixing some vertex $v \in \text{Vertices}(T)$ acts minimally on $\check{T}_v$ with finite edge stabilizers, with

$$\mathcal{H}_i = \{H \in \mathcal{H} \mid H \cap G_i \neq \{1\}\}$$

and $C_i = G_e$ for some $v \in e \in \text{Edges}(T)$ acting elliptically. It follows that G_i is not one-ended relative to $\mathcal{H}_i \cup \{C\}$. □

This proof is easily adapted to give:

Corollary 1.5. *The fundamental group G of a graph of groups with two-ended edge groups is one-ended* (*relative to a collection $\mathcal{H}$ of subgroups*) *if and only if every vertex group G_v is one-ended relative to the incident edge groups* (*and the collection* $\{H^g \cap G_v \mid g \in G, H \in \mathcal{H}\}$).

Using the full strength of Theorem 3.1, we also generalize a result of Swarup [1986] on the decomposition of free groups to virtually free groups. This result was already partially generalized by Cashen [2012] to decompositions of virtually free groups with virtually cyclic edge groups.

Theorem 1.6. *Let G be finitely generated and virtually free.*

(1) *If G splits as an amalgamated free product $G = A *_C B$ with C finitely generated and infinite*, *then there is some $C_1 \preccurlyeq C$ such that $C_1 \preccurlyeq A$ or $C_1 \preccurlyeq B$.*

(2) *If G splits as an HNN extension $G = A *_{C,t}$ with C finitely generated and infinite*, *then there is an infinite subgroup $C_1 \preccurlyeq C$ and a splitting Δ of A as a graph of groups with finite edge groups relative to $\{C_1, t^{-1}C_1t\}$ such that either C_1 or $t^{-1}C_1t$ is a vertex group of Δ.*

Unlike in Swarup's proof, we do not use homological methods. Our proof is more along the lines of the geometric arguments found in [Wilton 2012; Louder 2008; Bestvina and Feighn 1994; Diao and Feighn 2005] using graphs of spaces X with $\pi_1(X) = G$. The presence of torsion, however, can make the attaching maps in the graphs of spaces difficult to describe. By using the more abstract G-cocompact core of the product of two G-trees [Guirardel 2005], we sidestep these difficulties. The core has been used before to study pairs of group splittings. In particular, Fujiwara and Papasoglu [2006] use it to show the existence of QH subgroups for one-ended groups that have hyperbolic-hyperbolic pairs of slender splittings; this is the main technicality in constructing group theoretical JSJ decompositions. Although it could be noted that the action of our group on the core gives rise to a G-orbihedron à la [Haefliger 1991], we will not need this machinery; in fact, modulo classical Bass–Serre theory and Guirardel's Core Theorem for simplicial trees (Theorem 2.3, of which we sketch a proof), our argument is self-contained.

2. Preliminaries

Group actions. All group actions will be from the left. Let X be a G-set. If $S \subset X$ is a subset, we denote by G_S the (setwise) stabilizer $\{g \in G \mid gS = S\}$. If $S = \{x\}$ is a singleton, then we write G_x instead of $G_{\{x\}}$. We call a subset $S \subset X$ *G-regular* if for any $x, y \in S$ in the same G-orbit, there is some $g \in G_S$ such that $gx = y$. The following lemma is immediate.

Lemma 2.1. *Let X be a G-set. If $S \subset X$ is G-regular, then we have an embedding*

$$G_S \backslash S \hookrightarrow G \backslash X.$$

In this paper, all trees will be simplicial. In particular we consider them to be topological spaces, equipped with a CW-structure, which also makes them into graphs. We further metrize these graphs by viewing edges as real intervals of length 1.

All G-trees T will be *without inversions*, meaning that for any edge $e \in \mathrm{Edges}(T)$, if $ge = e$ then g fixes e pointwise. Equivalently, if $u, v \in \mathrm{Vertices}(T)$ are the vertices at the ends of the edge e, then we have inclusions

$$G_u \geq G_e \leq G_v.$$

We call vertex stabilizers *vertex groups*, and edge stabilizers *edge groups*. We assume the reader is familiar with Bass–Serre theory and we switch freely between G-trees and splittings as graphs of groups, viewing the two as being equivalent.

A G-tree T is *essential* if every edge of T divides it into two infinite components. T is *minimal* if there are no proper subtrees $S \subset T$ with $G_S = G$. T is *cocompact* if $G \backslash T$ is compact. An element g or a subgroup H of G are said to *act elliptically on T* if the groups $\langle g \rangle$ or H fix some $v \in \mathrm{Vertices}(T)$.

Products of trees, cores, and leaf spaces. If T_1 and T_2 are G-trees, then we have a natural induced action $G \curvearrowright T_1 \times T_2$. Since the trees T_1, T_2 are 1-dimensional CW complexes, we may consider their product $T_1 \times T_2$ as a *square complex*, i.e., a 2-dimensional CW complex whose cells consist of vertices, edges, and squares. There are natural projections $p_i : T_1 \times T_2 \twoheadrightarrow T_i$. The following lemma is immediate.

Lemma 2.2. *If the actions $G \curvearrowright T_1$ and $G \curvearrowright T_2$ are without inversions, then so is the action $G \curvearrowright T_1 \times T_2$, i.e., if $\sigma \supset \epsilon$ is an inclusion of cells (e.g., a square containing an edge), then $G_\sigma \leq G_\epsilon$.*

If $\mathcal{H}$ is a collection of subgroups acting elliptically on T_1 and T_2, then each subgroup in $\mathcal{H}$ fixes a vertex of $T_1 \times T_2$.

The action $G \curvearrowright T_1 \times T_2$ is not cocompact in general. It turns out, however, that we can extract a useful subset, namely Guirardel's cocompact core. We state the special case of his result applied to simplicial trees.

Theorem 2.3 (the Core Theorem, see [Guirardel 2005, Théorème principal and Corollaire 8.2]). *Let $G \curvearrowright T_1$, $G \curvearrowright T_2$ be two minimal actions of a finitely generated group G on simplicial trees T_1, T_2 with finitely generated edge stabilizers. Suppose furthermore that T_1, T_2 do not equivariantly collapse to a common nontrivial tree.*

Then there is a G-invariant subset $\mathcal{C} \subset T_1 \times T_2$, called the core of the action $G \curvearrowright T_1 \times T_2$, which is defined as the smallest connected G-invariant subset such that the restrictions of the projections $p_i|_{\mathcal{C}} : \mathcal{C} \twoheadrightarrow T_i$ have connected fibers. The quotient $\mathcal{S} = G\backslash\mathcal{C}$ is compact.

Suppose for the remainder of this section that T_1, T_2 satisfy the hypotheses of Theorem 2.3. The restrictions of the projections $p_i|_\sigma : \sigma \to T_i$ are well defined for each cell (i.e., a vertex, edge, square) $\sigma \subset T_1 \times T_2$. If σ is a square then the projection is onto an edge $p_i(\sigma) \in \mathrm{Edges}(T)_i$. If $\lambda_1, \lambda_2 \subset \sigma$ are two fibers of such a projection (see Figure 1), we can define a distance $d_i^\sigma(\lambda_1, \lambda_2)$ to be the distance

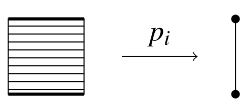

Figure 1. The projection of a square on an edge and some of its fibers.

in $p_i(\sigma)$ between the points $p_i(\lambda_1)$ $p_i(\lambda_2)$, thus putting a metric d_i^σ on the set of p_i-fibers in a cell σ. We now define the *i-leaf space* $\mathcal{L}_i$ of a subset $Z \subset T_1 \times T_2$ to be the set of connected unions of p_i-fibers of cells in Z, called *leaves*, so that we see Z as being *foliated* by the leaves in $\mathcal{L}_i$. $\mathcal{L}_i$ is a 1-complex with metrized edges; therefore we can endow $\mathcal{L}_i$ with the path metric d_i. As a consequence of the direct product structure we have the following.

Lemma 2.4. *If $Z \subset T_1 \times T_2$, then the leaf spaces $\mathcal{L}_i$ are forests* (see Figure 2).

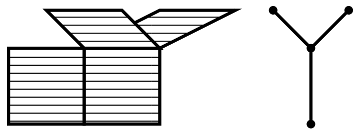

Figure 2. The i-leaves in a square complex and the resulting leaf space, which is a tree.

If $\mathscr{C} \subset T_1 \times T_2$ is a core then the leaf spaces $\mathscr{L}_i$ are homeomorphic to the trees T_i. Later, however, we will be performing operations that will alter the leaf spaces.

Induced splittings. Let $v \in \mathrm{Vertices}(T_i)$, $e \in \mathrm{Edges}(T_i)$ and let m_e be the midpoint of e. Let $\tau_v = p_i^{-1}(\{v\}) \cap \mathscr{C}$ and $\tau_e = p_i^{-1}(\{m_e\}) \cap \mathscr{C}$. By Theorem 2.3 the preimages τ_v, τ_e are connected and are therefore leaves in $\mathscr{L}_i$.

Since we have an action $G \curvearrowright \mathscr{C}$, since τ_v, τ_e are defined as T_i-point preimages via a G-equivariant map, and since G_v, G_e are exactly the stabilizers of these points v, m_e, the subsets $\tau_v, \tau_e \leq \mathscr{C}$ are G-regular. So, by Lemma 2.1 we have embeddings

$$G_v \backslash \tau_v \hookrightarrow G \backslash \mathscr{C} \hookleftarrow G_e \backslash \tau_e.$$

By Theorem 2.3, $G \backslash \mathscr{C}$ is compact so the quotients $G_v \backslash \tau_v$, $G_v \backslash \tau_v$ must be as well. Moreover, because τ_v, τ_e are contained in p_i-fibers, for $j \neq i$ the restrictions

$$p_j|_{\tau_v} : \tau_v \to T_j, \quad p_j|_{\tau_e} : \tau_e \to T_j$$

are injective. Finally, the projection $p_j|_{\mathscr{C}} : \mathscr{C} \twoheadrightarrow T_j$ is G-equivariant. We have shown the following.

Lemma 2.5. *If $v \in \mathrm{Vertices}(T_i)$, $e \in \mathrm{Edges}(T_i)$, $j \neq i$, then the fibers τ_v, τ_e are mapped injectively via p_j to subtrees that are G_v, G_e-invariant, respectively. Viewed as subsets of the core $\mathscr{C} \subset T_1 \times T_2$, τ_v and τ_e coincide with their j-leaf spaces.*

The actions $G_e \curvearrowright \tau_e$, $G_v \curvearrowright \tau_v$ are cocompact. Moreover τ_v, τ_e are infinite if and only if the actions of the subgroups $G_v \curvearrowright T_j$, $G_e \curvearrowright T_j$ are without global fixed points.

The G_v, G_e-trees τ_v, τ_e give splittings induced by the action on T_j. The blowups of Theorem 3.1 will be obtained by modifying the trees τ_v. For aficionados of CAT(0) cube complexes, it is worth remarking that the core $\mathscr{C}$ is a CAT(0) square complex, in fact a $\mathscr{V}\mathscr{H}$-complex, and that the set of fibers τ_e, $e \in \mathrm{Edges}(T_i)$ is the set of hyperplanes.

Spurs, free faces, and cleavings. In the previous subsection we obtained cocompact G_v, G_e-trees τ_v, τ_e. We say a tree has a *spur* if it has a vertex of degree 1. An edge adjacent to a spur is called a *hair*. We now give a shaving process.

Lemma 2.6. *Let T be a cocompact G-tree. T is minimal if T doesn't have any spurs. If T is not minimal, then we can obtain the minimal subtree $T(G)$ as the final term of a finite sequence*

$$T = T_0, \dots, T_k = T(G),$$

where T_{i+1} is obtained from T_i by G-equivariantly contracting one G-orbit of hairs to points.

Proof. Let $v \in \mathrm{Vertices}(T)$ be a spur adjacent to an edge $e \in \mathrm{Edges}(T)$ and let $u \in \mathrm{Vertices}(T)$ be the other endpoint of e. The map $T \to T$ obtained by G-equivariantly collapsing ge onto gu for $g \in G$ is a deformation retraction onto a proper G-invariant subtree, so T is not minimal.

Suppose now that T is not minimal. Then there is some proper G-invariant subtree $S \subset T$. Let K be the closure of some connected component of $T \setminus S$. Then $K \cap S = \{v\}$ for some $v \in \mathrm{Vertices}(S)$. Since S is G-invariant and connected, we must have $G_K \leq G_v$. It follows that for any $w \in \mathrm{Vertices}(K)$ and any $g \in G_K$ the distance $d_T(w, v) = d_T(gw, v)$, i.e., the action of G_K on K is the action on a rooted tree with root v. Since K is G-regular, we have an embedding $G_K \backslash K \hookrightarrow G \backslash T$ which is compact; thus K must have finite radius since G_K preserves distances from the root.

Since K is a rooted tree with finite diameter it must have a nonroot vertex of valence 1. By the argument at the beginning of the proof, we can G_K-equivariantly collapse hairs, and since $G_K \curvearrowright K$ is cocompact, after finitely many collapses we will have collapsed K to v. Again since $G \curvearrowright T$ is cocompact, there are only finitely many orbits of connected components of $T \setminus S$, so the result follows. $\square$

If σ is a square in some $Z \subset T_1 \times T_2$, then we say an edge $\epsilon \subset \sigma$ is a *free face* if it is only contained in one square. The following terminology is due to Wise [2004].

Definition 2.7. Let $e \in \mathrm{Edges}(T_i)$ and let $\tau_e \subset \mathscr{C}$ be the fiber of e as in Lemma 2.5. The *hypercarrier* $\mathscr{H}_{\mathscr{C}}(\tau_e)$ is the union of squares of $\mathscr{C}$ intersecting τ_e nontrivially.

We note that for $e \in T_i$, a hypercarrier is mapped to an edge of T_i and that $\mathscr{H}_{\mathscr{C}}(\tau_e)$ is homeomorphic to $\tau_e \times [-1, 1]$.

Definition 2.8. We say an edge ϵ in some $Z \subset T_1 \times T_2$ is i-transverse if it coincides with its i-leaf space, or equivalently if it is mapped monomorphically via $p_i|_\epsilon$, or equivalently if it is contained in a j-leaf.

An immediate consequence of Lemma 2.6 and Figure 3 is the following.

Lemma 2.9. *Let $e \in \mathrm{Edges}(T_i)$. If $G_e \curvearrowright \tau_e$ is not minimal then $\mathscr{H}_{\mathscr{C}}(\tau_e)$ has a square σ containing an i-transverse free face ϵ.*

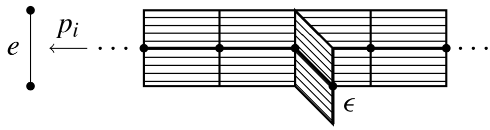

Figure 3. A spur of τ_e and the corresponding free face ϵ in the hypercarrier $\mathcal{H}_{\mathscr{C}}(\tau_e)$.

We now borrow some terminology from [Diao and Feighn 2005].

Definition 2.10. A simplicial map $S \to T$ between two trees that is obtained by identifying edges sharing a common vertex is called a *folding*. If T is obtained from S by a folding, then we say S is obtained from T by a *cleaving*.

The next lemma is now immediate (see Figure 4).

Lemma 2.11. *Let $\epsilon \subset Z \subset T_1 \times T_2$ be an i-transverse free face in a square σ. If we collapse σ onto the face opposite to ϵ, then the leaf space $\mathscr{L}_i$ is unchanged and the leaf space $\mathscr{L}_j$ gets cleaved.*

In fact this lemma can be used backwards to give a proof of Theorem 2.3. We will sketch it, leaving the details to an interested reader familiar with folding sequences [Bestvina and Feighn 1991; Stallings 1991; Dunwoody 1998; Kapovich et al. 2005].

Sketch of the proof of Theorem 2.3. Pick some vertex $v \in T_1 \times T_2$ and consider its G-orbit. We can add finitely many connected G-orbits of edges to get a connected G-complex $Gv \subset \mathscr{C}_1 \subset T_1 \times T_2$. $\mathscr{C}_1$ has leaf spaces $\mathscr{L}_1, \mathscr{L}_2$ which project onto T_1, T_2. The disconnectedness of the fibers of the projections $p_i|_{\mathscr{C}_1} : \mathscr{C}_1 \twoheadrightarrow T_i$ coincides with

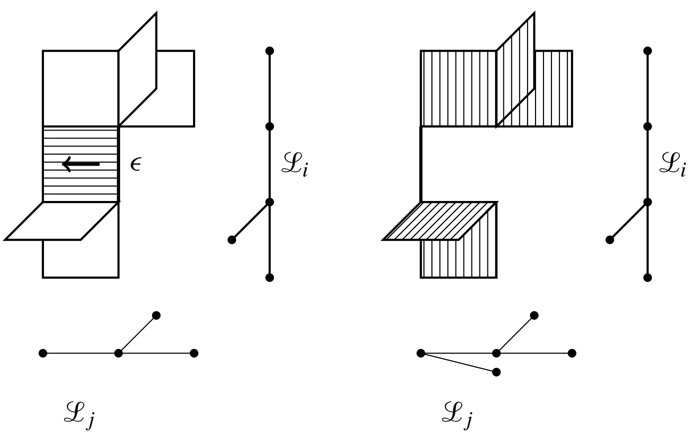

Figure 4. The effects of collapsing an i-transverse free face ϵ: the leaf space $\mathscr{L}_j$ gets cleaved, $\mathscr{L}_i$ remains unchanged. On the right the j-leaves are drawn.

the failure of injectivity of the projections $\mathscr{L}_i \twoheadrightarrow T_i$. By Lemma 2.11 (backwards) adding a square can give a folding of one of the leaf spaces. Since the edge groups of T_1, T_2 are finitely generated, and because adding all the squares of $T_1 \times T_2$ folds $\mathscr{L}_i$ to T_i, it follows that the leaf spaces $\mathscr{L}_i$ can be made to coincide with T_i after adding finitely many G-orbits of squares. □

3. The statement and proof of the main theorem

For this section we fix a collection $\mathscr{H}$ of subgroups of G. We let T_∞ and $T_{\mathscr{F}}$ be cocompact, minimal G-trees in which the subgroups in $\mathscr{H}$ act elliptically. We further require that edge groups of T_∞ are infinite and finitely generated and that edge groups of $T_{\mathscr{F}}$ are finite. Note that any nontrivial tree obtained by a collapse of T_∞ has infinite edge groups whereas any collapse of $T_{\mathscr{F}}$ has finite edge groups. It follows that T_∞ and $T_{\mathscr{F}}$, having no nontrivial common collapses, satisfy the hypotheses of Theorem 2.3.

Theorem 3.1 (Main Theorem). *Let $\mathscr{H}$ be a collection of subgroups of G and let T_∞ and $T_{\mathscr{F}}$ be cocompact, minimal G-trees in which the subgroups in $\mathscr{H}$ act elliptically. Suppose furthermore that the edge groups of $T_{\mathscr{F}}$ are finite and that the edge groups of T_∞ are infinite. Then there exists a vertex $v \in \mathrm{Vertices}(T_\infty)$ and a nontrivial, cocompact, minimal G_v-tree $\check{T}_v$ such that*

(i) *for every $f \in \mathrm{Edges}(T_\infty)$ incident to v the subgroups $G_f \leq G_v$ act elliptically on $\check{T}_v$, and*

(ii) *for every $H \in \mathscr{H}$ and $g \in G$ the subgroup $H^g \cap G_v \leq G_v$ acts elliptically on $\check{T}_v$.*

Moreover, either

(1) *every edge group of $\check{T}_v$ is finite, or*

(2) *there is some edge $e \in \mathrm{Edges}(T_\infty)$, incident to v, that not only satisfies (i), but also satisfies the following:*
   (a) *G_e splits essentially as an amalgamated free product or an HNN extension with finite edge group.*
   (b) *$G_e = G_{v_e}$ for some vertex $v_e \in \mathrm{Vertices}(\check{T}_v)$.*
   (c) *The edge stabilizers of $\check{T}_v$ are conjugate in G_v to the vertex group(s) of the splitting of G_e found in (a); in particular, the edge groups of $\check{T}_v$ are $\prec G_e$.*
   (d) *The vertex groups of $\check{T}_v$ that are not conjugate in G_v to G_e are also vertex groups of a one-edge splitting of G_v with a finite edge group; in particular these vertex groups of $\check{T}_v$ are $\prec G_v$.*

An example of what happens in situation (2) is shown in Figure 7.

Proof. Let $\mathscr{C}$ be the core of $T_\infty \times T_{\mathscr{F}}$. The ∞-leaf space $\mathscr{L}_\infty$ is the tree T_∞, and we can see $\mathscr{C}$ as a tree of spaces (see [Scott and Wall 1979] for details) which is a union of vertex spaces τ_v for $v \in \text{Vertices}(T_\infty)$ and edge spaces $\mathcal{H}_{\mathscr{C}}(\tau_e) = \tau_e \times [-1, 1]$ for $e \in \text{Edges}(T_\infty)$ attached to the τ_v along the subspaces $\tau_e \times \{\pm 1\}$.

It may be that for some $e \in \text{Edges}(T_\infty)$, the G_e-trees τ_e are not minimal. By Lemmas 2.9, 2.6, and 2.11, we can repeatedly G-equivariantly collapse ∞-transverse free faces, so that after finitely many steps we obtain a *shaved core* $\mathscr{C}'_s$ such that the $\tau_e \cap \mathscr{C}'_s$ are minimal G_e-trees. Although the $\mathscr{F}$-leaf space was cleaved repeatedly in the shaving process given by Lemma 2.6, the ∞-leaf space is unchanged. We still write $\mathscr{L}_\infty = T_\infty$.

We now construct a complex $\mathscr{C}_s \subset \mathscr{C}'_s \subset \mathscr{C}$, called the *$\infty$-minimal core*. Its principal feature is that for every $v \in \text{Vertices}(T_\infty)$ and $e \in \text{Edges}(T_\infty)$, the trees $\tau_v \cap \mathscr{C}_s$ and $\tau_e \cap \mathscr{C}_s$ are minimal G_v- and G_e-trees, respectively. Define $\mathcal{H}_{\mathscr{C}'_s}(\tau_e) = \mathcal{H}_{\mathscr{C}}(\tau_e) \cap \mathscr{C}'_s$. We call $\mathcal{H}_{\mathscr{C}'_s}(\tau_e)$ the *$\mathscr{C}'_s$-hypercarrier* attached to a vertex space τ_v in $\mathscr{C}'_s$. Note that $\tau_e \cap \mathscr{C}'_s$ naturally projects injectively into τ_v as a minimal G_e-invariant subtree where $G_e \le G_v$. If T is a G-tree and $H \le G$, denoting by $T(S)$ the minimal S-invariant subtree, we have $T(H) \subset T(G)$. It therefore follows that all the $\mathscr{C}'_s$-hypercarriers attached to τ_v are actually attached to the minimal G_v-invariant subtree of τ_v. By Lemma 2.6, after finitely many equivariant spur collapses we can make the vertex spaces τ_v into minimal G_v-trees. None of these collapses will affect the attached $\mathscr{C}'_s$-hypercarriers $\mathcal{H}_{\mathscr{C}'_s}(\tau_e)$, and the leaf space $\mathscr{L}_\infty = T_\infty$ is preserved. We have therefore constructed $\mathscr{C}_s$, the ∞-minimal core. Denote $\mathcal{H}_{\mathscr{C}_s}(\tau_e) = \mathcal{H}_{\mathscr{C}'_s}(\tau_e) \cap \mathscr{C}_s$. By what was written above, $\mathcal{H}_{\mathscr{C}_s}(\tau_e) = \mathcal{H}_{\mathscr{C}'_s}(\tau_e)$, and we now call $\mathcal{H}_{\mathscr{C}_s}(\tau_e)$ a *$\mathscr{C}_s$-hypercarrier*.

For every $k \in \text{Edges}(T_{\mathscr{F}})$, G_k is finite, therefore a minimal G_k-tree is a point; thus, by cocompactness and regularity, the trees $\tau_k \in \mathscr{C}$ have finite diameter and the same must be true of every connected component of $\tau_k \cap \mathscr{C}_s$. So, every connected component of $\tau_k \cap \mathscr{C}_s$ has a spur. It therefore follows that $\mathscr{C}_s$ must have an $\mathscr{F}$-transverse free face ϵ containing a spur of some connected component of $\tau_k \cap \mathscr{C}_s$ for some $k \in \text{Edges}(T_{\mathscr{F}})$. Furthermore, the stabilizer $G_\epsilon \le G_{p_{\mathscr{F}}(\epsilon)}$ is an edge stabilizer of $T_{\mathscr{F}}$, and therefore finite. This $\mathscr{F}$-transverse free face ϵ must be contained in some $\tau_v \cap \mathscr{C}_s$ for $v \in \text{Vertices}(T_\infty)$. Suppose first that ϵ was not contained in any $\mathscr{C}_s$-hypercarrier attached to $\tau_v \cap \mathscr{C}_s$. Then for every $e \ni v$ in $\text{Edges}(T_\infty)$, G_e fixes some $\mathscr{C}_s$-hypercarrier $\mathcal{H}_{\mathscr{C}_s}(\tau_e)$ such that $\mathcal{H}_{\mathscr{C}_s}(\tau_e) \cap \tau_v = \tau_e^+$ is contained in the complement $(\tau_v \cap \mathscr{C}_s) \setminus G_v\epsilon$.

Definition 3.2. Let T be a minimal G-tree and $e \in \text{Edges}(T)$. We denote by $C(T, e)$ the *non-e-collapse of* T, the tree whose edges are the orbit $Ge \subset T$ and whose vertices are the closures of the connected components of $T \setminus Ge$, with a vertex v adjacent to an edge e' in $C(T, e)$ if and only if, viewed as subsets of T, $e' \cap v \neq \varnothing$.

It therefore follows that $\check{T}_v = C(\tau_v \cap \mathscr{C}_s, \epsilon)$ is a tree with finite edge groups, in which each $G_e \le G_v$ ($e \in \text{Edges}(T_\infty)$) acts elliptically, and also conjugates of groups in $\mathcal{H}$ intersecting G_v act elliptically. Thus (i), (ii) and (1) are satisfied.

Otherwise, the free face $\epsilon \subset \tau_v \cap \mathscr{C}_s$ is, by definition of a free face, contained in *exactly one* $\mathscr{C}_s$-hypercarrier $\mathscr{H}_{\mathscr{C}_s}(\tau_e)$. We now construct the G_v-tree $\check{T}_v$ satisfying (2). This construction is illustrated in Figure 5. We first take the subset

$$Z = \Big(\tau_v \cup \bigcup_{e \ni v} \mathscr{H}_{\mathscr{C}_s}(\tau_e)\Big) \cap \mathscr{C}_s,$$

i.e., $\tau_v \cap \mathscr{C}_s$ to which we attach all adjacent $\mathscr{C}_s$-hypercarriers. Now the G_v-translates of ϵ are contained in the $\mathscr{C}_s$-hypercarriers $\mathscr{H}_{\mathscr{C}_s}(\tau_{ge})$ for $g \in G_v$. For each such $\mathscr{C}_s$-hypercarrier we denote by τ_{ge}^- the connected component of $\tau_e \times \{\pm 1\} \subset \mathscr{H}_{\mathscr{C}_s}(\tau_{ge})$ not contained in $\tau_v \cap \mathscr{C}_s$ (see the top of Figure 5).

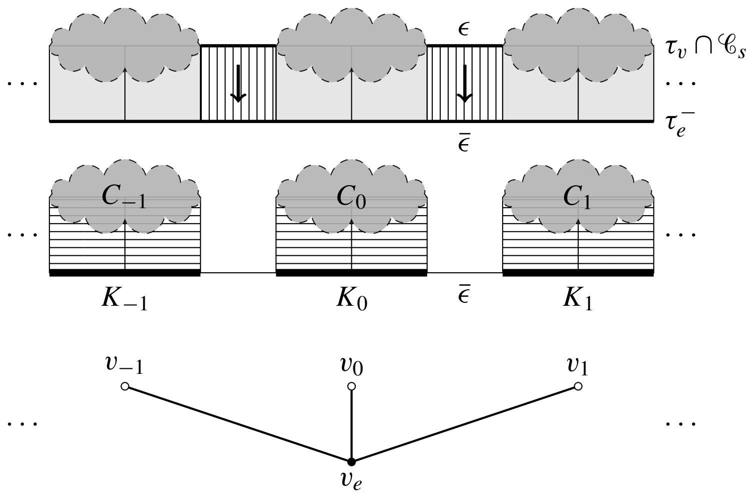

Figure 5. Constructing $\check{T}_v$. The top shows a portion of Z, the middle shows the result of equivariantly collapsing the free face ϵ, and the bottom shows the corresponding ∞-leaf space.

We now G_v-equivariantly collapse the square $\sigma \supset \epsilon$ onto the opposite side $\bar{\epsilon}$, obtaining a connected G_v-subset $Z_c \subset Z$ (see the middle of Figure 5). The resulting intersection $\tau_v \cap Z_c$ consists of a collection of connected components $\{C_i \mid i \in I\}$. Similarly, the G_e-translates of $\bar{\epsilon}$ give connected components $\{K_i \mid i \in I\}$ of $\tau_e \setminus G_e\bar{\epsilon}$. Because G_e acts on $C(\tau_e^-, \bar{\epsilon})$, and by minimality of $\tau_e \cap \mathscr{C}_s$, this action is also minimal with one edge orbit. This gives us (a).

For every edge $f \in \mathrm{Edges}(T_\infty)$ incident to v that is not in the G_v-orbit of e, the orbit $G_v\epsilon$ does not intersect $\mathscr{H}_{\mathscr{C}_s}(\tau_f) \cap \tau_v$. It follows that each such $G_f \leq G_v$ stabilizes some component C_i. We now detach from Z_c all $\mathscr{C}_s$-hypercarriers not stabilized by a G_v-conjugate of G_e, producing a G_v-complex $Z_c' \subset Z_c$, specifically

$$Z_c' = Z_c \cap \Big(\tau_v \cup \bigcup_{g \in G_v} \mathscr{H}_{\mathscr{C}_s}(\tau_{ge})\Big).$$

Next, to get the G_v-tree T_v, we collapse each G_v-translate of τ_e^- to a vertex v_e, collapse each component C_i to a vertex v_i, and collapse each connected component of G_v-translates of $\mathscr{H}_{\mathscr{C}_S}(\tau_e) \cap Z_c'$ onto an edge connecting v_e and the corresponding vertex v_i. This is illustrated at the bottom of Figure 5.

Equivalently, if we consider the ∞-leaf space corresponding to the union of the $\mathscr{C}_s$-hypercarriers $g\mathscr{H}_{\mathscr{C}_s}(\tau_e)$ attached to $\tau_v \cap \mathscr{C}_s$ for $g \in G_v$, then we have a tree of radius 1, which is G_v-isomorphic to $\{v\} \cup \left(\bigcup_{g \in G_v} ge\right) \subset T_\infty$. After equivariantly collapsing the free face ϵ, Lemma 2.11 gives us a cleaving of this radius 1 subtree to the infinite tree $\check{T}_v$ constructed above. See Figure 6. We note that if we took the ∞-leaf space of Z_c, i.e., had we not detached the other hypercarriers, the resulting leaf space would be a tree with many spurs. The tree $\check{T}_v$ we obtain is a minimal G_v-tree that satisfies (b) and (i).

Moreover, we note that by construction, every subgroup $H^g \cap G_v$, for $g \in G$ and $H \in \mathscr{H}$, acts elliptically on $\check{T}_v$. So (ii) is satisfied as well.

Since τ_e^- is G_v-regular, the vertex stabilizers of $C(\tau_e^-, \bar{\epsilon})$ coincide with the component stabilizers $(G_e)_{K_i} = (G_v)_{K_i}$. We also have $(G_v)_{C_i} \cap (G_v)_{\tau_e^-} = (G_v)_{K_i}$ (again referring to the middle of Figure 5). It now follows that the edge stabilizers of $\check{T}_v$ satisfy (c).

Finally note that the vertex groups of $\check{T}_v$ that are not stabilized by G_v-conjugates of G_e are also the vertex groups of $C(\tau_v, \epsilon)$ (see the top of Figure 5). Finally, since G_ϵ is finite, (d) follows. □

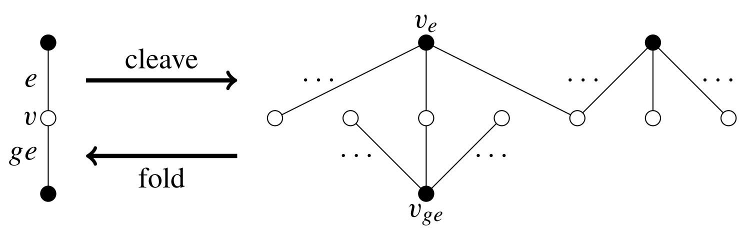

Figure 6. Equivariant collapsing free faces cleaves the leaf space of Z_C' to a tree $\check{T}_v$ with infinite diameter.

4. Splittings of virtually free groups

Another way to use Theorem 3.1 is to obtain cleavings of G-trees whose edge and vertex groups are “smaller”. This will be used as the inductive step in our proof of Theorem 1.6.

Corollary 4.1. *Let T be a G-tree in which the subgroups $\mathscr{H}$ act elliptically with infinite edge groups, and let G be many-ended relative to $\mathscr{H}$. Either some vertex*

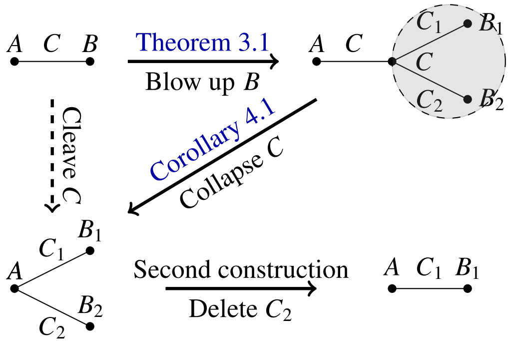

Figure 7. An example of the effects of Theorem 3.1, Corollary 4.1, and the second construction of the proof of Theorem 1.6 on a graph of groups. The vertices and edges are labeled by the corresponding vertex and edge groups. In all cases $B_i \prec B$ and $C_i \prec C$.

$v \in \mathrm{Vertices}(T)$ *can be blown up to a tree with finite edge groups*; *or*, *there is an edge* $e \in \mathrm{Edges}(T)$ *such that we can blow up* T, *relative to* $\mathcal{H}$, *to some tree* $\check{T}$, *and then collapse the edges in the orbit of* e *to points. The resulting tree* T' *can also be obtained from* T *by equivariantly cleaving some edge* e. *If* $e' \in \mathrm{Edges}(T')$ *is a new edge obtained by a cleaving of* e, *then* $G_{e'} \prec G_e$. *Also*, *for each new vertex* $v' \in \mathrm{Vertices}(T')$, *there is some* $v \in \mathrm{Vertices}(T)$ *that got cleaved such that* $G_{v'} \prec G_v$.

Furthermore, *in passing from* T *to* T' *the number of edge orbits and the number of vertex orbits does not decrease and increases by at most* 1.

Proof. Suppose we are in case (2) of Theorem 3.1. Then some vertex v gets blown up to $\check{T}_v$ and some vertex stabilizer of $\check{T}_v$ coincides with G_e. Specifically $\check{T}$ can be obtained by deleting each blown up vertex v from T and then equivariantly reattaching every edge incident to v to the corresponding vertex in $\check{T}_v$.

In particular, if $e \in \mathrm{Edges}(T)$ is an edge incident to v that satisfies (2) of Theorem 3.1, then it is attached to the vertex $v_e \in \mathrm{Vertices}(\check{T}_v)$. We obtain T' by collapsing the G-orbits of e to points. This amounts to identifying the vertex v_e with the vertex $u_e \in \mathrm{Vertices}(\check{T})$ that is the other endpoint of e. From Figure 6 it is clear that T' is obtained by cleaving T.

We finally note that in passing from T to $\check{T}$ and then from $\check{T}$ to T', the vertex and edge groups are nonincreasing. Otherwise, the required properties of T' are immediately satisfied by Theorem 3.1 (see Figure 7). □

Finally, we can give our description of the decompositions of virtually free groups as amalgamated free products or HNN extensions.

Proof of Theorem 1.6. We prove this result by successively applying Corollary 4.1 until some desirable terminating condition is met. Virtually free groups have no one-ended subgroups, so we will always be able to apply our corollary; furthermore, virtually free groups are finitely presented. It now follows by Dunwoody accessibility [Dunwoody 1985] that there are no infinite chains $C_1 \succ C_2 \succ \cdots$ of virtually free groups (recall Definition 1.3), and that all such chains must terminate with finite groups.

First construction (pass to relatively one-ended vertex subgroups): Let T be a G-tree with one edge orbit Ge with G_e infinite. By accessibility, we may pass to a tree $T^{(2)}$ obtained by blowing up some vertices v of T to trees $\breve{T}_v$ such that the vertex groups of $\breve{T}_v$ are either finite or one-ended relative to the stabilizers G_f of the incident edges $f \ni v$. If possible, we take $T^{(1)} \subset T^{(2)}$ to be an infinite connected subtree obtained by deleting edges with finite stabilizers, and we set $G^{(1)} = G_{T^{(1)}}$, the setwise stabilizer. Note that the vertex groups of $T^{(1)}$ are $\preccurlyeq$ the vertex groups of T, and vertex groups are one-ended relative to the incident edge groups.

Second construction (pass to smaller edge groups): The second construction utilizes Corollary 4.1. If T_i is a G_i-tree with one edge orbit whose vertex groups are one-ended relative to the incident edge groups, we first apply Theorem 3.1 to blow up a vertex $v \in \mathrm{Vertices}(T_i)$, and find ourselves in case (2) of the theorem. If $\breve{T}_v$ has a finite edge group then G_v is not one-ended relative to the incident edge groups, contradicting our assumption. By Corollary 4.1 we can collapse an edge of the blowup of T_i to get a cleaving T_i' that has at most two edge orbits, with edge groups $\prec$ the edge groups of T_i. The new vertex groups are also $\preccurlyeq$ the old vertex groups. If there are two edge orbits, then we obtain $T_{i+1} \subset T_i'$ as a maximal subtree containing only one edge orbit and set $G_{i+1} = (G_i)_{T_{i+1}}$, the setwise stabilizer. (See Figure 7.) If T' already has only one edge orbit then $T_{i+1} = T_i$ and $G_{i+1} = G_i$.

In both constructions, we pass to subgroups that split as graphs of groups such that the edge groups and vertex groups are $\preccurlyeq$ the edge and vertex groups of the original splitting of the overgroup.

We start with the amalgamated free product case. Let $T = T_0$ be the Bass–Serre tree corresponding to the splitting given in (1) of the statement of Theorem 1.6. Take the blowup $T_0^{(2)}$ obtained from the first construction. If one of the vertex groups of this blowup coincides with an incident edge group then we are done. Otherwise, we may pass to the $G^{(1)}$-tree $T_0^{(1)}$, which still has one edge orbit and two vertex orbits, and whose vertex groups are one-ended relative to the incident edge groups. Furthermore, because the new vertex groups are $\preccurlyeq$ the vertex groups of T, if the statement of the theorem holds for $G^{(1)}$ and the splitting corresponding to its action on $T_0^{(1)}$ (which is also an amalgamated free product), then the statement also holds for G and the splitting corresponding to its action on T.

We can now apply our second construction to the $G_0^{(1)}$-tree $T_0^{(1)}$ to obtain a G_1-tree T_1, which again must have one edge orbit and two vertex orbits. Furthermore, for the (conjugacy class of the) edge group, we have a proper containment $C_1 \prec C$. Again, because the vertex groups of T_1 are $\preccurlyeq$ the vertex groups of $T_0^{(1)}$, if the Theorem holds for this subgroup, it holds for G.

We repeatedly apply our construction, thus obtaining a sequence of groups that split as amalgamated free products. With each iteration of the second construction, we pass to a smaller edge group. Hence, by accessibility, eventually there is a subgroup G_i acting on $T_i^{(2)}$ (see the first construction) such that the vertex groups split as graphs of groups with finite edge groups and one of the incident edge groups coincides with the vertex group. Since $\preccurlyeq$ is transitive, (1) of Theorem 1.6 is satisfied.

We now tackle the HNN extension case. The proof proceeds in the same way. We repeatedly blow up, cleave, and pass to subtrees, the main difference being that the G-tree T has only one vertex orbit. If at some point one of the trees T_i or $T_i^{(1)}$ has two vertex orbits, then these vertex groups are vertex groups of a splitting of the vertex group of T_{i-1} with finite edge groups. It follows that if T_i satisfies (1) of Theorem 1.6, then T_{i-1} satisfies (2) of Theorem 1.6, and thus by transitivity of $\preccurlyeq$, so must our original splitting T. Otherwise, the proof goes through identically. □

Acknowledgements

I wish to thank John MacKay and Alessandro Sisto for asking me for a proof of Corollary 1.1. I had actually even given them what I thought to be a counterexample; their countercounterexample gave me ample motivation to investigate this problem further. This paper also would not have been possible without everything I learned from Lars Louder. The ideas of Section 2, especially the usefulness of the Core Theorem, arose from our discussions while working on strong accessibility. I am also grateful for the meticulous work of the anonymous referee who spotted many tiny mistakes, as well as a couple embarrassing ones, and gave suggestions that substantially improved the exposition. Finally, I thank Inna Bumagin. This paper was written while I was supported as a postdoctoral fellow by her NSERC grant.

References

[Bestvina and Feighn 1991] M. Bestvina and M. Feighn, "Bounding the complexity of simplicial group actions on trees", *Invent. Math.* **103**:3 (1991), 449–469. MR 92c:20044 Zbl 0724.20019

[Bestvina and Feighn 1994] M. Bestvina and M. Feighn, "Outer limits", unpublished, 1994, Available at http://andromeda.rutgers.edu/~feighn/papers/outer.pdf.

[Cashen 2012] C. H. Cashen, "A geometric proof of the structure theorem for cyclic splittings of free groups", preprint, 2012. arXiv 1208.3652

[Diao and Feighn 2005] G.-A. Diao and M. Feighn, "The Grushko decomposition of a finite graph of finite rank free groups: an algorithm", *Geom. Topol.* **9** (2005), 1835–1880. MR 2006i:20045 Zbl 1093.20022

[Dunwoody 1985] M. J. Dunwoody, "The accessibility of finitely presented groups", *Invent. Math.* **81**:3 (1985), 449–457. MR 87d:20037 Zbl 0572.20025

[Dunwoody 1998] M. J. Dunwoody, "Folding sequences", pp. 139–158 in *The Epstein Birthday Schrift*, edited by I. Rivin et al., Geom. Topol. Monogr. **1**, Geom. Topol. Publ., Coventry, 1998. MR 2000f:20037 Zbl 0927.20013

[Fujiwara and Papasoglu 2006] K. Fujiwara and P. Papasoglu, "JSJ-decompositions of finitely presented groups and complexes of groups", *Geom. Funct. Anal.* **16**:1 (2006), 70–125. MR 2007c:20100 Zbl 1097.20037

[Guirardel 2005] V. Guirardel, "Cœur et nombre d'intersection pour les actions de groupes sur les arbres", *Ann. Sci. École Norm. Sup.* (4) **38**:6 (2005), 847–888. MR 2007e:20055 Zbl 1110.20019

[Haefliger 1991] A. Haefliger, "Complexes of groups and orbihedra", pp. 504–540 in *Group theory from a geometrical viewpoint* (Trieste, 1990), edited by É. Ghys et al., World Sci. Publ., 1991. MR 93m:20048 Zbl 0858.57013

[Kapovich et al. 2005] I. Kapovich, R. Weidmann, and A. Miasnikov, "Foldings, graphs of groups and the membership problem", *Internat. J. Algebra Comput.* **15**:1 (2005), 95–128. MR 2005m:20081 Zbl 1089.20018

[Linnell 1983] P. A. Linnell, "On accessibility of groups", *J. Pure Appl. Algebra* **30**:1 (1983), 39–46. MR 85c:20021 Zbl 0545.20020

[Louder 2008] L. Louder, "Krull dimension for limit groups III: Scott complexity and adjoining roots to finitely generated groups", preprint, 2008. arXiv math/0612222

[Scott and Wall 1979] P. Scott and T. Wall, "Topological methods in group theory", pp. 137–203 in *Homological group theory* (Durham, 1997), edited by C. T. C. Wall, London Math. Soc. Lecture Note Ser. **36**, Cambridge Univ. Press, Cambridge, 1979. MR 81m:57002 Zbl 0423.20023

[Serre 1980] J.-P. Serre, *Trees*, Springer, New York, 1980. MR 82c:20083 Zbl 0548.20018

[Stallings 1971] J. Stallings, *Group theory and three-dimensional manifolds*, Yale Mathematical Monographs **4**, Yale University Press, New Haven, 1971. MR 54 #3705 Zbl 0241.57001

[Stallings 1991] J. Stallings, "Foldings of G-trees", pp. 355–368 in *Arboreal group theory* (Berkeley, 1988), edited by R. C. Alperin, Math. Sci. Res. Inst. Publ. **19**, Springer, New York, 1991. MR 92g:20040 Zbl 0782.20018

[Swarup 1986] G. A. Swarup, "Decompositions of free groups", *J. Pure Appl. Algebra* **40**:1 (1986), 99–102. MR 87e:20092 Zbl 0579.20021

[Wilton 2012] H. Wilton, "One-ended subgroups of graphs of free groups with cyclic edge groups", *Geom. Topol.* **16**:2 (2012), 665–683. MR 2928980 Zbl 1248.20047

[Wise 2004] D. T. Wise, "Cubulating small cancellation groups", *Geom. Funct. Anal.* **14**:1 (2004), 150–214. MR 2005c:20069 Zbl 1071.20038

Received September 2, 2014. Revised March 25, 2015.

NICHOLAS TOUIKAN
DEPARTMENT OF MATHEMATICAL SCIENCES
STEVENS INSTITUTE OF TECHNOLOGY
1 CASTLE POINT TERRACE
HOBOKEN, NJ 07030
UNITED STATES

nicholas.touikan@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.278.479

ON THE STRUCTURE OF VERTEX CUTS SEPARATING THE ENDS OF A GRAPH

GARETH R. WILKES

Dinic, Karzanov and Lomonosov showed that the minimal edge cuts of a finite graph have the structure of a cactus, a tree-like graph constructed from cycles. Evangelidou and Papasoglu extended this to minimal cuts separating the ends of an infinite graph. In this paper we show that minimal vertex cuts separating the ends of a graph can be encoded by a succulent, a mild generalisation of a cactus that is still tree-like. We go on to show that the earlier cactus results follow from our work.

1. Introduction and definitions

Lying on the boundaries of several topic areas, vertex and edge cuts of graphs have been considered by graph theorists, network theorists, topologists and geometric group theorists, and the study of their structure has led to applications ranging from algorithms to classical group theoretic propositions.

Vertex cut pairs of finite graphs were studied by Tutte [1984], who showed that a graph possessing such cuts can be modelled with a tree. This was extended to infinite, locally finite graphs in [Droms et al. 1995]. Dunwoody and Krön [2015] then extended this work to cuts of other cardinalities, using vertex cuts to associate structure trees to graphs in a more general context.

This process of finding trees associated to graphs gives a way into geometric group theory. If, for instance, we find a structure tree for the Cayley graph of a group, then in light of the work of Bass and Serre [Serre 1980], we can obtain information about the group from its action on the tree. An example is Stallings' theorem [1968] on the classification of groups with many ends. The work of Dunwoody and Krön [2015] and of Evangelidou and Papasoglu [2014] yields more proofs of Stallings' theorem along these lines.

Dinic, Karzanov and Lomonosov [Dinic et al. 1976] showed that minimal edge cuts of a finite graph have, in addition to a tree-like nature, the finer structure of a cactus graph. For a recent elementary proof, see [Fleiner and Frank 2009]. Evangelidou and Papasoglu [2014] extended this, encoding all minimal edge cuts

MSC2010: primary 05C40; secondary 20E08, 57M07.
Keywords: ends, cuts, cactus, succulent.

separating the ends of an infinite graph by a cactus. The important stages in these proofs involve showing that certain collections of "crossing" cuts have a circular structure. In this paper we switch our attention to vertex cuts, showing that we can encode all minimal vertex end cuts of a graph by a tree-like structure called a succulent, which is a mild generalisation of a cactus. A traditional cactus is composed of cycles joined together at vertices in a tree-like fashion. For our succulents, we also allow cycles to attach along a single edge, again in a tree-like way. Once again the key step is to show that crossing cuts have a cyclic nature.

We will also show how the earlier cactus theorems can be regarded as special cases of our work, and discuss an application to certain finite graphs.

Let $\Gamma = (V, E)$ be a connected graph. If $K \subseteq V$ is a set of vertices of the graph, we denote by $\Gamma - K$ the graph obtained from Γ by removing K and all edges incident to K. K is called a *vertex cut* if K is finite and $\Gamma - K$ is not connected. If $A, B \subseteq \Gamma$ then we say K *separates* A and B if any path joining a vertex of A to a vertex of B intersects K.

A *ray* of Γ is an infinite sequence of distinct consecutive vertices of Γ. We say that two rays r_1, r_2 are *equivalent* if for any vertex cut K, all vertices of $r_1 \cup r_2$ except finitely many are contained in the same component of $\Gamma - K$. The *ends* of Γ are equivalence classes of rays. If K is a vertex cut of Γ, we say K is an *end cut* if there are at least two components of $\Gamma - K$ which contain rays. We say that an end cut is a *mincut* if its cardinality is minimal amongst end cuts of Γ. A mincut is said to *separate* ends e_1, e_2 of a graph if there are rays r_1, r_2 representing e_1, e_2 respectively such that r_1, r_2 are contained in different components of $\Gamma - K$. A mincut gives a partition of the set $\mathcal{E}$ of ends of the graph. Two mincuts are called *equivalent* if they give the same partition of $\mathcal{E}$. We denote the equivalence class of a mincut K by $[K]$, and write $K \sim L$ if K, L are equivalent.

A *succulent* is a graph constructed from cycles by joining cycles together at vertices or at single edges, in a "tree-like" fashion. We give a more formal definition of this as Definition 8.1 below. An *end vertex* of a succulent is one incident to at most two edges. We now state the main theorem of the paper.

Theorem 8.2. *Let Γ be a connected graph such that there are vertex end cuts of Γ with finite cardinality. There is a succulent $\mathcal{S}$ with the following properties*:

(1) *There is a subset A of vertices of $\mathcal{S}$ called the* anchors *of $\mathcal{S}$. If two anchors are adjacent, one of them is an end vertex of the graph. Every vertex of $\mathcal{S}$ not in A is adjacent to an anchor. We define an anchor cut of $\mathcal{S}$ to be a vertex cut containing no anchors which separates some anchors of $\mathcal{S}$. We say anchor cuts are equivalent if they partition A in the same way.*

(2) *There is an onto map f from the ends of Γ to the union of the ends of $\mathcal{S}$ with the end vertices of $\mathcal{S}$ which are anchors.*

(3) *There is a bijective map g from equivalence classes of minimal end cuts of* Γ *to equivalence classes of minimal anchor cuts of* $\mathcal{S}$ *such that ends* e_1, e_2 *of* Γ *are separated by* $[K]$ *if and only if* $f(e_1), f(e_2)$ *are separated by* $g([K])$.

(4) *Any automorphism of* Γ *induces an automorphism of* $\mathcal{S}$.

2. Preliminaries

Definition 2.1. Given a mincut K, we call a component of $\Gamma - K$ *proper* if it contains an end, and a *slice* if not. Given a set of vertices C, its *boundary* ∂C is the set of those vertices not in C but adjacent to a vertex of C; and $C^* = V(\Gamma) - (C \cup \partial C)$.

It will be convenient to assume our graph contains no slices. In the following lemmas we show that we can do this by replacing Γ with another graph $\hat{\Gamma}$ which has the same ends and cuts, but no slices. The results in this section are adapted for our needs from more general results proved by Dunwoody and Krön [2015].

Lemma 2.2. *Let K, L be mincuts and C, D proper components of $\Gamma - K$, $\Gamma - L$. Suppose that both $C \cap D$ and $C^* \cap D^*$ contain an end. Then $\partial(C \cap D)$, $\partial(C^* \cap D^*)$ are mincuts*,

$$\partial(C \cap D) = (C \cap L) \cup (K \cap L) \cup (K \cap D),$$
$$\partial(C^* \cap D^*) = (C^* \cap L) \cup (K \cap L) \cup (K \cap D^*),$$

and

$$|C \cap L| = |K \cap D^*|,$$
$$|D \cap K| = |L \cap C^*|.$$

Proof. The boundaries $\partial(C \cap D)$, $\partial(C^* \cap D^*)$ are certainly end cuts, with

$$\partial(C \cap D) \subseteq (C \cap L) \cup (K \cap L) \cup (K \cap D),$$
$$\partial(C^* \cap D^*) \subseteq (C^* \cap L) \cup (K \cap L) \cup (K \cap D^*).$$

Consider the following diagram (see Figure 1), where a, b, c, d, u denote the cardinalities of the indicated sets. Let n be the cardinality of a mincut.

Then

$$a + c + u = n,$$
$$b + d + u = n.$$

Since $(C \cap L) \cup (K \cap L) \cup (K \cap D)$ is an end cut, we have

$$d + a + u \geq n,$$

and similarly

$$b + c + u \geq n.$$

Summing these and comparing with the equalities above, we find them to be equalities; it follows that $a = b$, $c = d$. □

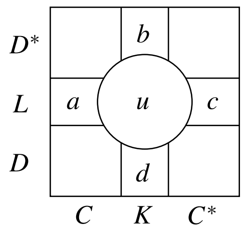

Figure 1. Diagram for Lemma 2.2.

An analogous result holds when $C^* \cap D$, $D^* \cap C$ both contain ends.

Lemma 2.3. *If C, D are proper components of cuts K, L then there is a proper component of $\Gamma - K$ containing $C^* \cap L$.*

Proof. Since K, L are end cuts, one of the pairs $\{C \cap D, C^* \cap D^*\}$, $\{C^* \cap D, C \cap D^*\}$ both contain ends. Let A be the appropriate one of $C^* \cap D^*$, $C^* \cap D$. Then using Lemma 2.2, ∂A is a mincut. Let E be a component of A containing an end; then $\partial E = \partial A$ is a mincut. Let C_0^* be the component of C^* containing E. By Lemma 2.2 every vertex $x \in C^* \cap L$ is adjacent to E, so $x \in C_0^*$ and $C^* \cap L \subseteq C_0^*$. □

Lemma 2.4. *A slice component of a mincut has empty intersection with each mincut. Distinct slices are disjoint. If Q is a slice, then no pair of elements of ∂Q are separated by any mincut.*

Proof. Let Q_1 be a slice component of $\Gamma - K$ for a mincut K and let L be a mincut, with a proper component D of $\Gamma - L$. By Lemma 2.3, there is a proper component of $\Gamma - K$ containing $C^* \cap L$, and C, a proper component, contains $C \cap L$. Q is disjoint from both of these, so $Q_1 \cap L = \varnothing$.

Suppose Q_2 is a slice component of $\Gamma - L$. We have $\partial Q_2 \subseteq L$, $\partial Q_1 \subseteq K$, and hence $Q_1 \cap \partial Q_2$, $Q_2 \cap \partial Q_1$ are both empty. The components Q_1, Q_2 are connected, so this implies that they are disjoint or equal.

Finally suppose $x, y \in \partial Q$ for a slice component Q of $\Gamma - K$ and x, y are separated by a mincut L. The slice Q is connected so there is a path in Q from x to y, which must intersect L, but we have seen this is impossible. □

We will now show how to replace Γ with another graph $\hat{\Gamma}$ which has the same ends and cuts, but no slices. The vertex set $\hat{V}$ of $\hat{\Gamma}$ consists of those vertices of Γ which are contained in no slice. Two vertices $u, v \in \hat{V}$ are joined by an edge in $\hat{\Gamma}$ if and only if they are joined by an edge in Γ or if u, v lie in the boundary of some slice of Γ.

Lemma 2.5. *The graph $\hat{\Gamma}$ is connected and the mincuts of $\hat{\Gamma}$ are the same as the mincuts of Γ. There are no slices in $\hat{\Gamma}$. The ends of $\hat{\Gamma}$ are in bijection with the ends of Γ.*

Proof. First we show that if K is a mincut and C is a proper component thereof, then $\partial \hat{C}$, the boundary of $\hat{C} = C \cap \hat{\Gamma}$ as a subset of $\hat{\Gamma}$, is equal to K.

Suppose there is $x \in \partial\hat{C} - K$. If $x \in C$ then $x \in \hat{C}$, so $x \in C^*$. Also $x \in \partial\hat{C}$ so there is $y \in \hat{C}$ adjacent to x in $\hat{\Gamma}$. Then there is an edge from x to y in $\hat{\Gamma}$, but not in Γ; so x, y lie in the boundary of some slice Q of Γ. By Lemma 2.4, $K \cap Q = \varnothing$. The slice Q is connected, and Q intersects C (at y), so $Q \subseteq C$. We then have a path from x to y in Γ which is contained in Q except for its endpoint x, which is a path from C^* to C not intersecting K, a contradiction.

Suppose there exists $x \in K - \partial\hat{C}$; x has a neighbour y in $C - \hat{C}$. Then y is contained in a slice component Q of $\Gamma - L$ for a mincut L. If $K = L$ then C, Q are disjoint; but $y \in C \cap Q$. So $K \neq L$ and since $Q \subseteq C$ (Q does not intersect K but does intersect C), there is $z \in C \cap \partial Q \subseteq C \cap L$. Now z is not in any slice, so $z \in \hat{C}$. Then x, z are adjacent in $\hat{\Gamma}$; but $z \in \hat{C}$, $x \notin \hat{C} \cup \partial\hat{C}$, a contradiction.

Let us discuss the ends of $\hat{\Gamma}$. By definition, slices contain no rays. Thus if r is any ray in Γ, we can form a new ray in $\hat{\Gamma}$ by deleting any vertices in a slice; the extra edges added in the construction of $\hat{\Gamma}$ will ensure that this is a bona fide ray. If two rays are separated by a (not necessarily minimal) end cut K in Γ, then the union of $K \cap V(\hat{\Gamma})$ with the boundaries of any slices intersecting K gives an end cut separating the images of the rays in $\hat{\Gamma}$. Similarly, if two rays in $\hat{\Gamma}$ are separated by an end cut K in $\hat{\Gamma}$, then taking the union of K with any slice boundaries intersecting K gives an end cut separating the same rays in Γ. It follows that the ends of $\hat{\Gamma}$ are in a natural bijection with those of Γ.

The end cuts of $\hat{\Gamma}$ inherited from mincuts of Γ are indeed the minimal end cuts of $\hat{\Gamma}$. Suppose K is an end cut of $\hat{\Gamma}$ which is not also an end cut of Γ. Then two proper components of $\hat{\Gamma} - K$ are connected in Γ. A path between them can only not intersect K if it passes through a slice Q; but points on the boundary of Q are connected in $\hat{\Gamma}$ so we get a path between the two components in $\hat{\Gamma}$ as well, a contradiction. So all mincuts of $\hat{\Gamma}$ are mincuts of Γ as well, and the notion of minimality carries over to $\hat{\Gamma}$ too.

Finally, there are no slices in $\hat{\Gamma}$. Let C be a component of $\hat{\Gamma} - K$ for a mincut K of $\hat{\Gamma}$ (equivalently of Γ). Let C' be the component of $\Gamma - K$ containing C. Now C' cannot be a slice as it intersects $V(\hat{\Gamma})$. So C' contains an end of Γ, whence from above so does C. So C is not a slice. □

For the rest of the paper we replace Γ with $\hat{\Gamma}$. As we have seen, the ends and cuts of the two graphs are the same, and this is all the structure with which we are concerned, so we lose nothing by doing this. All components of a cut are now proper.

We now start to prove some basic properties of mincuts, putting restrictions on cuts which "interact" with each other in some sense, and showing that a mincut does not interact with any but finitely many other mincuts. We first define what it means for cuts to not interact with each other. We are still following Dunwoody and Krön [2015] here, with some minor modifications.

Definition 2.6. Two cuts K, L are called *nested* if there are components E, F of $\Gamma - K$, $\Gamma - L$ respectively with $E \subseteq F$ or $F \subseteq E$.

Note that if K, L are nested and not equal with say $E \subseteq F$ then all components of $\Gamma - L$ except F are contained in the same component of $\Gamma - K$. This follows since there is an element of L in E^*, and by minimality all components of $\Gamma - L$ except F are connected to this vertex by paths which do not intersect $F \cup L$, and hence do not intersect $E \cup K$. Note also that these components are still connected in $\Gamma - K$ by similar reasoning. Conversely, all components of $\Gamma - K$ except one are contained in F.

Definition 2.7. A mincut is called an *A-cut* if it is nested with all other mincuts. It is called a *B-cut* if it separates Γ into exactly two components.

Lemma 2.8. *A mincut is either an A-cut or a B-cut.*

Proof. Let K be a mincut which is not an A-cut. Then there is a mincut L with which K is not nested. Let C be a (proper) component of $\Gamma - K$ and D a (proper) component of $\Gamma - L$. By Lemma 2.3, there is a component C_0^* of C^* containing $C^* \cap L$. We wish to show this is the only component of C^*. If there is another one C_1^* then $C_1^* \cap L$ is empty; C_1^* is connected so $C_1^* \subseteq D$ or $C_1^* \subseteq D^*$. In the first case, K, L are nested; so the second one happens no matter which component C_1^* we choose. So $D \cap C^* \subseteq C_0^*$. Also, $K = \partial C_1^*$ by minimality, so $C_1^* \subseteq D^*$ implies $K \cap D = \varnothing$. Then $D \subseteq C$ or $D \subseteq C^*$ (whence $D \subseteq C_0^*$); in either case, K and L are nested. This is a contradiction, so K is a B-cut. □

We call a set S of vertices a *tight x-y-separator* if $\Gamma - S$ has two distinct components A, B which are adjacent to all elements of S, with $x \in A$, $y \in B$.

Lemma 2.9. *For each integer k and every pair x, y of vertices of a graph, there are only finitely many tight x-y-separators of order k.*

Proof. We proceed by induction. If we take a path from x to y, any tight x-y-separator of order 1 would have to be a vertex on this path, so there are only finitely many of these.

Suppose the lemma holds for all tight x-y-separators of order k in all connected graphs. Take a path π from x to y in a graph Γ and suppose there are infinitely many tight x-y-separators of order $k+1 \geq 2$. Then there is a vertex $z \in \pi - \{x, y\}$ which is contained in infinitely many of these separators. If S_1, S_2 are distinct such tight x-y-separators of order $k+1$ in Γ then $S_1 - \{z\}$, $S_2 - \{z\}$ are distinct tight x-y-separators of order k in $\Gamma - \{z\}$, so there are infinitely many of these, giving the required contradiction. □

Lemma 2.10. *A mincut is nested with all but finitely many mincuts.*

Proof. Suppose K is a mincut and L is a mincut not nested with K. By Lemma 2.8 both K, L are B-cuts, with components C_1, C_2 of $\Gamma - K$ and D_1, D_2 of $\Gamma - L$. If $L \cap C_1$ was empty then by connectedness $C_1 \subseteq D_1$ or $C_1 \subseteq D_2$, both of which would imply that K, L were nested. Similarly none of $C_2 \cap L$, $D_1 \cap K$, $D_2 \cap K$ are empty. Then L is a tight x-y-separator for some $x \in K \cap D_1$, $y \in K \cap D_2$. There are only finitely many such separators for each pair x, y and only finitely many elements of K, so only finitely many such L are possible. □

3. Crossing cuts

The complexity in the structure of mincuts comes from so-called "crossing" cuts, which we now define.

Definition 3.1. Let K, L be mincuts. Let $\mathcal{E}$ be the set of ends of Γ, and let $\mathcal{E} = K^{(1)} \sqcup K^{(2)} \sqcup \cdots \sqcup K^{(r)}$, $\mathcal{E} = L^{(1)} \sqcup L^{(2)} \sqcup \cdots \sqcup L^{(s)}$ be the partitions of $\mathcal{E}$ given by K, L respectively. We say $[K], [L]$ *cross* if, possibly after relabelling, $K^{(i)} \cap L^{(j)} \neq \varnothing$ for $i, j = 1, 2$. We write $K + L$.

The following is a direct consequence of Lemma 2.8, having removed slices from our graph.

Lemma 3.2. *If* $[K], [L]$ *cross then* $\Gamma - K$, $\Gamma - L$ *have exactly two components.*

Later we will show that crossing mincuts possess a cyclic structure. Initially, however, we shall just consider two or three crossing cuts.

Lemma 3.3. *Let* $[K], [L]$ *be crossing classes of mincuts. Let* $\Gamma - K = C_1 \sqcup C_2$, $\Gamma - L = D_1 \sqcup D_2$. *Then* $|C_1 \cap L| = |C_2 \cap L| = |D_1 \cap K| = |D_2 \cap K|$; *i.e.*, $K \cup L$ *splits into four equal pieces, plus the "centre"* $U = K \cap L$.

Proof. This follows from two applications of Lemma 2.2. □

In the case of edge cuts, one can also show that the centre $K \cap L$ is empty, but in the case of vertex cuts this fails to be true. As we will show in Lemma 3.5 below, the centre is in some sense distinguished, but this result must wait until we have placed some restrictions on the division of a graph produced by three cuts.

Let K, L, M be mincuts with K crossing L and L crossing M, and let $C_1 \sqcup C_2$, $D_1 \sqcup D_2$, $E_1 \sqcup E_2$ be the components of $\Gamma - K$, $\Gamma - L$, $\Gamma - M$ respectively. A priori, these three cuts could divide Γ into eight components each containing an end. The natural diagram with which to illustrate this would be a suitably divided cube. To produce this in 2D we divide the cube into three slices as shown in Figure 2.

We now rule out certain arrangements of ends of the graph.

Lemma 3.4. *It is not possible for each of* $C_1 \cap D_1 \cap E_1$, $C_2 \cap D_1 \cap E_2$, $C_2 \cap D_2 \cap E_1$, $C_1 \cap D_2 \cap E_2$ *to contain an end* (*or any arrangement obtained from this by relabellings*).

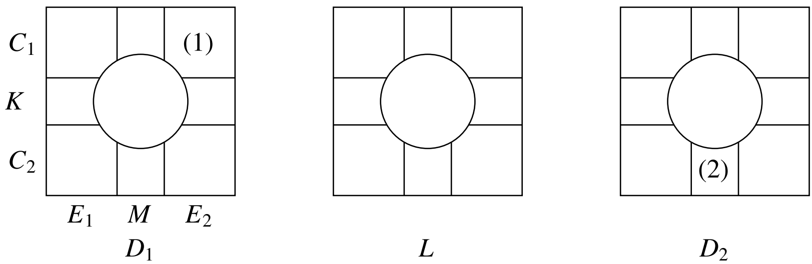

Figure 2. The notation shows that, for example, (1) is $C_1 \cap D_1 \cap E_2$ and (2) is $C_2 \cap D_2 \cap M$.

Proof. Denote by $a, \dots, u$ the cardinalities of the various subgraphs as shown in Figure 3; the ϵ_i indicate the presence of ends.

Let n be the cardinality of a mincut. Then $|K| = |L| = |M| = n$, so

$$n = a + c + l + j + m + p + r + t + u,$$
$$n = e + f + g + h + q + r + s + t + u,$$
$$n = b + d + i + k + m + p + q + s + u.$$

We also have an end cut separating each ϵ_i from the others; these yield, in order,

$$n \le a + b + e + m + q + t + u,$$
$$n \le c + d + g + m + r + s + u,$$
$$n \le k + l + h + p + s + t + u,$$
$$n \le i + j + f + p + q + r + u.$$

Summing these four, we have

$$\begin{aligned} 4n &\le (a + c + l + j + m + p + r + t + u) + (e + f + g + h + q + r + s + t + u) \\ &\qquad + (b + d + i + k + m + p + q + s + u) + u \\ &= 3n + u, \end{aligned}$$

whence $u = n$, everything else vanishes, and $K = L = M$, each separating the graph into at least four components, contradicting Lemma 3.4. □

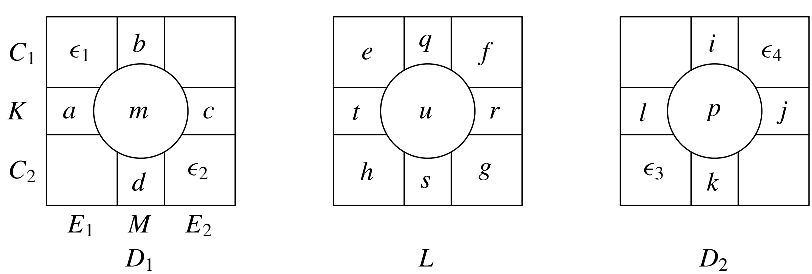

Figure 3. Diagram for Lemma 3.4.

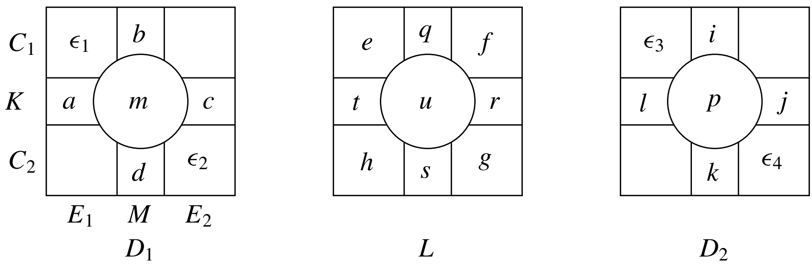

Figure 4. Diagram for Lemma 3.5.

Note that this result implies that the three cuts split the graph into at most six components containing ends. Since K crosses L, there are at least four such components. A quick exercise in filling in corners with ends subject to the crossings and Lemma 3.4 shows that, following relabellings, $C_1 \cap D_1 \cap E_1$, $C_2 \cap D_1 \cap E_2$, $C_1 \cap D_2 \cap E_1$, $C_2 \cap D_2 \cap E_2$ all contain ends (with possibly other corners also).

Lemma 3.5. *Let K, L, M be mincuts with K crossing L and L crossing M (in particular, if K is equivalent to M). Then $K \cap L = L \cap M$, and $L \cap C_1 = L \cap E_1$, $L \cap C_2 = L \cap E_2$.*

Proof. Retain the notations of the previous lemma and see Figure 4. Again since K, L, M are mincuts,

$$n = a + c + l + j + m + p + r + t + u,$$
$$n = e + f + g + h + q + r + s + t + u,$$
$$n = b + d + i + k + m + p + q + s + u,$$

and again, considering end cuts separating a corner containing an end ϵ_i from the others, we have

$$n \le a + b + e + m + q + t + u,$$
$$n \le c + d + g + m + r + s + u,$$
$$n \le i + l + e + p + q + t + u,$$
$$n \le j + k + g + p + r + s + u.$$

Summing these,

$$\begin{aligned}4n &\le (a + c + l + j + m + p + r + t + u) + (b + d + i + k + m + p + q + s + u) \\ &\qquad + 2e + 2g + q + r + s + t + 2u \\ &= 2(e + f + g + h + q + r + s + t + u) + 2n - 2f - 2h - q - r - s - t \\ &= 4n - (2f + 2h + q + r + s + t),\end{aligned}$$

whence $f = h = q = r = s = t = 0$, so that $K \cap L = L \cap M$. □

Lemma 3.6. *A cut is crossed by at most finitely many cuts.*

Proof. If two cuts cross they are not nested, so this lemma follows directly from Lemma 2.10. □

4. Half-cuts

It follows from the last section's results that each mincut in a crossing system can be decomposed into three pieces; two "half-cuts" and a "centre". We now prove some facts about these half-cuts, which enable us to arrange the half-cuts on a circle.

Definition 4.1. If K, M are mincuts (more properly, classes of mincuts under $\sim$, but we will often pass over this technicality), we write $K \# L$ if there are mincuts $K = L_0, L_1, \ldots, L_n = M$ such that $L_0 + L_1 + \cdots + L_n$; that is, L_0 crosses L_1, L_1 crosses L_2 and so on. L_0 may or may not cross L_2. Note that # is an equivalence relation on $\sim$-classes of mincuts, decomposing these into equivalence classes, which we call #-classes.

By Lemma 3.5, elements K of a #-class have a unique decomposition $K = K_1 \cup U \cup K_2$, where if $K + L$ then $K \cap L = U$ and K_1, K_2 are in different components of $\Gamma - L$. From the same lemma, this U is the same for all cuts in the #-class; we call it the *centre* of the #-class. Also, $|K_1| = |K_2|$ and this cardinality is again the same across the class. The K_i are called the *half-cuts* of the #-class. We now prove a series of lemmas clarifying the structure of a #-class and its half-cuts.

Lemma 4.2. *If K, M are mincuts in the same #-class then either $K + M$ or there is an L in this #-class such that $K + L + M$; that is, $K + L$ and $L + M$.*

Proof. By definition we have a sequence of cuts $K = L_0, L_1, \ldots, L_n = M$ such that $L_0 + L_1 + \cdots + L_n$. Take a shortest such sequence, and suppose $n \geq 3$. We will show we can find a shorter sequence. Without loss of generality we can assume $n = 3$. Let $\mathcal{E} = L_i^{(1)} \cup L_i^{(2)}$ be the partition induced by L_i. The fact that K does not cross L_2, and similar facts, give us, after relabelling,

$$K^{(1)} \subseteq L_2^{(1)}, \quad L_2^{(2)} \subseteq K^{(2)},$$
$$M^{(2)} \subseteq L_1^{(2)}, \quad L_1^{(1)} \subseteq M^{(1)},$$

whence the crossings give us that each of $M^{(2)} \cap K^{(2)}$, $M^{(1)} \cap K^{(2)}$, $M^{(1)} \cap K^{(1)}$ is nonempty. Hence $K + M$ unless $M^{(2)} \cap K^{(1)}$ is empty, and hence $K^{(1)} \subseteq M^{(1)}$. It is this that allows us to place the ends ϵ_3, ϵ_6 in Figure 5, and hence to conclude that $K + (L_{11} \cup U \cup L_{22}) + M$, where for instance L_{11} is the half-cut of L_1 lying in the $K^{(1)}$-component of $\Gamma - K$. □

Corollary 4.3. *There are only finitely many cuts in a #-class and hence only finitely many half-cuts.*

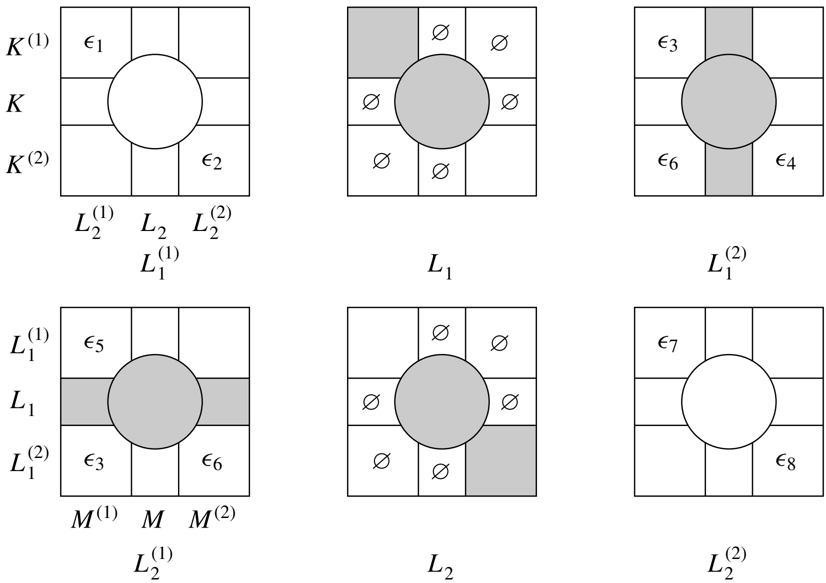

Figure 5. Diagram for Lemma 4.2. The cut $L_{11} \cup U \cup L_{22}$ is shown shaded in both diagrams.

Lemma 4.4. *Let K_1, M_1 be half-cuts in the same #-class. Then $K_1 \cup U \cup M_1$ is a mincut if and only if there are mincuts K', M' containing K_1, M_1 as half-cuts respectively such that $K' + M'$.*

Proof. One direction is clear. For the other, pick K_2, M_2 such that $K = K_1 \cup U \cup K_2$, $M = M_1 \cup U \cup M_2$ are cuts in this #-class. Then either $K + M$, in which case we are done, or there is L such that $K + L + M$. Now $K_1 \cup U \cup M_1$ is a cut; hence we have ϵ_5 in Figure 6. Then $K_1 \cup U \cup M_2$, $K_2 \cup U \cup M_1$ cross. $\square$

Definition 4.5. Two half-cuts K_1, L_1 in the same #-class are *equivalent* if whenever K_2 is a half-cut such that $K_1 \cup U \cup K_2$ is a mincut, then $L_1 \cup U \cup K_2$ is an equivalent cut and vice versa.

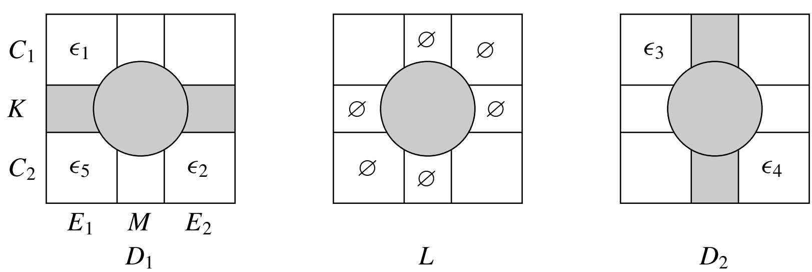

Figure 6. Diagram for Lemma 4.4. The cut $K_1 \cup U \cup M_2$ is shown shaded.

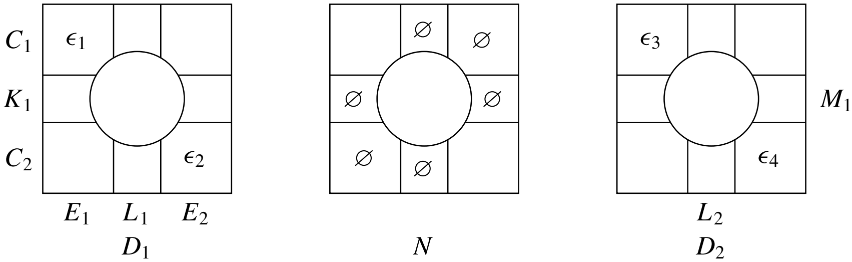

Figure 7. Diagram for Lemma 4.6.

Two half-cuts K_1, L_1 in the same #-class are *quasiequivalent* if there is a half-cut K_2 such that $K_1 \cup U \cup K_2$ is a mincut and $L_1 \cup U \cup K_2$ is an equivalent cut.

Lemma 4.6. *If two half-cuts K_1, M_1 form a cut then they are not quasiequivalent.*

Proof. Let $K = K_1 \cup U \cup M_1$ be the cut formed by hypothesis. Let L_1 be some other half-cut; we will show that $K_1 \cup U \cup L_1$ is not equivalent to $L_1 \cup U \cup M_1$ as cuts, and hence that K_1, M_1 are not quasiequivalent. Let L_2 be a half-cut such that $L = L_1 \cup U \cup L_2$ is in the #-class. If $L + K$ then the result is clear. If not, there is a mincut N such that $K + N + L$; without loss of generality take K_1, L_1 to be in the same component of $\Gamma - N$. Then $L_1 \cup U \cup M_1$ is a cut, and from Figure 7 we see that either $K_1 \cup U \cup L_1$ is not an end cut or it is not equivalent to $L_1 \cup U \cup M_1$. □

Lemma 4.7. *Let $K = K_1 \cup U \cup K_2$ be a cut in the #-class and let M_1 be a half-cut in the same class not quasiequivalent to either K_1, K_2. Then there is M_2 such that $M_1 \cup U \cup M_2$ is a cut crossing K. Hence $K_1 \cup U \cup M_1$ and $K_2 \cup U \cup M_2$ are cuts.*

Proof. Let M_2 be a half-cut such that $M = M_1 \cup U \cup M_2$ is a cut of the class; see Figure 8. If $K + M$, we are done. Otherwise, there is a cut L with $K + L + M$. After possibly relabelling the K_i, we can assume that K_1, M_1 are in the same component of $\Gamma - L$. If there is an end in $C_1 \cap D_1 \cap E_2$ then $M_1 \cup U \cup L_2$ is a cut crossing K. If there is an end in $C_2 \cap D_1 \cap E_1$ then $M_1 \cup U \cup L_1$ is a cut crossing K. If neither of these happens, then $K_2 \cup U \cup M_1$ is equivalent to $K_2 \cup U \cup K_1$, a contradiction

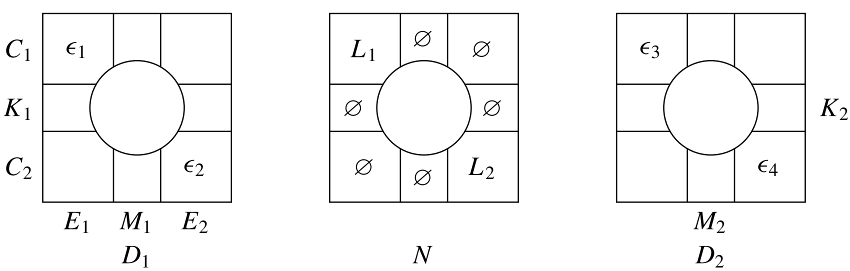

Figure 8. Diagram for Lemma 4.7.

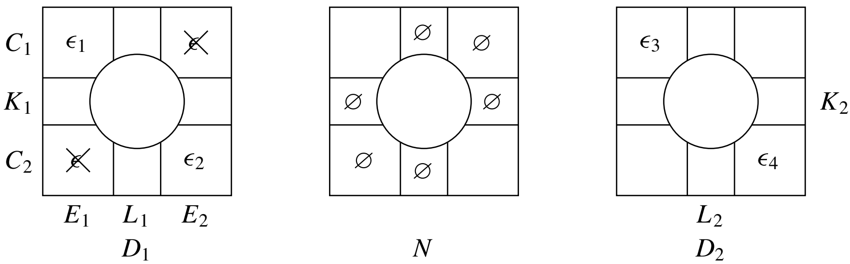

Figure 9. Diagram for Lemma 4.8.

(we note that these cuts are genuinely equivalent, since the presence of "links" such as L_1 guarantees that ends which appear to be connected up really are). □

Lemma 4.8. *Quasiequivalence is an equivalence relation. If K_1, L_1 are quasiequivalent and L_2 is a half-cut such that $L_1 \cup U \cup L_2$ is in the #-class then $K_1 \cup U \cup L_2 \sim L_1 \cup U \cup L_2$.*

Proof. Let K_2 be such that $K = K_1 \cup U \cup K_2$ is in the #-class; see Figure 9. The cut K does not cross $L = L_1 \cup U \cup L_2$ since in this case $K_1 \cup U \cup L_1$ would be a cut, so K_1, L_1 are not quasiequivalent by Lemma 4.6, giving a contradiction. Then there is an N such that $K + N + L$. Again, $K_1 \cup U \cup L_1$ is not a cut, so there are no ends in certain corners as indicated. Then $K_1 \cup U \cup L_2 \sim L_1 \cup U \cup L_2$ as required, noting again that ends which appear connected actually are so that the cuts are genuinely equivalent.

As for quasiequivalence being an equivalence relation, it is clearly symmetric and reflexive. If M_1 is another half-cut quasiequivalent to L_1, then by the above

$$K_1 \cup U \cup L_2 \sim L_1 \cup U \cup L_2 \sim M_1 \cup U \cup L_2,$$

so K_1, M_1 are quasiequivalent. □

Lemma 4.9. *Let $K = K_1 \cup U \cup K_2$ be a cut in the #-class and let L_1, M_1 be half-cuts in the same class not quasiequivalent to K_1, K_2. Then either $L_1 \cup U \cup M_1$ is a cut crossing K or L_1, M_1 are contained in the same component of $\Gamma - K$.*

Proof. By Lemma 4.7, we can complete L_1 to a cut crossing K, so that L_1 separates some ends of a component of $\Gamma - K$, and it works similarly for M_1. If the two half-cuts are in different components, then $L_1 \cup U \cup M_1$ is a cut crossing K provided it separates $\Gamma - K$ into two components, as indeed it must. □

5. Separation systems

We now turn our attention to demonstrating that the half-cuts of a system have a cyclic structure. We will do this by showing that they satisfy a certain axiomatic system, which implies that they can be arranged cyclically in a fashion compatible with their cut structure. This axiomatic structure is taken from [Huntington 1935].

Definition 5.1. A separation relation on a set Z is a relation $R \subseteq Z^4$ satisfying the following axioms. We write $abcd$ if $(a, b, c, d) \in R$.

(1) If $abcd$ then a, b, c, d are distinct.

(2) There are a, b, c, d such that $abcd$, i.e., $R \neq \varnothing$.

(3) If $abcd$ then $bcda$.

(4) If $abcd$ then $\neg(abdc)$.

(5) There are a, b, c, d such that $abcd$ and $dcba$.

(6) If $abcd$ and $x \in Z$ is another element then either $axcd$ or $abcx$.

Lemma 5.2. *Let Z be a set equipped with a separation relation. Then*

(1) *if $a, b, c, d \in Z$ are distinct, then at least one of the twenty-four tetrads $abcd, abdc, \ldots, dcba$ is true,*

(2) *if $abcd$ then $dcba$,*

(3) *if $abxc$ and $abcy$ then $abxy$,*

(4) *if $abcx$ and $abcy$ then $abxy$ or $abyx$,*

(5) *if $abcx$ and $abcy$ then $acxy$ or $acyx$,*

where in the last three statements distinct letters are assumed to represent different elements of Z.

Proofs can be found in [Huntington and Rosinger 1932] along with further similar propositions.

Lemma 5.3. *Let Z be a finite set with a separation relation. For each z, there are unique a, b such that for all $c \in Z - \{z, a, b\}$, we have $azbc$. We call these the elements adjacent to z.*

Proof. We approach existence by induction. For $|Z| = 4$, the result is trivial. Assume it is true for all separation relations with $|Z| = n$, and suppose $|Z| = n + 1$. Remove an element d of Z not equal to z to leave a smaller separation relation, and let a, b be the elements adjacent to z in this new relation, so that for all $c \in Z - \{z, a, b, d\}$, we have $azbc$.

By Lemma 5.2, one of $azbd$, $adzb$, $azdb$ holds. If $azbd$ holds then a, b are adjacent to z in Z. If not, without loss of generality, $azdb$. We claim a, d are adjacent to z in Z. By Lemma 5.2 above, if $c \in Z - \{z, a, d, b\}$ then $azdb$ and $azbc$ imply $azdc$.

For the uniqueness part, suppose there are two such pairs a_1, b_1, a_2, b_2. If any of these coincide we have an immediate contradiction to part (4) of the definition of the relation. So suppose they are all distinct. Then $a_1zb_1a_2$, $a_1zb_1b_2$ imply $a_1za_2b_2$ or $a_1zb_2a_2$, both of which contradict $a_2zb_2a_1$. □

Lemma 5.4. *Let Z be a finite set equipped with a separation relation. Then there is a map $F : Z \to S^1$ such that for $a, b, c, d \in Z$, $abcd$ if and only if $F(b)$ and $F(d)$ lie in different components of $S^1 - \{F(a), F(c)\}$; i.e., Z is isomorphic to a finite subset of the circle under its natural separation relation.*

Proof. We will proceed by induction. Pick an element $z \in Z$ and take a separation-preserving map $\bar{F} : Z - \{z\} \to S^1$. By the previous lemma, there are elements a, b of Z adjacent to z. We will map z to the circle by placing it between $\bar{F}(a)$, $\bar{F}(b)$, but first we must show these are adjacent on the circle. If not, there are c, d so that $\bar{F}(a)\bar{F}(c)\bar{F}(b)\bar{F}(d)$, whence $acbd$. But $azbc$, $azbd$ imply $abcd$ or $abdc$, both contradicting $acbd$. So $\bar{F}(a)$, $\bar{F}(b)$ are adjacent on the circle, and we can define $F : Z \to S^1$ by setting $F = \bar{F}$ on $Z - \{z\}$ and $F(z)$ to lie between $\bar{F}(a)$, $\bar{F}(b)$ on the circle.

A full proof that this F works would be lengthy and uninformative, so we just indicate the main steps; the remainder is just use of axioms and Lemma 5.2. We inherit from $\bar{F}$ that any relations not involving z are preserved. Let $zABC$ be another relation and suppose A, B, C are distinct from a, b; the other cases are easier. Then we have $azbA$, $azbB$, $azbC$, $zABC$ from which we deduce $aABC$, $bABC$. These relations carry over to the circle under F, as do $azbA$, $azbB$, $azbC$ by construction. From these relations on the circle, we then find $F(z)F(A)F(B)F(C)$. □

Definition 5.5. Let Z be the set of quasiequivalence classes of half-cuts of a #-class. We define a separation relation R on Z by setting $(a, b, c, d) \in R$ if and only if $ac + bd$, where ac denotes the cut $K_1 \cup U \cup L_1$ and K_1, L_1 are representatives of a, c and so on.

Lemma 5.6. *This is well-defined; i.e., it does not matter which representatives of quasiequivalence classes we choose. Furthermore, it is a bona fide separation relation.*

Proof. Well-definedness follows immediately from Lemma 4.8. Parts (1)–(5) of the definition of a separation relation are trivial. For part (6), by Lemma 4.9 either $abcx$ or b, x are in the same component of $\Gamma - ac$; and either $axcd$ or x, d are in the same component of $\Gamma - ac$. But b, d are in different components of $\Gamma - ac$ so one of $abcx$, $axcd$ holds. □

Hence we have:

Proposition 5.7. *To each #-class we can associate a cycle where each vertex represents a quasiequivalence class and each cut of the #-class is associated to a vertex cut of the cycle, with the notions of crossing preserved.*

6. The structure of a #-class

We are now in a position to characterize the structure of a #-class. Let $[K_1]_q$ denote the quasiequivalence class of a half-cut K_1. From Proposition 5.7, there are two quasiequivalence classes adjacent to $[K_1]_q$ in this #-class. If L_1 is a half-cut in the #-class not in $[K_1]_q$ or either quasiequivalence class adjacent to it, then Lemma 4.8 implies that

$$K_1 \cup U \cup L_1 \sim K_1' \cup U \cup L_1$$

for all $K_1' \in [K_1]_q$. So only the two quasiequivalence classes adjacent to $[K_1]_q$ can contain L_1 such that $K = K_1 \cup U \cup L_1$ and $K' = K_1' \cup U \cup L_1$ are not equivalent for $K_1' \in [K_1]_q$.

How can these cuts be nonequivalent? We recall that by minimality every component left by a mincut is connected to every element of that cut. Thus in the "larger component" left by the cut, i.e, the one containing half-cuts in the same class, every vertex is connected to the half-cuts in this "component", which is thus genuinely connected. Thus one part of the partitions induced by K, K' is the same. The others can only differ if at least one of the cuts splits Γ into more than two parts, and hence splits the "smaller" component into more than one part. Suppose K intersects one of the smaller components of $\Gamma - K'$. Then each end not in the larger component of $\Gamma - K'$ is connected to each vertex of the part of K in the smaller component; hence K' splits Γ into exactly two components. If K does not intersect one of the smaller components of $\Gamma - K'$, then since $K \neq K'$ and K, K' have the same cardinality, K' intersects one of the smaller components of $\Gamma - K$, whence K splits Γ into exactly two components.

Hence, having chosen L_1, there are at most two equivalence classes of cuts formed from L_1 and $[K_1]_q$. By symmetry, there are at most two equivalence classes of cuts formed from K_1 and $[L_1]_q$. From these discussions, it follows that for each quasiequivalence class adjacent to $[K_1]_q$, there are at most two equivalence classes of cuts formed by these two classes; one producing a split of Γ into two components, the other more. Hence there are at most four equivalence classes of half-cuts within $[K_1]_q$.

We now define the structure by which we model the #-class. For edge cuts this would be a simple cycle, but here we need extra complexity to deal with the possibility of splitting the graph into more than two components.

Definition 6.1. A *ring* is constructed as follows. Take a finite cycle of vertices and attach to each edge some number of triangles by identifying an edge of the triangle with the edge of the cycle. The vertex of the triangle not included in the original cycle is called an *anchor*; see Figure 11.

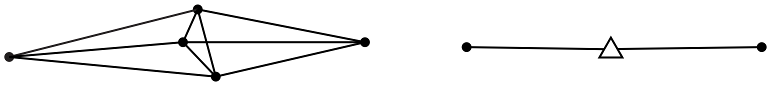

Figure 10. A 3-vertex connected to two 1-vertices, and the schematic representation of this.

Definition 6.2. An *n-vertex* will be a copy of the complete graph on n vertices; we will say it is connected to a vertex if there is an edge from the vertex to each constituent vertex of the n-vertex. We will depict a 3-vertex as a triangle and only draw one edge from it to each vertex to which it is connected; see Figure 10.

We now associate to each #-class an appropriate ring encoding the cuts formed by half-cuts in the class. First use Proposition 5.7 to form a cycle with one vertex for each quasiequivalence class. For each pair of adjacent quasiequivalence classes, find half-cuts in those classes separating Γ into as many components as possible, and attach one fewer anchors than this between the two classes in the cycle (one fewer to account for the "large" component). If a quasiequivalence class contains more than one equivalence class, insert an extra vertex into the cycle here. If we

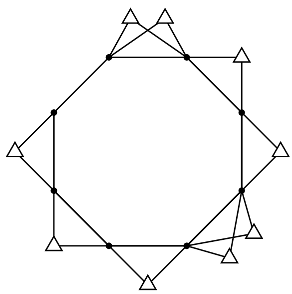

Figure 11. A ring, with the anchors replaced by 3-vertices.

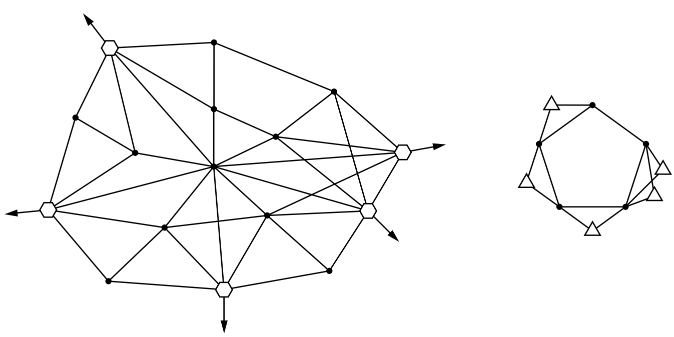

Figure 12. A #-class and its associated ring. Hexagons represent 6-vertices and arrows ends.

“thicken” up the anchors to 3-vertices to remove cut-points, there is now a bijective correspondence between equivalence classes of cuts formed from half-cuts of the #-class and equivalence classes of cuts of the ring, where we treat the anchors as ends for the purpose of equivalence etc.; see Figure 12.

7. Pretrees

We now proceed towards the central theorem of the paper. First we seek to impose a tree structure on the #-classes and the other cuts; then we will reintroduce the extra complexity. We will do this using pretrees, which we now define.

Definition 7.1. Let $\mathcal{P}$ be a set and let $R \subseteq \mathcal{P}^3$ be a ternary relation on $\mathcal{P}$. If $(x, y, z) \in R$ then we write xyz and say y is between x, z. A set $\mathcal{P}$ equipped with this relation is a *pretree* if the following hold:

(1) If xyz then $y \neq x, z$, and there are no x, y such that xyx.

(2) If xyz then zyx.

(3) For all x, y, z, if xyz then $\neg(xzy)$.

(4) If xzy and $w \neq x, y, z$ then xzw or yzw.

If there is no z such that xzy we say x, y are adjacent.

A pretree is called *discrete* if for any $x, y \in \mathcal{P}$ there are at most finitely many z such that xzy.

It should perhaps be noted that despite us using the word “between” this is not a betweenness relation in the usual sense of the word as, for example, in [Huntington 1935]. Let $\mathcal{P}$ be a discrete pretree. We will describe briefly how to pass from $\mathcal{P}$ to a tree; a fuller description may be found in [Bowditch 1999].

We call a subset H of $\mathcal{P}$ a *star* if all $a, b \in H$ are adjacent. We now define a tree T as follows:

$$V(T) = \mathcal{P} \cup \{\text{maximal stars of } \mathcal{P}\},$$
$$E(T) = \{(v, H) : v \in \mathcal{P}, v \in H, H \text{ a maximal star}\}.$$

We show that T is indeed a tree. If $x, y \in \mathcal{P}$ then by discreteness there are only finitely many z between x, y. From among these z we can then find $z_1, \ldots, z_n$ such that x is adjacent to z_1, z_i is adjacent to z_{i+1} and z_n is adjacent to y, giving a path in T from x to y. Hence T is connected.

If T contains a circuit then there are $x_1, \ldots, x_n$ in $\mathcal{P}$ such that x_i is adjacent to x_{i+1} but not to x_{i+2} for each $i \in \mathbb{Z}_n$. Then there is y such that $x_i y x_{i+2}$. If $y \neq x_{n+1}$ then either $x_i y x_{i+1}$ or $x_{i+1} y x_{i+2}$, both of which are forbidden. So $x_i x_{i+1} x_{i+2}$. We claim $x_1 x_i x_{i+1}$ holds for all $i \leq n$ by induction. Since $x_1 x_{i-1} x_i$ and $x_{i-1} \neq x_{i+1}$, either $x_1 x_{i-1} x_{i+1}$ holds or we have a contradiction. Since $x_{i-1} x_i x_{i+1}$ and $x_1 \neq x_i$,

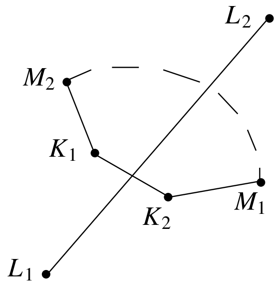

Figure 13. Diagram for Lemma 7.3.

either $x_1 x_i x_{i+1}$ or $x_{i-1} x_i x_1$; so to avoid contradiction, $x_1 x_i x_{i+1}$. But then we have $x_1 x_{n-1} x_n$; but x_1, x_n were supposed to be adjacent. The contradiction means T is a tree.

We now prove some lemmas which will allow us to define a pretree of cut classes.

Definition 7.2. We call a mincut *isolated* if it does not cross any mincut, and hence is not contained in any #-class.

A cut is a *corner cut* of a #-class if it is (equivalent to) a cut formed from two half-cuts of the class but is not itself in the class. We call a mincut *totally isolated* if it does not cross any mincut and is not a corner cut of any #-class.

Lemma 7.3. *Corner cuts are isolated.*

Proof. Let Q be a #-class and let $K = K_1 \cup U \cup K_2$ be a corner cut of Q. Suppose there is a cut L with $K + L$. Then L separates some ends of each component of $\Gamma - K$. Let M_1, M_2 be half-cuts in Q adjacent to K, with K_1 adjacent to M_2 and K_2 adjacent to M_1, with no quasiequivalences present; see Figure 13. Either L crosses $K_1 \cup U \cup M_1$ or all ends of the component of $\Gamma - K_1 \cup U \cup M_1$ containing M_2 are in the same component of $\Gamma - L$, whence L crosses $K_2 \cup U \cup M_2$. So L, hence K, are in the #-class Q, a contradiction. □

Each #-class Q induces two partitions

$$\mathcal{E} = Q^{(1)} \sqcup \cdots \sqcup Q^{(m)},$$
$$\mathcal{E} = \bar{Q}^{(1)} \sqcup \cdots \sqcup \bar{Q}^{(m')}$$

of the ends of Γ. In one partition, which we call the fine partition and denote without bars, each member of the partition corresonds to one of the anchors in the ring representing Q; and for each $Q^{(i)}$, there is a corner cut of Q separating $Q^{(i)}$ from all the other $Q^{(j)}$. For the other partition, the coarse partition, we identify those $Q^{(i)}$ together which lie between the same two adjacent half-cuts. Then in the coarse partition we can distinguish between members $\bar{Q}^{(i)}$ using only cuts properly in the #-class Q; for the fine partition, we may need corner cuts also. We recall also that a cut K also gives a partition of the ends of Γ:

$$\mathcal{E} = K^{(1)} \sqcup \cdots \sqcup K^{(n)}.$$

Lemma 7.4. *Given a cut K and a #-class Q, with K neither in Q nor a corner cut of it, there are i, j such that all $Q^{(k)}$ except $Q^{(i)}$ are contained in $K^{(j)}$; i.e.,*

$$\coprod_{k \neq i} Q^{(k)} \subseteq K^{(j)}$$

or

$$\coprod_{k \neq j} K^{(k)} \subseteq Q^{(i)}.$$

We say K divides *$Q^{(i)}$.*

Proof. Suppose K is an A-cut. Then it is nested with every cut and corner cut of Q, hence the result.

Otherwise K is a B-cut, separating Γ into two components. If the result is not true, then both $K^{(i)}$ intersect at least two $Q^{(i)}$.

Suppose a $Q^{(i)}$ intersects both $K^{(i)}$. Let M be the corner cut of Q splitting off $Q^{(i)}$. If M is a B-cut, then $K + M$, giving a contradiction. Otherwise, M is nested with K, whence a $K^{(i)}$ is contained in a $Q^{(i)}$, again giving a contradiction.

If both $K^{(i)}$ contain two $Q^{(i)}$ not between two adjacent half-cuts (in the ring representing Q), we can find a cut of Q crossing K. So for say $K^{(1)}$, all the $Q^{(i)}$ contained in $K^{(1)}$ lie between two adjacent half-cuts of Q. Let M be the corner cut corresponding to these half-cuts.

M is necessarily an A-cut, and hence is nested with K. As in the discussion of quasiequivalent cuts earlier, K can only intersect the "large" component of $\Gamma - M$, that containing the other half-cuts of Q; conversely M does not intersect the large component C_1 of $\Gamma - K$. Pick another half-cut L_1 in Q, with $L = M_1 \cup U \cup L_1$ a cut of Q. The cut L also only intersects the large component E_1 of $\Gamma - M$. With suitable labelling of the $L^{(i)}$, we have

$$L^{(1)} \subseteq M^{(1)}, \quad L^{(1)} \subseteq K^{(1)}, \quad K^{(2)} \subseteq L^{(2)}, \quad K^{(1)} = M^{(1)}.$$

Hence we have the arrangement shown in Figure 14.

Then, using the notations from Figure 14, we have

$$\begin{aligned} a+m+p+l+t+u &= n, \\ d+m+p+k+s+u &= n, \\ e+t+u+h+s &= n, \\ a+m+e+t+u &\geq n, \\ p+l+e+t+u &\geq n, \\ p+k+u+s &\geq n, \end{aligned}$$

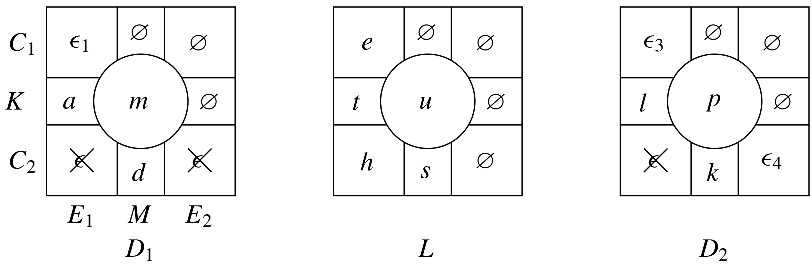

Figure 14. Diagram for Lemma 7.4.

where n is the cardinality of a mincut. Immediately $d = m = 0$. Furthermore, since L_2, M_1, M_2 are half-cuts in the same class, $s = e + t + h$. Then

$$\begin{aligned} 2n &\le (a+e+t+u)+(p+l+e+t+u) \\ &= (a+p+l+t+u)+(e+e+t+u) \\ &\le n+e+t+h+s+u \\ &= 2n. \end{aligned}$$

Then all the inequalities are equalities; hence $t = h = 0$, and K decomposes into U together with two equal half-cuts; and choosing L_1 appropriately we find that these half-cuts are quasiequivalent to the half-cuts of M, so K was a corner cut of Q. □

Lemma 7.5. *Given two #-classes Q, R, all cuts in R divide the same $Q^{(i)}$.*

Proof. Note first that the cuts in R do divide a $Q^{(i)}$ because they are not isolated, hence not corner cuts, and are not in Q. Suppose $K \in R$ divides $Q^{(i)}$ and $L \in R$ divides $Q^{(j)}$, with $i \neq j$; see Figure 15.

If there is one, take a cut M crossing K and L. We have $K^{(2)} \subseteq Q^{(i)}$ and $L^{(2)} \subseteq Q^{(j)}$, so M contradicts Lemma 7.4.

Then $K \# L$. Take a cut $M \in Q$ separating $Q^{(i)}$ from $Q^{(j)}$, and let $N = K_2 \cup U \cup L_2$. The cut N separates some ends of $Q^{(i)}$ and $Q^{(j)}$; it is not nested with M, and hence is a B-cut and crosses M, giving a contradiction. □

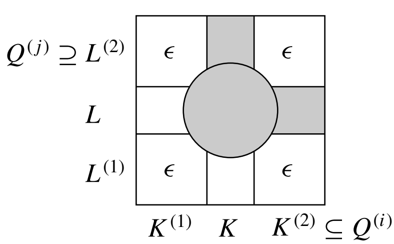

Figure 15. Diagram for Lemma 7.5. The cut N is shown shaded.

Lemma 7.6. *Given two totally isolated cuts K, L, we have that K divides only one $L^{(i)}$; i.e., there are i, j such that*

$$\coprod_{k\neq i} L^{(k)} \subseteq K^{(j)}, \quad \coprod_{k\neq j} K^{(k)} \subseteq L^{(i)}.$$

Proof. If K, L are nested, the result is immediate. If not, they are both B-cuts, and the result follows since they do not cross. □

We now define a pretree encoding the mincuts of Γ. Let $\mathcal{P}$ be the set of all #-classes of Γ and all equivalence classes of totally isolated cuts of Γ. Given $x, y, z \in \mathcal{P}$, we say y is between x, z if the cuts in x, z divide different elements of the coarse partition of $\mathcal{E}$ induced by y, and y is not equal to x, z.

Lemma 7.7. *This relation defines a pretree.*

Proof. Let

$$\mathcal{E} = x^{(1)} \sqcup \cdots \sqcup x^{(n_x)},$$
$$\mathcal{E} = y^{(1)} \sqcup \cdots \sqcup y^{(n_y)},$$
$$\mathcal{E} = z^{(1)} \sqcup \cdots \sqcup z^{(n_z)}$$

be the coarse partitions of the ends of Γ induced by x, y, z. First we check that the definition makes sense; i.e., given $x, y \in \mathcal{P}$, there are unique i, j with

$$\coprod_{k\neq i} x^{(k)} \subseteq y^{(j)}.$$

If one of x, y is an equivalence class of totally isolated cuts, then Lemmas 7.4 and 7.6 yield this. Suppose both are #-classes Q, R. By Lemma 7.5, given $K \in R$ there is $Q^{(i)}$ such that

$$K^{(2)} \subseteq Q^{(i)}.$$

$Q^{(i)}$ is contained in a $\bar{Q}^{(i)}$, so

$$K^{(2)} \subseteq \bar{Q}^{(i)},$$

and furthermore this $\bar{Q}^{(i)}$ is independent of the cut K chosen. For each j, j', we can find $K \in R$ with $\bar{R}^{(j)}$, $\bar{R}^{(j')}$ in different $K^{(k)}$ since we are using the coarse partition, whence one of $\bar{R}^{(j)}$, $\bar{R}^{(j')}$ is contained in $\bar{Q}^{(i)}$. Hence all but one $\bar{R}^{(j)}$ is contained in $\bar{Q}^{(i)}$; i.e.,

$$\coprod_{k\neq i} \bar{R}^{(k)} \subseteq \bar{Q}^{(i)}.$$

For part (1) of the definition of a pretree, note that if

$$\coprod_{k\neq i_1} x^{(k)} \subseteq y^{(j_1)}, \quad \coprod_{k\neq i_2} x^{(k)} \subseteq y^{(j_2)},$$

with $j_1 \neq j_2$, then since $y^{(j_1)}$, $y^{(j_2)}$ are disjoint, $n_x = 2 = n_y$ and x, y are equivalent cuts, and hence are equal as elements of $\mathcal{P}$. So xyx does not hold.

Part (2) is trivial. For part (3), after relabelling we have

$$\coprod_{k\neq 1} y^{(k)} \subseteq x^{(1)}, \quad \coprod_{k\neq 1} x^{(k)} \subseteq y^{(1)},$$
$$\coprod_{k\neq 2} y^{(k)} \subseteq z^{(1)}, \quad \coprod_{k\neq 1} z^{(k)} \subseteq y^{(2)}.$$

Then $y^{(1)} \cap y^{(2)} = \varnothing$ implies $x^{(i)} \cap z^{(j)} = \varnothing$ for $i, j \neq 1$. Hence

$$\coprod_{k\neq 1} x^{(k)} \subseteq z^{(1)},$$

so x, y divide the same $z^{(k)}$, and thus xzy does not hold.

For part (4), suppose that xzy so that

$$\coprod_{k\neq 1} z^{(k)} \subseteq x^{(1)}, \quad \coprod_{k\neq 2} z^{(k)} \subseteq y^{(1)};$$

i.e., x divides $z^{(1)}$, y divides $z^{(2)}$. If $w \neq z$ then w divides a unique $z^{(i)}$. If $i = 1$ then yzw. If not, then xzw. □

We recall the vertex version of Menger's theorem (see, for instance, [Bondy and Murty 2008, Theorem 9.1]):

Menger's theorem. *Let Γ be a graph and a, b be vertices of Γ. Then the minimum size of a vertex cut separating a, b is equal to the maximum number of vertex-independent simple paths joining a, b.*

Lemma 7.8. *This pretree is discrete.*

Proof. Let K, L, M be mincuts with M between K, L. Elements of $\mathcal{P}$ are of course not cuts; take a representative cut of any equivalence class or an appropriate corner cut of a #-class. By Lemma 2.10, only finitely many cuts are not nested with both K, L, so we need only consider the case when M is nested with both. Let C_i, D_i, E_i denote components of $\Gamma - K$, $\Gamma - L$, $\Gamma - M$ respectively. We have that K is nested with M, so (after relabelling if necessary) $C_1 \subseteq E_1$, and similarly $D_1 \subseteq E_2$. We have $E_1 \neq E_2$ as M is between K, L. By the remarks following Definition 2.6, E_1 is contained in a component D_i, whence K, L are nested.

If we now form a new graph by collapsing both of K, L to a single vertex and apply Menger's theorem in this graph, we obtain n vertex-independent paths from K to L, where n is the cardinality of a mincut. In the case when K, L are not disjoint, some of these paths collapse into points. The cut M must intersect each of

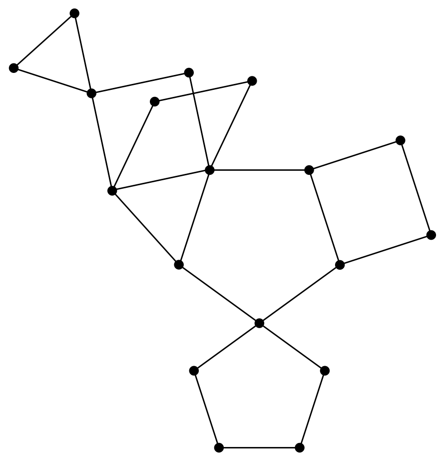

Figure 16. A succulent.

these paths as it separates K, L, and $|M| = n$, so M is contained in the union of these paths. Then there are only finitely many choices for M.

If we took different choices for K, L the only additional choices for M would be equivalent in $\mathcal{P}$ to some already considered. So the pretree is discrete. $\square$

We now have a discrete pretree $\mathcal{P}$, which as discussed above gives us a tree encoding the mincuts of Γ and how they interact with the ends of Γ.

8. Succulents

We have now obtained a tree encoding the cuts of the graph, with #-classes collapsed down to points. We now seek to reintroduce the cyclic structure of these in order to obtain the final "cactus" theorem. We will not be able to use cactus graphs as such; these work well for encoding edge cuts, but cannot represent a vertex cut yielding several components. We will therefore use a slightly more general structure which, for the sake of a horticultural joke, we call succulents.

Definition 8.1. A *succulent* is a connected graph built up from cycles (including possibly 2-cycles, consisting of two vertices joined by a double edge) in the following manner. Two cycles may be joined together either at a single vertex or along a single edge. The construction is tree-like in the sense that if we have a "cycle of cycles" $C_1, \dots, C_n$ with C_i attached to C_{i+1} (mod n) then all the C_i share a common edge/vertex. See Figure 16 for an example. The analogous property in a tree is that if we have a cycle of edges with each attached to the next one, they all meet at a common vertex. An *end vertex* of a succulent is one contained in only one cycle; a vertex of a succulent is an end vertex if it has at most two edges adjacent to it.

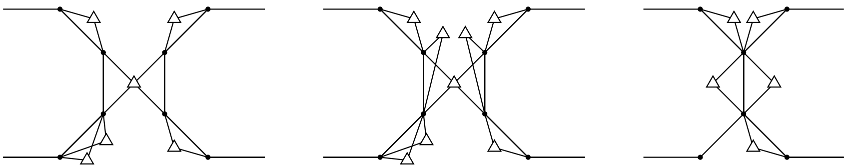

Figure 17. Diagrams illustrating how we connect rings around a star vertex if both corner cuts are B-cuts (left), if the corner cuts are nested and not equal (middle), and if they are equal (right).

We now construct a succulent encoding the mincuts of Γ. We already have the tree T from the previous section whose vertices are (equivalence classes of) totally isolated cuts and #-classes joined together via “star” vertices. There is at most one star for each corner cut of a #-class. If there is no corresponding star, then the components split off by this cut are not further subdivided by mincuts.

Before moving on further, we note that totally isolated cuts can be represented by a degenerate sort of ring, constructed by attaching triangles to a segment rather than to a cycle. So we can always talk about the anchors of a member of $\mathcal{P}$.

To form our succulent, we replace each member of $\mathcal{P}$ by its associated ring. We must now consider how we connect these; i.e., we need to consider the behaviour around each star vertex. Recall that if Q is a #-class attached to some star vertex, all members of $\mathcal{P}$ divide the same member of the (coarse) partition of the ends of Γ corresponding to Q, $\bar{Q}^{(1)}$ say, and that there is a corner cut of Q separating $\bar{Q}^{(1)}$ from the rest of the ends.

Suppose Q, R are #-classes adjacent to the same star vertex, so that there is no member of $\mathcal{P}$ between them. See Figure 17. Let K, L be the corresponding corner cuts and $\bar{Q}^{(1)}$, $\bar{R}^{(1)}$ the members of the coarse partition. If both K, L are B-cuts then each of $\bar{Q}^{(1)}$, $\bar{R}^{(1)}$ comprises only one $Q^{(i)}$, $R^{(i)}$ and there is only one member of the fine partitions divided by the other #-class. We join these classes by identifying the appropriate anchors. If there are no other elements of $\mathcal{P}$ joined to this star vertex then K, L are equivalent so we could further simplify things by removing the anchors and joining the cycles for Q, R together directly.

If one of K, L is not a B-cut then the two cuts are nested. Then either they are equal or all components except one of $\Gamma - L$ are contained in the same component of $\Gamma - K$ and vice versa. In the latter case, there is only one member of the fine partitions divided by the other #-class, so again we can represent this by identifying the appropriate anchors. If the corner cuts are equal, then we glue together the rings via the corner cuts. Of the anchors attached to each, one represents the other #-class and we simply delete this; the other anchors come in pairs, each representing the same set of ends but originating from different rings; we identify these together so we don’t get redundancy. Then to produce our succulent, we first glue together those rings sharing a corner cut, and then attach the other members of $\mathcal{P}$ adjacent to this star by identifying the appropriate anchors (for totally isolated cuts we simply note that the coarse and fine partitions coincide so there will be an obvious anchor to use and we have none of the issues above).

We must now show that this is a true succulent, that is, that we still have a tree-like structure. We inherit much of the tree-like nature from T; we only need check that no "cycles of cycles" form from the identifications made between rings all adjacent to the same star vertex. We will proceed by contradiction, supposing we have a shortest cycle of cycles $\mathcal{C}$ preventing our graph being a succulent. We can place limitations on which constituent cycles of the graph can be present in $\mathcal{C}$. First the cycles on which our rings are based do not appear. This is because any two cycles meeting one of these in the same $\bar{Q}^{(i)}$ intersect along an edge. So $\mathcal{C}$ consists of the triangles which contain anchors; these can be joined together either at an anchor or along the opposite edge. Because our cycle is shortest, we alternate between joins along edges and at anchors. Hence our cycle has at least four members. Let $T_1, \dots, T_4, \dots$ be the triangles in $\mathcal{C}$ with T_1, T_2 meeting at an anchor, T_2, T_3 at an edge and so on. By construction the points at the bases of the T_i represent cuts K_i of Γ partitioning the ends of Γ, and after suitable labelling we have

$$\coprod_{i\neq 1} K_1^{(1)} \subseteq K_2^{(1)} = K_3^{(1)},$$

$$\coprod_{i\neq 1} K_4^{(1)} \subseteq K_2^{(2)} = K_3^{(2)},$$

whence

$$\coprod_{i\neq 1} K_1^{(1)} \subseteq K_4^{(1)} \subseteq K_6^{(1)} \dots .$$

But $\mathcal{C}$ is a cycle, so we eventually come back to the start, whence all the inequalities become equalities, K_1 becomes a B-cut and

$$K_1^{(2)} = K_2^{(1)} = K_4^{(1)} = \dots .$$

We could have started at any other point, so the other K_i are also B-cuts and all of them are equivalent. Then $\mathcal{C}$ becomes trivial and we have indeed constructed a succulent.

We have now proved most of this:

Theorem 8.2. *Let Γ be a connected graph such that there are vertex end cuts of Γ with finite cardinality. There is a succulent $\mathcal{S}$ with the following properties*:

(1) *There is a subset A of vertices of $\mathcal{S}$ called the* anchors *of $\mathcal{S}$. If two anchors are adjacent, one of them is an end vertex of the graph. Every vertex of $\mathcal{S}$ not in A is adjacent to an anchor. We define an anchor cut of $\mathcal{S}$ to be a vertex cut*

containing no anchors which separates some anchors of $\mathcal{S}$. We say anchor cuts are equivalent if they partition A in the same way.

(2) *There is an onto map f from the ends of Γ to the union of the ends of $\mathcal{S}$ with the end vertices of $\mathcal{S}$ which are anchors.*

(3) *There is a bijective map g from equivalence classes of minimal end cuts of Γ to equivalence classes of minimal anchor cuts of $\mathcal{S}$ such that ends e_1, e_2 of Γ are separated by $[K]$ if and only if $f(e_1), f(e_2)$ are separated by $g([K])$.*

(4) *Any automorphism of Γ induces an automorphism of $\mathcal{S}$.*

Proof. We already have a succulent containing a representative of each mincut; i.e., we already have the map g. We now discuss how we modify the succulent to define the map f of the ends of Γ. Some issues arise because there may be ends of Γ which are distinguished from each other only by nonminimal cuts. If such ends exist, we will treat them as a single end for the present section; i.e., we will map them all to the same place using f. Let ϵ be an end of Γ. If there is a mincut K such that this end is the sole element of one of the sets $K^{(i)}$, then this mincut appears somewhere in the succulent either as a corner cut of a #-class or as a totally isolated cut and there is an end anchor of the succulent corresponding to this $K^{(i)}$; define $f(\epsilon)$ to be this anchor.

If not, there may be a sequence of $x_i \in \mathcal{P}$ with

$$x_1^{(1)} \supseteq x_2^{(1)} \supseteq \cdots \ni \epsilon.$$

This defines a ray in the tree T associated to $\mathcal{P}$, hence an end of that tree. There is a unique such end, since T is a tree so two ends can be separated using a single point, which we may take to be some $y \in \mathcal{P}$. But there is only one $y^{(i)}$ containing ϵ, so only one end will do. So we have an end of T, hence of the succulent, associated with ϵ; this is where we will map ϵ.

The remaining cases will correspond to ends which can only be split off by nonminimal cuts, which are not associated to some end of the tree T. To fit these into our succulent, we will essentially pretend that they can be split off by a mincut; we will add an element to $\mathcal{P}$ for each such end, inducing a partition

$$\mathcal{E} = \{\epsilon\} \cup (\mathcal{E} - \{\epsilon\}).$$

This member of $\mathcal{P}$ will not be between any two members of $\mathcal{P}$; and it does not disrupt the discreteness of $\mathcal{P}$ because an infinite set of betweenness in $\mathcal{P}$ would induce a descending sequence of partitions as above, so we would already have dealt with this end. So we have added an end vertex to the tree T. When constructing the succulent, the extra member of $\mathcal{P}$ will be modelled as two anchors joined by a double edge, one of which becomes attached to a relevant anchor in the succulent. The other anchor is an end anchor, which we define to be $f(\epsilon)$.

We now have the map f, which by construction interacts with g in the way stated; note that the extra anchors added in the third step above are never split off by an anchor cut of $\mathcal{S}$.

To see that f is onto, we note that any end anchors of $\mathcal{S}$ arise either as $f(\epsilon)$ in the third case above, or as part of a ring, where they correspond to some member of a partition of $\mathcal{E}$, whose members will be mapped there. Any ends of $\mathcal{S}$ arise from ends of T, hence from sequences of members of $\mathcal{P}$. From the vertices in the relevant cuts, we can construct a ray in Γ, giving an end that will be mapped to the end of $\mathcal{S}$.

Part (4) arises since an automorphism of Γ induces corresponding automorphisms of the cuts and ends of Γ, preserving crossings, nestings, equivalences; in short, all the information used to construct $\mathcal{S}$. $\square$

We make some remarks about the theorem. In part (3) we must say equivalence classes of cuts of $\mathcal{S}$ because we may have equivalent distinct cuts of $\mathcal{S}$; these arise if there are quasiequivalent, nonequivalent half-cuts in a #-class , whence there will be some equivalent cuts contained in the relevant ring; but this is not really a concern.

If we wish to obtain a graph in which we do not have to exclude anchors from cuts, we can replace each anchor with a 3-vertex and treat these as ends, so that the anchor cuts in the theorem become bona fide mincuts of the resulting graph $\mathcal{S}'$.

If we collapse the extra end anchors we added in the proof above onto the adjacent anchors, then we obtain a variant theorem:

Theorem 8.3. *Let Γ be a connected graph such that there are vertex end cuts of Γ with finite cardinality. There is a succulent $\mathcal{S}$ with the following properties*:

(1) *There is a subset A of vertices of $\mathcal{S}$ called the* anchors *of $\mathcal{S}$. No two anchors are adjacent, and every vertex of $\mathcal{S}$ not in A is adjacent to an anchor. We define an anchor cut of $\mathcal{S}$ to be a vertex cut containing no anchors which separates some anchors of $\mathcal{S}$. We say anchor cuts are equivalent if they partition A in the same way.*

(2) *There is a map f from the ends of Γ to the union of the ends of $\mathcal{S}$ with the anchors of $\mathcal{S}$.*

(3) *There is a bijective map g from equivalence classes of minimal end cuts of Γ to equivalence classes of minimal anchor cuts of $\mathcal{S}$ such that ends e_1, e_2 of Γ are separated by $[K]$ if and only if $f(e_1)$, $f(e_2)$ are separated by $g([K])$.*

(4) *Any automorphism of Γ induces an automorphism of $\mathcal{S}$.*

Consider a finite graph Γ. We call a set J of vertices of a graph Γ *n-inseparable* if $|J| \geq n+1$ and for any set K of vertices with $|K| \leq n$, J is contained in a single component of $\Gamma - K$. Let κ be the smallest integer for which there are κ-inseparable sets J_1, J_2 and a vertex cut K with $|K| = \kappa$ and J_1, J_2 in different components of $\Gamma - K$. We can consider the maximal κ-inseparable sets of Γ as ends of the graph;

or attach a sequence of ($\kappa+1$)-vertices to each to turn them into a bona fide end. The inseparability conditions ensure that this does not affect the cuts of Γ of size κ. Then the size-κ vertex cuts separating two inseparable sets become minimal end cuts of our graph, so we can obtain a succulent theorem for them:

Theorem 8.4. *Let Γ be a finite connected graph such that there exists κ for which there are κ-inseparable sets J_1, J_2 and a vertex cut K with $|K| = \kappa$ and J_1, J_2 in different components of $\Gamma - K$, and take the minimal such κ. There is a succulent $\mathcal{S}$ with the following properties*:

(1) *There is a subset A of vertices of $\mathcal{S}$ called the* anchors *of $\mathcal{S}$. No two anchors are adjacent*, *and every vertex of $\mathcal{S}$ not in A is adjacent to an anchor. We define an anchor cut of $\mathcal{S}$ to be a vertex cut containing no anchors which separates some anchors of $\mathcal{S}$. We say anchor cuts are equivalent if they partition A in the same way.*

(2) *There is a map f from the κ-inseparable sets of Γ to the anchors of $\mathcal{S}$.*

(3) *There is a bijective map g from equivalence classes of minimal cuts of Γ separating κ-inseparable sets to equivalence classes of minimal anchor cuts of $\mathcal{S}$ such that κ-inseparable sets J_1, J_2 of Γ are separated by $[K]$ if and only if $f(J_1)$, $f(J_2)$ are separated by $g([K])$.*

(4) *Any automorphism of Γ induces an automorphism of $\mathcal{S}$.*

Tutte [1984] produced structure trees for the cases $\kappa = 1, 2$, which Dunwoody and Krön [2015] then extended to higher κ. These trees were based on "optimally nested" cuts in the language of [loc. cit.], which in this case means A-cuts. Roughly speaking, the trees consist of the totally isolated cuts and corner cuts of our succulents, together with "blocks" which are not decomposed by the cuts in question; these include the maximal inseparable sets, and also sets broken up by cuts which are not optimally nested; these sets correspond to the #-classes. The structure trees can then be obtained from our succulents by replacing each ring with a star with one central vertex and one vertex joined to it for each corner cut. So these earlier results also follow from our work.

9. Applications

First we note that our work yields a proof of Stallings' theorem, based on the Bass–Serre theory of groups acting on trees (see [Serre 1980]).

Stallings' theorem. *Let G be a finitely generated group acting transitively on a graph Γ with more than two ends. Then G can be expressed as an amalgam $G = A *_F B$ or an HNN extension $G = A*_F$, where F has a finite index subgroup which is the stabilizer of a vertex of Γ.*

Proof. From the pretree $\mathcal{P}$ we obtain a tree T on which G acts. The tree T is nontrivial; the action is transitive and Γ has more than two ends so there are infinitely many ends and many inequivalent cuts. The action is without inversion since T is bipartite, formed of star vertices and elements of $\mathcal{P}$. Then G is isomorphic to the fundamental group of a certain graph of groups; G is finitely generated so this graph is finite. The action is nontrivial as G acts transitively on Γ, so it follows that G splits over the stabilizer of an edge of T. An element fixing an edge of T fixes the adjacent element of $\mathcal{P}$, and hence fixes either a #-class or an equivalence class of totally isolated cuts. A #-class contains finitely many vertices; and the transitivity of the action implies that there can only be finitely many cuts in each equivalence class, since we can find two cuts between which every cut of the class lies, and then apply the methods of Lemma 7.8. The result follows. □

Stallings' original theorem covers the two-ended case as well, but our tree is trivial here. The two-ended case can be covered by more elementary means, however.

We now discuss how earlier cactus theorems concerning edge cuts follow from ours. We turn a question about edge end cuts into a question about vertex end cuts as follows. First replace the graph Γ with its barycentric subdivision Γ^b. This is defined as follows:

$$V(\Gamma^b) = V(\Gamma) \cup E(\Gamma),$$
$$E(\Gamma^b) = \{(v, e) : v \in V(\Gamma), e \in E(\Gamma), v \text{ an endpoint of } e\}.$$

If the cardinality of a minimal edge end cut of Γ is n, we now "thicken up" each vertex of Γ^b that was a vertex of Γ by replacing it with an $(n+1)$-vertex (see Definition 6.1) to obtain a graph Γ^*. In this way, an edge cut of Γ separating some ends of Γ corresponds precisely with a vertex cut of Γ^* of the same cardinality. In Γ^*, because all the vertex cuts are essentially edge cuts, all of the minimal vertex cuts of Γ^* split the graph into precisely two pieces, each containing an end. So we do not need to remove slices from the graph, and all cuts are B-cuts. It follows that quasiequivalent half-cuts are equivalent, and each ring becomes simple enough to be replaced by a cycle, in which the anchors become the vertices and the other vertices become the edges. Our succulent from Theorem 8.3 can then be replaced with a cactus, so we have the cactus theorem for edge end cuts:

Theorem 9.1 [Evangelidou and Papasoglu 2014]. *Let Γ be a connected graph such that there are edge end cuts of Γ with finite cardinality. There is a cactus $\mathcal{C}$ with the following properties*:

(1) *There is a map f from the ends of Γ to the union of the ends of $\mathcal{C}$ with the vertices of $\mathcal{C}$.*

(2) *There is a bijective map g from equivalence classes of minimal end cuts of Γ*

to minimal edge cuts of $\mathcal{C}$ *such that ends* e_1, e_2 *of* Γ *are separated by* $[K]$ *if and only if* $f(e_1)$, $f(e_2)$ *are separated by* $g([K])$.

(3) *Any automorphism of* Γ *induces an automorphism of* $\mathcal{C}$.

To deal with the classical cactus theorem for edge cuts of finite graphs, we proceed as before to get the graph Γ^*. Then to each $(n+1)$-vertex we attach an infinite chain of $(n+1)$-vertices, so that a vertex in the original graph Γ becomes a de facto end of our new graph. "Equivalent cuts" in this graph correspond to the same cut of the original graph. Once again the succulent can be replaced with a cactus, so we have the cactus theorem of Dinic, Karzanov, and Lomonosov:

Theorem 9.2 [Dinic et al. 1976]. *Let* Γ *be a connected finite graph. There is a cactus* $\mathcal{C}$ *with the following properties*:

(1) *There is a map* f *from the vertices of* Γ *to the vertices of* $\mathcal{C}$.

(2) *There is a bijective map* g *from equivalence classes of minimal edge cuts of* Γ *to minimal edge cuts of* $\mathcal{C}$ *such that vertices* v_1, v_2 *of* Γ *are separated by* $[K]$ *if and only if* $f(v_1)$, $f(v_2)$ *are separated by* $g([K])$.

(3) *Any automorphism of* Γ *induces an automorphism of* $\mathcal{C}$.

Acknowledgements

The author would like to thank Panos Papasoglu for suggesting this problem, and Jonathan Bradford and Leo Wright for proofreading the paper. The author is also grateful to the EPSRC for funding this research.

References

[Bondy and Murty 2008] J. A. Bondy and U. S. R. Murty, *Graph theory*, Graduate Texts in Mathematics **244**, Springer, New York, 2008. MR 2009c:05001 Zbl 1134.05001

[Bowditch 1999] B. H. Bowditch, *Treelike structures arising from continua and convergence groups*, vol. 139, Mem. Amer. Math. Soc. **662**, Amer. Math. Soc., Providence, RI, 1999. MR 2000c:20061 Zbl 0961.20034

[Dinic et al. 1976] E. A. Dinic, A. V. Karzanov, and M. V. Lomonosov, "The structure of a system of minimal edge cuts of a graph", pp. 290–306 in Исследования по дискретной оптимизации (*Studies in discrete optimization*), edited by A. A. Fridman, Nauka, Moscow, 1976. In Russian. MR 58 #16395

[Droms et al. 1995] C. Droms, B. Servatius, and H. Servatius, "The structure of locally finite two-connected graphs", *Electron. J. Combin.* **2** (1995), Research Paper 17. MR 96f:05124 Zbl 0829.05041

[Dunwoody and Krön 2015] M. J. Dunwoody and B. Krön, "Vertex Cuts", *J. Graph Theory* **80**:2 (2015), 136–171.

[Evangelidou and Papasoglu 2014] A. Evangelidou and P. Papasoglu, "A cactus theorem for end cuts", *Internat. J. Algebra Comput.* **24**:1 (2014), 95–112. MR 3189668 Zbl 1286.05078

[Fleiner and Frank 2009] T. Fleiner and A. Frank, "A quick proof for the cactus representation of mincuts", technical report QP-2009-03, Egerváry Research Group, Budapest, 2009, available at http://www.cs.elte.hu/egres.

[Huntington 1935] E. V. Huntington, "Inter-relations among the four principal types of order", *Trans. Amer. Math. Soc.* **38**:1 (1935), 1–9. MR 1501800 Zbl 0012.14601

[Huntington and Rosinger 1932] E. V. Huntington and K. E. Rosinger, "Postulates for separation of point-pairs (reversible order on a closed line)", *Proc. Amer. Acad. Arts Sci.* **67** (1932), 61–145.

[Serre 1980] J.-P. Serre, *Trees*, Springer, Berlin, 1980. MR 82c:20083 Zbl 0548.20018

[Stallings 1968] J. R. Stallings, "On torsion-free groups with infinitely many ends", *Ann. of Math.* (2) **88** (1968), 312–334. MR 37 #4153 Zbl 0238.20036

[Tutte 1984] W. T. Tutte, *Graph theory*, Encyclopedia of Mathematics and its Applications **21**, Addison-Wesley, Reading, MA, 1984. MR 87c:05001 Zbl 0554.05001

Received October 9, 2014. Revised January 30, 2015.

GARETH R. WILKES
MATHEMATICAL INSTITUTE
UNIVERSITY OF OXFORD
ANDREW WILES BUILDING, RADCLIFFE OBSERVATORY QUARTER
WOODSTOCK ROAD
OXFORD
OX2 6GG
UNITED KINGDOM
gareth.wilkes@sjc.ox.ac.uk

CONTENTS

Volume 278, no. 1 and no. 2

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

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 will 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, which should be used sparingly.

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.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 278 No. 2 December 2015
