Pacific Journal of Mathematics

Volume 276 No. 2 August 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. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.257

FREE EVOLUTION ON ALGEBRAS WITH TWO STATES, II

MICHAEL ANSHELEVICH

Denote by $\mathcal{J}$ the operator of coefficient stripping. We show that for any free convolution semigroup $\{\mu_t : t \geq 0\}$ with finite variance, applying a single stripping produces semicircular evolution with nonzero initial condition, $\mathcal{J}[\mu_t] = \rho \boxplus \sigma_{\beta,\gamma}^{\boxplus t}$, where $\sigma_{\beta,\gamma}$ is the semicircular distribution with mean β and variance γ. For more general freely infinitely divisible distributions τ, expressions of the form $\tilde{\rho} \boxplus \tau^{\boxplus t}$ arise from stripping $\tilde{\mu}_t$, where $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$ forms a semigroup under the operation of two-state free convolution. The converse to this statement holds in the algebraic setting. Numerous examples illustrating these constructions are computed. Additional results include the formula for generators of such semigroups.

1. Introduction

A probability measure μ on $\mathbb{R}$ with finite moments can be described by two sequences of Jacobi parameters

$$J(\mu) = \begin{pmatrix} \beta_0, & \beta_1, & \beta_2, & \beta_3, & \dots \\ \gamma_0, & \gamma_1, & \gamma_2, & \gamma_3, & \dots \end{pmatrix}.$$

For example, its Cauchy transform

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-x} \, d\mu(x)$$

(which determines the measure) has the continued fraction expansion

$$G_\mu(z) = \cfrac{1}{z - \beta_0 - \cfrac{\gamma_0}{z - \beta_1 - \cfrac{\gamma_1}{z - \beta_2 - \cfrac{\gamma_2}{z - \beta_3 - \cfrac{\gamma_3}{z - \cdots}}}}}$$

This work was supported in part by NSF grants DMS-0900935 and DMS-1160849.
MSC2010: 46L54.
Keywords: free convolution semigroups, two-state free convolution, coefficient stripping, subordination distribution.

Define new measures $\Phi[\mu]$ and $\mathcal{J}[\mu]$ by the right and left shifts on Jacobi parameters

$$J(\Phi[\mu]) = \begin{pmatrix} 0, & \beta_0, & \beta_1, & \beta_2, & \dots \\ 1, & \gamma_0, & \gamma_1, & \gamma_2, & \dots \end{pmatrix}$$

and

$$J(\mathcal{J}[\mu]) = \begin{pmatrix} \beta_1, & \beta_2, & \beta_3, & \beta_4, & \dots \\ \gamma_1, & \gamma_2, & \gamma_3, & \gamma_4, & \dots \end{pmatrix}.$$

$\mathcal{J}$ is sometimes called coefficient stripping. Actually, both Φ and $\mathcal{J}$ can be defined more generally: Φ for any probability measure, and $\mathcal{J}$ for any probability measure with finite variance. See Definition 2.

Denote by

$$d\sigma_{\beta,\gamma}(x) = \frac{1}{2\pi\gamma}\sqrt{4\gamma - (x-\beta)^2}\,dx$$

the semicircular distribution with mean β and variance γ, $\sigma = \sigma_{0,1}$ the standard semicircular distribution, and $\boxplus$ the operation of free convolution. The semicircular family $\{\sigma_{\beta t,\gamma t} = \sigma_{\beta,\gamma}^{\boxplus t} : t \geq 0\}$ forms a free convolution semigroup. General free convolution semigroups

$$\{\mu_t : t \geq 0\}$$

with mean 0 and variance t are indexed by probability measures ρ. In Proposition 9 of [Anshelevich 2013], we showed that for any such free convolution semigroup,

$$\mathcal{J}[\mu_t] = \rho \boxplus \sigma^{\boxplus t},$$

so that the "once-stripped" free convolution semigroup is always a "free heat evolution" started at ρ. Needless to say, this statement has no analog for semigroups with respect to usual convolution. In the first result of the paper, we extend this formula to the case of general finite variance: for a free convolution semigroup $\{\mu_t\}$ with mean βt and nonzero variance γt,

$$\mathcal{J}[\mu_t] = \rho \boxplus \sigma_{\beta,\gamma}^{\boxplus t}. \tag{1}$$

Since any free convolution semigroup, when stripped, always gives a semicircular evolution, it is natural to ask for which families of measures $\{\tilde{\mu}_t : t \geq 0\}$ is

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \tau^{\boxplus t} \tag{2}$$

for other measures τ. The main result of the article is that if this is the case, there exists a free convolution semigroup $\{\mu_t : t \geq 0\}$ such that the family of pairs of measures $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$ forms a semigroup under the operation $\boxplus_c$ of *two-state free convolution*. Note that formula (2) can sometimes be assigned a meaning even if τ is not freely infinitely divisible. For example, if $\tilde{\rho} = \nu \boxplus \tau$ for some ν, then for general probability measures τ, ν, there exists a family of measures forming the first component of the two-state free convolution semigroup such that

$$\mathcal{J}[\tilde{\mu}_t] = \nu \boxplus \tau^{\boxplus(1+t)}$$

(recall that in free probability, $\tau^{\boxplus(1+t)}$ is well-defined for any τ as long as $t \geq 0$). The most general case covered by the main theorem of the article (Theorem 11) is that for some semigroups,

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \omega^{\boxplus(t/p)},$$

where $\tau = \omega^{\boxplus(1/p)}$ need not even be a positive measure, but where the *subordination distribution* $\omega \boxright \tilde{\rho}$ is freely infinitely divisible. It is unclear at this point whether every two-state free convolution semigroup (with finite variance) is of this form. Nevertheless, a large group of examples fit into this framework: free convolution semigroups, Boolean convolution semigroups, two-state free Brownian motions, and two-state free Meixner distributions. Moreover, in the last section of the paper we show that in the algebraic setting, when $(\tilde{\mu}_t, \mu_t)$ are linear functionals on polynomials but do not necessarily come from positive measures, formula (2) always holds for some (not necessarily positive) τ. In that section we also prove a basic formula for the moment-generating function of the multivariate subordination distribution (see below), which really belongs on the long list of properties of that distribution proven in [Nica 2009].

The other aspects of two-state free convolution semigroups are investigated at the end of Section 3. We compute the two-state version of Voiculescu's evolution equation for the Cauchy transform. Then we combine it with the preceding results to find the formula for the generators of two-state free convolution semigroups with finite variance.

Finally, we would like to explain the connection between this article and part I of the same title [Anshelevich 2010]. Belinschi and Nica [2008; 2009] proved that the eponymous family of transformations $\{\mathbb{B}_t : t \geq 0\}$, is related to the free heat evolution via

$$\mathbb{B}_t[\Phi[\rho]] = \Phi[\rho \boxplus \sigma^{\boxplus t}]. \tag{3}$$

Equation (1) follows from this observation after only a small amount of work. In part I, we constructed a two-variable map $\Phi[\cdot, \cdot]$ and proved that

$$\mathbb{B}_t[\Phi[\tau, \tilde{\rho}]] = \Phi[\tau, \tilde{\rho} \boxplus \tau^{\boxplus t}]. \tag{4}$$

Moreover, the transformation $\Phi[\cdot, \cdot]$, as defined in [Anshelevich 2010], also comes from two-state free probability theory. Nica [2009] observed that $\Phi[\tau, \tilde{\rho}]$ is closely related to the subordination distribution $\tau \boxright \tilde{\rho}$, which is a more important object in free probability, and so will be used in computations in this paper.

At this point the evolution formula (4) is only proven for measures with finite moments, while we are interested in a more general class of measures with finite variance. Moreover, the derivation of (1) from (3) does not generalize to a derivation of (2) from (4); the proof of (2) is quite different. Nevertheless, both this article and part I involve two-state free probability theory and generalization of semicircular evolution to more general free convolution semigroups.

2. Background

Notation 1. Denote by $m[\mu]$ and $\mathrm{Var}[\mu]$ the mean and variance of μ,

$$\begin{aligned}\mathcal{P} &= \{\text{probability measures on } \mathbb{R}\},\\ \mathcal{P}_2 &= \{\mu \in \mathcal{P} : \mathrm{Var}[\mu] < \infty\},\\ \mathcal{P}_{0,1} &= \{\mu \in \mathcal{P}_2 : m[\mu] = 0, \mathrm{Var}[\mu] = 1\},\\ \mathcal{ID}^{\boxplus} &= \{\mu \in \mathcal{P} : \mu \text{ is } \boxplus\text{-infinitely divisible}\}.\end{aligned}$$

For a probability measure μ on $\mathbb{R}$, its Cauchy transform is

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-x}\, d\mu(x),$$

and its F-transform is

$$F_\mu(z) = \frac{1}{G_\mu(z)}$$

(for a function f, f^{-1} will denote its compositional rather than a multiplicative inverse).

2.1. ***Convolutions.*** For $\mu \in \mathcal{P}$, define its Voiculescu transform ϕ_μ by

$$(\phi_\mu \circ F_\mu)(z) + F_\mu(z) = z.$$

See [Bercovici and Voiculescu 1993; Voiculescu et al. 1992]. The free convolution of two measures $\mu \boxplus \nu$ is determined by the equality

$$\phi_{\mu\boxplus\nu} = \phi_\mu + \phi_\nu$$

on a domain. A free convolution semigroup is a weakly continuous family $\{\mu_t : t \geq 0\} \subset \mathcal{P}$ satisfying

$$\mu_t \boxplus \mu_s = \mu_{t+s}.$$

In this case, we denote $\mu_t = \mu^{\boxplus t}$. A measure μ is $\boxplus$-infinitely divisible if $\mu = \mu_1$ for some free convolution semigroup. A fundamental result in [Nica and Speicher 1996], extended to measures with unbounded support in [Belinschi and Bercovici 2004], is that for any $\mu \in \mathcal{P}$, $\mu^{\boxplus t}$ is defined for $t \geq 1$.

We will refer to the set

$$\{(\beta, \gamma, \rho) : \beta \in \mathbb{R}, \gamma > 0, \rho \in \mathcal{P}\} \cup \{(\beta, 0, \cdot) : \beta \in \mathbb{R}\}$$

as the set of *canonical triples*. By a result of Maassen [1992], $\boxplus$-convolution semigroups with finite variance

$$\{\mu_t : t \geq 0, \mathrm{Var}[\mu_1] < \infty\}$$

are in bijection with canonical triples, with the bijection given by

$$\phi_{\mu_t}(z) = \beta t + \gamma t G_\rho(z). \tag{5}$$

Here $\beta = m[\mu_1]$ and $\gamma = \mathrm{Var}[\mu_1]$. The $\boxplus$-convolution semigroups with zero variance are of the form $\mu_t = \delta_{\beta t}$, and so correspond to $(\beta, 0, \cdot)$ with $\gamma = 0$ and ρ undefined.

Similarly, for $\tilde{\mu}, \mu \in \mathcal{P}$, define the two-state Voiculescu transform $\phi_{\tilde{\mu},\mu}$ by

$$(\phi_{\tilde{\mu},\mu} \circ F_\mu)(z) + F_{\tilde{\mu}}(z) = z. \tag{6}$$

See [Krystek 2007; Wang 2011]. The two-state free convolution of two pairs of measures

$$(\rho, \mu \boxplus \nu) = (\tilde{\mu}, \mu) \boxplus_c (\tilde{\nu}, \nu)$$

is determined by the equality

$$\phi_{\rho,\mu\boxplus\nu} = \phi_{\tilde{\mu},\mu} + \phi_{\tilde{\nu},\nu}$$

on a domain. A two-state free convolution semigroup is a componentwise weakly continuous family $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$ satisfying

$$(\tilde{\mu}_t, \mu_t) \boxplus_c (\tilde{\mu}_s, \mu_s) = (\tilde{\mu}_{t+s}, \mu_{t+s}).$$

In this case, we denote $(\tilde{\mu}_t, \mu_t) = (\tilde{\mu}, \mu)^{\boxplus_c t}$. The pair $(\tilde{\mu}, \mu)$ is $\boxplus_c$-infinitely divisible if $(\tilde{\mu}, \mu) = (\tilde{\mu}_1, \mu_1)$ for some two-state free convolution semigroup.

For a fixed free convolution semigroup $\{\mu_t : t \geq 0\}$, the $\boxplus_c$-convolution semigroups $\{(\tilde{\mu}_t, \mu_t)\}$ such that $\tilde{\mu}_1$ has finite variance are in bijection with (relative) canonical triples $(\tilde{\beta}, \tilde{\gamma}, \tilde{\rho})$, with the bijection given by

$$\phi_{\tilde{\mu}_t,\mu_t}(z) = \tilde{\beta} t + \tilde{\gamma} t G_{\tilde{\rho}}(z). \tag{7}$$

Here $\tilde{\beta} = m[\tilde{\mu}_1]$ and $\tilde{\gamma} = \mathrm{Var}[\tilde{\mu}_1]$. This does not appear to be stated explicitly, but follows from the description of general two-state freely infinitely divisible distributions in Theorem 4.1 of [Wang 2011]. Again, the case $\mathrm{Var}[\tilde{\mu}_1] = 0$ can be included by setting $\tilde{\gamma} = 0$ and leaving $\tilde{\rho}$ undefined.

The Boolean convolution $\mu \uplus \nu$ is defined by

$$(\mu, \delta_0) \boxplus_c (\nu, \delta_0) = (\mu \uplus \nu, \delta_0).$$

More explicitly, $\phi_{\mu,\delta_0}(z) = z - F_\mu(z)$, so

$$z - F_{\mu \uplus \nu}(z) = (z - F_\mu(z)) + (z - F_\nu(z)).$$

Any distribution is $\uplus$-infinitely divisible, so $\mu^{\uplus t}$ is always defined for any $t \geq 0$.

Finally, a few arguments in the article simplify with the use of the monotone convolution $\mu \rhd \nu$, defined by

$$F_{\mu \rhd \nu} = F_\mu \rhd F_\nu.$$

Definition 2. For measures with finite moments, the transformations Φ and $\mathcal{J}$ were defined in the introduction. Here are the more general definitions. Φ is the bijection

$$\Phi : \mathcal{P} \to \mathcal{P}_{0,1}$$

defined by

$$F_{\Phi[\nu]}(z) = z - G_\nu(z).$$

See [Belinschi and Nica 2008]. For $\mu \in \mathcal{P}_2$ with $m[\mu] = \beta$ and $\mathrm{Var}[\mu] = \gamma > 0$, define $\mathcal{J}[\mu]$ by

$$F_\mu(z) = z - \beta - \gamma G_{\mathcal{J}[\mu]}(z).$$

Then

$$\mathcal{J} : \mathcal{P}_2 \to \mathcal{P},$$

and $\mathcal{J} \circ \Phi$ is the identity map, while $\Phi \circ \mathcal{J}$ is the identity on $\mathcal{P}_{0,1}$.

Definition 3. Recall that all probability measures are infinitely divisible in the Boolean sense. The Boolean-to-free bijection of Bercovici and Pata [1999, Section 6]

$$\mathbb{B} : \mathcal{P} \to \mathcal{ID}^{\boxplus}$$

is defined by

$$\phi_{\mathbb{B}[\mu]}(z) = z - F_\mu(z).$$

More generally, define the Belinschi–Nica transformations [2008] $\{\mathbb{B}_t : t \geq 0\}$ on $\mathcal{P}$ by

$$\mathbb{B}_t[\mu] = \left(\mu^{\boxplus(1+t)}\right)^{\uplus(1/1+t)}.$$

These transformations form a semigroup under composition, and $\mathbb{B}_1 = \mathbb{B}$.

Remark 4. Note that

$$\phi_{\mathbb{B}[\Phi[\rho]]}(z) = z - F_{\Phi[\rho]}(z) = G_\rho(z).$$

So for a free convolution semigroup $\{\mu_t : t \geq 0\}$, equation (5) is equivalent to

$$\mu_t = \delta_{\beta t} \boxplus \mathbb{B}[\Phi[\rho]]^{\boxplus \gamma t}. \tag{8}$$

Definition 5. For $\mu, \nu \in \mathcal{P}$, the subordination distribution [Lenczewski 2007; Nica 2009] $\mu \boxright \nu$ is the unique probability measure such that

$$G_{\mu \boxplus \nu}(z) = G_\nu(F_{\mu \boxright \nu}(z)).$$

Here $F_{\mu \boxright \nu}$ is the corresponding subordination function of $\mu \boxplus \nu$ with respect to ν. If $\mu \boxright \nu \in \mathcal{ID}^{\boxplus}$, we may define (cf. [Anshelevich 2010])

$$\Phi[\mu, \nu] = \mathbb{B}^{-1}[\mu \boxright \nu].$$

Lemma 6. *On a common domain*,

$$\phi_{\mu \boxright \nu}(z) = (\phi_\mu \circ F_\nu)(z).$$

Also, *whenever* $\Phi[\mu, \nu]$ *is defined*,

$$z - F_{\Phi[\mu,\nu]}(z) = (\phi_\mu \circ F_\nu)(z)$$

and

$$\phi_\mu = \phi_{\Phi[\mu,\nu],\nu}.$$

Proof. We compute

$$\begin{aligned}\phi_{\mu \boxright \nu}(z) &= F^{-1}_{\mu \boxright \nu}(z) - z = (F^{-1}_{\mu \boxplus \nu} \circ F_\nu)(z) - z \\ &= \big(\phi_{\mu \boxplus \nu}(F_\nu(z)) + F_\nu(z)\big) - \big(\phi_\nu(F_\nu(z)) + F_\nu(z)\big) = (\phi_\mu \circ F_\nu)(z).\end{aligned}$$

The second property follows by combining this with the definition of $\mathbb{B}$. Finally,

$$(\phi_\mu \circ F_\nu)(z) + F_{\Phi[\mu,\nu]}(z) = z,$$

which implies the third property after comparison with (6). $\square$

The following result is the analog of Corollary 4.13 in [Nica 2009] for single-variable, unbounded distributions.

Lemma 7. *If* $\mu \in \mathcal{ID}^{\boxplus}$, *or if* $\nu = \mu \boxplus \nu'$, *then* $\mu \boxright \nu \in \mathcal{ID}^{\boxplus}$.

Proof. If $\mu \in \mathcal{ID}^{\boxplus}$, then for any $t \geq 0$,

$$\phi_{\mu^{\boxplus t} \boxright \nu}(z) = \phi_{\mu^{\boxplus t}}(F_\nu(z)) = t\phi_\mu(F_\nu(z)) = \phi_{(\mu \boxright \nu)^{\boxplus t}}(z),$$

and so $(\mu \boxright \nu)^{\boxplus t} = \mu^{\boxplus t} \boxright \nu$ is well-defined.

If $\nu = \mu \boxplus \nu'$, then

$$\begin{aligned}\phi_{\mu \boxright \nu}(z) &= \phi_{\mu \boxright (\mu \boxplus \nu')}(z) = \phi_\mu(F_{\mu \boxplus \nu'}(z)) \\ &= \phi_{\mu \boxplus \nu'}(F_{\mu \boxplus \nu'}(z)) - \phi_{\nu'}(F_{\mu \boxplus \nu'}(z)) = z - F_{\mu \boxplus \nu'}(z) - \phi_{\nu'}(F_{\mu \boxplus \nu'}(z)) \\ &= z - F^{-1}_{\nu'}(F_{\mu \boxplus \nu'}(z)) = z - F_{\mu \boxright \nu'}(z) = \phi_{\mathbb{B}[\mu \boxright \nu']}(z),\end{aligned}$$

and so $\mu \boxright \nu = \mathbb{B}[\mu \boxright \nu'] \in \mathcal{ID}^{\boxplus}$. $\square$

Lemma 8. *For a canonical triple* (β, γ, ρ) *and* $t \geq 0$,

$$\mathbb{B}_t[\delta_\beta \uplus \Phi[\rho]^{\uplus\gamma}] = \delta_\beta \uplus \Phi[\rho \boxplus \delta_{\beta t} \boxplus \sigma^{\boxplus \gamma t}]^{\uplus\gamma}.$$

Proof. For $\gamma = 0$, the identity reduces to $\mathbb{B}_t[\delta_\beta] = \delta_\beta$. The argument for $\gamma > 0$ is a slight modification of Remark 4.4 (proof of Theorem 1.6) from [Belinschi and Nica 2008]. Following that paper, denote by θ the subordination function of $\rho \boxplus \delta_{\beta t} \boxplus \sigma^{\boxplus \gamma t}$ with respect to ρ, and by ω the subordination function of $(\delta_\beta \uplus \Phi[\rho]^{\uplus\gamma})^{\boxplus(t+1)}$ with respect to $(\delta_\beta \uplus \Phi[\rho]^{\uplus\gamma})$. On the one hand,

$$G_{\rho \boxplus \delta_{\beta t} \boxplus \sigma^{\boxplus \gamma t}}(z) = G_\rho(\theta(z))$$

and

$$z - F_{\delta_\beta \uplus \Phi[\rho]^{\uplus\gamma}}(z) = \beta + \gamma G_\rho(z).$$

Therefore

$$\theta(z) - F_{\delta_\beta \uplus \Phi[\rho]^{\uplus\gamma}}(\theta(z)) = \beta + \gamma G_{\rho \boxplus \delta_{\beta t} \boxplus \sigma^{\boxplus \gamma t}}(z). \tag{9}$$

On the other hand, denoting by $\tilde{\theta}$ the subordination function of $\rho \boxplus \sigma^{\boxplus \gamma t}$ with respect to ρ, by equation (4.8) in [Belinschi and Nica 2008],

$$\tilde{\theta}(z) = z - \gamma t G_{\rho \boxplus \sigma^{\boxplus \gamma t}}(z).$$

But

$$G_{\rho \boxplus \delta_{\beta t} \boxplus \sigma^{\boxplus \gamma t}}(z) = G_{\rho \boxplus \sigma^{\boxplus \gamma t}}(z - \beta t) = G_\rho(\tilde{\theta}(z - \beta t)) = G_\rho(\theta(z)).$$

Thus

$$\theta(z) = \tilde{\theta}(z - \beta t) = z - \beta t - \gamma t G_{\rho \boxplus \sigma^{\boxplus \gamma t}}(z - \beta t) = z - \beta t - \gamma t G_{\rho \boxplus \delta_{\beta t} \boxplus \sigma^{\boxplus \gamma t}}(z).$$

Combining this with (9), we see that

$$t\theta(z) - tF_{\delta_\beta \uplus \Phi[\rho]^{\uplus\gamma}}(\theta(z)) = z - \theta(z)$$

and

$$\theta(z) = \frac{1}{t+1} z + \left(1 - \frac{1}{t+1}\right) F_{\delta_\beta \uplus \Phi[\rho]^{\uplus\gamma}}(\theta(z)).$$

Then (see [Belinschi and Nica 2008]) it follows that $\theta = \omega$, and so the argument concludes as in that paper:

$$\begin{aligned} z - F_{\mathbb{B}_t[\delta_\beta \uplus \Phi[\rho]^{\uplus\gamma}]}(z) &= z - \left(\left(1 - \frac{1}{t}\right) z + \frac{1}{t}\omega(z)\right) \\ &= \frac{1}{t}(z - \omega(z)) = \frac{1}{t}(z - \theta(z)) \\ &= \beta + \gamma G_{\rho \boxplus \delta_{\beta t} \boxplus \sigma^{\gamma t}}(z) \\ &= z - F_{\delta_\beta \uplus \Phi[\rho \boxplus \delta_{\beta t} \boxplus \sigma^{\boxplus \gamma t}]^{\uplus\gamma}}(z). \end{aligned}$$

□

3. Single-variable, complex-analytic results

Proposition 9. *For any a canonical triple* (β, γ, ρ), *the corresponding free convolution semigroup is*

$$\mu_t = \delta_{\beta t} \uplus \Phi[\rho \boxplus \sigma_{\beta,\gamma}^{\boxplus t}]^{\uplus \gamma t}.$$

In particular, for any free convolution semigroup with nonzero, finite variance,

$$\mathcal{J}[\mu_t] = \rho \boxplus \sigma_{\beta,\gamma}^{\boxplus t}.$$

Proof. A free convolution semigroup with finite variance $\{\mu_t\}$ can be rewritten as

$$\begin{aligned}
\mu_t &= \delta_{\beta t} \boxplus \mathbb{B}[\Phi[\rho]]^{\boxplus \gamma t} && \text{(by the Maassen representation (8))}\\
&= \mathbb{B}_{t-1}[\delta_\beta \boxplus \mathbb{B}[\Phi[\rho]]^{\gamma}]^{\uplus t} && \text{(by definition of } \mathbb{B}_{t-1})\\
&= \mathbb{B}_t[\delta_\beta \uplus \Phi[\rho]^{\uplus \gamma}]^{\uplus t} && \text{(by definition of } \mathbb{B} = \mathbb{B}_1)\\
&= \delta_{\beta t} \uplus \Phi[\rho \boxplus \delta_{\beta t} \boxplus \sigma^{\boxplus \gamma t}]^{\uplus \gamma t} && \text{(by Lemma 8)}\\
&= \delta_{\beta t} \uplus \Phi[\rho \boxplus \sigma_{\beta,\gamma}^{\boxplus t}]^{\uplus \gamma t} && \text{(by definition of } \sigma_{\beta,\gamma}).
\end{aligned}$$

For $\gamma = 0$, we have $\mu_t = \delta_{\beta t} = \sigma_{\beta t,0}$, so the equation still holds. □

Lemma 10. *A family* $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$ *is the two-state free convolution semigroup with the relative canonical triple* $(\tilde{\beta}, \tilde{\gamma}, \tilde{\rho})$ *if and only if* $\{\mu_t : t \geq 0\}$ *forms a free convolution semigroup and*

$$\tilde{\mu}_t = \delta_{\tilde{\beta} t} \uplus \Phi[\tilde{\rho} \rhd \mu_t]^{\uplus \tilde{\gamma} t}.$$

In particular, whenever $\tilde{\gamma} > 0$, *such a family satisfies*

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \rhd \mu_t.$$

Proof. Using the properties of Boolean and monotone convolutions and the definition of Φ,

$$z - F_{\delta_{\tilde{\beta} t} \uplus \Phi[\tilde{\rho} \rhd \mu_t]^{\uplus \tilde{\gamma} t}}(z) = \tilde{\beta} t + \tilde{\gamma} t G_{\tilde{\rho} \rhd \mu_t}(z) = \tilde{\beta} t + \tilde{\gamma} t G_{\tilde{\rho}}(F_{\mu_t}(z)).$$

On the other hand, by formulas (6) and (7), for the two-state free convolution semigroup with the relative canonical triple $(\tilde{\beta}, \tilde{\gamma}, \tilde{\rho})$,

$$z - F_{\tilde{\mu}_t}(z) = (\phi_{\tilde{\mu}_t, \mu_t} \circ F_{\mu_t})(z) = \tilde{\beta} t + \tilde{\gamma} t G_{\tilde{\rho}}(F_{\mu_t}(z)).$$

Comparing these, we obtain the result. □

Theorem 11. *Fix* $\tilde{\beta} \in \mathbb{R}$, $\tilde{\gamma} > 0$, *and* $p > 0$. *Let* $\omega, \tilde{\rho} \in \mathcal{P}$ *be measures such that*

$$\omega \boxplus \tilde{\rho} \in \mathcal{ID}^{\boxplus}.$$

(a) *For any* $t \geq 0$,

$$\phi_{\tilde{\rho}} + (t/p)\phi_\omega$$

is a Voiculescu transform of a probability measure, and so

$$\tilde{\rho} \boxplus \omega^{\boxplus(t/p)}$$

is well-defined.

(b) *Define*

$$\mu = (\omega \boxminus \tilde{\rho})^{\boxplus(1/p)},$$

$\mu_t = \mu^{\boxplus t}$, *and*

$$\tilde{\mu}_t = \delta_{\tilde{\beta} t} \uplus \Phi[\tilde{\rho} \boxplus \omega^{\boxplus(t/p)}]^{\uplus \tilde{\gamma} t}.$$

Then $\tilde{\mu}_1$ *has finite, nonzero variance, the family* $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$ *forms a two-state free convolution semigroup, and*

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \omega^{\boxplus(t/p)}.$$

Proof. For part (a), using Lemma 6,

$$\begin{aligned}
\phi_{\tilde{\rho}}(z) + (t/p)\phi_\omega(z) &= \phi_{\tilde{\rho}}(z) + (t/p)\phi_{\omega \boxminus \tilde{\rho}}(F_{\tilde{\rho}}^{-1}(z)) \\
&= F_{\tilde{\rho}}^{-1}(z) - z + \phi_{(\omega \boxminus \tilde{\rho})^{\boxplus(t/p)}}(F_{\tilde{\rho}}^{-1}(z)) \\
&= F^{-1}_{(\omega \boxminus \tilde{\rho})^{\boxplus(t/p)}}(F_{\tilde{\rho}}^{-1}(z)) - z \\
&= F^{-1}_{\tilde{\rho} \rhd (\omega \boxminus \tilde{\rho})^{\boxplus(t/p)}}(z) - z \\
&= \phi_{\tilde{\rho} \rhd (\omega \boxminus \tilde{\rho})^{\boxplus(t/p)}}(z).
\end{aligned} \tag{10}$$

Since the monotone convolution is known to preserve positivity, this implies part (a). Next, it is clear that in part (b), $\{\mu_t : t \geq 0\}$ forms a free convolution semigroup. From (10), it follows that

$$\tilde{\rho} \boxplus \omega^{\boxplus(t/p)} = \tilde{\rho} \rhd \mu_t.$$

Part (b) now follows from Lemma 10. □

See Proposition 33 for a partial converse to the theorem.

The following corollary is an immediate consequence of Lemma 7.

Corollary 12. *The assumptions of Theorem 11 are satisfied in the following cases:*

(a) $\tilde{\rho} \in \mathcal{P}$ *is arbitrary and* $\omega = \tau \in \mathcal{ID}^{\boxplus}$. *In this case one can, without loss of generality, take* $p = 1$.

(b) $\omega \in \mathcal{P}$ *is arbitrary, and* $\tilde{\rho} = \nu \boxplus \omega$ *for some* $\nu \in \mathcal{P}$.

In particular, for any $\tilde{\rho} \in \mathcal{P}$ and $\tau \in \mathcal{ID}^{\boxplus}$, there exists a two-state free convolution semigroup $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$ such that

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \tau^{\boxplus t}.$$

I am grateful to Serban Belinschi for a discussion leading to the following example.

Example 13. Recall that the analytic R-transform is defined by $R_\mu(z) = \phi_\mu(1/z)$. Let

$$\tau_\varepsilon = \tfrac{1}{2}(\delta_{-\varepsilon} + \delta_\varepsilon).$$

Then

$$R_{\tau_\varepsilon}(z) = \frac{2\varepsilon^2 z}{\sqrt{1 + 4\varepsilon^2 z^2} + 1}$$

is analytic for $|z| < (2\varepsilon)^{-1}$ and grows as $|R_{\tau_\varepsilon}(z)| \approx \varepsilon^2 |z|$. It follows from Theorem 2 of [Bercovici and Voiculescu 1995] that for sufficiently small ε,

$$z + t R_{\tau_\varepsilon}(z)$$

is an R-transform of a positive measure for all $t \in [0, 1]$. On the other hand, $\tau_\varepsilon^{\boxplus t}$ is well-defined and positive for all $t \geq 1$. It follows that $\sigma \boxplus \tau_\varepsilon^{\boxplus t}$ is well-defined and positive for all $t \geq 0$. However, $\tau_\varepsilon \notin \mathcal{ID}^{\boxplus}$, and $\sigma \neq \nu \boxplus \tau_\varepsilon^{\boxplus p}$ for any $p > 0$, so this family is not covered by the preceding corollary. Nevertheless, $F_\sigma(\mathbb{C}^+) = \mathbb{C}^+ \setminus \{z : |z| \leq 2\}$, and

$$\phi_{\tau_\varepsilon}(z) = \frac{\sqrt{z^2 + 4\varepsilon^2} - z}{2}$$

is analytic on this image for $\varepsilon < 1$. It follows that

$$\phi_{\tau_\varepsilon \boxplus \sigma} = \phi_{\tau_\varepsilon} \circ F_\sigma$$

analytically extends to $\mathbb{C}^+$, and so $\tau_\varepsilon \boxplus \sigma \in \mathcal{ID}^{\boxplus}$. So this family is still covered by the preceding theorem.

Question. Can the hypothesis of Theorem 11 be weakened to the assumption in the following proposition? In other words, does this assumption imply that the (equivalent) statements in the following proposition necessarily hold?

Proposition 14. *Let $\tilde{\rho} \in \mathcal{P}$, $\tau \in \mathcal{P}$, and suppose that $\tilde{\rho} \boxplus \tau^{\boxplus t}$ is defined for all $t \geq 0$. The following are equivalent*:

(a) $\tau \boxplus \tilde{\rho} \in \mathcal{ID}^{\boxplus}$.

(b) *$F_{\tilde{\rho} \boxplus \tau^{\boxplus t}}$ is subordinate to $F_{\tilde{\rho}}$ for all $t \geq 0$, in the sense that there exist analytic transformations $\theta_t : \mathbb{C}^+ \to \mathbb{C}^+$ such that $F_{\tilde{\rho} \boxplus \tau^{\boxplus t}}(z) = F_{\tilde{\rho}}(\theta_t(z))$.*

(c) $\{\Phi[\tilde{\rho} \boxplus \tau^{\boxplus t}] : t \geq 0\}$ *is the first component of a two-state free convolution semigroup.*

Proof. Calculations in the proof of Theorem 11 show that if θ_t exists, then $\theta_t = F_{(\tau \boxvdash \tilde{\rho})^{\boxplus t}}$. This shows that (a) and (b) are equivalent. The same calculations also imply that if $\{(\tilde{\mu}_t = \Phi[\tilde{\rho} \boxplus \tau^{\boxplus t}], \mu_t) : t \geq 0\}$ is a two-state free convolution semigroup, then $\mu_t = (\tau \boxvdash \tilde{\rho})^{\boxplus t}$. Thus (a) and (c) are equivalent. $\square$

Lemma 15. *Subordination distributions have the following properties*:

$$
\begin{aligned}
(\mu \boxplus \nu) \boxvdash \rho &= (\mu \boxvdash \rho) \boxplus (\nu \boxvdash \rho),\\
\sigma \boxvdash \mu &= \mathbb{B}[\Phi[\mu]],\\
\mu \boxvdash \delta_0 &= \mu,\\
\mu \boxvdash \mu &= \mathbb{B}[\mu],\\
\delta_a \boxvdash \mu &= \delta_a.
\end{aligned}
$$

There is a corresponding list of properties for $\Phi[\cdot,\cdot]$.

Proof. All of these properties follow immediately from

$$\phi_{\mu \boxvdash \nu}(z) = (\phi_\mu \circ F_\nu)(z). \qquad \square$$

In a number of the following examples, free convolution semigroups $\{\mu_t : t \geq 0\}$ will have finite variance, and so will be associated with canonical triples (β, γ, ρ); in all cases, the relative canonical triple is $(\tilde{\beta}, \tilde{\gamma}, \tilde{\rho})$.

Example 16. Let $\tilde{\rho} = \rho \in \mathcal{P}$. Then the first component of the corresponding two-state free convolution semigroup satisfies

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \sigma_{\beta,\gamma}^{\boxplus t},$$

so that in Corollary 12, $\tau = \sigma_{\beta,\gamma} \in \mathcal{ID}^{\boxplus}$ is a semicircular distribution. Indeed,

$$\sigma_{\beta,\gamma} \boxvdash \tilde{\rho} = (\delta_\beta \boxplus \sigma^{\boxplus \gamma}) \boxvdash \tilde{\rho} = \delta_\beta \boxplus \mathbb{B}[\Phi[\tilde{\rho}]]^{\boxplus \gamma} = \mu.$$

In the particular case when $\tilde{\beta} = \beta$ and $\tilde{\gamma} = \gamma$, it follows that $\tilde{\mu}_t = \mu_t$ form a free convolution semigroup, and we are in the Belinschi–Nica setting of Proposition 9.

Example 17. Let $\tilde{\rho} = \delta_0$ and $\rho \in \mathcal{P}$. Then the first component of the corresponding two-state free convolution semigroup satisfies

$$\mathcal{J}[\tilde{\mu}_t] = \mu_t,$$

so that in Corollary 12, $\tau = \mu \in \mathcal{ID}^{\boxplus}$ is arbitrary. Indeed,

$$\mu \boxvdash \delta_0 = \mu.$$

These are the (distributions of) two-state free Brownian motions (in [Anshelevich 2011], they were called algebraic two-state free Brownian motions).

Example 18. Let $\tilde{\rho} \in \mathcal{P}$ and $\gamma = 0$, so that $\mu_t = \delta_{\beta t}$. Then the first component of the corresponding two-state free convolution semigroup satisfies

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \delta_{\beta t},$$

so that in Corollary 12, $\tau = \delta_\beta \in \mathcal{ID}^{\boxplus}$. Indeed,

$$\delta_\beta \boxplus \tilde{\rho} = \delta_\beta = \mu.$$

For general β and measures with finite moments,

$$\tilde{\mu}_t = \delta_{\tilde{\beta} t} \uplus \Phi[\tilde{\rho} \boxplus \delta_{\beta t}]^{\uplus \tilde{\gamma} t}$$

are precisely the families constructed in [Anshelevich and Młotkowski 2012, Proposition 7]. For $\beta = 0$, this is a Boolean convolution semigroup, and an arbitrary Boolean convolution semigroup (with finite variance) arises in this way.

On the other hand, if $\tilde{\mu}_t = \delta_{\tilde{\beta} t}$, for any free convolution semigroup $\{\mu_t : t \geq 0\}$, the measures $(\delta_{\tilde{\beta} t}, \mu_t)$ form a two-state free convolution semigroup.

Example 19. Let $\tilde{\rho}, \rho \in \mathcal{P}$ such that $\mathcal{J}[\tilde{\rho}] = \rho$. That is, for some $\tilde{b}$ and $\tilde{c} > 0$,

$$\tilde{\rho} = \delta_{\tilde{b}} \uplus \Phi[\rho]^{\uplus \tilde{c}}.$$

Let

$$p = \tilde{c}/\gamma, \quad u = \tilde{b} - \beta \tilde{c}/\gamma.$$

Then the first component of the corresponding two-state free convolution semigroup satisfies

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \omega^{\boxplus p t},$$

where in Corollary 12, $\omega = \delta_{-u} \boxplus \tilde{\rho} \in \mathcal{P}$ but, in general, is not freely infinitely divisible. Indeed,

$$\begin{aligned}((\delta_{-u} \boxplus \tilde{\rho}) \boxplus \tilde{\rho})^{\boxplus(1/p)} &= \delta_{-(u/p)} \boxplus \mathbb{B}[\tilde{\rho}]^{\boxplus(1/p)} = \mathbb{B}[\delta_{-(u/p)} \uplus \tilde{\rho}^{\uplus(1/p)}] \\ &= \mathbb{B}\big[\delta_{-(u/p)} \uplus (\delta_{\tilde{b}} \uplus \Phi[\rho]^{\uplus \tilde{c}})^{\uplus(1/p)}\big] = \mathbb{B}[\delta_\beta \uplus \Phi[\rho]^{\uplus \gamma}] = \mu.\end{aligned}$$

If $\tau = \omega^{\boxplus(\gamma/\tilde{c})} \in \mathcal{P}$, then

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \tau^{\boxplus t}.$$

Remark 20. A free Meixner distribution $\mu_{b,c,\beta,\gamma}$ with parameters $b, \beta \in \mathbb{R}$, $c + \gamma, \gamma \geq 0$ is the probability measure with Jacobi parameters

$$J(\mu_{b,c,\beta,\gamma}) = \begin{pmatrix} \beta, & b+\beta, & b+\beta, & b+\beta, & \dots \\ \gamma, & c+\gamma, & c+\gamma, & c+\gamma, & \dots \end{pmatrix}.$$

For other values of c, γ, these Jacobi parameters determine a unital, linear, but not positive definite functional. Normalized free Meixner distributions $\mu_{b,c} = \mu_{b,c,0,1}$ have mean 0 and variance 1, and are positive for $c \geq -1$.

Free Meixner distributions form a two-parameter semigroup with respect to $\boxplus$:

$$\mu_{b,c,\beta,\gamma} \boxplus \mu_{b,c,\beta',\gamma'} = \mu_{b,c,\beta+\beta',\gamma+\gamma'}.$$

See Definition 2 of [Anshelevich and Młotkowski 2012]. In particular,

$$\mu_{b,c,\beta,\gamma}^{\boxplus t} = \mu_{b,c,\beta t,\gamma t}.$$

Also,

$$\mathbb{B}_t[\mu_{b,c,\beta,\gamma}] = \mu_{b+\beta t,c+\gamma t,\beta,\gamma}.$$

Lemma 21. *The subordination distribution of two Meixner distributions with special parameters*

$$\mu_{b,c,\beta',\gamma'} \boxright \mu_{b,c,\beta,\gamma} = \mu_{b+\beta,c+\gamma,\beta',\gamma'}$$

is again a free Meixner distribution.

Proof. Using Lemma 15 and the properties from the preceding remark, we compute

$$\begin{aligned}
\mu_{b,c,\beta',\gamma'} \boxright \mu_{b,c,\beta,\gamma} &= \left(\delta_{\beta'-\beta\gamma'/\gamma} \boxplus \mu_{b,c,\beta,\gamma}^{\boxplus(\gamma'/\gamma)}\right) \boxright \mu_{b,c,\beta,\gamma} \\
&= \delta_{\beta'-\beta\gamma'/\gamma} \boxplus \mu_{b,c,\beta,\gamma} \boxright \mu_{b,c,\beta,\gamma}^{\boxplus(\gamma'/\gamma)} \\
&= \delta_{\beta'-\beta\gamma'/\gamma} \boxplus \mathbb{B}[\mu_{b,c,\beta,\gamma}]^{\boxplus(\gamma'/\gamma)} \\
&= \delta_{\beta'-\beta\gamma'/\gamma} \boxplus \mu_{b+\beta,c+\gamma,\beta,\gamma}^{\boxplus(\gamma'/\gamma)} \\
&= \delta_{\beta'-\beta\gamma'/\gamma} \boxplus \mu_{b+\beta,c+\gamma,\beta\gamma'/\gamma,\gamma'} \\
&= \mu_{b+\beta,c+\gamma,\beta',\gamma'}. \qquad \square
\end{aligned}$$

Remark 22. Since $\nu \rhd (\mu \boxright \nu) = \mu \boxplus \nu$, the preceding lemma implies a monotone convolution identity

$$\mu_{b,c,\beta,\gamma} \rhd \mu_{b+\beta,c+\gamma,\beta',\gamma'} = \mu_{b,c,\beta+\beta',\gamma+\gamma'}.$$

This result can also be proved directly using the F-transforms, but the computation is rather surprising. Since $\mu_{0,0,\beta,\gamma}$ are semicircular, $\mu_{b,0,\beta,\gamma}$ are free Poisson, $\mu_{b,-\gamma,\beta,\gamma}$ are Bernoulli, and $\mu_{0,-\gamma,0,2\gamma}$ are arcsine distributions, we get various identities between them involving the monotone convolution. For example,

$$\mu_{b,c} \rhd \mu_{b,c+1} = \mu_{b,c}^{\boxplus 2},$$

which for $b = 0$, $c = -1$ gives Bernoulli $\rhd$ semicircle $=$ arcsine. See [Młotkowski 2010] for related results.

Example 23. For a particular case of Example 19, let $\tilde{c} > 0$ and $c \geq 0$. The two-state free Meixner semigroups from [Anshelevich and Młotkowski 2012] satisfy

$$J(\tilde{\mu}_t) = \begin{pmatrix} \tilde{\beta}t, & \tilde{b} + \beta t, & b + \beta t, & b + \beta t, & \dots \\ \tilde{\gamma}t, & \tilde{c} + \gamma t, & c + \gamma t, & c + \gamma t, & \dots \end{pmatrix}$$

and

$$J(\mu_t) = \begin{pmatrix} \beta t, & b + \beta t, & b + \beta t, & \dots \\ \gamma t, & c + \gamma t, & c + \gamma t, & \dots \end{pmatrix}.$$

Thus

$$J(\mathcal{J}[\tilde{\mu}_t]) = \begin{pmatrix} \tilde{b} + \beta t, & b + \beta t, & b + \beta t, & \dots \\ \tilde{c} + \gamma t, & c + \gamma t, & c + \gamma t, & \dots \end{pmatrix},$$

so $\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \omega^{\boxplus(\gamma/\tilde{c})t}$, where

$$J(\tilde{\rho}) = \begin{pmatrix} \tilde{b}, & b, & b, & b, & \dots \\ \tilde{c}, & c, & c, & c, & \dots \end{pmatrix}$$

and

$$J(\omega) = \begin{pmatrix} \beta\tilde{c}/\gamma, & \beta\tilde{c}/\gamma + b - \tilde{b}, & \beta\tilde{c}/\gamma + b - \tilde{b}, & \beta\tilde{c}/\gamma + b - \tilde{b}, & \dots \\ \tilde{c}, & c, & c, & c, & \dots \end{pmatrix}.$$

In particular, $\tilde{\rho} = \delta_{\tilde{b} - \beta\tilde{c}/\gamma} \boxplus \omega$. Note that both $\tilde{\rho}$ and ω are free Meixner distributions. Also,

$$J(\rho) = \begin{pmatrix} b, & b, & b, & b, & \dots \\ c, & c, & c, & c, & \dots \end{pmatrix},$$

so $\rho = \sigma_{b,c} = \delta_b \boxplus \sigma^{\boxplus c}$ and $\mathcal{J}[\tilde{\rho}] = \rho$. Finally, for $\tau = \omega^{\boxplus(\gamma/\tilde{c})}$,

$$J(\tau) = \begin{pmatrix} \beta, & \beta + b - \tilde{b}, & \beta + b - \tilde{b}, & \beta + b - \tilde{b}, & \dots \\ \gamma, & \gamma + c - \tilde{c}, & \gamma + c - \tilde{c}, & \gamma + c - \tilde{c}, & \dots \end{pmatrix}.$$

So $\tau \in \mathcal{ID}^{\boxplus}$ for $c \geq \tilde{c}$, $\tau \in \mathcal{P}$ for $\gamma + c \geq \tilde{c}$, and for $\gamma + c < \tilde{c}$, τ is not a positive functional.

Proposition 24. *Let $(\tilde{\mu}_t, \mu_t)$ be a general two-state free convolution semigroup. Then we have two evolution equations,*

$$\partial_t F_{\tilde{\mu}_t} = \phi_\mu(F_{\mu_t}) - \phi_{\tilde{\mu},\mu}(F_{\mu_t}) - \phi_\mu(F_{\mu_t})\partial_z F_{\tilde{\mu}_t} \tag{11}$$

and

$$\partial_t F_{\mu_t} = -\phi_\mu(F_{\mu_t})\partial_z F_{\mu_t}.$$

Proof. The second equation is standard; see equation (3.18) in [Voiculescu et al. 1992]. Using (6),

$$\begin{aligned} &\partial_t F_{\tilde{\mu}_t} + \phi_{\tilde{\mu},\mu}(F_{\mu_t}) + t\phi'_{\tilde{\mu},\mu}(F_{\mu_t})\partial_t F_{\mu_t} \\ &\qquad = \partial_t F_{\tilde{\mu}_t} + \phi_{\tilde{\mu},\mu}(F_{\mu_t}) - t\phi'_{\tilde{\mu},\mu}(F_{\mu_t})\phi_\mu(F_{\mu_t})\partial_z F_{\mu_t} = 0 \end{aligned}$$

and

$$\partial_z F_{\tilde{\mu}_t} + t\phi'_{\tilde{\mu},\mu}(F_{\mu_t})\partial_z F_{\mu_t} = 1.$$

Plugging in, we get

$$\begin{aligned}\partial_t F_{\tilde{\mu}_t} &= -\phi_{\tilde{\mu},\mu}(F_{\mu_t}) + \frac{1-\partial_z F_{\tilde{\mu}_t}}{\phi'_{\tilde{\mu},\mu}(F_{\mu_t})\partial_z F_{\mu_t}}\phi'_{\tilde{\mu},\mu}(F_{\mu_t})\phi_\mu(F_{\mu_t})\partial_z F_{\mu_t} \\ &= \phi_\mu(F_{\mu_t}) - \phi_{\tilde{\mu},\mu}(F_{\mu_t}) - \phi_\mu(F_{\mu_t})\partial_z F_{\tilde{\mu}_t}.\end{aligned}$$

□

Definition 25. The functional L_t is the *generator* of the family $\{\mu_t : t \geq 0\}$ of functionals at time t, with domain $\mathcal{D}$, if for any $f \in \mathcal{D}$,

$$\langle L_t, f\rangle = \frac{d}{dt}\langle \mu_t, f\rangle.$$

Proposition 26. *Let $(\tilde{\mu}_t, \mu_t)$ be a general two-state free convolution semigroup with finite variance, with canonical triples $\{(\tilde{\beta}, \tilde{\gamma}, \tilde{\rho}), (\beta, \gamma, \rho)\}$. Let $\mathcal{J}[\tilde{\mu}_t] = \tilde{\nu}_t$ and $\mathcal{J}[\mu_t] = \nu_t$. Note that*

$$\nu_t = \rho \boxplus \sigma_{\beta,\gamma}^{\boxplus t},$$

and for measures covered in Theorem 11, $\tilde{\nu}_t = \tilde{\rho} \boxplus \tau^{\boxplus t}$.

Then the generators of the families $\{\tilde{\mu}_t\}$ and $\{\mu_t\}$ with domain

$$\mathcal{D} = \mathrm{Span}\left(\left\{\frac{1}{z-x} : z \in \mathbb{C}\setminus\mathbb{R}\right\}\right)$$

are, respectively,

$$\begin{aligned}\tilde{L}_t = \tilde{\gamma}(\tilde{\mu}_t \otimes \tilde{\mu}_t \otimes \tilde{\nu}_t)\partial^2 - \gamma(\tilde{\mu}_t \otimes \tilde{\mu}_t \otimes \nu_t)\partial^2 \\ + (\tilde{\beta} - \beta)(\tilde{\mu}_t \otimes \tilde{\mu}_t)\partial + \gamma(\tilde{\mu}_t \otimes \nu_t)(\partial_x \otimes 1)\partial + \beta\tilde{\mu}_t\partial_x\end{aligned}$$

and

$$L_t = \gamma(\mu_t \otimes \nu_t)(\partial_x \otimes 1)\partial + \beta\mu_t\partial_x.$$

Here $\partial : \mathcal{D} \to \mathcal{D} \otimes \mathcal{D}$ is the difference quotient operation

$$(\partial f)(x, y) = \frac{f(x) - f(y)}{x - y}.$$

Proof. Note first that

$$(\phi_{\tilde{\mu},\mu} \circ F_{\mu_t})(z) = \frac{1}{t}(\phi_{\tilde{\mu}_t,\mu_t} \circ F_{\mu_t})(z) = \frac{1}{t}(z - F_{\tilde{\mu}_t}(z)) = \tilde{\beta} + \tilde{\gamma}G_{\tilde{\nu}_t}(z),$$

and similarly $\phi_\mu \circ F_{\mu_t} = \beta + \gamma G_{\nu_t}$. Therefore in this case, (11) gives

$$\partial_t F_{\tilde{\mu}_t} = \beta + \gamma G_{\nu_t} - \tilde{\beta} + \tilde{\gamma}G_{\tilde{\nu}_t} - (\beta + \gamma G_{\nu_t})\partial_z F_{\tilde{\mu}_t}.$$

Equivalently,

$$\partial_t G_{\tilde{\mu}_t} = -(\beta + \gamma G_{\nu_t} - \tilde{\beta} + \tilde{\gamma} G_{\tilde{\nu}_t}) G_{\tilde{\mu}_t}^2 - (\beta + \gamma G_{\nu_t}) \partial_z G_{\tilde{\mu}_t}.$$

In other words,

$$\begin{aligned}\partial_t \Big\langle \tilde{\mu}_t, \frac{1}{z-x} \Big\rangle = -\gamma \Big\langle \tilde{\mu}_t \otimes \tilde{\mu}_t \otimes \nu_t, \partial^2 \frac{1}{z-x} \Big\rangle - \beta \Big\langle \tilde{\mu}_t \otimes \tilde{\mu}_t, \partial \frac{1}{z-x} \Big\rangle \\ + \tilde{\gamma} \Big\langle \tilde{\mu}_t \otimes \tilde{\mu}_t \otimes \tilde{\nu}_t, \partial^2 \frac{1}{z-x} \Big\rangle + \tilde{\beta} \Big\langle \tilde{\mu}_t \otimes \tilde{\mu}_t, \partial \frac{1}{z-x} \Big\rangle \\ + \gamma \Big\langle \tilde{\mu}_t \otimes \nu_t, (\partial_x \otimes 1) \partial \frac{1}{z-x} \Big\rangle + \beta \Big\langle \tilde{\mu}_t, \partial_x \frac{1}{z-x} \Big\rangle.\end{aligned}$$

The formula for the generator $\widetilde{L}_t$ of $\{\tilde{\mu}_t\}$ on the span of such functions follows. The formula for L_t follows by setting $\tilde{\mu}_t = \mu_t$. □

Remark 27. Setting $t = 0$ in the preceding proposition, $\mu_0 = \tilde{\mu}_0 = \delta_0$, $\tilde{\nu}_0 = \tilde{\rho}$, and $\nu_0 = \rho$. Thus

$$\begin{aligned}L_0 f &= \gamma \langle \delta_0 \otimes \rho, (\partial_x \otimes 1) \partial f \rangle + \beta \langle \delta_0, \partial_x f \rangle \\ &= \gamma \int_{\mathbb{R}} \frac{f(y) - f(0) - y f'(0)}{y^2} \, d\rho(y) + \beta f'(0),\end{aligned}$$

and

$$\begin{aligned}\widetilde{L}_0 f &= \tilde{\gamma} \langle \delta_0 \otimes \delta_0 \otimes \tilde{\rho}, \partial^2 f \rangle - \gamma \langle \delta_0 \otimes \delta_0 \otimes \rho, \partial^2 f \rangle + (\tilde{\beta} - \beta) \langle \delta_0 \otimes \delta_0, \partial f \rangle \\ &\qquad + \gamma \langle \delta_0 \otimes \rho, (\partial_x \otimes 1) \partial f \rangle + \beta \langle \delta_0, \partial_x f \rangle \\ &= \int_{\mathbb{R}} \frac{f(y) - f(0) - y f'(0)}{y^2} \, d(\tilde{\gamma} \tilde{\rho} - \gamma \rho)(y) + (\tilde{\beta} - \beta) f'(0) \\ &\qquad + \int_{\mathbb{R}} \frac{f(y) - f(0) - y f'(0)}{y^2} \, d(\gamma \rho)(y) + \beta f'(0) \\ &= \tilde{\gamma} \int_{\mathbb{R}} \frac{f(y) - f(0) - y f'(0)}{y^2} \, d\tilde{\rho}(y) + \tilde{\beta} f'(0)\end{aligned}$$

has exactly the same form as in Proposition 3 of [Anshelevich 2013]; see also Remark 11 of that paper.

Remark 28. Boolean evolution corresponds to $\beta = \gamma = 0$, $\mu_t = \delta_0$. Then

$$\partial_t F_{\tilde{\mu}_t}(z) = -\phi_{\tilde{\mu}, \delta_0}(z).$$

In fact, since $\phi_{\tilde{\mu}, \delta_0}(z) = z - F_{\tilde{\mu}}(z)$, this is easy to see directly. It follows that in this case,

$$\widetilde{L}_t = \tilde{\gamma} (\tilde{\mu}_t \otimes \tilde{\mu}_t \otimes \tilde{\nu}_t) \partial^2 + \tilde{\beta} (\tilde{\mu}_t \otimes \tilde{\mu}_t) \partial.$$

For $t = 0$, we again get the formula from the preceding remark.

Similarly, distributions of analytic two-state free Brownian motions correspond to $\tilde{\gamma} = \gamma = 1$, $\tilde{\beta} = 0$, $\tilde{\mu} = \Phi[\mu]$ and $\mu = \delta_\beta \boxplus \sigma$, so that $\tilde{\nu}_t = \nu_t = \mu_t$. Then the generator formula reduces to

$$\begin{aligned} \widetilde{L}_t &= -\beta(\tilde{\mu}_t \otimes \tilde{\mu}_t)\partial + (\tilde{\mu}_t \otimes \mu_t)(\partial_x \otimes 1)\partial + \beta \tilde{\mu}_t \partial_x \\ &= \tilde{\mu}_t\big(-\beta(1 \otimes \tilde{\mu}_t)\partial + \partial_x(1 \otimes \mu_t)\partial + \beta \partial_x\big), \end{aligned}$$

consistent with the result of Proposition 24 in [Anshelevich 2013].

4. Background, II

4.1. ***Multivariate polynomials.*** The number $d \in \mathbb{N}$ will be fixed throughout the remainder of the article. Denote by

$$\boldsymbol{x} = (x_1, x_2, \dots, x_d)$$

a d-tuple of variables, and define $\boldsymbol{z}$, etc., similarly. Let

$$\mathbb{C}\langle \boldsymbol{x} \rangle = \mathbb{C}\langle x_1, x_2, \dots, x_d \rangle$$

be the algebra of polynomials in d noncommuting variables. For $k \geq 1$ and

$$\vec{u} = (u(1), u(2), \dots, u(k)) \in \{1, \dots, d\}^k$$

a multi-index, let

$$x_{\vec{u}} = x_{u(1)} x_{u(2)} \cdots x_{u(k)}.$$

Let

$$\mathcal{D}_{\mathrm{alg}}(d) = \{\mu : \mathbb{C}\langle x_1, x_2, \dots, x_d \rangle \to \mathbb{C} \text{ unital, linear functionals}\}.$$

For $\beta \in \mathbb{R}^d$, the element $\delta_\beta \in \mathcal{D}_{\mathrm{alg}}(d)$ is

$$\delta_\beta[x_{\vec{u}}] = \beta_{\vec{u}}.$$

4.2. ***Free, Boolean, and two-state free convolutions.*** Let $\mu \in \mathcal{D}_{\mathrm{alg}}(d)$. Denote its moment-generating function by

$$M^\mu(\boldsymbol{z}) = \sum_{\vec{u}} \mu[x_{\vec{u}}] z_{\vec{u}}.$$

The (combinatorial) R-transform R^μ of μ is determined by

$$R^\mu\big(z_1(1 + M^\mu(\boldsymbol{z})), \dots, z_d(1 + M^\mu(\boldsymbol{z}))\big) = M^\mu(\boldsymbol{z}). \tag{12}$$

See Lecture 16 of [Nica and Speicher 2006]. The free convolution of two functionals $\mu \boxplus \nu$ is determined by the equality

$$R^{\mu \boxplus \nu} = R^\mu + R^\nu.$$

In the algebraic setting, any functional is $\boxplus$-infinitely divisible.

Similarly, the η-transform η^μ is

$$\eta^\mu(\mathbf{z}) = (1 + M^\mu(\mathbf{z}))^{-1} M^\mu(\mathbf{z})$$

(for a multivariate power series F, F^{-1} will denote its multiplicative inverse). The Boolean convolution of two functionals $\mu \uplus \nu$ is determined by the equality

$$\eta^{\mu \uplus \nu} = \eta^\mu + \eta^\nu.$$

Finally, for $\tilde{\mu}, \mu \in \mathcal{D}_{\text{alg}}(d)$, the two-state R-transform $R^{\tilde{\mu},\mu}$ is determined by

$$\eta^{\tilde{\mu}}(\mathbf{z}) = R^{\tilde{\mu},\mu}\big(z_1(1 + M^\mu(\mathbf{z})), \dots, z_d(1 + M^\mu(\mathbf{z}))\big)(1 + M^\mu(\mathbf{z}))^{-1},$$

and the two-state free convolution of two pairs of functionals

$$(\rho, \mu \boxplus \nu) = (\tilde{\mu}, \mu) \boxplus_c (\tilde{\nu}, \nu)$$

is determined by the equality

$$R^{\rho,\mu \boxplus \nu} = R^{\tilde{\mu},\mu} + R^{\tilde{\nu},\nu}.$$

See Section 2.5 of [Anshelevich 2010].

If $d = 1$ and μ is a compactly supported probability measure on $\mathbb{R}$, it can be identified with an element of $\mathcal{D}_{\text{alg}}(1)$. In this case, the complex function transforms from Section 2 have power series expansions related to the power series from this section by

$$1 + M^\mu(z) = \frac{1}{z} G_\mu\left(\frac{1}{z}\right), \quad R^\mu(z) = z R_\mu(z) = z\phi_\mu\left(\frac{1}{z}\right),$$
$$\eta^\mu(z) = \frac{1}{z} - F_\mu\left(\frac{1}{z}\right), \quad R^{\tilde{\mu},\mu}(z) = z\phi_{\tilde{\mu},\mu}\left(\frac{1}{z}\right).$$

4.3. ***Transformations.*** For $\nu \in \mathcal{D}_{\text{alg}}(d)$, the functional $\Phi[\nu]$ is determined by

$$\eta^{\Phi[\nu]}(\mathbf{z}) = \sum_{i=1}^{d} z_i(1 + M^\nu(\mathbf{z}))z_i.$$

See [Belinschi and Nica 2009; Anshelevich 2009].

In the algebraic setting, $\mathbb{B}$ is a bijection from $\mathcal{D}_{\text{alg}}(d)$ to itself determined by

$$R^{\mathbb{B}[\mu]} = \eta^\mu.$$

Finally, for $\mu, \nu \in \mathcal{D}_{\text{alg}}(d)$, the multivariate subordination distribution $\mu \boxplus \nu \in \mathcal{D}_{\text{alg}}(d)$ is defined via

$$R^{\mu \boxplus \nu}(\mathbf{z}) = R^\mu\big(z_1(1 + M^\nu(\mathbf{z})), \dots, z_d(1 + M^\nu(\mathbf{z}))\big)(1 + M^\nu(\mathbf{z}))^{-1}. \tag{13}$$

See Definition 1.1 in [Nica 2009].

5. Multivariate, algebraic results

The following proposition is the analog of the single-variable relation $G_{\mu \boxplus \nu}(z) = G_\nu(F_{\mu \boxminus \nu}(z))$.

Proposition 29. *The subordination distribution $\mu \boxminus \nu$ satisfies*

$$1 + M^{\mu \boxplus \nu}(\boldsymbol{z}) \\ = (1 + M^{\mu \boxminus \nu}(\boldsymbol{z}))\big(1 + M^\nu\big(z_1(1 + M^{\mu \boxminus \nu}(\boldsymbol{z})), \dots, z_d(1 + M^{\mu \boxminus \nu}(\boldsymbol{z}))\big)\big).$$

Consequently, for a fixed ν, the map $\mu \mapsto \mu \boxminus \nu$ is a bijection on $\mathcal{D}_{\mathrm{alg}}(d)$.

Proof. Note first that the equation

$$1 + M^{\mu \boxplus \nu}(\boldsymbol{z}) = (1 + M^\lambda(\boldsymbol{z}))\big(1 + M^\nu\big(z_1(1 + M^\lambda(\boldsymbol{z})), \dots, z_d(1 + M^\lambda(\boldsymbol{z}))\big)\big)$$

has a unique solution λ. Indeed,

$$(\mu \boxplus \nu)[x_{\vec{u}}] = \lambda[x_{\vec{u}}] + \nu[x_{\vec{u}}] + P_{\vec{u}}(\lambda[x_{\vec{v}}], \nu[x_{\vec{v}}] : |\vec{v}| < |\vec{u}|)$$

for some polynomial $P_{\vec{u}}$.

Let $w_i = z_i(1 + M^\lambda(\boldsymbol{z}))$. Then

$$\begin{aligned} M^{\mu \boxplus \nu}(\boldsymbol{z}) &= R^{\mu \boxplus \nu}\big(z_1(1 + M^{\mu \boxplus \nu}(\boldsymbol{z})), \dots, z_d(1 + M^{\mu \boxplus \nu}(\boldsymbol{z}))\big) \\ &= R^{\mu \boxplus \nu}\big(w_1(1 + M^\nu(\boldsymbol{w})), \dots, w_d(1 + M^\nu(\boldsymbol{w}))\big) \\ &= R^\mu\big(w_1(1 + M^\nu(\boldsymbol{w})), \dots, w_d(1 + M^\nu(\boldsymbol{w}))\big) + M^\nu(\boldsymbol{w}). \end{aligned}$$

On the other hand,

$$\begin{aligned} M^{\mu \boxplus \nu}(\boldsymbol{z}) &= (1 + M^\lambda(\boldsymbol{z}))(1 + M^\nu(\boldsymbol{w})) - 1 \\ &= M^\lambda(\boldsymbol{z})(1 + M^\nu(\boldsymbol{w})) + M^\nu(\boldsymbol{w}) \\ &= R^\lambda(\boldsymbol{w})(1 + M^\nu(\boldsymbol{w})) + M^\nu(\boldsymbol{w}). \end{aligned}$$

Combining these two equations,

$$R^\mu\big(w_1(1 + M^\nu(\boldsymbol{w})), \dots, w_d(1 + M^\nu(\boldsymbol{w}))\big) = R^\lambda(\boldsymbol{w})(1 + M^\nu(\boldsymbol{w})).$$

Comparing with (13), we see that $\lambda = \mu \boxminus \nu$.

The equation in the proposition shows that given ν and λ, $\mu \boxplus \nu$, and consequently μ, is uniquely determined. Conversely, the uniqueness statement above shows that given ν and μ, $\mu \boxminus \nu$ is uniquely determined. $\square$

In the multivariate, algebraic setting, all the results in Lemma 15 were proved in [Nica 2009]; see Remark 1.2, Theorem 1.8, equation (5.7), and Proposition 5.3. We will use them without proof.

Proposition 30. *Let $\tilde{\beta} \in \mathbb{R}^d$, $\tilde{\gamma} > 0$, $\tilde{\rho} \in \mathcal{D}_{\mathrm{alg}}(d)$, and $\{\mu_t : t \geq 0\} \subset \mathcal{D}_{\mathrm{alg}}(d)$ be a free convolution semigroup. Define a two-state free convolution semigroup $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$ by*

$$R^{\tilde{\mu}_t,\mu_t}(z) = t\tilde{\beta} \cdot z + t\tilde{\gamma} \sum_{i=1}^{d} z_i(1 + M^{\tilde{\rho}}(z))z_i.$$

Define $\tau \in \mathcal{D}_{\mathrm{alg}}(d)$ via

$$\mu = \mu_1 = \tau \boxplus \tilde{\rho}.$$

Then

$$\tilde{\mu}_t = \delta_{t\tilde{\beta}} \uplus \Phi[\tilde{\rho} \boxplus \tau^{\boxplus t}]^{\uplus \tilde{\gamma} t}.$$

Proof. By the preceding proposition, for $\nu = \tau^{\boxplus t}$,

$$1+M^{\nu\boxplus\tilde{\rho}}(z)=(1+M^{\nu\boxplus\tilde{\rho}}(z))\big(1+M^{\tilde{\rho}}\big(z_1(1+M^{\nu\boxplus\tilde{\rho}}(z)),\dots,z_d(1+M^{\nu\boxplus\tilde{\rho}}(z))\big)\big).$$

Since

$$\mu_t = \tau^{\boxplus t} \boxplus \tilde{\rho} = \nu \boxplus \tilde{\rho},$$

we have

$$1+M^{\tilde{\rho}\boxplus\tau^{\boxplus t}}(z)=(1+M^{\mu_t}(z))\big(1+M^{\tilde{\rho}}\big(z_1(1+M^{\mu_t}(z)),\dots,z_d(1+M^{\mu_t}(z))\big)\big).$$

On the other hand,

$$\begin{aligned}\eta^{\tilde{\mu}_t}(z) &= R^{\tilde{\mu}_t,\mu_t}\big(z_1(1 + M^{\mu_t}(z)),\dots,z_d(1 + M^{\mu_t}(z))\big)(1 + M^{\mu_t}(z))^{-1}\\ &= t\tilde{\beta} \cdot z + t\tilde{\gamma} \sum_{i=1}^{d} z_i(1 + M^{\mu_t}(z))\\ &\qquad \big(1 + M^{\tilde{\rho}}\big(z_1(1 + M^{\mu_t}(z)),\dots,z_d(1 + M^{\mu_t}(z))\big)\big)z_i.\end{aligned}$$

Combining these two equations, it follows that

$$\eta^{\tilde{\mu}_t}(z) = t\tilde{\beta} \cdot z + t\tilde{\gamma} \sum_{i=1}^{d} z_i(1 + M^{\tilde{\rho}\boxplus\tau^{\boxplus t}}(z))z_i$$

and

$$\tilde{\mu}_t = \delta_{t\tilde{\beta}} \uplus \Phi[\tilde{\rho} \boxplus \tau^{\boxplus t}]^{\uplus \tilde{\gamma} t}. \qquad \square$$

I am grateful to Hari Bercovici for a discussion leading to the following observations.

Corollary 31. *Let $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$ be a two-state free convolution semigroup of compactly supported probability measures such that $\tilde{\mu}_1$ has nonzero variance. Then*

$$\mu = \tau \boxplus \tilde{\rho}$$

and

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \tau^{\boxplus t}$$

for some $\tilde{\rho}$ a compactly supported probability measure, and τ a unital, not necessarily positive linear functional with nonnegative variance such that $|\tau[x^n]| \leq C^n$ for some C.

Proof. The result follows by applying the preceding proposition in the case $d = 1$, when every compactly supported two-state free convolution semigroup is of the form specified in that proposition. Since for each t,

$$\mathrm{Var}[\tilde{\rho}] + t\,\mathrm{Var}[\tau] = \mathrm{Var}[\mathcal{J}[\tilde{\mu}_t]] \geq 0,$$

it follows that $\mathrm{Var}[\tau] \geq 0$. The positivity of $\tilde{\rho}$ follows from the positivity of $(\tilde{\mu}_t, \mu_t)$, and the compact support of $\tilde{\rho}$ and the growth conditions on τ follow from the compact support of $(\tilde{\mu}_t, \mu_t)$. ☐

Lemma 32. *Let τ be a unital linear functional with positive variance such that $|\tau[x^n]| \leq C^n$ for some C. Then $\tau^{\boxplus t}$ is positive definite, and so can be identified with a compactly supported measure, for sufficiently large t.*

Proof. Without loss of generality, we may assume that τ has mean 0 and variance 1. By assumption, the moments of τ, and so also its free cumulants, grow no faster than exponentially. Therefore the R-transform of τ can be identified with an analytic function whose power series expansion at zero starts with z. It follows that for sufficiently large t, the R-transform of $D_{1/\sqrt{t}}\tau^{\boxplus t}$ satisfies the conditions of Theorem 2 of [Bercovici and Voiculescu 1995]. Applying that theorem, we conclude that $D_{1/\sqrt{t}}\tau^{\boxplus t}$, and so also $\tau^{\boxplus t}$, can be identified with a positive measure. ☐

Proposition 33. *Let $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$ be a two-state free convolution semigroup of* compactly supported *measures such that $\tilde{\mu}_1$ has nonzero variance. Then there exist $\tilde{\beta} \in \mathbb{R}$, $\tilde{\gamma} > 0$, $p > 0$ and $\omega, \tilde{\rho} \in \mathcal{P}$ such that $\omega \boxplus \tilde{\rho} \in \mathcal{ID}^{\boxplus}$, $\tilde{\rho} \boxplus \omega^{\boxplus(t/p)}$ is well-defined (in the sense of part (a) of Theorem 11) for all $t \geq 0$, and*

$$\mu_t = (\omega \boxplus \tilde{\rho})^{\boxplus(t/p)}$$

and

$$\tilde{\mu}_t = \delta_{\tilde{\beta}t} \uplus \Phi[\tilde{\rho} \boxplus \omega^{\boxplus(t/p)}]^{\uplus \tilde{\gamma} t},$$

so that in particular,

$$\mathcal{J}[\tilde{\mu}_t] = \tilde{\rho} \boxplus \omega^{\boxplus(t/p)}.$$

Proof. First suppose that $\mu_t = \delta_{\beta t}$ has zero variance. Then all the relations hold if we set $p = 1$, $\omega = \delta_\beta$, and $\tilde{\rho}$ to be the measure in the relative canonical triple of $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$.

In the remainder of the argument, we assume that μ_1 has nonzero variance. Let $\tilde{\rho}$ be the measure in the relative canonical triple of $\{(\tilde{\mu}_t, \mu_t) : t \geq 0\}$. By Corollary 31, there exists a linear, unital, not necessarily positive functional τ such that

$$\tau \boxplus \tilde{\rho} = \mu \tag{14}$$

is $\boxplus$-infinitely divisible, and

$$\tilde{\rho} \boxplus \tau^{\boxplus p} = \mathcal{J}[\tilde{\mu}_t]$$

can be identified with a positive measure. Moreover, it follows from (14) that $\mathrm{Var}[\tau] = \mathrm{Var}[\mu] > 0$. So by Lemma 32, for sufficiently large p, $\omega = \tau^{\boxplus p}$ can itself be identified with a positive measure. The result follows. □

Acknowledgments

The author is grateful to Dan Voiculescu for asking the question which led to Proposition 24, to Hari Bercovici, Serban Belinschi, and Wojtek Młotkowski for discussions leading to Lemma 32 and Example 13, and to the referee for numerous helpful comments.

References

[Anshelevich 2009] M. Anshelevich, "Appell polynomials and their relatives, III: Conditionally free theory", *Illinois J. Math.* **53**:1 (2009), 39–66. MR 2011m:46117 Zbl 1185.46044

[Anshelevich 2010] M. Anshelevich, "Free evolution on algebras with two states", *J. Reine Angew. Math.* **638** (2010), 75–101. MR 2012d:46155 Zbl 1195.46069

[Anshelevich 2011] M. Anshelevich, "Two-state free Brownian motions", *J. Funct. Anal.* **260**:2 (2011), 541–565. MR 2012c:46147 Zbl 1221.46066

[Anshelevich 2013] M. Anshelevich, "Generators of some non-commutative stochastic processes", *Probab. Theory Related Fields* **157**:3-4 (2013), 777–815. MR 3129803 Zbl 1290.46057

[Anshelevich and Młotkowski 2012] M. Anshelevich and W. Młotkowski, "Semigroups of distributions with linear Jacobi parameters", *J. Theoret. Probab.* **25**:4 (2012), 1173–1206. MR 2993018 Zbl 1276.46053

[Belinschi and Bercovici 2004] S. T. Belinschi and H. Bercovici, "Atoms and regularity for measures in a partially defined free convolution semigroup", *Math. Z.* **248**:4 (2004), 665–674. MR 2006i:46095 Zbl 1065.46045

[Belinschi and Nica 2008] S. T. Belinschi and A. Nica, "On a remarkable semigroup of homomorphisms with respect to free multiplicative convolution", *Indiana Univ. Math. J.* **57**:4 (2008), 1679–1713. MR 2009f:46087 Zbl 1165.46033

[Belinschi and Nica 2009] S. T. Belinschi and A. Nica, "Free Brownian motion and evolution towards $\boxplus$-infinite divisibility for k-tuples", *Internat. J. Math.* **20**:3 (2009), 309–338. MR 2010g:46108 Zbl 1173.46306

[Bercovici and Pata 1999] H. Bercovici and V. Pata, "Stable laws and domains of attraction in free probability theory", *Ann. of Math.* (2) **149**:3 (1999), 1023–1060. MR 2000i:46061 Zbl 0945.46046

[Bercovici and Voiculescu 1993] H. Bercovici and D. V. Voiculescu, "Free convolution of measures with unbounded support", *Indiana Univ. Math. J.* **42**:3 (1993), 733–773. MR 95c:46109 Zbl 0806.46070

[Bercovici and Voiculescu 1995] H. Bercovici and D. V. Voiculescu, "Superconvergence to the central limit and failure of the Cramér theorem for free random variables", *Probab. Theory Related Fields* **103**:2 (1995), 215–222. MR 96k:46115 Zbl 0831.60036

[Krystek 2007] A. D. Krystek, "Infinite divisibility for the conditionally free convolution", *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **10**:4 (2007), 499–522. MR 2009d:46118 Zbl 1147.46041

[Lenczewski 2007] R. Lenczewski, "Decompositions of the free additive convolution", *J. Funct. Anal.* **246**:2 (2007), 330–365. MR 2008d:28009 Zbl 1129.46055

[Maassen 1992] H. Maassen, "Addition of freely independent random variables", *J. Funct. Anal.* **106**:2 (1992), 409–438. MR 94g:46069 Zbl 0784.46047

[Młotkowski 2010] W. Młotkowski, "Fuss–Catalan numbers in noncommutative probability", *Doc. Math.* **15** (2010), 939–955. MR 2012c:46170 Zbl 1213.44004

[Nica 2009] A. Nica, "Multi-variable subordination distributions for free additive convolution", *J. Funct. Anal.* **257**:2 (2009), 428–463. MR 2010j:46121 Zbl 1186.46068

[Nica and Speicher 1996] A. Nica and R. Speicher, "On the multiplication of free n-tuples of noncommutative random variables", *Amer. J. Math.* **118**:4 (1996), 799–837. MR 98i:46069 Zbl 0856.46035

[Nica and Speicher 2006] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, London Mathematical Society Lecture Note Series **335**, Cambridge University Press, 2006. MR 2008k:46198 Zbl 1133.60003

[Voiculescu et al. 1992] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables: a noncommutative probability approach to free products with applications to random matrices, operator algebras, and harmonic analysis on free groups*, CRM Monograph Series **1**, American Mathematical Society, Providence, RI, 1992. MR 94c:46133 Zbl 0795.46049

[Wang 2011] J.-C. Wang, "Limit theorems for additive conditionally free convolution", *Canad. J. Math.* **63**:1 (2011), 222–240. MR 2012b:46145 Zbl 1218.46041

Received April 1, 2012. Revised June 24, 2014.

MICHAEL ANSHELEVICH
DEPARTMENT OF MATHEMATICS
TEXAS A&M UNIVERSITY
MAILSTOP 3368
COLLEGE STATION, TX 77843-3368
UNITED STATES

manshel@math.tamu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.281

SYSTEMS OF PARAMETERS AND HOLONOMICITY OF A-HYPERGEOMETRIC SYSTEMS

CHRISTINE BERKESCH ZAMAERE, STEPHEN GRIFFETH AND EZRA MILLER

We give an elementary proof of holonomicity for A-hypergeometric systems, with no requirements on the behavior of their singularities, a result originally due to Adolphson (1994) after the regular singular case by Gelfand and Gelfand (1986). Our method yields a direct de novo proof that A-hypergeometric systems form holonomic families over their parameter spaces, as shown by Matusevich, Miller, and Walther (2005).

Dedication. Every now and then Andrei Zelevinsky had occasion to write a short and in many ways elementary paper with deep consequences. Particularly close to our hearts are his paper on graded nilpotent classes [Zelevinsky 1985] and his paper with Gelfand and Graev on hypergeometric systems [Gelfand et al. 1987]; both of these had enormous impact on our mathematical careers. It is in that spirit that we dedicate to Andrei this elementary perspective on topics that he influenced substantially for many years.

Introduction

An A-hypergeometric system is the D-module counterpart of a toric ideal. Solutions to A-hypergeometric systems are functions, with a fixed infinitesimal homogeneity, on an affine toric variety. The solution space of an A-hypergeometric system behaves well in part because the system is holonomic, which in particular implies that the vector space of germs of analytic solutions at any nonsingular point has finite dimension.

This note provides an elementary proof of holonomicity for arbitrary A-hypergeometric systems, relying only on the statement that a module over the Weyl algebra in n variables is holonomic precisely when its characteristic variety has dimension at most n (see [Gabber 1981] or [Borel et al. 1987, Theorem 1.12]), along with standard facts about transversality of subvarieties and about Krull

EM had support from NSF grant DMS-1001437. SG acknowledges the financial support of Fondecyt Proyecto Regular 1110072.
MSC2010: primary 33C70; secondary 13N10, 14M25, 16S32, 33C99.
Keywords: hypergeometric, differential operators, toric, holonomic.

dimension. In particular, our proof requires no assumption about the singularities of the A-hypergeometric system; equivalently, the associated toric ideal need not be standard graded. Holonomicity was proved in the regular singular case by Gelfand and Gelfand [1986], and later by Adolphson [1994, §3] regardless of the behavior of the singularities of the system. Adolphson's proof relies on careful algebraic analysis of the coordinate rings of a collection of varieties whose union is the characteristic variety of the system. Another proof of the holonomicity of an A-hypergeometric system, by Schulze and Walther [2008], yields a more general result: for a weight vector L from a large family of possibilities, the L-characteristic variety for the L-filtration is a union of conormal varieties and hence has dimension n. Holonomicity follows when $L = (0, \dots, 0, 1, \dots, 1)$ induces the order filtration on the Weyl algebra. The L-filtration method uses an explicit combinatorial interpretation of initial ideals of toric ideals, which requires a series of technical lemmas.

Holonomicity of A-hypergeometric systems forms part of the statement and proof, by Matusevich, Miller, and Walther [2005], that A-hypergeometric systems determine holonomic families over their parameter spaces. The new proof of that statement here serves as a model suitable for generalization to hypergeometric systems for reductive groups, in the sense of Kapranov [1998].

The main step (Theorem 1.2) in our proof is an easy geometric argument showing that the Euler operators corresponding to the rows of an integer matrix A form part of a system of parameters on the product $\Bbbk^n \times X_A$, where $\Bbbk$ is any algebraically closed field and X_A is the toric variety over $\Bbbk$ determined by A. This observation leads quickly in Section 2 to the conclusion that the characteristic variety of the associated A-hypergeometric system has dimension at most n, and hence that the system is holonomic. Since the algebraic part of the proof holds when the entries of the parameter β are considered as independent variables that commute with all other variables, the desired stronger consequence is immediate: the A-hypergeometric system forms a holonomic family over its parameter space (Theorem 2.1).

1. Systems of parameters via transversality

Fix a field $\Bbbk$. Let $x = x_1, \dots, x_n$ and $\xi = \xi_1, \dots, \xi_n$ be sets of coordinates on $\Bbbk^n$ and let $x\xi$ denote the column vector with entries $x_1\xi_1, \dots, x_n\xi_n$. Given a rectangular matrix L with n columns, write $Lx\xi$ for the vector of bilinear forms given by multiplying L times $x\xi$.

Lemma 1.1. *Let $\Bbbk^{2n} = \Bbbk^n_x \times \Bbbk^n_\xi$ have coordinates (x, ξ) and let $X \subseteq \Bbbk^n_\xi$ be a subvariety. If L is an $\ell \times n$ matrix with entries in $\Bbbk$, then the variety* $\mathrm{Var}(Lx\xi)$ *of $Lx\xi$ in $\Bbbk^{2n}$ is transverse to $\Bbbk^n \times X$ at any smooth point of $\Bbbk^n \times X$ whose ξ-coordinates are all nonzero.*

Proof. It suffices to prove the statement after passing to the algebraic closure of $\Bbbk$, so assume $\Bbbk$ is algebraically closed. Let (p, q) be a smooth point of $\Bbbk^n \times X$ that lies in $\mathrm{Var}(Lx\xi)$ and has all coordinates of q nonzero. The tangent space to $\Bbbk^n \times X$ at (p, q) contains $\Bbbk^n \times \{0\}$. The tangent space $T_{(p,q)}$ to $\mathrm{Var}(Lx\xi)$ is the kernel of the $\ell \times 2n$ matrix $[L(q)\ L(p)]$, where $L(p)$ (respectively, $L(q)$) is the $\ell \times n$ matrix that results after multiplying each column of L by the corresponding coordinate of p (respectively, q). Since the q coordinates are all nonzero, $T_{(p,q)}$ projects surjectively onto the last n coordinates; indeed, if $\eta \in \Bbbk^n_\xi$ is given, then taking $y_i = -p_i\eta_i/q_i$ yields $y \in \Bbbk^n_x$ with $L(q)y + L(p)\eta = 0$. Thus the tangent spaces at (p, q) sum to the ambient space, so the intersection is transverse. □

The next result applies the lemma to an affine toric variety X. A fixed $d \times n$ integer matrix $A = [a_1\ a_2\ \cdots\ a_{n-1}\ a_n]$ defines an action of the algebraic torus $T = (\Bbbk^*)^d$ on $\Bbbk^n_\xi$ by

$$t \cdot \xi = (t^{a_1}\xi_1, \dots, t^{a_n}\xi_n).$$

The orbit $\mathrm{Orb}(A)$ of the point $\mathbf{1} = (1, \dots, 1) \in \Bbbk^n$ is the image of an algebraic map $T \to \Bbbk^n$ that sends $t \mapsto t \cdot \mathbf{1}$. The closure of $\mathrm{Orb}(A)$ in $\Bbbk^n$ is the affine toric variety $X_A = \mathrm{Var}(I_A)$ cut out by the *toric ideal*

$$I_A = \langle \xi^u - \xi^v \mid Au = Av \rangle \subseteq \Bbbk[\xi]$$

of A in the polynomial ring $\Bbbk[\xi] = \Bbbk[\xi_1, \dots, \xi_n]$. The T-action induces an A-*grading* on $\Bbbk[\xi]$ via $\deg(\xi_i) = a_i$, and the semigroup ring $S_A = \Bbbk[\xi]/I_A$ is A-graded [Miller and Sturmfels 2005, Chapters 7–8].

For any face τ of the real cone $\mathbb{R}_{\geq 0}A$ generated by the columns of A, write $\tau \preceq A$ and let $\mathbf{1}^\tau \in \{0, 1\}^n \subset \Bbbk^n$ be the vector with nonzero entry $\mathbf{1}^\tau_i = 1$ precisely when A has a nonzero column $a_i \in \tau$. The variety X_A decomposes as a finite disjoint union $X_A = \bigsqcup_{\tau \preceq A} \mathrm{Orb}(\tau)$ of orbits, where $\mathrm{Orb}(\tau) = T \cdot \mathbf{1}^\tau$. Each orbit has dimension $\dim \mathrm{Orb}(\tau) = \mathrm{rank}(A_\tau)$, where A_τ is the submatrix of A consisting of those columns lying in τ, and $\dim X_A = \mathrm{rank}(A)$.

Theorem 1.2. *The ring $\Bbbk[x, \xi]/(I_A + \langle Ax\xi \rangle)$ has Krull dimension n. In particular, if A has rank d then the forms $Ax\xi$ are part of a system of parameters for $\Bbbk[x] \otimes_\Bbbk S_A$.*

Proof. Let $\Bbbk^\tau \subseteq \Bbbk^n$ be the subspace consisting of vectors with 0 in coordinate i if $a_i \notin \tau$, and let $|\tau|$ be its dimension. Since $\Bbbk[x, \xi]/I_A = \Bbbk[x] \otimes_\Bbbk S_A$ has dimension $n + \mathrm{rank}(A)$ and the number of $\Bbbk$-linearly independent generators of $\langle Ax\xi \rangle$ is $\mathrm{rank}(A)$, the Krull dimension in question is at least n. Hence it suffices to prove that $(\Bbbk^n \times \mathrm{Orb}(\tau)) \cap \mathrm{Var}(Ax\xi) \subseteq \Bbbk^n \times \Bbbk^\tau$ has dimension at most n. Let x_τ and ξ_τ denote the subsets corresponding to τ in the variable sets x and ξ, respectively. The projection of the intersection onto the subspace $\Bbbk^\tau \times \Bbbk^\tau$ has image contained in

$$(\Bbbk^\tau \times \mathrm{Orb}(\tau)) \cap \mathrm{Var}(A_\tau x_\tau \xi_\tau) \subseteq \Bbbk^\tau \times \Bbbk^\tau.$$

It therefore suffices to show that the dimension of this latter intersection is at most $|\tau|$. By Lemma 1.1, the intersection is transverse in $\Bbbk^\tau \times \Bbbk^\tau$. But the dimension of $\mathrm{Orb}(\tau)$ is the codimension of $\mathrm{Var}(A_\tau x_\tau \xi_\tau)$ in $\Bbbk^\tau \times \Bbbk^\tau$, which completes the proof. □

2. Hypergeometric holonomicity

In this section, the matrix A is a $d \times n$ integer matrix of full rank d. Let

$$D = \mathbb{C}\langle x, \partial \mid [\partial_i, x_j] = \delta_{ij} \text{ and } [x_i, x_j] = 0 = [\partial_i, \partial_j]\rangle$$

denote the Weyl algebra over the complex numbers $\mathbb{C}$, where $x = x_1, \dots, x_n$ and ∂_i corresponds to $\partial/\partial x_i$. This is the ring of $\mathbb{C}$-linear differential operators on $\mathbb{C}[x]$.

For $\beta \in \mathbb{C}^d$, the *A-hypergeometric system* with parameter β is the left D-module

$$M_A(\beta) = D/D \cdot (I_A^\partial, \{E_i - \beta_i\}_{i=1}^d),$$

where $I_A^\partial = \langle \partial^u - \partial^v \mid Au = Av\rangle \subseteq \mathbb{C}[\partial]$ is the toric ideal associated to A and

$$E_i - \beta_i = \sum_{j=1}^n a_{ij} x_j \partial_j - \beta_i$$

are *Euler operators* associated to A.

The *order filtration* F filters D by order of differential operators. The *symbol* of an operator P is its image $\mathrm{in}(P) \in \mathrm{gr}^F D$. Writing $\xi_i = \mathrm{in}(\partial_i)$, this means $\mathrm{gr}^F D$ is the commutative polynomial ring $\mathbb{C}[x, \xi]$. The *characteristic variety* of a left D-module M is the variety in $\mathbb{A}^{2n}$ of the associated graded ideal $\mathrm{gr}^F \mathrm{ann}(M)$ of the annihilator of M. A nonzero D-module is *holonomic* if its characteristic variety has dimension n; this is equivalent to requiring that the dimension be at most n (see [Gabber 1981] or [Borel et al. 1987, Theorem 1.12]). The *rank* of a holonomic D-module M is the (always finite) dimension of $\mathbb{C}(x) \otimes_{\mathbb{C}[x]} M$ as a vector space over $\mathbb{C}(x)$; this number equals the dimension of the vector space of germs of analytic solutions of M at any nonsingular point in $\mathbb{C}^n$ [Saito et al. 2000, Theorem 1.4.9].

Viewing the A-hypergeometric system $M_A(\beta)$ as having a varying parameter $\beta \in \mathbb{C}^d$, the rank of $M_A(\beta)$ is upper semicontinuous as a function of β [Matusevich et al. 2005, Theorem 2.6]. This follows by viewing $M_A(\beta)$ as a *holonomic family* [ibid., Definition 2.1] parametrized by $\beta \in \mathbb{C}^d$. By definition, this means not only that $M_A(\beta)$ is holonomic for each β, but also that it satisfies a coherence condition over $\mathbb{C}^d$: after replacing β with variables $b = b_1, \dots, b_d$, the module $\mathbb{C}(x) \otimes_{\mathbb{C}[x]} M_A(b)$ is finitely generated over $\mathbb{C}(x)[b]$. (The definition of holonomic family cited above allows sheaves of D-modules over arbitrary complex base schemes, but that generality is not needed here.)

The derivation of the holonomic family property for $M_A(b)$ from the holonomicity of the A-hypergeometric system is more or less the same as [ibid., Theorem 7.5], which was phrased in the generality of Euler–Koszul homology of toric modules. The brief deduction here isolates the steps necessary for A-hypergeometric systems; its brevity stems from the special status of affine semigroup rings among all toric modules [ibid., Definition 4.5]. Note further that this proof does not require technical combinatorial arguments using standard pairs, as in [Saito et al. 2000]; indeed, $\mathrm{in}(I_A)$ need not be a monomial ideal.

Theorem 2.1. *The module $M_A(b)$ forms a holonomic family over $\mathbb{C}^d$ with coordinates b. In more detail, as a $D[b]$-module the parametric A-hypergeometric system $M_A(b)$ satisfies*:

(1) *the fiber $M_A(\beta) = M_A(b) \otimes_{\mathbb{C}[b]} \mathbb{C}[b]/\langle b-\beta\rangle$ is holonomic for all β*; *and*

(2) *the module $\mathbb{C}(x) \otimes_{\mathbb{C}[x]} M_A(b)$ is finitely generated over $\mathbb{C}(x)[b]$.*

Proof. Since $R = \mathbb{C}[x,\xi]/\langle \mathrm{in}(I_A), Ax\xi\rangle$ surjects onto $\mathrm{gr}^F M_A(\beta)$, it is enough to show that the ring R has dimension n. If $M_A(\beta)$ is standard $\mathbb{Z}$-graded (equivalently, the rowspan of A over the rational numbers contains the row $[1\ 1\ \cdots\ 1\ 1]$ of length n), then $\mathrm{in}(I_A) = I_A \subseteq \mathbb{C}[\xi]$, and the result follows from Theorem 1.2.

When $M_A(\beta)$ is not standard $\mathbb{Z}$-graded, let $\hat{A}$ be the $(d+1)\times(n+1)$ matrix obtained by adding a row of 1's across the top of A and then adding as the leftmost column $(1,0,\ldots,0)$. If ξ_0 denotes a new variable corresponding to the leftmost column of $\hat{A}$, and $\hat{\xi} = \{\xi_0\}\cup\xi$, then $\mathbb{C}[\xi]/\mathrm{in}(I_A) \cong \mathbb{C}[\hat{\xi}]/\langle I_{\hat{A}}, \xi_0\rangle$. In particular,

$$\frac{\mathbb{C}[\hat{x},\xi]}{\langle \mathrm{in}(I_A), Ax\xi\rangle} \cong \frac{\mathbb{C}[\hat{x},\hat{\xi}]}{\langle I_{\hat{A}}, \xi_0, \hat{A}\hat{x}\hat{\xi}\rangle},$$

where $\hat{x} = \{x_0\}\cup x$. Since $\langle I_{\hat{A}}, \xi_0\rangle$ is $\hat{A}$-graded and $\hat{A}$ has a row $[1\ 1\ \cdots\ 1\ 1]$, we have reduced to the case where $M_A(\beta)$ is $\mathbb{Z}$-graded, completing part (1).

With R as in part (1), the ring $R[b]$ surjects onto $\mathrm{gr}^F M_A(b)$, so it suffices for part (2) to show that $R[b]$ becomes finitely generated over $\mathbb{C}(x)[b]$ upon inverting all nonzero polynomials in x. Since the ideal $\langle \mathrm{in}(I_A), Ax\xi\rangle$ has no generators involving b variables, it suffices to show that $R(x)$ itself has finite dimension over $\mathbb{C}(x)$. The desired result follows from the statement proved for part (1): any scheme of dimension n has finite degree over $\mathbb{C}^n_x$. □

References

[Adolphson 1994] A. Adolphson, "Hypergeometric functions and rings generated by monomials", *Duke Math. J.* **73**:2 (1994), 269–290. MR 96c:33020 Zbl 0804.33013

[Borel et al. 1987] A. Borel, P.-P. Grivel, B. Kaup, A. Haefliger, B. Malgrange, and F. Ehlers, *Algebraic D-modules*, Perspectives in Mathematics **2**, Academic Press, Boston, 1987. MR 89g:32014 Zbl 0642.32001

[Gabber 1981] O. Gabber, "The integrability of the characteristic variety", *Amer. J. Math.* **103**:3 (1981), 445–468. MR 82j:58104 Zbl 0492.16002

[Gelfand and Gelfand 1986] I. M. Gelfand and S. I. Gelfand, "Generalized hypergeometric equations", *Dokl. Akad. Nauk SSSR* **288**:2 (1986), 279–283. In Russian; translated in *Sov. Math., Dokl.* **33** (1986), 643–646. MR 87h:22013 Zbl 0634.58030

[Gelfand et al. 1987] I. M. Gelfand, M. I. Graev, and A. V. Zelevinsky, "Holonomic systems of equations and series of hypergeometric type", *Dokl. Akad. Nauk SSSR* **295**:1 (1987), 14–19. In Russian; translated in *Soviet Math. Dokl.* **36**:1 (1988), 5–10. MR 88j:58118 Zbl 0661.22005

[Kapranov 1998] M. Kapranov, "Hypergeometric functions on reductive groups", pp. 236–281 in *Integrable systems and algebraic geometry* (Kobe/Kyoto, 1997), edited by M.-H. Saito et al., World Sci. Publ., River Edge, NJ, 1998. MR 2000e:33018 Zbl 0987.33008

[Matusevich et al. 2005] L. F. Matusevich, E. Miller, and U. Walther, "Homological methods for hypergeometric families", *J. Amer. Math. Soc.* **18**:4 (2005), 919–941. MR 2007d:13027 Zbl 1095.13033

[Miller and Sturmfels 2005] E. Miller and B. Sturmfels, *Combinatorial commutative algebra*, Graduate Texts in Mathematics **227**, Springer, New York, 2005. MR 2006d:13001 Zbl 1090.13001

[Saito et al. 2000] M. Saito, B. Sturmfels, and N. Takayama, *Gröbner deformations of hypergeometric differential equations*, Algorithms and Computation in Mathematics **6**, Springer, Berlin, 2000. MR 2001i:13036 Zbl 0946.13021

[Schulze and Walther 2008] M. Schulze and U. Walther, "Irregularity of hypergeometric systems via slopes along coordinate subspaces", *Duke Math. J.* **142**:3 (2008), 465–509. MR 2009b:13067 Zbl 1144.13012

[Zelevinsky 1985] A. V. Zelevinsky, "Two remarks on graded nilpotent classes", *Uspekhi Mat. Nauk* **40**:1(241) (1985), 199–200. In Russian; translated in *Russ. Math. Surv.* **40**:1 (1985), 249–250. MR 86e:14027 Zbl 0577.20032

Received January 10, 2015. Revised February 19, 2015.

CHRISTINE BERKESCH ZAMAERE
SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455
UNITED STATES
cberkesc@math.umn.edu

STEPHEN GRIFFETH
INSTITUTO DE MATEMÁTICA Y FÍSICA
UNIVERSIDAD DE TALCA
TALCA
CHILE
sgriffeth@inst-mat.utalca.cl

EZRA MILLER
MATHEMATICS DEPARTMENT
DUKE UNIVERSITY
DURHAM, NC 27708
UNITED STATES
ezra@math.duke.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.287

COMPLEX INTERPOLATION AND TWISTED TWISTED HILBERT SPACES

FÉLIX CABELLO SÁNCHEZ, JESÚS M. F. CASTILLO AND NIGEL J. KALTON

We show that Rochberg's generalized interpolation spaces $\mathscr{X}^{(n)}$ arising from analytic families of Banach spaces form exact sequences $0 \to \mathscr{X}^{(n)} \to \mathscr{X}^{(n+k)} \to \mathscr{X}^{(k)} \to 0$. We study some structural properties of those sequences; in particular, we show that nontriviality, having strictly singular quotient map, or having strictly cosingular embedding depend only on the basic case $n = k = 1$. If we focus on the case of Hilbert spaces obtained from the interpolation scale of ℓ_p spaces, then $\mathscr{X}^{(2)}$ becomes the well-known Kalton–Peck space Z_2; we then show that $\mathscr{X}^{(n)}$ is (or embeds in, or is a quotient of) a twisted Hilbert space only if $n = 1, 2$, which solves a problem posed by David Yost; and that it does not contain ℓ_2 complemented unless $n = 1$. We construct another nontrivial twisted sum of Z_2 with itself that contains ℓ_2 complemented.

1. Introduction

Kalton and Peck [1979, Section 4] developed a method to produce nontrivial self-extensions of most quasi-Banach spaces with unconditional basis, including all Banach spaces apart from c_0; see [Cabello Sánchez et al. 2012, Theorem 1]. The most shining examples are perhaps the so-called Z_p spaces, which are twisted sums of the ℓ_p spaces. If, however, one wants to construct twisted sums of Z_p, the Kalton–Peck method simply does not work because of their poor unconditional structure. On the other hand, the existence of such twisted sums is guaranteed by the local theory of exact sequences, at least when $p > 1$; see, e.g., [Cabello Sánchez and Castillo 2004]. Our starting goal with this paper was to develop a method to obtain twisted sums of twisted sum spaces, keeping the Z_p spaces as the control case.

The path connecting interpolation theory and twisted sums was opened by Rochberg and Weiss, who introduce [1983] certain spaces which naturally arise in the study of "analytic families" of Banach spaces and that turn out to be twisted sums of the "intermediate" spaces. Actually, if $\mathscr{F}$ is the usual Calderón space of analytic

This research has been supported in part by project MTM2013-45643-C2-1-P.
MSC2010: 46B20, 46B70, 46M18, 46M35.
Keywords: Complex interpolation, twisted sums of Banach spaces.

functions on the strip $0 < \Re z < 1$ associated to the couple (ℓ_∞, ℓ_1) in the complex interpolation method, then, as is well known, $[\ell_\infty, \ell_1]_\theta = \{f(\theta) : f \in \mathcal{F}\} = \ell_p$, where $p = 1/\theta$ for $0 < \theta < 1$ and Z_p is isomorphic to

$$\{(f'(\theta), f(\theta)) : f \in \mathcal{F}\}, \tag{1}$$

with the quotient norm inherited from $\mathcal{F}$ — though this is not made explicit in [loc. cit.].

Nothing seems to prevent one from adding more derivatives to (1) and figuring out that the resulting space represents the iterated twisted sum spaces. Rochberg [1996] did just this, in the broader setting of analytic families of Banach spaces. Accomplishing that is not, by far, as simple as it sounds; and perhaps the turning point in Rochberg's approach is the use of Taylor coefficients instead of merely derivatives, as is suggested in [Kalton and Montgomery-Smith 2003, Section 10, p. 1161].

In this paper, we adopt such an approach, which can be considered a spin-off from that of Rochberg [1996], with several variations, the first of which is the use of admissible spaces of analytic functions instead of analytic families, which makes "reiteration" both unavailable and unnecessary. Thus, given an admissible space of analytic functions $\mathcal{F}$, we consider the space $\mathcal{X}^{(n)}$ of all possible lists of length n of Taylor coefficients of functions in $\mathcal{F}$ — at a fixed point z, which is understood from now on — endowed with the obvious infimum norm on it. Then we observe that if $m = n + k$, there is an exact sequence

$$0 \longrightarrow \mathcal{X}^{(n)} \longrightarrow \mathcal{X}^{(m)} \longrightarrow \mathcal{X}^{(k)} \longrightarrow 0 \tag{2}$$

and so $\mathcal{X}^{(m)}$ is a twisted sum of $\mathcal{X}^{(n)}$ and $\mathcal{X}^{(k)}$. The key nontrivial step here is obtaining the right form of the embedding. To this we devote Section 3 in which we obtain two (equivalent) representations for the embedding, depending on the representation of the spaces. Regarding the sequences themselves, we will show that many properties — such as nontriviality, having strictly singular quotient map, or having strictly cosingular embedding — depend only on the seed case $n = k = 1$. The nontriviality of this case has to be worked apart.

We then focus on the case in which $\mathcal{F}$ is the Calderón space of the couple (ℓ_∞, ℓ_1). If we fix $z = \frac{1}{2}$, then $\mathcal{X}^{(1)} = \ell_2$ and $\mathcal{X}^{(2)}$ is the Kalton–Peck space Z_2 [1979]. The space $\mathcal{X}^{(3)}$ is both a twisted sum of ℓ_2 with Z_2 and a twisted sum of Z_2 with ℓ_2, and $\mathcal{X}^{(4)}$ is, among other possibilities, a twisted sum of Z_2 with itself, as desired. We then pass to establish structural properties of the spaces $\mathcal{X}^{(n)}$ and of the sequences (2). Regarding the spaces, we will show that $\mathcal{X}^{(n)}$ is (or embeds in, or is a quotient of) a twisted Hilbert space only if $n = 1, 2$ — which solves a problem posed by David Yost — and that it does not contain ℓ_2 complemented unless $n = 1$. To put this result in perspective, we will construct a nontrivial twisted sum of Z_2 with itself that contains ℓ_2 complemented.

2. Preliminaries

We warmly recommend the reader who is not familiar with [Kalton and Peck 1979] or [Rochberg 1996] to postpone reading this article until getting acquainted with them. Perusing the papers [Cwikel et al. 1989; Rochberg 2007], the article [Kalton and Montgomery-Smith 2003], and the monograph [Castillo and González 1997] can help with the background. Anyway, the basic ingredients to read this paper are described next.

2.1. ***Exact sequences.*** A short sequence of Banach spaces and (linear, bounded) operators

$$0 \longrightarrow A \xrightarrow{I} B \xrightarrow{Q} C \longrightarrow 0 \tag{3}$$

is said to be exact if the kernel of each arrow equals the image of the preceding one. As $I(A)$ is closed in B the operator I embeds A as a subspace of B and C is isomorphic to the quotient $B/I(A)$, by the open mapping theorem. For this reason one often says that B is a twisted sum of A with C (in that order); the whole sequence (3) is said to be an extension of C by A (the order was reversed for "functorial" reasons).

The extension (3) is said to be trivial if there is an operator $P : B \to A$ such that $P \circ I = \mathbf{1}_A$, i.e., $I(A)$ is complemented in B; equivalently, there is an operator $J : C \to B$ such that $Q \circ J = \mathbf{1}_C$. In this case $P \times Q : B \to A \times C$ is an isomorphism, with inverse $I \oplus J$, and thus the "twisted sum" B is (isomorphic to) the direct sum $A \oplus C = A \times C$.

2.2. ***Admissible spaces of analytic functions.*** We will work within the framework of an admissible space of analytic functions as defined in [Kalton and Montgomery-Smith 2003, Section 10]. So, let U be an open set of $\mathbb{C}$, conformally equivalent to the disc $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$, and let W be a complex Banach space. A Banach space $\mathscr{F}$ of analytic functions $F : U \to W$ is said to be admissible provided:

(a) For each $z \in U$, the evaluation map $\delta_z : \mathscr{F} \to W$ is bounded.

(b) If $\varphi : U \to \mathbb{D}$ is a conformal equivalence, then $F \in \mathscr{F}$ if and only if $\varphi \cdot F \in \mathscr{F}$ and $\|\varphi \cdot F\|_{\mathscr{F}} = \|F\|_{\mathscr{F}}$.

For each $z \in U$ we define $X_z = \{x \in W : x = F(z) \text{ for some } F \in \mathscr{F}\}$ with the norm $\|x\| = \inf\{\|F\|_{\mathscr{F}} : x = F(z)\}$ so that X_z is isometric to $\mathscr{F}/\ker \delta_z$ with the quotient norm. One often says that $(X_z)_{z \in U}$ is an analytic family of Banach spaces. The simplest examples arise from complex interpolation. Indeed, let (X_0, X_1) be a Banach couple and take $W = X_0 + X_1$ and U the strip $0 < \Re z < 1$. Let $\mathscr{F} = \mathscr{C}(X_0, X_1)$ be the Calderón space of those continuous functions $F : \overline{U} \to W$ which are analytic on U and satisfy the boundary conditions: for $k = 0, 1$, one has

$F(k+ti) \in X_k$ and $\|F\|_{\mathscr{C}} = \sup\{\|F(k+ti)\|_{X_k} : t \in \mathbb{R}, k = 0, 1\} < \infty$. Then $\mathscr{F}$ is admissible and $X_z = [X_0, X_1]_\theta$, with $\theta = \Re z$, is an analytic family.

It is important now to realize that when $\mathscr{F}$ is admissible, the map $\delta_z^n : \mathscr{F} \to W$, evaluation of the n-th derivative at z, is bounded for all $z \in U$ and all $n \in \mathbb{N}$ by an iterated use of (a), the definition of derivative, and the principle of uniform boundedness. Thus, it makes sense to consider the Banach spaces

$$\mathscr{F} \Big/ \bigcap_{i<n} \ker \delta_z^i \quad \text{for } n \in \mathbb{N}. \tag{4}$$

3. Exact sequences of derived spaces

3.1. ***Lists of Taylor coefficients.*** Following Rochberg, let us fix $z \in U$ and consider the following spaces:

$$\mathscr{X}_z^{(n)} = \{(x_{n-1}, \ldots, x_0) \in W^n : x_i = \hat{f}[i; z] \text{ for some } f \in \mathscr{F} \text{ and all } 0 \le i < n\},$$

where $\hat{f}[i; z] = f^{(i)}(z)/i!$ is the i-th Taylor coefficient of f at z. Thus, the elements of $\mathscr{X}_z^{(n)}$ are (truncated) sequences of Taylor's coefficients (at z) of functions in $\mathscr{F}$ arranged in decreasing order. Here, we deviate from Rochberg notation in two points: first, the superscript (n) refers to the "number of variables" and not to the highest derivative, and second, we have arranged Taylor coefficients decreasingly in order to match with the usual notation for twisted sums, with the subspace on the left and the quotient on the right. If we equip $\mathscr{X}_z^{(n)}$ with the obvious quotient norm, then it is isometric to $\mathscr{F}/\bigcap_{i<n} \ker \delta_z^i$ via Taylor coefficients and so it is complete. From now on we shall omit the base point z, which is understood to be fixed.

3.2. ***Operators.*** Next, we introduce certain "natural" operators linking the various spaces $\mathscr{X}^{(n)}$ as n varies. Those operators will be used to construct the exact sequences we want.

To this end, for $1 \le n, k < m$ we denote by $\imath_{n,m} : W^n \to W^m$ the inclusion on the left given by $\imath_{n,m}(x_n, \ldots, x_1) = (x_n, \ldots, x_1, 0 \ldots, 0)$ and by $\pi_{m,k} : W^m \to W^k$ the projection on the right given by $\pi_{m,k}(x_m, \ldots, x_k, \ldots, x_1) = (x_k, \ldots, x_1)$. While $\pi_{m,k}$ is obviously a quotient map from $\mathscr{X}^{(m)}$ onto $\mathscr{X}^{(k)}$, it is not clear at all that $\imath_{n,m}$ maps $\mathscr{X}^{(k)}$ to $\mathscr{X}^{(n)}$, let alone that it is continuous. To prove that this is indeed the case, we need to do some extra work.

Observe that if φ is as in (b) and if Π is a "polynomial" in φ; that is, $\Pi = \sum_i a_i \varphi^i$ for some finite sequence of complex numbers (a_i), then $\Pi \cdot f \in \mathscr{F}$ for each $f \in \mathscr{F}$ and $\|\Pi \cdot f\|_{\mathscr{F}} \le (\sum_i |a_i|) \|f\|_{\mathscr{F}}$.

Lemma 1. *Let $\varphi : U \to \mathbb{D}$ be a conformal equivalence vanishing at z. Then, for $0 \le k \le m$, there is a polynomial P of degree at most m such that*

$$\widehat{P \circ \varphi}[i; z] = \delta_{ik} \quad \textit{for every } 0 \le i \le m.$$

Proof. If $f : U \to \mathbb{C}$ is holomorphic, then $f \circ \varphi^{-1}$ is holomorphic on the disk and

$$f(\varphi^{-1}(w)) = \sum_{n=0}^{\infty} a_n w^n, \quad |w| < 1,$$

where a_n is the n-th Taylor coefficient of $f \circ \varphi^{-1}$ at the origin. In particular, $f \circ \varphi^{-1}$ has a contact of order m with the polynomial defined by $P(w) = \sum_{n=0}^{m} a_n w^n$ at the origin. As φ is a conformal equivalence, we have that $f = f \circ \varphi^{-1} \circ \varphi$ has a contact of order m with the function

$$P \circ \varphi = \sum_{n=0}^{m} a_n \varphi^n \quad \text{at } z = \varphi^{-1}(0).$$

In particular, the first m derivatives of f and $\sum_{n=0}^{m} a_n \varphi^n$ agree at z. The lemma follows by applying this construction to the function $f(w) = (w - z)^k$. □

The following is a slight generalization of [Rochberg 1996, Proposition 3.1]:

Proposition 2. *Suppose* $1 \le n, k < m$. *Then*,

(a) *The map* $\imath_{n,m} : \mathscr{X}^{(n)} \to \mathscr{X}^{(m)}$ *is bounded.*

(b) *The map* $\pi_{m,k} : \mathscr{X}^{(m)} \to \mathscr{X}^{(k)}$ *is an "isometric" quotient.*

Proof. Part (b) is obvious. To prove (a) we must prove that there is a constant M such that if $(x_n, \dots, x_1)$ is the list of Taylor coefficients of some $f \in \mathscr{F}$, then there is another $g \in \mathscr{F}$ whose coefficients are $(x_n, \dots, x_1, 0, \dots, 0)$ with $\|g\|_{\mathscr{F}} \le M \|f\|_{\mathscr{F}}$. Set $k = m - n$ and apply Lemma 1 to get a polynomial $\Pi = \sum_{i=0}^{m-1} a_i \varphi^i$ such that $\widehat{\Pi}[i; z] = \delta_{ik}$ for $0 \le i < n + k$ and take $M = \sum_{i=0}^{n+k-1} |a_i|$. Now, if $f \in \mathscr{F}$ and $g = \Pi f$, then $\|g\|_{\mathscr{F}} \le M \|f\|_{\mathscr{F}}$. Moreover, for $i \in [0, n + k)$,

$$\hat{g}[i; z] = (\widehat{\Pi f})[i; z] = \sum_{j=0}^{i} \widehat{\Pi}[j; z] \cdot \hat{f}[i - j; z],$$

by the Leibniz rule. Hence, $\hat{g}[i; z] = 0$ if $i < k$ and $\hat{g}[i; z] = \hat{f}[i - k; z]$ if $i \ge k$, as required. □

3.3. ***Exactness.*** From now on, we will omit the names $\imath_{m,n}$ and $\pi_{m,k}$, so unlabeled arrows $\mathscr{X}^{(n)} \to \mathscr{X}^{(m)}$ must be understood to be $\imath_{n,m}$ if $n \le m$ and $\pi_{n,m}$ if $n \ge m$, unless otherwise declared. With these conventions, the aim of this section is to prove that, given integers n and k, the "obvious" sequence $0 \longrightarrow \mathscr{X}^{(n)} \longrightarrow \mathscr{X}^{(n+k)} \longrightarrow \mathscr{X}^{(k)} \longrightarrow 0$ is exact.

First of all, observe that the various possible sequences passing through a given $\mathscr{X}^{(m)}$ are compatible in the sense that if $m = k + n = i + j$, with $k < i$ say, then the following diagram is commutative:

$$
\begin{tikzcd}
\mathscr{X}^{(j)} \arrow[r, equal] \arrow[d] & \mathscr{X}^{(j)} \arrow[d] & \\
\mathscr{X}^{(n)} \arrow[r] \arrow[d] & \mathscr{X}^{(m)} \arrow[r] \arrow[d] & \mathscr{X}^{(k)} \arrow[d, equal] \\
\mathscr{X}^{(n-j)} \arrow[r] & \mathscr{X}^{(i)} \arrow[r] & \mathscr{X}^{(k)}
\end{tikzcd}
\tag{5}
$$

The key point is isolated in the next lemma.

Lemma 3. *If* $(x, 0, \dots, 0) \in \mathscr{X}^{(k+1)}$, *then* $x \in \mathscr{X}^{(1)}$.

Proof. Pick $x \in W$ and suppose $(x, 0, \dots, 0) \in \mathscr{X}^{(k+1)}$. Let us take $f \in \mathscr{F}$ such that $\hat{f}[i; z] = 0$ for $i < k$ and $x = \hat{f}[k; z]$. Then f has a zero of order $k-1$ at z and it can be written as $f = \varphi^k g$, where $g : U \to W$ is analytic. It follows from (b) that $g \in \mathscr{F}$ and $\|g\|_{\mathscr{F}} = \|f\|_{\mathscr{F}}$. But,

$$x = \hat{f}[k; z] = (\widehat{\varphi^k g})[k; z] = \sum_{i=0}^{k} (\widehat{\varphi^k})[i; z] \cdot \hat{g}[k-i; z] = \frac{(\varphi^k)^{(k)}(z) g(z)}{k!} = \varphi'(z)^k g(z)$$

and so $x \in \mathscr{X}^{(1)}$. $\square$

Theorem 4. *The sequence* $0 \longrightarrow \mathscr{X}^{(n)} \longrightarrow \mathscr{X}^{(n+k)} \longrightarrow \mathscr{X}^{(k)} \longrightarrow 0$ *is exact.*

Proof. The proof proceeds by induction on $m = n + k$. The previous lemma shows that for every $m \in \mathbb{N}$, the sequence $0 \longrightarrow \mathscr{X}^{(1)} \longrightarrow Z^{(m)} \longrightarrow \mathscr{X}^{(m-1)} \longrightarrow 0$ is exact. By the induction hypothesis, the sequence $0 \longrightarrow \mathscr{X}^{(n-1)} \longrightarrow \mathscr{X}^{(m-1)} \longrightarrow \mathscr{X}^{(k)} \longrightarrow 0$ is also exact. The compatibility of such sequences yields the commutative diagram:

$$
\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & & \\
 & \mathscr{X}^{(1)} \arrow[r, equal] \arrow[d] & \mathscr{X}^{(1)} \arrow[d] & & \\
 & \mathscr{X}^{(n)} \arrow[r] \arrow[d] & \mathscr{X}^{(m)} \arrow[r] \arrow[d] & \mathscr{X}^{(k)} \arrow[d, equal] & \\
0 \arrow[r] & \mathscr{X}^{(n-1)} \arrow[r] \arrow[d] & \mathscr{X}^{(m-1)} \arrow[r] \arrow[d] & \mathscr{X}^{(k)} \arrow[r] & 0 \\
 & 0 & 0 & &
\end{tikzcd}
$$

and a simple arrow chase shows that the middle sequence must also be exact. $\square$

Corollary 5. *If* $(x_n, \dots, x_1, 0, \dots, 0) \in \mathscr{X}^{(n+k)}$, *then* $(x_n, \dots, x_1) \in \mathscr{X}^{(n)}$.

This implies that $\|(x_n, \dots, x_1, 0, \dots, 0)\|_{\mathscr{X}^{(n+k)}}$ is equivalent to $\|(x_n, \dots, x_1)\|_{\mathscr{X}^{(n)}}$, although we will not pursue any bound here.

A new look can be had now at (5) to exploit its form to study the splitting of the exact sequences it contains. After Theorem 4 the diagram has become:

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & & \\
 & \mathscr{X}^{(j)} \arrow[r, equal] \arrow[d] & \mathscr{X}^{(j)} \arrow[d] & & \\
0 \arrow[r] & \mathscr{X}^{(n)} \arrow[r] \arrow[d] & \mathscr{X}^{(m)} \arrow[r] \arrow[d] & \mathscr{X}^{(k)} \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \mathscr{X}^{(n-j)} \arrow[r] \arrow[d] & \mathscr{X}^{(i)} \arrow[r] \arrow[d] & \mathscr{X}^{(k)} \arrow[r] & 0 \\
 & 0 & 0 & &
\end{tikzcd} \tag{6}$$

Thus, if the middle horizontal sequence splits, then so does the lower one, and if the middle vertical sequence splits, then so does the vertical sequence on the left. Putting together these two pieces, one gets the commutative diagram:

$$\begin{tikzcd}
0 \arrow[r] & \mathscr{X}^{(n)} \arrow[r] \arrow[d] & \mathscr{X}^{(n+k)} \arrow[r] \arrow[d] & \mathscr{X}^{(k)} \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \mathscr{X}^{(1)} \arrow[r] \arrow[d, equal] & \mathscr{X}^{(k+1)} \arrow[r] & \mathscr{X}^{(k)} \arrow[r] & 0 \\
0 \arrow[r] & \mathscr{X}^{(1)} \arrow[r] & \mathscr{X}^{(2)} \arrow[r] \arrow[u] & \mathscr{X}^{(1)} \arrow[r] \arrow[u] & 0
\end{tikzcd}$$

which immediately yields the following corollary.

Corollary 6. *If the sequence* $0 \longrightarrow \mathscr{X}^{(n)} \longrightarrow \mathscr{X}^{(n+k)} \longrightarrow \mathscr{X}^{(k)} \longrightarrow 0$ *is nontrivial for* $n = k = 1$, *then it is nontrivial for all integers* n *and* k.

3.4. ***An isometric variant.*** There is another form for the short exact sequences $0 \to \mathscr{X}^{(n)} \to \mathscr{X}^{(n+k)} \to \mathscr{X}^{(k)} \to 0$ which is even easier to describe in abstract terms. Consider again the quotient spaces

$$Q_z^{(n)} = \mathscr{F} \Big/ \bigcap_{i<n} \ker \delta_z^i \quad \text{for } n \in \mathbb{N}.$$

These spaces are isometric to the corresponding $\mathscr{X}_z^{(n)}$ via Taylor coefficients, but we do not need this fact at this moment. Fix integers n and k. It is clear that there is a natural quotient map from $Q_z^{(n+k)}$ onto $Q_z^{(k)}$ that we shall not even label. Less obvious is that the kernel of this map is isometric to $\mathscr{X}_z^{(n)}$, although this time the isometry is not "natural". To see this, let us fix a conformal equivalence $\varphi : U \to \mathbb{D}$ having a (single) zero at z. (We observe that if Π is another conformal equivalence with $\Pi(z) = 0$, then $\Pi = \lambda\varphi$, where $\lambda \in \mathbb{T}$; thus, φ is unique if we insist that $\varphi'(0)$ is real and positive.)

Now, recall that $f \in \bigcap_{i<k} \ker \delta_z^i$ if and only if there is a (necessarily unique) $g \in \mathscr{F}$ such that $f = \varphi^k g$ and one has $\|f\|_{\mathscr{F}} = \|g\|_{\mathscr{F}}$, by (b). Therefore, it is clear that the map $f \in \mathscr{F} \mapsto \varphi^k f \in \mathscr{F}$ induces an isometry of $Q_z^{(n)}$ into $Q_z^{(n+k)}$ whose range is

$$\ker(Q_z^{(n+k)} \to Q_z^{(k)}).$$

As a consequence, the space $Q_z^{(n+k)}$ is an "isometric" twisted sum of $Q_z^{(n)}$ and $Q_z^{(k)}$. More precisely, the short sequence

$$0 \longrightarrow Q_z^{(n)} \xrightarrow{\varphi^k \cdot} Q_z^{(n+k)} \longrightarrow Q_z^{(k)} \longrightarrow 0 \tag{7}$$

is exact.

From now on, we will omit the base point z, which is understood. As before, the decompositions of a given $Q_z^{(m)}$ into a twisted sum of the preceding spaces $Q_z^{(n)}$ are all compatible in the sense that if $m = k + n = i + j$, with $k < i$, then the following diagram is commutative:

$$\begin{array}{ccccc} Q^{(j)} & = & Q^{(j)} & & \\ \Big\downarrow{\scriptstyle \varphi^{j-n}\cdot} & & \Big\downarrow{\scriptstyle \varphi^{i}\cdot} & & \\ Q^{(n)} & \xrightarrow{\varphi^k \cdot} & Q^{(m)} & \longrightarrow & Q^{(k)} \\ \Big\downarrow & & \Big\downarrow & & \Big\| \\ Q^{(j-n)} & \xrightarrow{\varphi^k \cdot} & Q^{(i)} & \longrightarrow & Q^{(k)} \end{array}$$

It is interesting to compare the sequence (7) to that appearing in Theorem 4. To this end, we observe that, after identifying $\mathscr{X}^{(m)}$ and $Q^{(m)}$ through Taylor coefficients, the operator $\mathscr{X}^{(n)} \to \mathscr{X}^{(n+k)}$ which corresponds to $\imath_{n,n+k} : \mathscr{X}^{(n)} \to \mathscr{X}^{(n+k)}$ is just multiplication by Π, where Π is the polynomial appearing in the proof of Proposition 2(a), that is,

$$\Pi = \sum_{0 \le i < n+k} a_i \varphi^i, \quad \text{with } \widehat{\Pi}[i; z] = \delta_{ik} \text{ for } 0 \le i < n+k.$$

Clearly, $a_i = 0$ for $0 \le i < k$, so $\Pi = \varphi^k \psi$, where $\psi = \sum_{k \le i < n+k} a_i \varphi^{i-k}$. Thus, the following diagram is commutative:

$$\begin{tikzcd}
Q^{(n)} \arrow[r, "\varphi^k \cdot"] \arrow[d, equal] & Q^{(n+k)} \arrow[r] \arrow[d, "\psi \cdot"'] & Q^{(k)} \arrow[d, equal] \\
Q^{(n)} \arrow[r, "\Pi \cdot"] \arrow[d, "\hat{\cdot}"'] & Q^{(n+k)} \arrow[r] \arrow[d, "\hat{\cdot}"'] & Q^{(k)} \arrow[d, "\hat{\cdot}"'] \\
\mathscr{X}^{(n)} \arrow[r, "\iota_{n,n+k}"] & \mathscr{X}^{(n+k)} \arrow[r] & \mathscr{X}^{(k)}
\end{tikzcd}$$

It follows from the 3-lemma (see, for instance, [Hilton and Stammbach 1971, Lemma 1.1]) and the open mapping theorem that multiplication by ψ induces an automorphism of $Q^{(n+k)}$. So, in the preceding diagram, the first row is equivalent to the second one, and both are "isomorphically equivalent" (in the language of [Castillo and Moreno 2004, p. 256]) to the third one, which means that the three sequences have the same "isomorphic" properties.

3.5. ***The space $\mathscr{X}^{(n+k)}$ as a twisted sum of $\mathscr{X}^{(n)}$ and $\mathscr{X}^{(k)}$.*** It is a part of the (by now) classical theory of twisted sums as developed by Kalton [1978, Proposition 3.3; Kalton and Peck 1979, Theorem 2.4] that if A and C are Banach or quasi-Banach spaces, then every short exact sequence $0 \to A \to B \to C \to 0$ arises, up to equivalence, from a quasilinear map from C to A. Thus, in view of Theorem 4, given integers k and n, there must be some quasilinear map $\Omega_{k,n}$ associated to the exact sequence $0 \to \mathscr{X}^{(n)} \to \mathscr{X}^{(n+k)} \to \mathscr{X}^{(k)} \to 0$. From an abstract point of view, the description of $\Omega_{k,n}$ is rather easy: fix some (small) $\varepsilon > 0$. Now, given $x = (x_{k-1}, \dots, x_0) \in \mathscr{X}^{(k)}$, select (homogeneously) $f \in \mathscr{F}$ such that $\|f\| \le (1+\varepsilon)\|x\|_{\mathscr{X}_{(n)}}$ and $\hat{f}[i; z] = x_i$ for $0 \le i < k$, and define $\Omega_{k,n} : \mathscr{X}^{(k)} \to W^n$ by letting

$$\Omega_{k,n}(x) = (\hat{f}[n+k-1; z], \dots, \hat{f}[k; z]). \tag{8}$$

Following the uses of the theory, the twisted sum space (sometimes known as the derived space) is then defined by

$$\mathscr{X}^{(n)} \oplus_{\Omega_{k,n}} \mathscr{X}^{(k)} = \{(y, x) \in W^n \times W^k : x \in \mathscr{X}^{(k)},\ y - \Omega_{k,n}(x) \in \mathscr{X}^{(n)}\},$$

endowed with the quasinorm

$$\|(y, x)\|_{\Omega_{k,n}} = \|y - \Omega_{k,x}(x)\|_{\mathscr{X}^{(n)}} + \|x\|_{\mathscr{X}^{(k)}}. \tag{9}$$

Of course, it has not yet been proved either that $\Omega_{k,n}$ is quasilinear or that (9) defines a quasinorm. We may skip these steps since we have the following.

Proposition 7. *The spaces $\mathscr{X}^{(n)} \oplus_{\Omega_{k,n}} \mathscr{X}^{(k)}$ and $\mathscr{X}^{(n+k)}$ agree. Their quasinorms are equivalent.*

Proof. Suppose $(y, x) = (y_{n-1}, \dots, y_0, x_{k-1}, \dots, x_0) \in \mathscr{X}^{(n+k)}$ so that there is $F \in \mathscr{F}$ whose list of Taylor coefficients begins with (y, x). Then $x \in \mathscr{X}^{(k)}$ and $(\Omega_{k,n}(x), x) \in \mathscr{X}^{(n+k)}$, so $(y, x) - (\Omega_{k,n}(x), x) = (y - x, \Omega_{k,n}(x), 0)$ belongs to $\mathscr{X}^{(n+k)}$ and by Lemma 3 we have $y - \Omega_{k,n}(x) \in \mathscr{X}^{(n)}$. Regarding the involved norms,

$$\|y - \Omega_{k,x}(x)\|_{\mathscr{X}^{(n)}} \le C\big(\|(y, x)\|_{\mathscr{X}^{(n+k)}} - \|(\Omega_{k,n}(x), x)\|_{\mathscr{X}^{(n+k)}}\big) \le (C+1)\,\|(y, x)\|_{\mathscr{X}^{(n+k)}},$$

where C is the constant implicit in Lemma 3. Hence,

$$\|(y, x)\|_{\Omega_{k,n}} \le (C+2)\,\|(y, x)\|_{\mathscr{X}^{(n+k)}}.$$

As for the other containment, suppose $(y, x) \in \mathscr{X}^{(n)} \oplus_{\Omega_{k,n}} \mathscr{X}^{(k)}$, that is, $x \in \mathscr{X}^{(k)}$ and $y - \Omega_{k,n}(x) \in \mathscr{X}^{(n)}$. Then if f is the function associated to $\Omega_{k,n}(x)$ as in (8) and $g \in \mathscr{F}$ is almost optimal for $y - \Omega_{k,n}(x) \in \mathscr{X}^{(n)}$, taking Π as in Lemma 1, we have that (y, x) is the list of Taylor coefficients of $F = f + \Pi \cdot g$, so $(y, x) \in \mathscr{X}^{(n+k)}$ and

$$\|(y, x)\|_{\mathscr{X}^{(n+k)}} \le \|f + \Pi \cdot g\|_{\mathscr{F}} \le (1+\varepsilon)\big(M\|y - \Omega_{k,n}(x)\|_{\mathscr{X}^{(n)}} + \|x\|_{\mathscr{X}^{(k)}}\big),$$

where M is as in the proof of Proposition 2(a). □

4. Singularity of the exact sequences of derived spaces

Recall that an operator is said to be strictly singular if its restriction to an infinite dimensional subspace of its domain is never an isomorphism; and that an operator $u : A \to B$ is strictly cosingular if for every infinite codimensional subspace C of B the composition $\pi \circ u : A \to B \to B/C$ fails to be onto; equivalently, $u^* : B^* \to A^*$ is not an isomorphism when restricted to any weakly- closed infinite-dimensional subspace of B^*. Strictly singular operators were introduced by Kato [1958] and strictly cosingular operators by Pełczyński [1965].

An exact sequence is said to be singular when the quotient map is strictly singular and will be called cosingular when the embedding is strictly cosingular. We refer the reader to [Castillo and Moreno Salguero 2007; Cabello Sánchez et al. 2012] for some steps into the theory of singular and cosingular sequences. The Kalton–Peck sequences $0 \to \ell_p \to Z_p \to \ell_p \to 0$ are singular for all $p \in (0, \infty)$ and cosingular at least for $p \in (1, \infty)$. We need the following result.

Lemma 8. *Assume one has a commutative diagram*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{I} & B & \xrightarrow{Q} & C & \longrightarrow & 0 \\ & & {\scriptstyle t}\downarrow & & {\scriptstyle T}\downarrow & & \| & & \\ 0 & \longrightarrow & D & \longrightarrow & E & \longrightarrow & C & \longrightarrow & 0 \end{array}$$

with exact rows. If both Q and t are strictly singular, then T is strictly singular.

Proof. We need the following characterization of strictly singular quotient maps. Let B be a Banach space and A a closed subspace of B. Then the quotient map $Q : B \to B/A$ is strictly singular if and only if for every infinite-dimensional subspace $B' \subset B$ there is an infinite-dimensional $A' \subset A$ and a compact (actually nuclear) operator $K : A' \to B$ such that $I + K$ embeds A' into B' isomorphically. This maybe folklore; see [Castillo et al. 2012, Proposition 3.2] for an explicit proof. A certainly classical result establishes that an operator $t : A \to D$ is strictly singular if given any infinite dimensional subspace $A' \subset A$ and $\varepsilon > 0$ there is a further infinite dimensional subspace $A'' \subset A'$ such that $\|t|_{A''}\| < \varepsilon$. Both things together yield that given $B' \subset B$ there is $A'' \subset A' \subset A$ such that $I + K : A'' \to B'$ is an into isomorphism and $\|t|_{A''}\| < \varepsilon$. There is no loss of generality assuming that $\|K|_{A''}\| < \varepsilon$. Therefore, $\|T|_{(I+K)(A'')}\| = \|t|_{A''} + TK|_{A''}\| < (1 + \|T\|)\varepsilon$. □

We thus obtain the "strictly singular counterpart" to Corollary 6:

Proposition 9. *If the natural quotient map $\mathscr{X}^{(2)} \to \mathscr{X}^{(1)}$ is strictly singular, then so is $\mathscr{X}^{(n)} \to \mathscr{X}^{(k)}$ for every $n > k$.*

Proof. Note that if $n > m > k$, then $\mathscr{X}^{(n)} \to \mathscr{X}^{(k)}$ is $\mathscr{X}^{(n)} \to \mathscr{X}^{(m)}$ followed by $\mathscr{X}^{(m)} \to \mathscr{X}^{(k)}$. As the composition of a strictly singular operator with any operator is again strictly singular, we have that the proposition is trivial if $k = 1$ and also that one can assume $n = k + 1$. We shall prove that $\mathscr{X}^{(k+1)} \to \mathscr{X}^{(k)}$ is strictly singular by induction on $k \in \mathbb{N}$. There is nothing to prove for $k = 1$, so assume $k > 1$. Since one has the commutative diagram

$$\begin{tikzcd}
 & \mathscr{X}^{(1)} \arrow[r, equal] \arrow[d] & \mathscr{X}^{(1)} \arrow[d] & & \\
0 \arrow[r] & \mathscr{X}^{(k)} \arrow[r] \arrow[d, "\pi_{k,k-1}"'] & \mathscr{X}^{(k+1)} \arrow[r, "\pi_{k+1,1}"] \arrow[d, "\pi_{k+1,k}"'] & \mathscr{X}^{(1)} \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \mathscr{X}^{(k-1)} \arrow[r] & \mathscr{X}^{(k)} \arrow[r] & \mathscr{X}^{(1)} \arrow[r] & 0
\end{tikzcd}$$

But $\pi_{k+1,1}$ is strictly singular and so is $\pi_{k,k-1}$, by the induction hypothesis. Thus, the result follows from Lemma 8. □

We omit the proofs of the dual results:

Lemma 10. *Assume one has a commutative diagram*

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r, "I"] & B \arrow[r, "Q"] & C \arrow[r] & 0 \\
0 \arrow[r] & A \arrow[r] \arrow[u, equal] & D \arrow[r] \arrow[u, "T"'] & E \arrow[r] \arrow[u, "t"'] & 0
\end{tikzcd}$$

with exact rows. If both I and t are strictly cosingular, then T is strictly cosingular.

Corollary 11. *If the inclusion map $\mathscr{X}^{(1)} \to \mathscr{X}^{(2)}$ is strictly cosingular, then so is $\mathscr{X}^{(k)} \to \mathscr{X}^{(n)}$ for every $k < n$.*

5. Applications to Hilbert spaces

5.1. ***The quasilinear map associated to twisted Kalton–Peck spaces.*** Some results in this section are, essentially, in [Rochberg 1996, Section 6B]. Let us consider the following variation of the Calderón space associated to the Banach couple (ℓ_∞, ℓ_1) which is designed to simplify the computation of extremals. Take $U = \mathbb{S} = \{z \in \mathbb{C} : 0 < \Re z < 1\}$, with $W = \ell_\infty$, and let $\mathscr{F}$ be the space of analytic functions $F : \mathbb{S} \to \ell_\infty$ having the following properties:

(1) F extends to a $\sigma(\ell_\infty, \ell_1)$ continuous function on $\overline{\mathbb{S}}$ that we denote again by F.

(2) $\|F\|_{\mathscr{F}} = \sup\{\|F(it)\|_\infty, \|F(1+it)\|_1 : t \in \mathbb{R}\} < \infty$.

Let $(\mathfrak{X}_z)_{z\in\mathbb{S}}$ denote the analytic family induced by $\mathscr{F}$. Then, of course, $\mathfrak{X}_z = [\ell_\infty, \ell_1]_\theta = \ell_p$, where $\theta = \Re z$ and $p = \frac{1}{\theta}$ for $\theta \in (0, 1)$, and, in particular, $\mathfrak{X}_z = \ell_2$ for $z = \frac{1}{2}$. For the remainder of this section, we fix $z = \frac{1}{2}$ as the base point.

If x is normalized in ℓ_2, then $F_x(z) = u|x|^{2z}$ is normalized in $\mathscr{F}$ (although it does not belong to $\mathscr{C}(\ell_\infty, \ell_1)$ in general), and one has $F_x\big(\frac{1}{2}\big) = x$, where $x = u|x|$ is the "polar decomposition" of x. Now,

$$F_x = u|x||x|^{2z-1} = x|x|^{2(z-1/2)} = x\sum_{n=0}^{\infty} \frac{2^n \log^n|x|}{n!}\big(z - \tfrac{1}{2}\big)^n,$$

and

$$\hat{F}_x\big[n; \tfrac{1}{2}\big] = \frac{2^n x \log^n|x|}{n!},$$

if $\|x\|_2 = 1$. For arbitrary $x \in \ell_2$, we have, by homogeneity,

$$\hat{F}_x\big[n; \tfrac{1}{2}\big] = \frac{2^n x}{n!} \log^n \frac{|x|}{\|x\|_2}. \tag{10}$$

In particular,

$$\begin{aligned}\Omega_{1,n}(x) &= \big(\hat{F}_x\big[n; \tfrac{1}{2}\big], \dots, \hat{F}_x\big[1; \tfrac{1}{2}\big]\big)\\ &= x\left(\frac{2^n}{n!}\log^n\frac{|x|}{\|x\|_2}, \dots, 2\log^2\frac{|x|}{\|x\|_2}, 2\log\frac{|x|}{\|x\|_2}\right),\end{aligned}$$

which allows us to describe the corresponding spaces $\mathfrak{X}^{(n)}$, for small n, as follows. First, we have

$$\mathfrak{X}^{(2)} \approx \ell_2 \oplus_{\Omega_{1,1}} \ell_2 = \left\{(y, x) : \left\| y - 2x\log\frac{|x|}{\|x\|_2}\right\|_2 + \|x\|_2 < \infty\right\},$$

which is isomorphic to the Kalton–Peck space Z_2 [1979, Section 6], which was defined there by the quasilinear map $x \mapsto x\log(\|x\|_2/|x|) = -\frac{1}{2}\Omega_{1,1}(x)$. Also,

$$\mathscr{X}^{(3)} \approx \mathscr{X}^{(2)} \oplus_{\Omega_{1,2}} \ell_2 \approx (\ell_2 \oplus_{\Omega_{1,1}} \ell_2) \oplus_{\Omega_{1,2}} \ell_2,$$

and the norm of $\mathscr{X}^{(3)}$ is equivalent to

$$\|(z,y,x)\|_{\Omega_{1,2}} = \left\| \left(z - 2x\log^2 \frac{|x|}{\|x\|_2},\; y - 2x\log \frac{|x|}{\|x\|_2} \right) \right\|_{\Omega_{1,1}} + \|x\|_2. \tag{11}$$

We will also finally display the quasilinear map $\Omega_{2,2}$ that allows one to represent $\mathscr{X}^{(4)}$ as a twisted sum of $\mathscr{X}^{(2)}$ with itself. After all, this was the starting point of this research. Let $\varphi : \mathbb{S} \to \mathbb{D}$ be conformal equivalence vanishing at $z_0 = \frac{1}{2}$ and let $\Pi = \sum_{1\le i\le 3} a_i \varphi^i$ be such that $\widehat{\Pi}\big[i; \frac{1}{2}\big] = \delta_{1i}$ for $0 \le i \le 3$. Given $(y,x) \in \mathscr{X}^{(2)}$, we construct an allowable $F_{(y,x)} \in \mathscr{F}$ as follows. Let F_x and $F_{(y-\Omega(x))}$ be extremals for x and $y - \Omega(x)$, respectively, where $\Omega(x) = \Omega_{1,1}(x) = F_x'\big(\frac{1}{2}\big) = 2x\log(|x|/\|x\|_2)$. Put

$$F_{(y,x)} = \Pi \cdot F_{(y-\Omega(x))} + F_x.$$

Then, $F_{(y,x)}\big(\frac{1}{2}\big) = x$, $F'_{(y,x)}\big(\frac{1}{2}\big) = y$, and

$$\|F_{(y,x)}\|_{\mathscr{F}} \le \|\Pi \cdot F_{(y-\Omega(x))}\|_{\mathscr{F}} + \|F_x\|_{\mathscr{F}} \le \|\Pi\|_\infty (\|y - \Omega(x)\|_2 + \|x\|_2),$$

where $\|\Pi\|_\infty \le |a_1| + |a_2| + |a_3|$, and we may define

$$\Omega_{2,2}(y,x) = \big(\hat{F}_{(y,x)}\big[3; \tfrac{1}{2}\big],\; \hat{F}_{(y,x)}\big[2; \tfrac{1}{2}\big]\big).$$

By the construction of Π, we have

$$\begin{aligned} \hat{F}_{(y,x)}\big[2; \tfrac{1}{2}\big] &= \hat{F}_{(y-\Omega(x))}\big[1; \tfrac{1}{2}\big] + \hat{F}_x\big[2; \tfrac{1}{2}\big], \\ \hat{F}_{(y,x)}\big[3; \tfrac{1}{2}\big] &= \hat{F}_{(y-\Omega(x))}\big[2; \tfrac{1}{2}\big] + \hat{F}_x\big[3; \tfrac{1}{2}\big], \end{aligned}$$

and thus

$$\begin{aligned} \Omega_{2,2}(y,x) = 2\bigg((y - \Omega x)\log^2 \frac{|y-\Omega x|}{\|y-\Omega x\|_2} + \frac{2x}{3}\log^3 \frac{|x|}{\|x\|_2}, \\ (y-\Omega x)\log \frac{|y-\Omega x|}{\|y-\Omega x\|_2} + x\log^2 \frac{|x|}{\|x\|_2} \bigg). \end{aligned}$$

At this point we cannot help to mention a few very accurate comments of the referee: "Speaking of computational details [...], we have that $\mathscr{X}^{(4)}$ is equivalent to both a twisted sum of $\mathscr{X}^{(2)}$ with $\mathscr{X}^{(2)}$ and to a twisted sum of $\mathscr{X}^{(3)}$ with $\mathscr{X}^{(1)}$. Presumably that can be seen in the case of ℓ_2 by working directly with the explicit formulas in Section 5.1 and doing some estimation. In fact, given that the twisted direct sum of $\mathscr{X}^{(2)}$ with $\mathscr{X}^{(1)}$ is not automatically equivalent to the twisted sum of $\mathscr{X}^{(1)}$ with $\mathscr{X}^{(2)}$, the fact that these are equivalent would also be visible through the formulas in [that section]. If the authors have done the computations or if they choose to do them, and if the result seems to carry any intuitive insight, they would make a good inclusion." To this plot the referee foresees: "My first thought is that the answer is 'Yes, the computations can be done. Yes, the result is a mess. No, it doesn't seem to give any insights'."

Indeed, the computations were done a long time ago and are a mess, generally speaking. However, the "3D-case" is crystalline since in $\mathscr{Z}^{(3)} = \ell_2 \oplus_{\Omega_{2,1}} \mathscr{Z}^{(2)} = \mathscr{Z}^{(2)} \oplus_{\Omega_{1,2}} \ell_2$ the quasinorms induced by $\Omega_{2,1}$ and $\Omega_{1,2}$ actually agree. To check this first observe that, letting $\Lambda(x) = \log(|x|/\|x\|_2)$, one can rewrite (11) as

$$\|(z,y,x)\|_{\Omega_{1,2}} = \|z - \Omega(x)\Lambda(x) - \Omega(y - \Omega(x))\|_2 + \|y - \Omega(x)\|_2 + \|x\|_2 \tag{12}$$

To compute $\|(z,y,x)\|_{\Omega_{2,1}}$ let us take a look at $\Omega_{2,2}$ and notice that $\Omega_{2,1}$ just "forgets" the first coordinate. Hence,

$$\Omega_{2,1}(y,x) = \Omega(y - \Omega(x)) + \Omega(x)\Lambda(x)$$

and

$$\begin{aligned}\|(z,y,x)\|_{\Omega_{2,1}} &= \|z - \Omega_{2,1}(y,x)\|_2 + \|(y,x)\|_{\Omega}\\ &= \|z - \Omega(y - \Omega(x)) - \Omega(x)\Lambda(x)\|_2 + \|y - \Omega(x)\|_2 + \|x\|_2\\ &= \|(z,y,x)\|_{\Omega_{1,2}}.\end{aligned}$$

5.2. ***The 3-space problem for twisted Hilbert spaces.*** We are now ready for the first concrete application. Recall that a twisted Hilbert space is a twisted sum of Hilbert spaces.

Proposition 12. *The space $\mathscr{Z}^{(n)}$ is a twisted Hilbert space if and only if $n = 1, 2$.*

Proof. The n-th type 2 constant $a_{n,2}(X)$ of a (quasi-)Banach space X is defined as the infimum of those C such that for every $x_1, \dots x_n \in X$ one has

$$\left(\int_0^1 \left\|\sum_{i=1}^n r_i(t)x_i\right\|^2 dt\right)^{\frac{1}{2}} \le C\left(\sum_{i=1}^n \|x_i\|^2\right)^{\frac{1}{2}},$$

where (r_n) is the sequence of Rademacher functions.

To prove that $\mathscr{Z}^{(3)}$ does not embed in any twisted Hilbert space, we will work with the equivalent quasinorm given by (11). Let (e_i) be the unit basis of ℓ_2 and take $x_i = (0, 0, e_i)$. These are normalized vectors, which makes $\left(\sum_{i=1}^n \|x_i\|^2\right)^{1/2} = \sqrt{n}$. On the other hand,

$$\left\|\sum_{i=1}^n \pm x_i\right\|_{\Omega_{1,2}} = \sqrt{n}(1 + \log^2 n).$$

Hence, the type 2 constants of $\mathscr{Z}^{(3)}$ cannot satisfy $a_{n,2} \le K \log n$, and this estimate holds in any twisted Hilbert space by [Kalton and Peck 1979, Theorem 6.2(a)]. □

Corollary 13. *"To be a twisted Hilbert space" is not a* 3*-space property.*

This corollary answers a question posed to us by David Yost a long time ago [Castillo and González 1997, p. 95]; the first author showed [Cabello Sánchez 1999] that "to be a subspace of a twisted Hilbert space" is not a 3-space property. Since $\mathfrak{X}^{(n)}$ is isomorphic to its dual [Rochberg 1996, Section 4] and the dual of any twisted Hilbert space is again a twisted Hilbert space, we see that $\mathfrak{X}^{(n)}$ is a quotient of a twisted Hilbert space if and only if $n = 1, 2$.

Thus, in the situation described in Section 5.1, recall that for $\mathscr{F} = \mathscr{F}(\ell_\infty, \ell_1)$ one gets $\mathfrak{X}^{(1)} = \ell_2$ and $\mathfrak{X}^{(2)}$ is isomorphic to the Kalton–Peck space Z_2, and, actually, the extension $0 \to \ell_2 \to \mathfrak{X}^{(2)} \to \ell_2 \to 0$ is isomorphically — and even "projectively", compare [Kalton and Peck 1979, Definition 2.1(b)] — equivalent to the Kalton–Peck sequence $0 \to \ell_2 \to Z_2 \to \ell_2 \to 0$, which has strictly singular quotient map and strictly cosingular inclusion (see Theorem 6.4 of the same reference). Thus applying the results of Section 4 we obtain the following.

Proposition 14. *The exact sequences* $0 \to \mathfrak{X}^{(k)} \to \mathfrak{X}^{(n+k)} \to \mathfrak{X}^{(n)} \to 0$ *are singular and cosingular for all integers* n, k.

As a direct application we get:

Proposition 15. *If* $k > 1$, *then* $\mathfrak{X}^{(k)}$ *does not contain complemented copies of* ℓ_2.

Proof. By [Kalton and Peck 1979, Corollary 6.7], $\mathfrak{X}^{(2)} = Z_2$ has no complemented subspaces isomorphic to ℓ_2. Now, if one has an exact sequence

$$0 \longrightarrow A \xrightarrow{\ I\ } B \xrightarrow{\ Q\ } C \longrightarrow 0$$

with Q strictly singular and A not containing ℓ_2 complemented, then B does not contain ℓ_2 complemented: assume otherwise that B has a subspace B' which is isomorphic to ℓ_2 and is complemented in B through a projection P. (Without loss of generality we may assume that $A = \ker Q$ and that I is the inclusion map.) Since Q is strictly singular, there exists an infinite dimensional subspace $A' \subset A$ and a nuclear operator $K : A' \to B$ such that $I - K : A' \to B'$ is an embedding. Passing to a further subspace if necessary, we may assume that the nuclear norm of K is strictly less than 1. Let N be a nuclear endomorphism of B extending K and having the same nuclear norm as K. Then, $\|N : B \to B\| < 1$ and $\mathbf{1}_B - N$ is invertible, with $(\mathbf{1}_B - N)^{-1} = \sum_{k \ge 0} N^k$ with convergence in the operator norm. Now, it is easily seen that

$$(\mathbf{1}_B - N) \circ P \circ (\mathbf{1}_B - N)^{-1}$$

is a projection of B (hence of A) onto A'. □

The proof also works replacing ℓ_2 by any other "complementably minimal" space (those Banach spaces, all of whose infinite dimensional closed subspaces contain subspaces isomorphic to and complemented in the whole space) such as ℓ_p for $1 < p < \infty$. This implies that Proposition 15 extends almost verbatim to $1 < p < \infty$.

5.3. ***A twisted sum of Z_2 containing ℓ_2 complemented.*** It is quite surprising that there exists a twisted sum of Z_2 containing complemented copies of ℓ_2. But they do exist:

Proposition 16. *There is a* (*nontrivial*) *exact sequence*

$$0 \longrightarrow Z_2 \longrightarrow \ell_2 \oplus \mathscr{X}^{(3)} \longrightarrow Z_2 \longrightarrow 0.$$

Proof. Recall from [Castillo and Moreno 2004, p. 257] the construction of the so-called diagonal pushout sequence: in a commutative diagram

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "\imath"] \arrow[d, "u"'] & B \arrow[r] \arrow[d, "v"'] & E \arrow[r] \arrow[d, equal] & 0 \\ 0 \arrow[r] & C \arrow[r, "\jmath"] & D \arrow[r] & E \arrow[r] & 0 \end{tikzcd}$$

the following sequence is exact

$$0 \longrightarrow A \xrightarrow{\imath \times u} B \oplus C \xrightarrow{v \oplus -\jmath} D \longrightarrow 0,$$

where $(\imath \times u)(a) = (\imath(a), u(a))$ and $(v \oplus -\jmath)(b, c) = v(b) - \jmath(c)$. Thus, taking $n = i = 1$ and $k = j = 2$ in (6) for $\mathscr{F} = \mathscr{F}(\ell_\infty, \ell_1)$, one gets a commutative diagram

$$\begin{tikzcd} 0 \arrow[r] & \mathscr{Z}^{(2)} \arrow[r] \arrow[d] & \mathscr{Z}^{(3)} \arrow[r] \arrow[d] & \ell_2 \arrow[r] \arrow[d, equal] & 0 \\ 0 \arrow[r] & \ell_2 \arrow[r] & \mathscr{Z}^{(2)} \arrow[r] & \ell_2 \arrow[r] & 0 \end{tikzcd}$$

from which, recalling that $\mathscr{X}^{(2)} = Z_2$, one obtains an exact sequence

$$0 \longrightarrow Z_2 \longrightarrow \ell_2 \oplus \mathscr{Z}^{(3)} \longrightarrow Z_2 \longrightarrow 0,$$

which is not trivial since, otherwise $Z_2 \simeq Z_2 \oplus Z_2 = \ell_2 \oplus \mathscr{Z}^{(3)}$; this is impossible since Z_2 does not contain ℓ_2 complemented. □

Therefore $\ell_2 \oplus \mathscr{Z}^{(3)}$ is a twisted sum of Z_2, which contains complemented Hilbert subspaces. We cannot resist remarking that while nobody knows whether Z_2 is isomorphic to its hyperplanes, it is obvious that $\ell_2 \oplus \mathscr{Z}^{(3)}$ is isomorphic to its own hyperplanes.

6. Open ends

6.1. ***On the splitting of the first extension.*** Very little is known about the splitting of the "first" exact sequence $0 \to \mathscr{Z}^{(1)} \to \mathscr{Z}^{(2)} \to \mathscr{Z}^{(1)} \to 0$ outside of the case in which it is induced by a couple of Banach lattices. On the other hand, Corollary 6 shows that once the first exact sequence obtained in an interpolation schema is nontrivial, the same happens to all the rest. Is the reciprocal true? That is, suppose that $\mathfrak{X}^{(2)}$ is a trivial self-extension of $\mathfrak{X}^{(1)}$. Does it follow that the extensions $0 \to \mathfrak{X}^{(k)} \to \mathfrak{X}^{(n+k)} \to \mathfrak{X}^{(n)} \to 0$ are trivial for all values of n and k?

6.2. ***Other twisted Hilbert spaces.*** Suppose we have a Banach space X_0 with a normalized basis that we use to consider X_0 inside ℓ_∞. Take $X_1 = \overline{X_0'}$ the complex conjugate of the closure X_0' of the subspace spanned by the coordinate functionals in X_0^*. Then (X_0, X_1) is a Banach couple, $[X_0, X_1]_{1/2}$ is a Hilbert space [Pisier and Xu 2003, p. 1471], and thus $\mathfrak{X}^{(2)}$ is a twisted Hilbert space. We believe that $\mathfrak{X}^{(2)}$ is a Hilbert space if and only if $X_0 = \ell_2$.

6.3. ***Other interpolation methods.*** It is the feeling of the authors that most of the work done in this paper could be reproduced for real interpolation by either the K or J methods with a careful analysis of the work done in [Carro et al. 1995]. It would be interesting to know to what extent the same occurs for other interpolation methods.

6.4. ***About the vanishing of* Ext²*.*** A problem on the horizon, for us, is whether the second derived functor Ext^2 vanishes on Hilbert spaces, which can be understood as a twisted reading of a question of Palamodov for Fréchet spaces [1971, Section 12, Problem 6].

Given Banach spaces A and D, one considers the set of all possible four-term exact sequences

$$0 \longrightarrow A \xrightarrow{I} B \xrightarrow{U} C \xrightarrow{Q} D \longrightarrow 0. \tag{13}$$

Under a certain equivalence relation, which is not necessary to define here, the set of such four-term exact sequences becomes a linear space denoted by $\mathrm{Ext}^2(D, A)$, whose zero is (the class of all exact sequences equivalent to)

$$0 \longrightarrow A = A \xrightarrow{0} D = D \longrightarrow 0.$$

It is important to realize that if we are given a short exact sequence of the form

$$0 \longrightarrow A \xrightarrow{I} B \xrightarrow{P} E \longrightarrow 0 \tag{14}$$

and another sequence of the form

$$0 \longrightarrow E \xrightarrow{J} C \xrightarrow{Q} D \longrightarrow 0 \tag{15}$$

then we may form a four-term sequence

$$0 \longrightarrow A \xrightarrow{I} B \xrightarrow{U} C \xrightarrow{Q} D \longrightarrow 0$$

just taking $U = J \circ P$. This resulting "long" sequence will be zero in $\mathrm{Ext}^2(D, A)$ if and only if (14) and (15) fit inside a commutative diagram

$$\begin{tikzcd}
 & & 0 & 0 & \\
 & & D \arrow[u] \arrow[r, equal] & D \arrow[u] & \\
0 \arrow[r] & A \arrow[r] & F \arrow[u] \arrow[r] & C \arrow[u] \arrow[r] & 0 \\
0 \arrow[r] & A \arrow[u, equal] \arrow[r] & B \arrow[u] \arrow[r] & E \arrow[u] \arrow[r] & 0 \\
 & & 0 \arrow[u] & 0 \arrow[u] &
\end{tikzcd} \tag{16}$$

whose rows and columns are exact.

The skeptical reader will wonder how this is related to the main subject of the paper. Let μ be a σ-finite measure on a measure space S and let L_0 be the space of all (complex) measurable functions on S, where we identify two functions if they agree almost everywhere. If X is a Köthe space on μ, then a centralizer on X is a homogeneous mapping $\Omega : X \to L_0$ having the following property: there is a constant $C = C(\Omega)$ such that, for every $f \in X$ and every $a \in L_\infty$, the difference $\Omega(af) - a\Omega(f)$ belongs to X and

$$\|\Omega(af) - a\Omega(f)\|_X \le C\,\|a\|_\infty\,\|f\|_X.$$

Every centralizer is quasilinear, so it induces a twisted sum

$$X \oplus_\Omega X = \{(y, x) : x,\ y - \Omega(x) \in X\}$$

which is quasinormed by the functional $\|(y, x)\|_\Omega = \|y - \Omega(x)\|_X + \|x\|_X$. A widely ignored result by Kalton states that if X is superreflexive then one can construct an admissible space of analytic functions $\mathscr{F}$ on a disc centered at the origin such that:

- $X = X_0^{(1)}$ (evaluation at 0) up to equivalent norm;
- $\Omega \approx \Omega_{1,1}$, where $\Omega_{1,1}$ is the corresponding "derivation" (see Section 3.5).

This means that for every $x \in X$ the difference $\Omega(x) - \Omega_{1,1}(x)$ falls in X, and one has the estimate $\|\Omega(x) - \Omega_{1,1}(x)\|_X \le K\|x\|_X$ for some constant K and every $x \in X$. Actually one can construct $\mathscr{F}$ by using no more than three Köthe spaces on the boundary of the disc [Kalton 1992, Theorem 7.9]; if Ω is "real" in the sense that it takes real functions into real functions, then two Köthe spaces on a strip suffice [op. cit., Theorem 7.6]. In particular since $X \oplus_\Omega X = X \oplus_{\Omega_{1,1}} X = X_0^{(2)}$, up to equivalent (quasi-)norms, we see that the self-extension induced by Ω fits into the commutative diagram (the operators $\imath_{n,k}$ are those appearing in Proposition 2)

$$\begin{tikzcd}
0 \arrow[r] & X \arrow[r] & X_0^{(3)} \arrow[r] & X \oplus_\Omega X \arrow[r] & 0 \\
0 \arrow[r] & X \arrow[r] \arrow[u, equal] & X \oplus_\Omega X \arrow[r] \arrow[u, "\imath_{2,3}"] & X \arrow[r] \arrow[u, "\imath_{1,2}"] & 0
\end{tikzcd}$$

which, when completed, has the same form as (16), witnessing that the juxtaposition of two copies of the extension induced by Ω, namely

$$0 \longrightarrow X \longrightarrow X \oplus_\Omega X \longrightarrow X \oplus_\Omega X \longrightarrow X \longrightarrow 0,$$

is zero in $\mathrm{Ext}^2(X, X)$. We do not know what happens with two different centralizers; more specifically, we ask the following: Let Ω and Φ be centralizers on a super-reflexive Köthe space X and consider the twisted sums $X \oplus_\Omega X$ and $X \oplus_\Phi X$. If, as before, we set $I(x) = (x, 0)$, $U(x, y) = (y, 0)$, and $Q(x, y) = y$, can the exact sequence

$$0 \longrightarrow X \xrightarrow{I} X \oplus_\Omega X \xrightarrow{U} X \oplus_\Phi X \xrightarrow{Q} X \longrightarrow 0$$

be nonzero in $\mathrm{Ext}^2(X, X)$?

References

[Cabello Sánchez 1999] F. Cabello Sánchez, "Twisted Hilbert spaces", *Bull. Austral. Math. Soc.* **59**:2 (1999), 177–180. MR 99m:46030 Zbl 0958.46003

[Cabello Sánchez and Castillo 2004] F. Cabello Sánchez and J. M. F. Castillo, "Uniform boundedness and twisted sums of Banach spaces", *Houston J. Math.* **30**:2 (2004), 523–536. MR 2005f:46032 Zbl 1063.46012

[Cabello Sánchez et al. 2012] F. Cabello Sánchez, J. M. F. Castillo, and J. Suárez, "On strictly singular nonlinear centralizers", *Nonlinear Anal.* **75**:7 (2012), 3313–3321. MR 2891170 Zbl 1248.46013

[Carro et al. 1995] M. J. Carro, J. Cerdà, and J. Soria, "Higher order commutators in interpolation theory", *Math. Scand.* **77**:2 (1995), 301–319. MR 97d:46094 Zbl 0852.46058

[Castillo and González 1997] J. M. F. Castillo and M. González, *Three-space problems in Banach space theory*, Lecture Notes in Mathematics **1667**, Springer, Berlin, 1997. MR 99a:46034 Zbl 0914.46015

[Castillo and Moreno 2004] J. M. F. Castillo and Y. Moreno, "On the Lindenstrauss–Rosenthal theorem", *Israel J. Math.* **140** (2004), 253–270. MR 2005d:46044 Zbl 1063.46003

[Castillo and Moreno Salguero 2007] J. M. F. Castillo and Y. Moreno Salguero, "Singular and cosingular exact sequences of quasi–Banach spaces", *Arch. Math.* (*Basel*) **88**:2 (2007), 123–132. MR 2008b:46004 Zbl 1116.46012

[Castillo et al. 2012] J. M. F. Castillo, M. Simoes, and J. Suárez de la Fuente, "On a question of Pełczyński about strictly singular operators", *Bull. Pol. Acad. Sci. Math.* **60**:1 (2012), 27–36. MR 2901385 Zbl 1244.46002

[Cwikel et al. 1989] M. Cwikel, B. Jawerth, M. Milman, and R. Rochberg, "Differential estimates and commutators in interpolation theory", pp. 170–220 in *Analysis at Urbana, II* (Urbana, IL, 1986–1987), edited by E. Berkson and T. Peck, London Math. Soc. Lecture Note Ser. **138**, Cambridge Univ. Press, 1989. MR 90k:46157 Zbl 0696.46050

[Hilton and Stammbach 1971] P. J. Hilton and U. Stammbach, *A course in homological algebra*, Graduate Texts in Mathematics **4**, Springer, New York-Berlin, 1971. MR 49 #10751 Zbl 0238.18006

[Kalton 1978] N. J. Kalton, "The three space problem for locally bounded F-spaces", *Compositio Math.* **37**:3 (1978), 243–276. MR 80j:46005 Zbl 0395.46003

[Kalton 1992] N. J. Kalton, "Differentials of complex interpolation processes for Köthe function spaces", *Trans. Amer. Math. Soc.* **333**:2 (1992), 479–529. MR 92m:46111 Zbl 0776.46033

[Kalton and Montgomery-Smith 2003] N. Kalton and S. Montgomery-Smith, "Interpolation of Banach spaces", pp. 1131–1175 in *Handbook of the geometry of Banach spaces, II*, edited by W. B. Johnson and J. Lindenstrauss, North-Holland, Amsterdam, 2003. MR 2006c:46061 Zbl 1041.46012

[Kalton and Peck 1979] N. J. Kalton and N. T. Peck, "Twisted sums of sequence spaces and the three space problem", *Trans. Amer. Math. Soc.* **255** (1979), 1–30. MR 82g:46021 Zbl 0424.46004

[Kato 1958] T. Kato, "Perturbation theory for nullity, deficiency and other quantities of linear operators", *J. Analyse Math.* **6** (1958), 261–322. MR 21 #6541 Zbl 0090.09003

[Palamodov 1971] V. P. Palamodov, "Homological methods in the theory of locally convex spaces", *Uspehi Mat. Nauk* **26**:1(157) (1971), 3–65. In Russian; translated in *Russ. Math. Surv.* **26(1971)**:1 (1974), 1–64. MR 45 #2442 Zbl 0247.46070

[Pełczyński 1965] A. Pełczyński, "On strictly singular and strictly cosingular operators, I: Strictly singular and strictly cosingular operators in $C(S)$-spaces", *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* **13** (1965), 31–36. MR 31 #1563 Zbl 0138.38604

[Pisier and Xu 2003] G. Pisier and Q. Xu, "Non-commutative L^p-spaces", pp. 1459–1517 in *Handbook of the geometry of Banach spaces, II*, edited by W. B. Johnson and J. Lindenstrauss, North-Holland, Amsterdam, 2003. MR 2004i:46095 Zbl 1046.46048

[Rochberg 1996] R. Rochberg, "Higher order estimates in complex interpolation theory", *Pacific J. Math.* **174**:1 (1996), 247–267. MR 97i:46134 Zbl 0866.46047

[Rochberg 2007] R. Rochberg, "Uses of commutator theorems in analysis", pp. 277–295 in *Interpolation theory and applications*, edited by L. De Carli and M. Milman, Contemp. Math. **445**, Amer. Math. Soc., Providence, RI, 2007. MR 2009b:46053 Zbl 1143.46306

[Rochberg and Weiss 1983] R. Rochberg and G. Weiss, "Derivatives of analytic families of Banach spaces", *Ann. of Math.* (2) **118**:2 (1983), 315–347. MR 86a:46099 Zbl 0539.46049

Received October 1, 2013.

FÉLIX CABELLO SÁNCHEZ
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD DE EXTREMADURA
AVENIDA DE ELVAS
06011 BADAJOZ
SPAIN

fcabello@unex.es

JESÚS M. F. CASTILLO
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD DE EXTREMADURA
AVENIDA DE ELVAS
06011 BADAJOZ
SPAIN
castillo@unex.es

NIGEL J. KALTON
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MISSOURI
COLUMBIA, MO 65211
UNITED STATES
nigel@math.missouri.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.309

THE RAMIFICATION GROUP FILTRATIONS OF CERTAIN FUNCTION FIELD EXTENSIONS

JEFFREY A. CASTAÑEDA AND QINGQUAN WU

We investigate the ramification group filtration of a Galois extension of function fields, if the Galois group satisfies a certain intersection property. For finite groups, this property is implied by having only elementary abelian Sylow p-subgroups. Note that such groups could be nonabelian. We show how the problem can be reduced to the totally wild ramified case on a p-extension. Our methodology is based on an intimate relationship between the ramification groups of the field extension and those of all degree-p subextensions. Not only do we confirm that the Hasse–Arf property holds in this setting, but we also prove that the Hasse–Arf divisibility result is the best possible by explicit calculations of the quotients, which are expressed in terms of the different exponents of all those degree-p subextensions.

1. Introduction

When investigating algebraic number fields and function fields, Hilbert ramification theory is a convenient tool, especially in the study of wild ramifications. Fix a function field K over a perfect constant field k with a place $\mathcal{P}$, and let L be a Galois extension of K with a place $\mathfrak{P}$ lying over $\mathcal{P}$. We investigate how the ramification group filtration of $\mathfrak{P}|\mathcal{P}$ is related to the ramification group filtration of $\mathfrak{P}_{\mathfrak{m}}|\mathcal{P}$, where $\mathfrak{P}_{\mathfrak{m}}$ is a place of some intermediate field $K \subseteq M \subseteq L$, so that $\mathfrak{P}$ lies over $\mathfrak{P}_{\mathfrak{m}}$ and $[M : K] = p$ for some prime number p.

We first analyze how and why we can simplify the problem to the setting when $\mathfrak{P}|\mathcal{P}$ is totally wildly ramified, i.e., $[L : K] = p^n$, where $p > 0$ is the characteristic of k, n is some positive integer, and the ramification index $e(\mathfrak{P}|\mathcal{P}) = p^n$.

Next we study how the ramification group filtration of $\mathfrak{P}|\mathcal{P}$ is closely related to the ramification group filtration of $\mathfrak{P}_{\mathfrak{m}}|\mathcal{P}$ for all those intermediate fields M such that $[M : K] = p$, for various degrees p. This relation is close if an intersection property (2-5) is assumed about $\mathrm{Gal}(L/K)$, which is satisfied by many abelian and nonabelian groups.

The second author is supported by the University Research Grant of the Texas A&M International University.
MSC2010: primary 11R58; secondary 11R32, 11S15.
Keywords: function fields, ramification groups, filtrations.

To prove such a relationship, we first prove a preliminary result, which states that the number of jumps in the ramification group filtration is equal to the number of pairwise distinct different exponents of the corresponding place extensions over all possible degree-p intermediate extensions M/K. This equality is significant since the degree-p intermediate extensions are considerably easier to investigate than the whole extension L/K. Nonetheless, we will show that these different exponents are closely related to those quotients given in the Hasse–Arf property, which we will show to be true.

We also study the relationship by applying the equality given by the transitivity of differents, where the different exponents are computed via Hilbert's different formulae applied to various field extension settings. These equalities lead to linear equations on the indices where the jumps of the ramification group filtration on L/K occur. With the intersection property assumption, we show that the number of such linear equations is equal to the number of such indices as the variables of these equations. Hence we can expect a unique solution. In fact, we can solve these linear equations explicitly to give closed-form formulae for the indices since the coefficient matrix of the linear equations is triangular.

The academic literature on ramification group filtration is extensive. A good introduction is [Serre 1979], where Herbrand's upper numbering is introduced. See [Fesenko and Vostokov 2002] for an introduction without the use of cohomologies. The ramification groups are studied in [Sen and Tate 1963] using class field theory. For an approach using Herbrand functions and without using class field theory, see [Wyman 1969]. Maus [1968] showed certain properties of a group filtration that are sufficient to guarantee it to be the ramification group filtration of a certain extension of complete discrete valuation fields. In [Maus 1972], the asymptotic behavior of quotients given by the Hasse–Arf property is studied. The paper [Maus 1971] is a collection of many results from Maus's Ph.D. thesis, without proofs.

The ramification group filtration is known to satisfy the Hasse–Arf property [Hasse 1930; 1934; Arf 1939] if the Galois group is abelian. However, the property may fail if the Galois group is not abelian. One such example is the Galois closure of a cyclotomic field over the rationals [Viviani 2004]. In [Doud 2003], it is shown that the ramification group filtration of a wildly ramified prime $\mathfrak{p}$ is uniquely determined by the $\mathfrak{p}$-adic valuation of the discriminant of the field extension L/K, when both the field extension degree and the residue characteristic of $\mathfrak{p}$ are equal to a prime number. When the Galois group is elementary abelian, the Galois module structure of certain ideals is related to the ramification group filtration, see [Byott and Elder 2002; 2005; 2009]. Such a relation is investigated when the Galois group is quaternion [Elder and Hooper 2007], and hence nonabelian.

For the function field extension setting, the wildly ramified case was studied in Artin–Schreier–Witt extensions, see [Thomas 2005]. The elementary abelian extension of Galois group $\mathbb{Z}_p \times \mathbb{Z}_p$ is investigated in [Anbar et al. 2009] and [Wu and Scheidler 2010]. It should be mentioned that the idea of utilizing transitivity of differents and Hilbert's different formula to investigate the ramification groups is used in [Garcia and Stichtenoth 2008], where the Hasse–Arf property for elementary abelian extensions of function fields is proved. In Roberts' review [2009] of the latter paper, it is shown that the proof can be very short if "some upper numbering system and its basic formalism" is applied. We take an approach similar to Garcia and Stichtenoth's, but we further explore the arithmetic and linear algebra provided by the application of transitivity of differents and Hilbert's different formula. Our objective in this paper is to generalize these results to function field extensions with Galois groups satisfying a certain intersection property which is true for elementary abelian groups.

2. Notation

A good introduction for the notation can be found in [Rosen 2002] or [Stichtenoth 2009]. Throughout this paper, we use the following notation:

- k is a perfect field of characteristic $p > 0$;
- K is a function field with constant field k;
- $\mathcal{P}$ is a place of K;
- $v_{\mathcal{P}} : K \to \mathbb{Z} \cup \{\infty\}$ is the (surjective) discrete valuation corresponding to $\mathcal{P}$;
- $\mathcal{O}_{\mathcal{P}} = \{\alpha \in K \mid v_{\mathcal{P}}(\alpha) \geq 0\}$ is the valuation ring corresponding to $\mathcal{P}$.

For any extension L of K and any place $\mathfrak{P}$ of L lying above $\mathcal{P}$, we write $\mathfrak{P}|\mathcal{P}$. Let $e(\mathfrak{P}|\mathcal{P})$ and $d(\mathfrak{P}|\mathcal{P})$ be the ramification index and different exponent of $\mathfrak{P}|\mathcal{P}$, respectively. If L/K is a Galois extension, the ramification groups of $\mathfrak{P}|\mathcal{P}$ are given by

$$G_i = G_i(\mathfrak{P}|\mathcal{P}) = \{\sigma \in \mathrm{Gal}(L/K) \mid v_{\mathfrak{P}}(t^\sigma - t) \geq i + 1 \text{ for all } t \in \mathcal{O}_{\mathfrak{P}}\} \tag{2-1}$$

for $i \geq 0$. The connection between these groups and the different exponent is shown in Hilbert's different formula (see for example Theorem 3.8.7, p. 136, of [Stichtenoth 2009]):

$$d(\mathfrak{P}|\mathcal{P}) = \sum_{i=0}^{\infty} (\#G_i(\mathfrak{P}|\mathcal{P}) - 1). \tag{2-2}$$

We also recall the transitivity of the ramification index and the different exponent. If $K \subseteq F \subseteq L$ are function fields, $\mathfrak{P}$ a place of L, $\mathfrak{P}_{\mathfrak{F}} = \mathfrak{P} \cap F$, and $\mathcal{P} = \mathfrak{P}_{\mathfrak{F}} \cap K$, then

$$e(\mathfrak{P}|\mathcal{P}) = e(\mathfrak{P}|\mathfrak{P}_{\mathfrak{F}})e(\mathfrak{P}_{\mathfrak{F}}|\mathcal{P}), \tag{2-3}$$

and we have transitivity of differents:

$$d(\mathfrak{P}|\mathcal{P}) = e(\mathfrak{P}|\mathfrak{P}_{\mathfrak{F}})d(\mathfrak{P}_{\mathfrak{F}}|\mathcal{P}) + d(\mathfrak{P}|\mathfrak{P}_{\mathfrak{F}}). \tag{2-4}$$

Henceforth, we assume that all nontrivial Sylow p-subgroups H_p of the Galois group $\mathrm{Gal}(L/K)$ satisfy the following intersection property.

Assume that $\#H_p = p^n > 1$. Then, for all proper subgroups $F \subsetneq H_p$, the intersection of all order p^{n-1} subgroups of H_p containing F is simply F. That is to say,

$$\bigcap_{\substack{H \supseteq F \\ \#H = p^{n-1}}} H = F. \tag{2-5}$$

It is easy to verify that all elementary abelian p-groups of order p^n satisfy this intersection property.

3. Reduction to the totally wildly ramified case

Let L be a Galois extension field of K, $\mathfrak{P}$ a place of L, and $\mathcal{P} = \mathfrak{P} \cap K$. Our goal in this section is to reduce the ramification group $G_i(\mathfrak{P}|\mathcal{P})$ calculation to the case that $\mathfrak{P}/\mathcal{P}$ is totally wildly ramified and the Galois group Gal (L/K) is a p-group for a certain prime number p.

Lemma 3.1. *Let L/K be a Galois extension of a function field, $\mathfrak{P}$ a place of L, $\mathcal{P} = \mathfrak{P} \cap K$, and $\mathfrak{P}_{\mathrm{m}} = \mathfrak{P} \cap M$, where M is the inertia field of $\mathcal{P}$ in L/K. Then, $G_i(\mathfrak{P}|\mathcal{P}) = G_i(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}})$ for every $i \geq 0$.*

Proof. Applying (2-4) to the field extension tower $L/M/K$, we have

$$d(\mathfrak{P}|\mathcal{P}) = e(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}})d(\mathfrak{P}_{\mathrm{m}}|\mathcal{P}) + d(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}}). \tag{3-1}$$

However, $d(\mathfrak{P}_{\mathrm{m}}|\mathcal{P}) = 0$ since $\mathcal{P}$ is unramified in M/K, so $d(\mathfrak{P}|\mathcal{P}) = d(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}})$. Now (2-2) yields

$$d(\mathfrak{P}|\mathcal{P}) = \sum_{i=0}^{\infty}(\#G_i(\mathfrak{P}|\mathcal{P}) - 1) \tag{3-2}$$

and

$$d(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}}) = \sum_{i=0}^{\infty}(\#G_i(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}}) - 1). \tag{3-3}$$

By definition (2-1), it is easy to check that $G_i(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}})$ is the intersection of $G_i(\mathfrak{P}|\mathcal{P})$ and the Galois group of L/M. In particular, we have $\#G_i(\mathfrak{P}|\mathcal{P}) \geq \#G_i(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}})$ for all i. Note that $d(\mathfrak{P}|\mathcal{P}) = d(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}})$, which implies that $\#G_i(\mathfrak{P}|\mathcal{P}) = \#G_i(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}})$ for all $i \geq 0$ by (3-2) and (3-3). Thus, $G_i(\mathfrak{P}|\mathcal{P}) = G_i(\mathfrak{P}|\mathfrak{P}_{\mathrm{m}})$ for all $i \geq 0$. □

Next, we want to reduce to the totally wildly ramified case, i.e., $[L : K] = e(\mathfrak{P}|\mathcal{P}) = p^m$, where p is the characteristic of K.

Proposition 3.2. *Let $p > 0$ be the characteristic of the Galois extension of function field L/K, $\mathfrak{P}$ a place of L, $\mathcal{P} = \mathfrak{P} \cap K$, N the inertia field of $\mathcal{P}$ in L/K, $\mathfrak{P}_{\mathfrak{n}} = \mathfrak{P} \cap N$, M the intermediate field of L/N corresponding to a Sylow p-subgroup of* $\mathrm{Gal}(L/N)$ *under Galois correspondence, and $\mathfrak{P}_{\mathfrak{m}} = \mathfrak{P} \cap M$. Then, $\mathfrak{P}_{\mathfrak{m}}$ is totally wildly ramified in the p-extension L/M, and $G_i(\mathfrak{P}|\mathcal{P}) = G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}})$ for every $i \geq 1$.*

Proof. By Lemma 3.1, we have $G_i(\mathfrak{P}|\mathcal{P}) = G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{n}})$ for every $i \geq 0$. It suffices to show that $G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{n}}) = G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}})$ for every $i \geq 1$. Assume that $[L : N] = p^m q$ and $\gcd(p, q) = 1$. We have $d(\mathfrak{P}_{\mathfrak{m}}|\mathfrak{P}_{\mathfrak{n}}) = q - 1$ since $\mathfrak{P}_{\mathfrak{m}}/\mathfrak{P}_{\mathfrak{n}}$ is totally tamely ramified, and we also have $e(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}}) = p^m$. By (3-1), we have

$$d(\mathfrak{P}|\mathfrak{P}_{\mathfrak{n}}) = p^m(q-1) + d(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}}). \tag{3-4}$$

Clearly $\#G_0(\mathfrak{P}|\mathfrak{P}_{\mathfrak{n}}) = p^m q$ and $\#G_0(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}}) = p^m$. Hence,

$$d(\mathfrak{P}|\mathfrak{P}_{\mathfrak{n}}) = p^m q - 1 + \sum_{i=1}^{\infty}(\#G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{n}}) - 1)$$

and

$$d(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}}) = p^m - 1 + \sum_{i=1}^{\infty}(\#G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}}) - 1)$$

by (3-2) and (3-3). Substituting these two equalities into (3-4), we have

$$\sum_{i=1}^{\infty}(\#G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{n}}) - 1) = \sum_{i=1}^{\infty}(\#G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}}) - 1),$$

which implies $G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{n}}) = G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}})$ for $i \geq 1$ since $G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{n}}) \geq G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}})$ for all i. □

4. Main results

Henceforth, let L/K be a Galois extension whose Galois group is a p-group, where p is the characteristic of K. Set t to be the number of distinct $d(\mathfrak{P}_{\mathfrak{m}}|\mathcal{P})$, where M runs through all degree-p intermediate fields M of L/K, and $\mathfrak{P}_{\mathfrak{m}} = \mathfrak{P} \cap M$. We assume that the ramification groups of $\mathfrak{P}|\mathcal{P}$ are

$$\begin{aligned} G_0 = \cdots = G_{m_0} &\supsetneq G_{m_0+1} = \cdots = G_{m_1} \\ &\supsetneq G_{m_1+1} = \cdots = G_{m_{l-1}} \supsetneq G_{m_{l-1}+1} = \{\mathrm{Id}\}. \end{aligned} \tag{4-1}$$

We let $\#G_{m_i} = p^{n_i}$ for $0 \leq i \leq l-1$. Then $p^{n_0} = p^n = [L : K]$. In order to investigate G_i, we need to know l, n_i, and m_i.

First, we claim that the number of jumps in the ramification groups G_i is the number of distinct different exponents $d(\mathfrak{P}_\mathfrak{m}|\mathcal{P})$; this is, $l = t$. Note that a jump means an index where a group in the ramification filtration contains the next one properly. Before we prove the claim, we need a lemma.

Lemma 4.1. *Let G be a p-group of order $p^n > 1$, $H < G$ a subgroup of order p^{n-1}, and $H' < G$ a subgroup such that $H \not\supseteq H'$. Then $\#(H \cap H') = \#H'/p$.*

Proof. By Theorem 4.7, page 39 of [Hungerford 1974], we have $\#(HH') = \#H\,\#H'/\#(H \cap H')$. In particular, $\#(HH')$ is a power of p. Since $H \not\supseteq H'$, we know that $p^{n-1} < \#(HH') \le p^n$. Hence, $\#(HH') = p^n$. The result follows by substituting $\#(HH') = p^n$ and $\#H = p^{n-1}$ into the equality $\#(HH') = \#H\,\#H'/\#(H \cap H')$. □

Furthermore, by (2-4) we know that $d(\mathfrak{P}|\mathcal{P}) = d(\mathfrak{P}_\mathfrak{m}|\mathcal{P})p^{n-1} + d(\mathfrak{P}|\mathfrak{P}_\mathfrak{m})$, where t distinct $d(\mathfrak{P}_\mathfrak{m}|\mathcal{P})$ values imply t distinct $d(\mathfrak{P}|\mathfrak{P}_\mathfrak{m})$. Let $d(\mathfrak{P}|\mathfrak{P}_\mathfrak{m}) = \sum_{i=0}^{\infty}(\#G_i(\mathfrak{P}|\mathfrak{P}_\mathfrak{m}) - 1)$, where $G_i(\mathfrak{P}|\mathfrak{P}_\mathfrak{m}) = H_m \cap G_i(\mathfrak{P}|\mathcal{P})$ such that $\#H_m = p^{n-1}$.

Now, for every jump $G_{m_i} \supsetneq G_{m_i+1}$ for $0 \le i \le l-1$, we want to know if there exists $H_m < G$ such that $\#H_m = p^{n-1}$ and $H_m \supseteq G_{m_i+1}$, but $H_m \not\supseteq G_{m_i}$. In other words, we want to know if there exists an order-p^{n-1} subgroup H_m which faithfully reveals a jump wherever it occurs in the ramification group filtrations. This is not a trivial question since a jump can be hidden if no such H_m can be found. The question is clarified by the following lemma.

Lemma 4.2. *Let G be a p-group satisfying property* (2-5). *Then for any two subgroups $F_1 \supsetneq F_2$ of G, there exists a subgroup H of G such that $\#H = p^{n-1}$ and $H \supseteq F_2$, but $H \not\supseteq F_1$.*

Proof. By way of contradiction, assume that the result is false. Then we can find subgroups $F_1 \supsetneq F_2$, and for all subgroups H of order p^{n-1}, $H \supseteq F_2$ implies $H \supseteq F_1$. Thus, the set $\{H < G \mid \#H = p^{n-1}, H \supseteq F_1\} = \{H < G \mid \#H = p^{n-1}, H \supseteq F_2\}$. Hence,

$$F_1 = \bigcap_{\substack{H \supseteq F_1 \\ \#H = p^{n-1}}} H = \bigcap_{\substack{H \supseteq F_2 \\ \#H = p^{n-1}}} H = F_2,$$

a contradiction by (2-5). □

Now we are ready to prove the result $l = t$.

Proposition 4.3. *Let L/K be a Galois extension of function fields whose Galois group is a p-group satisfying* (2-5), *where p is the characteristic of K. Let $\mathfrak{P} \mid \mathfrak{P}_\mathfrak{m} \mid \mathcal{P}$ be a tower of places, $\mathcal{P} \subseteq K$, $\mathfrak{P}_\mathfrak{m} \subseteq M$, $\mathfrak{P} \subseteq L$, where M is an intermediate field of degree p over K. Then the number of jumps in the ramification group filtration $G_i(\mathfrak{P}|\mathcal{P})$ is the number of distinct different exponents $d(\mathfrak{P}_\mathfrak{m}|\mathcal{P})$, where M runs through all intermediate fields of L/K of degree p over K.*

Proof. Fix $\mathfrak{P}$ and $\mathcal{P}$, and let M run through all possible degree-p intermediate fields of L/K. By (3-1), t is equal to the number of distinct values of $d(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}})$ since $d(\mathfrak{P}|\mathcal{P})$ and $e(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}}) = p^{n-1}$ are independent of the choice of M.

For $0 \le i \le l-1$, the i-th jump occurs at index m_i; that is, $G_{m_i} \supsetneq G_{m_i+1}$. By Lemma 4.2, there exists an order p^{n-1} subgroup H_i of G_0, so that $H_i \supseteq G_{m_i+1}$, but $H_i \not\supseteq G_{m_i}$. By Lemma 4.1, we have $\#(H_i \cap G_{m_i}) = \#G_{m_i}/p$. Similarly, $\#(H_i \cap G_j) = \#G_j/p$ for all $j \le m_i$ since G_j is decreasing. Hence, $\#(H_i \cap G_j) = \#G_j/p$ for $0 \le j \le m_i$, and $\#(H_i \cap G_j) = \#G_j$ for $j > m_i$.

Let M_i be the intermediate field of L/K corresponding to H_i under Galois correspondence, and set $\mathfrak{P}_i = \mathfrak{P} \cap M_i$. Since $G_j(\mathfrak{P}|\mathfrak{P}_{\mathrm{i}}) = H_i \cap G_j(\mathfrak{P}|\mathcal{P}) = H_i \cap G_j$ for all j, we have

$$\text{(4-2)} \qquad d(\mathfrak{P}|\mathfrak{P}_i) = \sum_{j=0}^{\infty}(\#G_j(\mathfrak{P}|\mathfrak{P}_i) - 1) = \sum_{j=0}^{\infty}(\#(H_i \cap G_j) - 1)$$
$$= \sum_{j=0}^{m_i}\Big(\frac{\#G_j}{p} - 1\Big) + \sum_{j=m_i+1}^{\infty}(\#G_j - 1).$$

The right-hand side of (4-2) is strictly decreasing with i. Hence, we find l pairwise distinct values of $d(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}})$. Note that any order p^{n-1} subgroup H of G_0 contains $G_{m_l+1} = \{\mathrm{Id}\}$ but not $G_{m_0} = G_0$. Hence, for any such H, there exists an i such that $H \supseteq G_{m_i+1}$, but $H \not\supseteq G_{m_i}$. In other words, H is one of the H_i by the previous analysis of the choice of H_i. Thus, there are exactly l pairwise distinct values of $d(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}})$ when M runs over all possible degree-p intermediate fields of L/K. Hence, there are exactly l pairwise distinct values of $d(\mathfrak{P}_{\mathfrak{m}}|\mathcal{P})$, i.e., $l = t$. □

By Proposition 3.7.8, p. 127 of [Stichtenoth 2009], $d(\mathfrak{P}_{\mathfrak{m}}|\mathcal{P})$ is a multiple of $p-1$ for any M. Hence,

$$\text{(4-3)} \qquad d_i = \frac{d(\mathfrak{P}_i|\mathcal{P})}{p-1} - 1$$

is an integer for all $0 \le i \le l-1$. By (4-2), $d(\mathfrak{P}|\mathfrak{P}_i)$ is strictly decreasing with i. Hence, d_i is strictly increasing with i by (4-3). Now we are ready for the main result.

Theorem 4.4. *Let L/K be a Galois extension of a function fields whose Galois group is a p-group satisfying* (2-5), *where p is the characteristic of K. Let $\mathfrak{P}|\mathcal{P}$ be places, $\mathcal{P} \subseteq K$, $\mathfrak{P} \subseteq L$, m_i as in* (4-1) *for $0 \le i$, and d_j as in* (4-3) *for $j \le l-1$. Then*

$$m_i = d_0 + \sum_{j=1}^{i} p^{n-n_j}(d_j - d_{j-1}) \quad \text{for } 0 \le i \le l-1.$$

Proof. By applying (3-2) to L/K, we have

$$d(\mathfrak{P}|\mathcal{P}) = \sum_{j=0}^{l-1}(p^{n_j}-1)(m_j - m_{j-1}), \tag{4-4}$$

where $m_{-1} = -1$.

For $0 \le i \le l-1$ and $\#G_j = p^{n_i}$ for $m_{i-1} < j \le m_i$, then (4-2) yields

$$d(\mathfrak{P}|\mathfrak{P}_i) = \sum_{j=0}^{i}(p^{n_j-1}-1)(m_j - m_{j-1}) + \sum_{j=i+1}^{l-1}(p^{n_j}-1)(m_j - m_{j-1}). \tag{4-5}$$

Now substituting (4-4), (4-5), $e(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}}) = p^{n-1}$, and (4-3) into (3-1) for the case $\mathfrak{P}_{\mathfrak{m}} = \mathfrak{P}_i$, we have

$$\begin{aligned}&\sum_{j=0}^{l-1}(p^{n_j}-1)(m_j - m_{j-1})\\ &= p^{n-1}(d_i+1)(p-1)+\sum_{j=0}^{i}(p^{n_j-1}-1)(m_j-m_{j-1})+\sum_{j=i+1}^{l-1}(p^{n_j}-1)(m_j-m_{j-1}).\end{aligned}$$

Hence, we have

$$\sum_{j=0}^{i}(p^{n_j}-1)(m_j - m_{j-1}) = (p^n - p^{n-1})(d_i+1) + \sum_{j=0}^{i}(p^{n_j-1}-1)(m_j - m_{j-1}),$$

which implies

$$\sum_{j=0}^{i}(p^{n_j}-p^{n_j-1})(m_j - m_{j-1}) = (p^n - p^{n-1})(d_i+1). \tag{4-6}$$

When $i = 0$, (4-6) yields

$$(p^n - p^{n-1})(m_0+1) = (p^{n_0} - p^{n_0-1})(m_0+1) = (p^n - p^{n-1})(d_0+1),$$

which implies $m_0 = d_0$. Thus, the formula in Theorem 4.4 is true when $i = 0$. Now we induct on i. By (4-6), we have

$$\begin{aligned}(p^{n_i}-p^{n_i-1})m_i - (p^{n_i}-p^{n_i-1})m_{i-1} + \sum_{j=0}^{i-1}(p^{n_j}-p^{n_j-1})(m_j - m_{j-1})&\\ = (p^n - p^{n-1})(d_i+1).&\end{aligned}$$

It follows that

$$(p^{n_i}-p^{n_i-1})m_i-(p^{n_i}-p^{n_i-1})m_{i-1}+(p^n-p^{n-1})(d_{i-1}+1)=(p^n-p^{n-1})(d_i+1)$$

by applying (4-6) to the case $i-1$, which implies

$$(p^{n_i}-p^{n_i-1})m_i=(p^n-p^{n-1})(d_i-d_{i-1})+(p^{n_i}-p^{n_i-1})m_{i-1}.$$

By the induction hypothesis, it follows that

$$\begin{aligned}(p^{n_i}-p^{n_i-1})m_i\\&=(p^n-p^{n-1})(d_i-d_{i-1})+(p^{n_i}-p^{n_i-1})\biggl(d_0+\sum_{j=1}^{i-1}p^{n-n_j}(d_j-d_{j-1})\biggr).\end{aligned}$$

Dividing both sides by $p^{n_i}-p^{n_i-1}$, we obtain

$$m_i=d_0+p^{n-n_i}(d_i-d_{i-1})+\sum_{j=1}^{i-1}p^{n-n_j}(d_j-d_{j-1})=d_0+\sum_{j=1}^{i}p^{n-n_j}(d_j-d_{j-1}).$$

Our result follows by induction. □

The formula in Theorem 4.4 can be reformulated to be easily compared to the Hasse–Arf property.

Corollary 4.5. *With notation as in Theorem 4.4, and setting $m_{-1}=-1$, we have $m_i-m_{i-1}=p^{n-n_i}(d_i-d_{i-1})$ for $0\le i\le l-1$.*

Proof. This is immediate by applying the formula in Theorem 4.4 to the cases i and $i-1$. □

5. The Hasse–Arf property

The formula in Corollary 4.5 is expected due to the well-known Hasse–Arf property, see [Arf 1939]. It claims that the distance between two consecutive jumps in a ramification group filtration is divisible by the index of the group at the jump in the first group of the filtration. The Hasse–Arf property is true when the Galois group is abelian yet not always true otherwise.

In our setting, the Hasse–Arf property translates to $p^{n-n_j}\mid m_j-m_{j-1}$. So, according to Corollary 4.5, not only do we verify that it is true, we also know the quotient to be d_i-d_{i-1}. An advantage of knowing the quotient explicitly is that we can discuss whether the Hasse–Arf property can be improved or not. In fact, we can construct an example where the group index is a power of p, and no higher power of p can divide m_j-m_{j-1} than the power guaranteed by the Hasse–Arf Property. Notice that the strictly increasing property and $d_i\not\equiv 0 \pmod p$ are the only two restrictions on the sequence d_i of positive integers. See [Anbar et al. 2009] or [Wu and Scheidler 2010] for a discussion of the type of the extension L/K. Although an explicit construction is not given there, the extension L/K herein is of the same type as described in those two papers. Actually, we can construct Artin–Schreier extensions M_i over the same base field K with any prescribed different exponent $(d_i+1)(p-1)$; then we can construct L to be the composite of those M_i. That is to say, for any strictly increasing sequence of nonnegative integers d_i of length l, we can construct a Galois extension L/K of function fields and corresponding extension of places $\mathfrak{P}$ lying over $\mathcal{P}$, so that there are exactly l jumps in the ramification group filtration of $\mathfrak{P}|\mathcal{P}$ and exactly l pairwise distinct values of the different exponents $(d_0+1)(p-1) < (d_1+1)(p-1) < \cdots < (d_{l-1}+1)(p-1)$ for $d(\mathfrak{P}_{\mathfrak{m}}|\mathcal{P})$ when M ranges over all degree-p intermediate fields of L/K. In particular, we can require that $d_i = d_{i-1}+1$ for all $1 \le i \le l-1$. With this example, we know that there is no way to improve the Hasse–Arf divisibility result. On the other hand, $d_i - d_{i-1}$ can be any prescribed positive integer, so it is possible to strengthen the Hasse–Arf divisibility result arbitrarily under certain circumstances.

Now we want to analyze the Hasse–Arf property under a more general assumption; that is to say, to remove the totally wildly ramified assumption. First, we consider the not totally ramified case. With notation as in Theorem 4.4 and M as the inertia field of the place extension $\mathfrak{P}|\mathcal{P}$, we know that $G_i(\mathfrak{P}|\mathcal{P}) = G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}})$ for all $i \ge 0$ by Lemma 3.1. Hence, the Hasse–Arf property is true for the partially ramified case with identical parameters and formulae to those in the totally ramified case.

However, the situation changes when we move to the tamely ramified case. For that purpose, let $[L:K] = p^m q$ such that $p \nmid q$, M the intermediate field of L/K corresponding to a Sylow p-subgroup of $\mathrm{Gal}(L/K)$ under Galois correspondence, and $\mathfrak{P}_{\mathfrak{m}} = \mathfrak{P} \cap M$. By Corollary 4.5, we have $m_i - m_{i-1} = p^{n-n_i}(d_i - d_{i-1})$ for $0 \le i \le l-1$, where m_i and d_i are defined for the ramification filtrations $G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}})$. By Proposition 3.2, we know $G_i(\mathfrak{P}|\mathcal{P}) = G_i(\mathfrak{P}|\mathfrak{P}_{\mathfrak{m}})$ for all $i \ge 1$. Hence, the i-th jump in the ramification filtrations of $G_i(\mathfrak{P}|\mathcal{P})$ is equal to $p^{n-n_i}(d_i - d_{i-1})$ for $1 \le i \le l-1$. Therefore, the tamely ramified case does not satisfy the Hasse–Arf property in general since the distance needs to be divisible by $p^{n-n_i}q$, not p^{n-n_i}. Noticeably, it violates the Hasse–Arf property simply because it has an unexpected leading element G_0 in the ramification filtration. Hence, this is a removable violation. An easy way to address this is to manually modify the group index assumption from G_0 to a Sylow p-subgroup of $\mathrm{Gal}(L/K)$ under Galois correspondence. As a consequence, the Hasse–Arf property is true in the case that L/K is not necessarily assumed to be totally wildly ramified.

6. Conclusion

We analyzed the ramification group filtrations of a Galois function field extension, and reduced the investigation to the totally wildly ramified case. It turns out that the result is explicit. An explanation of why we can obtain such an explicit formula as in Theorem 4.4 is as follows. From (4-6), we have exactly l linear equations for the l variables m_i for $0 \le l \le l-1$. Since the coefficient matrix is triangular in addition to being nonsingular, we can expect that the solution not only exists and is unique but also could be expressed explicitly.

Due to the explicit nature of Corollary 4.5, we can discuss the Hasse–Arf property of such extensions and explore whether it can be strengthened or not. The general answer is no, and there exist examples to show that Hasse–Arf is the best possible divisibility result. Although we can discuss the ramification groups under the totally wild ramified assumption without loss of generality, we discussed whether or not the Hasse–Arf property is true under the general assumptions. The answer is yes, but we have to slightly modify the formulation of the Hasse–Arf property to apply it to the tamely ramified case.

Acknowledgments

We thank Rachel Pries for indicating a result similar to Corollary 4.5. It is likely that G. Griffith Elder anticipated it too, judging by some personal communications with him. We also thank the anonymous referee for her or his comments, which greatly improved the quality of the paper.

References

[Anbar et al. 2009] N. Anbar, H. Stichtenoth, and S. Tutdere, "On ramification in the compositum of function fields", *Bull. Braz. Math. Soc.* (*N.S.*) **40**:4 (2009), 539–552. MR 2011c:14081 Zbl 1187.14029

[Arf 1939] C. Arf, "Untersuchungen über reinverzweigte Erweiterungen diskret bewerteter perfekter Körper", *J. Reine Angew. Math.* **181** (1939), 1–44. MR 1,4a Zbl 0021.20201

[Byott and Elder 2002] N. P. Byott and G. G. Elder, "Biquadratic extensions with one break", *Canad. Math. Bull.* **45**:2 (2002), 168–179. MR 2003a:11149 Zbl 1033.11054

[Byott and Elder 2005] N. P. Byott and G. G. Elder, "New ramification breaks and additive Galois structure", *J. Théor. Nombres Bordeaux* **17**:1 (2005), 87–107. MR 2006b:11149 Zbl 1162.11394

[Byott and Elder 2009] N. P. Byott and G. G. Elder, "On the necessity of new ramification breaks", *J. Number Theory* **129**:1 (2009), 84–101. MR 2009j:11186 Zbl 1196.11156

[Doud 2003] D. Doud, "Wild ramification in number field extensions of prime degree", *Arch. Math.* (*Basel*) **81**:6 (2003), 646–649. MR 2004j:11130 Zbl 1056.11066

[Elder and Hooper 2007] G. G. Elder and J. J. Hooper, "On wild ramification in quaternion extensions", *J. Théor. Nombres Bordeaux* **19**:1 (2007), 101–124. MR 2008m:11230 Zbl 1123.11037

[Fesenko and Vostokov 2002] I. B. Fesenko and S. V. Vostokov, *Local fields and their extensions*, 2nd ed., Translations of Mathematical Monographs **121**, Amer. Math. Soc., Providence, RI, 2002. MR 2003c:11150 Zbl 1156.11046

[Garcia and Stichtenoth 2008] A. Garcia and H. Stichtenoth, "Some remarks on the Hasse–Arf theorem", pp. 141–146 in *Finite fields and applications*, edited by G. L. Mullen et al., Contemp. Math. **461**, Amer. Math. Soc., Providence, RI, 2008. MR 2010a:11229 Zbl 1236.11100

[Hasse 1930] H. Hasse, "Führer, Diskriminante und Verzweigungskörper relativ–Abelscher Zahlkörper", *J. Reine Angew. Math.* **162** (1930), 169–184. JFM 56.0166.03

[Hasse 1934] H. Hasse, "Normenresttheorie galoisscher Zahlkörper mit Anwendungen auf Führer und Diskriminante abelscher Zahlkörper", *J. Fac. Sci., Univ. Tokyo, Sect. I* **2** (1934), 477–498. Zbl 0009.04906

[Hungerford 1974] T. W. Hungerford, *Algebra*, Graduate Texts in Mathematics **73**, Springer, New York, 1974. MR 82a:00006 Zbl 0442.00002

[Maus 1968] E. Maus, "Die gruppentheoretische Struktur der Verzweigungsgruppenreihen", *J. Reine Angew. Math.* **230** (1968), 1–28. MR 37 #1356 Zbl 0165.35703

[Maus 1971] E. Maus, "On the jumps in the series of ramifications groups", pp. 127–133 in *Colloque de Théorie des Nombres* (Univ. Bordeaux, Bordeaux, 1969), Mém. Soc. Math. France **25**, Soc. Math. France, Paris, 1971. MR 51 #449 Zbl 0245.12014

[Maus 1972] E. Maus, "Über die Verteilung der Grundverzweigungszahlen von wild verzweigten Erweiterungen $\mathfrak{p}$-adischer Zahlkörper", *J. Reine Angew. Math.* **257** (1972), 47–79. MR 50 #7099 Zbl 0263.12007

[Roberts 2009] D. P. Roberts, "Review of [Garcia and Stichtenoth 2008]", review, MathSciNet, 2009, Available at http://www.ams.org/mathscinet-getitem?mr=2436331.

[Rosen 2002] M. Rosen, *Number theory in function fields*, Graduate Texts in Mathematics **210**, Springer, New York, 2002. MR 2003d:11171 Zbl 1043.11079

[Sen and Tate 1963] S. Sen and J. Tate, "Ramification groups of local fields", *J. Indian Math. Soc. (N.S.)* **27** (1963), 197–202. MR 34 #1303 Zbl 0136.02702

[Serre 1979] J.-P. Serre, *Local fields*, Graduate Texts in Mathematics **67**, Springer, New York-Berlin, 1979. MR 82e:12016 Zbl 0423.12016

[Stichtenoth 2009] H. Stichtenoth, *Algebraic function fields and codes*, 2nd ed., Graduate Texts in Mathematics **254**, Springer, Berlin, 2009. MR 2010d:14034 Zbl 1155.14022

[Thomas 2005] L. Thomas, "Ramification groups in Artin–Schreier–Witt extensions", *J. Théor. Nombres Bordeaux* **17**:2 (2005), 689–720. MR 2007a:11157 Zbl 1207.11109

[Viviani 2004] F. Viviani, "Ramification groups and Artin conductors of radical extensions of $\mathbb{Q}$", *J. Théor. Nombres Bordeaux* **16**:3 (2004), 779–816. MR 2006j:11148 Zbl 1075.11073

[Wu and Scheidler 2010] Q. Wu and R. Scheidler, "The ramification groups and different of a compositum of Artin–Schreier extensions", *Int. J. Number Theory* **6**:7 (2010), 1541–1564. MR 2011k:11161 Zbl 1225.11148

[Wyman 1969] B. F. Wyman, "Wildly ramified gamma extensions", *Amer. J. Math.* **91** (1969), 135–152. MR 39 #2726 Zbl 0188.11003

Received May 22, 2014. Revised December 10, 2014.

JEFFREY A. CASTAÑEDA
DEPARTMENT OF MATHEMATICS AND PHYSICS
TEXAS A&M INTERNATIONAL UNIVERSITY
LAREDO, TX 78041-1900
UNITED STATES
jcasta1@dusty.tamiu.edu

QINGQUAN WU
DEPARTMENT OF MATHEMATICS AND PHYSICS
TEXAS A&M INTERNATIONAL UNIVERSITY
LAREDO, TX 78041-1900
UNITED STATES
qingquan.wu@tamiu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.321

A MEAN FIELD TYPE FLOW II: EXISTENCE AND CONVERGENCE

JEAN-BAPTISTE CASTÉRAS

This paper is the continuation of (Castéras 2015), in which we investigated a gradient flow related to the mean field type equation. First, we show that this flow exists for all time. Next, using the compactness result of Castéras (2015), we prove, under a suitable hypothesis on its energy, the convergence of the flow to a solution of the mean field type equation. We also get a divergence result if the energy of the initial data is largely negative.

Introduction

Let (M, g) be a compact Riemannian surface without boundary. We will study an evolution problem associated to a mean field type equation

$$-\Delta v + Q = \rho \frac{e^v}{\int_M e^v \, dV}, \tag{0-1}$$

where ρ is a real number, $Q \in C^\infty(M)$ is a given function such that $\int_M Q \, dV = \rho$ and Δ is the Laplacian with respect to the metric g. Equation (0-1) is equivalent to the mean field equation

$$-\Delta u + \rho \left(\frac{-f e^u}{\int_M f e^u \, dV} + \frac{1}{|M|} \right) = 0, \tag{0-2}$$

where $|M|$ stands for the volume of M with respect to the metric g and $f \in C^\infty(M)$ is a positive function. Indeed, if v is a solution of (0-1), by setting $v = u + \log f$, we recover that u is a solution of (0-2) with $Q = \rho/|M| + \Delta \log f$.

The mean field equation appears in conformal geometry but also in statistical mechanics from Onsager's vortex model for turbulent Euler flows. More precisely, in this setting the solution u of the mean field equation is the stream function in the infinite vortex limit (see [Caglioti et al. 1992]). It also arises in the abelian Chern–Simons–Higgs model (see for example [Caffarelli and Yang 1995; Han 2003; Tarantello 1996; Yang 2001]).

The author was supported by the CNPq (Brazil) project 501559/2012-4.
MSC2010: 35B33, 35J20, 53C44, 58E20.
Keywords: mean field equation, blow-up analysis, geometric flow.

Equation (0-2) has a variational structure and its solutions can be found as the critical points of the functional

$$(0\text{-}3)\quad I_\rho(u)=\frac{1}{2}\int_M |\nabla u|^2\,dV+\frac{\rho}{|M|}\int_M u\,dV-\rho\log\biggl(\int_M fe^u\,dV\biggr),\ u\in H^1(M).$$

When $\rho<8\pi$, from the Moser–Trudinger inequality one can easily prove that the functional I_ρ is bounded from below and coercive; thus one can find a solution of (0-2) by minimizing I_ρ. The existence of solutions becomes more delicate if $\rho\geq 8\pi$. When $\rho=8\pi$, I_ρ admits a lower bound but is no longer coercive, while for $\rho>8\pi$, I_ρ is unbounded from below and above. The existence of solutions of (0-1) has been widely studied in recent decades. Many partial existence results have been obtained for $\rho\neq 8k\pi,\ k\in\mathbb{N}^*$, and according to the Euler characteristic of M (see for example [Brezis and Merle 1991; Chen and Lin 2003; Ding et al. 1999; Li 1999; Li and Shafrir 1994; Malchiodi 2008; Struwe and Tarantello 1998]). Finally, when $\rho\neq 8k\pi$, $k\in\mathbb{N}^*$, Djadli [2008] has generalized the previous results, establishing the existence of solutions for all surfaces M by studying the topology of sublevels $\{I_\rho\leq -C\}$ to achieve a min-max scheme (already introduced in [Djadli and Malchiodi 2008]).

In this paper, we consider the evolution problem associated to (0-1), namely the equation

$$(0\text{-}4)\qquad \begin{cases}\dfrac{\partial}{\partial t}e^v=\Delta v-Q+\rho\dfrac{e^v}{\int_M e^v\,dV},\\ v(x,0)=v_0(x),\end{cases}$$

with initial data $v_0\in C^{2+\alpha}(M)$, $\alpha\in(0,1)$ and a function $Q\in C^\infty(M)$ such that $\int_M Q dV=\rho$. It is a gradient flow with respect to the following functional, which will be called energy:

$$(0\text{-}5)\quad J_\rho(v)=\frac{1}{2}\int_M|\nabla v|^2\,dV+\int_M Qv(t)\,dV-\rho\ln\biggl(\int_M e^v\,dV\biggr),\quad v\in H^1(M).$$

This functional is unbounded from below (except in the case $\rho<8\pi$) and above. The interest of this flow is that it satisfies some important geometrical properties useful for its convergence (see in particular estimate (2-2) of Section 2). When Q is a constant equal to the scalar curvature of M with respect to the metric g, the flow (0-4) (normalized) has been studied by Struwe [2002] (we note that in this case $\rho=\int_M Q\,dV\leq 8\pi$). For other curvature flows, we refer to [Baird et al. 2004; Brendle 2003; 2005; 2006; Castéras 2013; Hamilton 1988; Lamm et al. 2009; Malchiodi and Struwe 2006; Schwetlick and Struwe 2003; Struwe 2005] and the references therein.

We begin by studying the global existence of the flow (0-4). We prove:

Theorem 0.1. *For all initial data* $v_0 \in C^{2+\alpha}(M)$, $\alpha \in (0, 1)$, *all* $\rho \in \mathbb{R}$ *and all functions* $Q \in C^\infty(M)$ *such that* $\int_M Q\, dV = \rho$, *there exists a unique global solution* $v \in C^{2+\alpha, 1+\alpha/2}_{\mathrm{loc}}(M \times [0, +\infty))$ *of* (0-4).

Next, we investigate the convergence of the flow. Let $v(t) : M \to \mathbb{R}$ denote the function defined by $v(t)(x) = v(x, t)$. We show that if the energy $J_\rho(v(t))$ of the global solution is bounded from below uniformly in time (when $\rho > 8\pi$), then as $t \to +\infty$, $v(t)$ converges to a function v_∞ which is a solution of (0-1). More precisely, we have:

Theorem 0.2. *Let* $v(t)$ *be the solution of* (0-4).

(i) *If* $\rho < 8\pi$, *then* $v(t)$ *converges in* $H^2(M)$ *to a solution* $v_\infty \in C^\infty(M)$ *of*

$$-\Delta v_\infty + Q = \rho \frac{e^{v_\infty}}{\int_M e^{v_\infty}\, dV}.$$

(ii) *If* $\rho > 8\pi$, $\rho \neq 8k\pi$, $k \in \mathbb{N}^*$, *and if there exists a constant* $C > 0$ *not depending on* t *such that for all* $t \geq 0$,

$$J_\rho(v(t)) \geq -C, \tag{0-6}$$

then $v(t)$ *converges in* $H^2(M)$ *to a solution* $v_\infty \in C^\infty(M)$ *of*

$$-\Delta v_\infty + Q = \rho \frac{e^{v_\infty}}{\int_M e^{v_\infty}\, dV}.$$

Moreover, we prove that there exist initial data $v_0 \in C^\infty(M)$ such that the energy of the global solution $v(t)$ of the flow, with $v(0)(x) = v_0(x)$ for all $x \in M$, stays uniformly bounded from below, and hence, thanks to Theorem 0.2, such that the flow converges.

Theorem 0.3. *Let* $\rho \neq 8k\pi$, $k \in \mathbb{N}^*$. *There exist initial data* $v_0 \in C^\infty(M)$ *such that the global solution* $v(t)$ *of* (0-4) *with* $v(x, 0) = v_0(x)$, *for all* $x \in M$, *satisfies* (0-6), *i.e., such that the global solution* $v(t)$ *of* (0-4) *converges in* $H^2(M)$ *to a solution* $v_\infty \in C^\infty(M)$ *of* (0-1):

$$-\Delta v_\infty + Q = \rho \frac{e^{v_\infty}}{\int_M e^{v_\infty}\, dV}.$$

Finally, we show that if the energy of the initial data v_0 of (0-4) is largely negative then the flow diverges when $t \to +\infty$.

Theorem 0.4. *Let* $\rho \in (8k\pi, 8(k+1)\pi)$, $k \geq 1$. *There exists a constant* $C > 0$ *depending on* M, Q *and* ρ *such that for all* $v_0 \in C^{2+\alpha}(M)$ *satisfying* $J_\rho(v_0) \leq -C$, *the global solution* $v(t)$ *of* (0-4) *satisfies*

$$J_\rho(v(t)) \underset{n \to +\infty}{\longrightarrow} -\infty.$$

To prove these convergence results, we use the compactness result of [Castéras 2015]. There we studied the compactness property of solutions $(v_n)_n \subseteq H^2(M)$ of the perturbed elliptic mean field type equation

$$-\Delta v_n = Q_0 + h_n e^{v_n} + \rho e^{v_n}, \tag{0-7}$$

where $\rho > 0$, $Q_0 \in C^0(M)$ and $(h_n)_n \subseteq C^0(M)$. The term h_n corresponds to the parabolic term of (0-4). We also assume that there exists a constant $C > 0$ not depending on n such that

$$\begin{aligned} &\text{(i)}\ \lim_{n\to+\infty} \int_M h_n^2 e^{v_n}\, dV = 0,\\ &\text{(ii)}\ h_n(x)e^{v_n(x)} + \rho e^{v_n(x)} \geq -C, \quad \forall x \in M, \quad \forall n \geq 0. \end{aligned} \tag{0-8}$$

We will see that these conditions are satisfied by the solution of the flow (0-4). We have established in [Castéras 2015] the following compactness result:

Theorem 0.5. *Let $(v_n)_n \subseteq H^2(M)$ be a sequence of solutions of* (0-7) *such that $\int_M e^{v_n}\, dV = 1$ for all $n \geq 0$, and satisfying* (0-8). *If $\rho \neq 8k\pi$, $k \in \mathbb{N}^*$, then there exists a constant C not depending on n such that*

$$\|v_n\|_{H^2(M)} \leq C.$$

The paper is organized as follows. In Section 1, we prove the global existence of a solution of (0-4). We also show the continuity of the flow with respect to its initial data. In Section 2, we study the convergence of the flow (0-4). We begin by proving Theorem 0.2. We first show that the global solution $v(t)$ of (0-4) is uniformly (with respect to t) bounded in $H^1(M)$ when $v(t)$ satisfies condition (0-6). The proof involves the compactness result obtained in Theorem 0.5. We point out that, when $\rho < 8\pi$, condition (0-6) is always satisfied. Then, we show that the parabolic term of (0-4), $\partial v(t)/\partial t$, tends to 0 as $t \to +\infty$ in the $L^2(M)$ norm with respect to the metric $g_1(t) = e^{v(t)}g$. This implies that $v(t)$ is uniformly bounded in $H^2(M)$. Next we prove Theorem 0.3, i.e., there exists initial data in $C^\infty(M)$ such that condition (0-6) is satisfied. Our proof is based on the study of the topology of the level set

$$\{v \in X \ :\ J_\rho(v) \leq -L\},$$

where X is the space of $C^\infty(M)$ functions endowed with the $C^{2+\alpha}(M)$ norm, $\alpha \in (0, 1)$. The end of Section 2 is devoted to the proof of Theorem 0.4.

1. Global existence of the flow

We begin by noticing that since the flow is parabolic, standard methods (see for example [Friedman 1964]) provide short time existence. Thus, there exists $T_1 > 0$ such that $v \in C^{2+\alpha,1+\alpha/2}(M \times [0, T_1])$ is a solution of (0-4). We give two basic properties of the flow: the conservation of the volume of M endowed with the metric $g_1(t) = e^{v(t)} g$, and the decreasing along the flow of the functional $J_\rho(v(t))$.

Proposition 1.1. (i) *For all* $t \in [0, T_1]$, *we have*

$$\int_M e^{v(t)}\, dV = \int_M e^{v_0}\, dV. \tag{1-1}$$

(ii) *If* $0 \le t_0 \le t_1 \le T_1$, *we have*

$$J_\rho(v(t_1)) \le J_\rho(v(t_0)). \tag{1-2}$$

Proof. To see that (1-1) holds, it is sufficient to integrate (0-4) on M. Differentiating $J_\rho(v(t))$ with respect to t and integrating by parts, one finds, for all $t \in [0, T_1]$,

$$\frac{\partial}{\partial t} J_\rho(v(t)) = -\int_M \Big(\frac{\partial v(t)}{\partial t}\Big)^2 e^{v(t)}\, dV \le 0. \tag{1-3}$$

This implies (1-2). □

Proof of Theorem 0.1. To prove the global existence of the flow, we set

$$T = \sup\big\{\overline{T} > 0 : C^{2+\alpha,1+\alpha/2}(M \times [0, \overline{T}]) \text{ contains a solution } v \text{ of (0-4)}\big\},$$

and suppose that $T < +\infty$. From the definition of T, we must have a solution $v \in C^{2+\alpha,1+\alpha/2}_{\mathrm{loc}}(M \times [0, T))$. We show that there exists a constant $\widetilde{C}_T > 0$, depending on T, M, Q, ρ, α and $\|v_0\|_{C^{2+\alpha}(M)}$, such that

$$\|v\|_{C^{2+\alpha,1+\alpha/2}(M\times[0,T))} \le \widetilde{C}_T. \tag{1-4}$$

This estimate allows us to extend v beyond T, contradicting the definition of T.

In the following, C denotes constants depending on M, Q, ρ, α and $\|v_0\|_{C^{2+\alpha}(M)}$, while C_T represents constants depending on M, Q, ρ, α, $\|v_0\|_{C^{2+\alpha}(M)}$ and T. They are allowed to vary from line to line.

Proposition 1.2. *For all* $\rho \in \mathbb{R}$, *there exists a constant* $\widetilde{C}_{T,1}$ *depending on* M, Q, ρ, $\|v_0\|_{C^{2+\alpha}(M)}$ *and* T, *such that*

$$\|v(t)\|_{H^1(M)} \le \widetilde{C}_{T,1}, \quad \forall t \in [0, T). \tag{1-5}$$

Moreover, if $\rho < 8\pi$, *then there exists a constant* $\widetilde{C}_1$ *depending on* M, Q, ρ *and* $\|v_0\|_{C^{2+\alpha}(M)}$ *but not on* T *such that*

$$\|v(t)\|_{H^1(M)} \le \widetilde{C}_1, \quad \forall t \in [0, T). \tag{1-6}$$

Proof. We decompose the proof into three steps.

Step 1. Let $\rho \geq 8\pi$. There exists a constant C'_T, depending on M, Q, ρ, $\|v_0\|_{C^{2+\alpha}(M)}$ and T, such that

$$v(x,t) \leq C'_T, \quad \forall x \in M, \quad \forall t \in [0,T). \tag{1-7}$$

Proof of Step 1. Define $v_{\max}(t) = \max_{x \in M} v(x,t) = v(x_t,t)$ where $x_t \in M$. Consider the upper derivative of $v_{\max}(t)$, i.e.,

$$\frac{\partial}{\partial t} v_{\max}(t) = \limsup_{h \to 0^+} \frac{v(x_{t+h}, t+h) - v(x_t, t)}{h}. \tag{1-8}$$

We can assume that $v_{\max}(t)$ is differentiable. By the maximum principle, and since v satisfies (0-4), we find

$$\frac{\partial}{\partial t} e^{v_{\max}(t)} \leq \frac{\rho}{\int_M e^{v_0}\,dV}\left(\|Q\|_{L^\infty(M)} \frac{\int_M e^{v_0}\,dV}{\rho} + e^{v_{\max}(t)}\right),$$

where we use the fact that $\int_M e^{v(t)}\,dV = \int_M e^{v_0}\,dV$ for all $t \in [0,T)$. Integrating this last inequality between 0 and t, we get

$$e^{v_{\max}(t)} + \|Q\|_{L^\infty(M)} \frac{\int_M e^{v_0}\,dV}{\rho} \leq \left(e^{v_{\max}(0)} + \|Q\|_{L^\infty(M)} \frac{\int_M e^{v_0}\,dV}{\rho}\right) e^{\frac{\rho t}{\int_M e^{v_0}\,dV}},$$

and (1-7) follows.

Step 2. Let $\rho \geq 8\pi$. There exists a subset A of M, with volume satisfying $|A| > C_T$ for some constant $C_T > 0$, and a constant δ depending on M, Q, ρ, $\|v_0\|_{C^{2+\alpha}(M)}$ and T, such that

$$|v(x,t)| \leq \delta, \quad \forall x \in A, \quad \forall t \in [0,T). \tag{1-9}$$

Proof of Step 2. Fix $t \in [0,T)$ and set

$$M_\varepsilon = \{x \in M : e^{v(x,t)} < \varepsilon\},$$

where $\varepsilon > 0$ is a real number which will be determined later. Setting $\int_M e^{v_0}\,dV = a$, by the conservation of the volume and (1-7), we have

$$a = \int_M e^{v(t)}\,dV = \int_{M_\varepsilon} e^{v(t)}\,dV + \int_{M \setminus M_\varepsilon} e^{v(t)}\,dV \leq \varepsilon |M_\varepsilon| + e^{C'_T} |M \setminus M_\varepsilon|.$$

Taking $\varepsilon = \dfrac{a}{2|M|}$, we find

$$|M \setminus M_\varepsilon| \geq \frac{a}{2} e^{-C'_T} > 0. \tag{1-10}$$

Setting $A = M \setminus M_\varepsilon$, by definition of M_ε we have $v(x,t) \geq \ln \varepsilon = \ln(a/2|M|)$ for all $x \in A$ and $t \in [0,T)$. On the other hand, by Step 1, $v(x,t) \leq C'_T$ for all $x \in M$ and $t \in [0,T)$. Therefore we find that there exists a constant δ such that

$$|v(x,t)| \leq \delta, \quad \forall x \in A, \quad \forall t \in [0,T).$$

Step 3. Let $\rho \geq 8\pi$. For all $t \in [0, T)$, we have

$$\int_M v^2(t)\,dV \leq C_1 \int_M |\nabla v(t)|^2\,dV + C_2, \tag{1-11}$$

where C_1, C_2 are constants depending on T, Q, $|v_0\|_{C^{2+\alpha}(M)}$, M and A (where A is the set defined in Step 2).

Proof of Step 3. By Poincaré's inequality,

$$\int_M v^2(t)\,dV \leq \frac{1}{\lambda_1} \int_M |\nabla v(t)|^2\,dV + \frac{1}{|M|}\left(\int_M v(t)\,dV\right)^2, \tag{1-12}$$

where λ_1 is the first eigenvalue of the Laplacian. Now, using Young's inequality and (1-9), we find

$$\begin{aligned} \frac{1}{|M|}\left(\int_M v(t)\,dV\right)^2 &= \frac{1}{|M|}\left(\int_A v(t)\,dV\right)^2 \\ &\quad + \frac{1}{|M|}\left(\int_{M\setminus A} v(t)\,dV\right)^2 + \frac{2}{|M|}\left(\int_A v(t)\,dV\right)\left(\int_{M\setminus A} v(t)\,dV\right) \\ &\leq \frac{\delta^2|A|^2}{|M|} + \frac{1}{|M|}\left(\int_{M\setminus A} v(t)\,dV\right)^2 + \frac{2\delta^2|A|^2}{\varepsilon|M|} + \frac{2\varepsilon}{|M|}\left(\int_{M\setminus A} v(t)\,dV\right)^2, \end{aligned} \tag{1-13}$$

where ε is a positive constant which will be determined later. By the Cauchy–Schwarz inequality,

$$\left(\int_{M\setminus A} v(t)\,dV\right)^2 \leq |M\setminus A| \int_{M\setminus A} v^2(t)\,dV. \tag{1-14}$$

Thus, (1-12), (1-13) and (1-14) yield

$$\begin{aligned} \int_M v^2(t)\,dV \\ \leq \frac{1}{\lambda_1} \int_M |\nabla v(t)|^2\,dV + \left(1 - \frac{|A|}{|M|} + \frac{2\varepsilon}{|M|}|M\setminus A|\right) \int_M v^2(t)\,dV + \widetilde{C}, \end{aligned} \tag{1-15}$$

where

$$\widetilde{C} = \frac{\delta^2|A|^2}{|M|} + \frac{2\delta^2|A|^2}{\varepsilon|M|}.$$

Choosing ε such that the factor in parentheses in (1-15) equals $\alpha < 1$, we deduce

$$(1-\alpha)\int_M v^2(t)\,dV \leq \frac{1}{\lambda_1} \int_M |\nabla v(t)|^2\,dV + \widetilde{C},$$

establishing (1-11).

Proof of Proposition 1.2. We consider separately the cases $\rho < 8\pi$ and $\rho \geq 8\pi$.

In the first case, we prove that the constant $\widetilde{C}_1$ of estimate (1-6) is independent of T. Using Poincaré's and Young's inequalities, we have

$$C\int_M |v(t)-\bar{v}(t)|\,dV \leq \varepsilon\int_M |\nabla v(t)|^2\,dV + C,$$

where $\varepsilon > 0$ is a small constant to be chosen later. This implies that

$$\begin{aligned}(1\text{-}16)\quad J_\rho(v(t)) &= \\ &\frac{1}{2}\int_M |\nabla v(t)|^2\,dV + \int_M Q(v(t)-\bar{v}(t))\,dV - \rho\log\Bigl(\int_M e^{v(t)-\bar{v}(t)}\,dV\Bigr) \\ &\geq \Bigl(\frac{1}{2}-\varepsilon\Bigr)\int_M |\nabla v(t)|^2\,dV - C - \rho\log\Bigl(\int_M e^{v(t)-\bar{v}(t)}\,dV\Bigr).\end{aligned}$$

By Jensen's inequality, we have

$$\log\Bigl(\int_M e^{v(t)-\bar{v}(t)}\,dV\Bigr) \geq C, \quad \forall t \in [0, T), \tag{1-17}$$

where $\bar{v}(t) = (\int_M v(t)\,dV)/|M|$. Hence, using (1-16) and (1-17), and setting $\rho_1 = \max\{\rho, 0\}$, we deduce that

$$J_\rho(v(t)) \geq \Bigl(\frac{1}{2}-\varepsilon\Bigr)\int_M |\nabla v(t)|^2\,dV - C - \rho_1\log\Bigl(\int_M e^{v(t)-\bar{v}(t)}\,dV\Bigr).$$

By the Moser–Trudinger inequality (see [Moser 1970/71; Trudinger 1967]), one has

$$\log\int_M e^{(v(t)-\bar{v}(t))}\,dV \leq \frac{1}{16\pi}\int_M |\nabla v(t)|^2\,dV + C. \tag{1-18}$$

Therefore

$$J_\rho(v(t)) \geq \Bigl(\frac{1}{2}-\frac{\rho_1}{16\pi}-\varepsilon\Bigr)\int_M |\nabla v(t)|^2\,dV - C.$$

Thus, by taking $\varepsilon = (8\pi-\rho_1)/32\pi$ and using the fact that $J_\rho(v(t)) \leq J_\rho(v_0)$ for all $t \in [0, T)$, we find that

$$\int_M |\nabla v(t)|^2\,dV \leq C, \quad \forall \rho < 8\pi. \tag{1-19}$$

Now, using (1-19) and Poincaré's inequality, we obtain

$$\|v(t)-\bar{v}(t)\|_{H^1(M)} \leq C, \quad \forall \rho < 8\pi. \tag{1-20}$$

Since $\int_M e^{v(t)}\,dV = \int_M e^{v_0}\,dV$ for all $t \in [0, T)$, using Jensen's inequality (1-17), the Moser–Trudinger inequality (1-18) and (1-19), we deduce that

$$|\bar{v}(t)| \leq C.$$

Finally, from (1-20) and the previous inequality, we find that for all $\rho < 8\pi$, there exists a constant $\widetilde{C}_1$ independent of T such that

$$\|v(t)\|_{H^1(M)} \le \widetilde{C}_1, \quad \forall\, 0 \le t < T. \tag{1-21}$$

We now consider the second case, $\rho \ge 8\pi$. Since $\int_M e^{v(t)}\,dV = \int_M e^{v_0}$ for all $t \in [0, T)$, by Young's inequality we have

$$\begin{aligned} J_\rho(v_0) \ge J_\rho(v) &\ge \frac{1}{2}\int_M |\nabla v|^2\,dV + \int_M Qv\,dV - C \\ &\ge \frac{1}{2}\int_M |\nabla v|^2\,dV - \varepsilon\int_M v^2(t)\,dV - C, \end{aligned} \tag{1-22}$$

where ε is a positive constant which will be chosen later. Thanks to estimate (1-11) of Step 3, inequality (1-22) leads to

$$\frac{1}{2}\int_M |\nabla v(t)|^2\,dV \le C + C_1\varepsilon\int_M |\nabla v(t)|^2\,dV.$$

Choosing ε such that $1/2 - \varepsilon C_1 > 0$, we find that for all $t \in [0, T)$, there exists a constant $C_T > 0$ such that

$$\int_M |\nabla v(t)|^2\,dV \le C_T. \tag{1-23}$$

Combining (1-11) and (1-23), we obtain $\int_M v^2(t)\,dV \le C_T$. Finally, for all $\rho \in \mathbb{R}$, there exists a constant $\widetilde{C}_{T,1} > 0$ such that

$$\|v(t)\|_{H^1(M)} \le \widetilde{C}_{T,1}, \quad \forall t \in [0, T). \qquad \square$$

Proposition 1.3. *There is a constant $\widetilde{C}_{T,2} > 0$, depending on M, Q, ρ, $\|v_0\|_{C^{2+\alpha}(M)}$ and T, such that*

$$\|v(t)\|_{H^2(M)} \le \widetilde{C}_{T,2}, \quad \forall\, 0 \le t < T.$$

Proof. Since $\|v\|_{H^1(M)} \le \widetilde{C}_{T,1}$, we just need to bound $\int_M (\Delta v(t))^2\,dV$ for all $t \in [0, T)$. To this purpose, set

$$w(t) = \frac{\partial v(t)}{\partial t} e^{v(t)/2}.$$

By differentiating with respect to t and integrating by parts on M, we have

$$\begin{aligned} \frac{1}{2}\frac{\partial}{\partial t}\int_M (\Delta v(t))^2\,dV &= \int_M \Big(w(t)e^{v(t)/2} + Q - \frac{\rho e^{v(t)}}{\int_M e^{v_0}\,dV}\Big)\Delta\big(w(t)e^{-v(t)/2}\big)\,dV \\ &= -\int_M |\nabla w(t)|^2\,dV + \frac{1}{4}\int_M w^2(t)|\nabla v(t)|^2\,dV + \int_M \Delta Q\big(w(t)e^{-v(t)/2}\big)\,dV \\ &\quad + \frac{\rho}{\int_M e^{v_0}\,dV}\Big(\int_M \nabla v(t)\nabla w(t)e^{v(t)/2}\,dV - \frac{1}{2}\int_M w(t)e^{v(t)/2}|\nabla v(t)|^2\,dV\Big). \end{aligned}$$

Since $Q \in C^\infty(M)$ and $w(t) = \dfrac{\partial v(t)}{\partial t} e^{v(t)/2}$, we find

$$
\begin{aligned}
(1\text{-}24)\quad & \frac{1}{2}\frac{\partial}{\partial t}\int_M (\Delta v(t))^2\, dV \\
&\le -\int_M |\nabla w(t)|^2\, dV + \frac{1}{4}\int_M w^2(t)|\nabla v(t)|^2\, dV + C\Big\|\frac{\partial v(t)}{\partial t}\Big\|_{L^1(M)} \\
&\qquad + C\Big(\int_M e^{v(t)/2}\big(|\nabla w(t)||\nabla v(t)| + |\nabla v(t)|^2|w(t)|\big)\, dV\Big).
\end{aligned}
$$

We now estimate the positive terms on the right of (1-24). From the Gagliardo–Nirenberg inequality (see for example [Brouttelande 2003]), for all $f \in H^1(M)$,

$$\|f\|^2_{L^4(M)} \le C\|f\|_{L^2(M)}\|f\|_{H^1(M)}.$$

Using the Cauchy–Schwarz inequality and (1-5), we have

$$
\begin{aligned}
(1\text{-}25)\qquad \int_M w^2(t)|\nabla v(t)|^2\, dV &\le \|w(t)\|^2_{L^4(M)}\|\nabla v(t)\|^2_{L^4(M)} \\
&\le C_T\|w(t)\|_{L^2(M)}\|w(t)\|_{H^1(M)}\|v(t)\|_{H^2(M)}.
\end{aligned}
$$

Using (1-5) and the Moser–Trudinger inequality (1-18), we deduce that there exists a constant C_T such that, for all $t \in [0, T)$ and $p \in \mathbb{R}$,

$$(1\text{-}26)\qquad \int_M e^{pv(t)}\, dV \le C_T.$$

By the same reasoning used to prove (1-25), from (1-5) and (1-26) we have

$$
\begin{aligned}
(1\text{-}27)\quad & \int_M |\nabla v(t)|^2|w(t)|e^{v(t)/2}\, dV \\
&\le \Big(\int_M |\nabla v(t)|^4\, dV\Big)^{1/2}\Big(\int_M w^4(t)\, dV\Big)^{1/4}\Big(\int_M e^{2v(t)}\, dV\Big)^{1/4} \\
&\le C_T\|v(t)\|_{H^2(M)}\|w(t)\|^{1/2}_{H^1(M)}\|w(t)\|^{1/2}_{L^2(M)},
\end{aligned}
$$

$$
\begin{aligned}
(1\text{-}28)\quad & \int_M |\nabla w(t)||\nabla v(t)|e^{v(t)/2}\, dV \\
&\le \Big(\int_M |\nabla w(t)|^2\, dV\Big)^{1/2}\Big(\int_M |\nabla v(t)|^4\, dV\Big)^{1/4}\Big(\int_M e^{2v(t)}\, dV\Big)^{1/4} \\
&\le C_T\|w(t)\|_{H^1(M)}\|v(t)\|^{1/2}_{H^2(M)},
\end{aligned}
$$

$$
\begin{aligned}
(1\text{-}29)\quad \int_M \Big|\frac{\partial v(t)}{\partial t}\Big|\, dV &\le \Big(\int_M \Big(\frac{\partial v(t)}{\partial t}\Big)^2 e^{v(t)}\, dV\Big)^{1/2}\Big(\int_M e^{-v(t)}\, dV\Big)^{1/2} \\
&\le C_T\|w(t)\|_{L^2(M)}.
\end{aligned}
$$

Finally, putting (1-25), (1-27), (1-28) and (1-29) in (1-24), we obtain

$$\begin{aligned}\frac{1}{2}\frac{\partial}{\partial t}\int_M (\Delta v(t))^2\,dV &\le -\int_M |\nabla w(t)|^2\,dV \\ &\quad + C_T\|w(t)\|_{H^1(M)}\|w(t)\|_{L^2(M)}\|v(t)\|_{H^2(M)} + C_T\|w(t)\|_{L^2(M)} \\ &\quad + C_T\big(\|w(t)\|_{H^1(M)}\|v(t)\|^{1/2}_{H^2(M)} + \|w(t)\|^{1/2}_{H^1(M)}\|w(t)\|^{1/2}_{L^2(M)}\|v(t)\|_{H^2(M)}\big).\end{aligned}$$

Using Young's inequality, we get

$$\frac{\partial}{\partial t}\bigg(\int_M (\Delta v(t))^2\,dV + 1\bigg) \le C_T\bigg(\int_M (\Delta v(t))^2\,dV + 1\bigg)\big(\|w(t)\|^2_{L^2(M)} + 1\big). \tag{1-30}$$

On the other hand, by (1-3), we have for all $t \in [0, T)$

$$\begin{aligned}\int_0^t \|w(s)\|^2_{L^2(M)}\,ds &= \int_0^t\int_M \Big(\frac{\partial v(s)}{\partial s}\Big)^2 e^{v(s)}\,dV\,ds \\ &= -\int_0^t \frac{\partial}{\partial s}J_\rho(v(s))\,ds = J_\rho(v_0) - J_\rho(v(t)) \le C_T,\end{aligned} \tag{1-31}$$

where we use the fact that $\|v(t)\|_{H^1(M)} \le \widetilde{C}_{T,1}$ from Proposition 1.2. Integrating (1-30) with respect to t and using (1-31), we have

$$\int_M (\Delta v(t))^2\,dV \le C_T, \quad \forall t \in [0, T).$$

Since $\|v(t)\|_{H^1(M)} \le \widetilde{C}_{T,1}$, we deduce that there exists a constant $\widetilde{C}_{T,2}$ such that

$$\|v(t)\|_{H^2(M)} \le \widetilde{C}_{T,2}, \quad \forall t \in [0, T). \qquad \square$$

Proof of Theorem 0.1. We recall that to prove the global existence of the flow it is sufficient to prove (1-4), i.e., there exists a constant $\widetilde{C}_T$ depending on T and $\alpha \in (0, 1)$ such that

$$\|v\|_{C^{2+\alpha,1+\alpha/2}(M\times[0,T))} \le \widetilde{C}_T.$$

First, we claim that for all $\alpha \in (0, 1)$, there exists a constant C_T such that

$$|v(x_1, t_1) - v(x_2, t_2)| \le C_T(|t_1 - t_2|^{\alpha/2} + |x_1 - x_2|^\alpha), \tag{1-32}$$

for all $x_1, x_2 \in M$ and $t_1, t_2 \in [0, T)$. Here $|x_1 - x_2|$ stands for the geodesic distance from x_1 to x_2 with respect to the metric g. From Proposition 1.3, for all $t \in [0, T)$ we have $\|v(t)\|_{H^2(M)} \le \widetilde{C}_{T,2}$. Thus, by Sobolev's embedding theorem (see [Hebey 1997]), we find for $\alpha \in (0, 1)$, $v(t) \in C^\alpha(M)$ and for all $x, y \in M$,

$$|v(x, t) - v(y, t)| \le C_T|x - y|^\alpha. \tag{1-33}$$

If $t_2 - t_1 \geq 1$, using (1-33) it is easy to see that (1-32) holds. Thus, from now on we assume that $0 < t_2 - t_1 < 1$. Since $v(t)$ is a solution of (0-4) and $\|e^{v(t)}\|_{C^\alpha(M)} \leq C_T$, for all $t \in [0, T)$ one has

$$\left|\frac{\partial v(t)}{\partial t}\right|^2 \leq C_T |\Delta v(t)|^2 + C_T.$$

Integrating on M, we obtain for all $t \in [0, T)$

$$\int_M \left|\frac{\partial v(t)}{\partial t}\right|^2 dV \leq C_T \|v(t)\|^2_{H^2(M)} + C_T \leq C_T. \tag{1-34}$$

Now, we write

$$\begin{aligned} |v(x, t_1) - v(x, t_2)| &= \frac{1}{|B_{\sqrt{t_2 - t_1}}(x)|} \int_{B_{\sqrt{t_2-t_1}}(x)} |v(x, t_1) - v(x, t_2)| \, dV(y)\,, \\ &\leq P_1 + P_2 + P_3 \end{aligned} \tag{1-35}$$

where $B_{\sqrt{t_2-t_1}}(x)$ stands for the geodesic ball of center x and radius $\sqrt{t_2 - t_1}$ and

$$P_1 = \frac{C}{t_2 - t_1} \int_{B_{\sqrt{t_2-t_1}}(x)} |v(x, t_1) - v(y, t_1)| \, dV(y),$$

$$P_2 = \frac{C}{t_2 - t_1} \int_{B_{\sqrt{t_2-t_1}}(x)} |v(y, t_1) - v(y, t_2)| \, dV(y),$$

$$P_3 = \frac{C}{t_2 - t_1} \int_{B_{\sqrt{t_2-t_1}}(x)} |v(y, t_2) - v(x, t_2)| \, dV(y),$$

Using (1-33), we obtain

$$P_1 \leq \frac{C_T}{t_2 - t_1} \int_{B_{\sqrt{t_2-t_1}}(x)} |x - y|^\alpha \, dV(y) \leq C_T (t_2 - t_1)^{\alpha/2}. \tag{1-36}$$

In the same way, we have

$$P_3 \leq C_T (t_2 - t_1)^{\alpha/2}. \tag{1-37}$$

From Hölder's inequality and (1-34) it follows that

$$\begin{aligned} P_2 &\leq C \sup_{t_1 \leq \tau \leq t_2} \int_{B_{\sqrt{t_2-t_1}}(x)} \left|\frac{\partial v}{\partial s}\right|(y, \tau) \, dV(y) \\ &\leq C\sqrt{t_2 - t_1} \sup_{t_1 \leq \tau \leq t_2} \left(\int_{B_{\sqrt{t_2-t_1}}(x)} \left|\frac{\partial v}{\partial s}\right|^2 (y, \tau) \, dV(y) \right)^{1/2} \\ &\leq C_T \sqrt{t_2 - t_1}. \end{aligned} \tag{1-38}$$

Putting (1-36), (1-37) and (1-38) in (1-35), and noticing that $\sqrt{t_2 - t_1} \leq (t_2 - t_1)^{\alpha/2}$

for all $0 < t_2 - t_1 < 1$, we find

$$|v(x,t_1) - v(x,t_2)| \le C_T (t_2 - t_1)^{\alpha/2}. \tag{1-39}$$

Therefore, from (1-33) and (1-39), we see that (1-32) holds. In view of (1-32), we may apply the standard regularity theory for parabolic equations (see for example [Friedman 1964]) to derive the existence of a constant $\widetilde{C}_T$ depending on T and $\alpha \in (0,1)$ such that

$$\|v\|_{C^{2+\alpha,1+\alpha/2}(M\times[0,T))} \le \widetilde{C}_T.$$

This establishes the existence part of Theorem 0.1. The uniqueness follows from Proposition 1.5. ☐

Remark 1.4. Following the proof of Theorem 0.1, we see that, for all $T > 0$ fixed, if $\|u_0\|_{C^{2+\alpha}(M)} \le K$ for some constant $K > 0$, then there exists a constant $C_T > 0$ depending on K and T such that

$$\|u\|_{C^{2+\alpha,1+\alpha/2}(M\times[0,T])} \le C_T.$$

Continuity of the flow with respect to its initial data. We now state the continuity of the flow with respect to its initial data, which will be useful for the proof of Theorem 0.3 (see Section 2). The proof is standard and we omit it.

Proposition 1.5. *Let $u, v \in C^{2+\alpha,1+\alpha/2}_{\mathrm{loc}}(M\times[0,+\infty))$, $\alpha \in (0,1)$ be solutions of*

$$\begin{cases} \dfrac{\partial}{\partial t} e^v = \Delta v - Q + \rho \dfrac{e^v}{\int_M e^v \, dV}, \\ v(x,0) = v_0(x), \end{cases}$$

and

$$\begin{cases} \dfrac{\partial}{\partial t} e^u = \Delta u - Q + \rho \dfrac{e^u}{\int_M e^u \, dV}, \\ u(x,0) = u_0(x), \end{cases}$$

where $u_0, v_0 \in C^{2+\alpha}(M)$. Then for all $T > 0$, there exists a constant $C_T > 0$, depending on $\|u_0\|_{C^{2+\alpha}(M)}$, $\|v_0\|_{C^{2+\alpha}(M)}$ and T, such that

$$\|u - v\|_{C^{2+\alpha,1+\alpha/2}(M\times[0,T])} \le C_T \|u_0 - v_0\|_{C^{2+\alpha}(M)}. \tag{1-40}$$

Remark 1.6. One can also prove that, for all $T > 0$ fixed, if $\|u_0\|_{C^{2+\alpha}(M)} \le K_1$ and $\|v_0\|_{C^{2+\alpha}(M)} \le K_2$ for some constants $K_1, K_2 > 0$, then there exists a constant $C_T > 0$ depending on K_1, K_2 and T such that

$$\|u - v\|_{C^{2+\alpha,1+\alpha/2}(M\times[0,T])} \le C_T \|u_0 - v_0\|_{C^{2+\alpha}(M)}.$$

2. Convergence of the flow

This section is devoted to the proof of Theorems 0.2, 0.3 and 0.4.

Proof of Theorem 0.2. Let $v : M \times [0, +\infty) \to \mathbb{R}$ be the global solution of (0-4). Throughout this subsection, we assume without loss of generality that $\int_M e^{v(t)}\, dV = 1$ for all $t \geq 0$. C will denote constants not depending on t.

In order to prove Theorem 0.2, we need to bound $\|v(t)\|_{H^2(M)}$, $t \geq 0$ uniformly in time. For this, we first bound $\|v(t)\|_{H^1(M)}$, $t \geq 0$ uniformly in time. To bound $\|v(t)\|_{H^1(M)}$, we use the compactness result of Theorem 0.5. More precisely, using Theorem 0.5, we first prove that there exists a sequence $(t_n)_n$ with $\lim_{n\to+\infty} t_n = +\infty$ such that

$$\|v(t_n)\|_{H^2(M)} \leq C, \quad \forall n \geq 0.$$

Therefore we aim to prove that there exists a sequence $(t_n)_n$, $\lim_{n\to+\infty} t_n = +\infty$, such that, setting $v_n = v(t_n)$ and $h_n = -(\partial v/\partial t)(t_n)$, the sequence $(v_n)_n \subseteq H^2(M)$ satisfies conditions (0-8) of Theorem 0.5. First, we show that there exists a sequence $(t_n)_n$, $\lim_{n\to+\infty} t_n = +\infty$, such that (0-8)(i) is satisfied for $v_n = v(t_n)$. Recall that for all $T > 0$,

$$\int_0^T \int_M \Big(\frac{\partial v(t)}{\partial t}\Big)^2 e^{v(t)}\, dV dt = J_\rho(v(0)) - J_\rho(v(T)).$$

Using hypothesis (2-5), we deduce that there exists a sequence $(t_n)_n$ such that $n \leq t_n \leq n+1$, for all $n \in \mathbb{N}$, and

$$\lim_{n\to+\infty} \int_M \Big|\frac{\partial v(t_n)}{\partial t}\Big|^2 e^{v(t_n)}\, dV = 0. \tag{2-1}$$

The next proposition shows that condition (0-8)(ii) of Theorem 0.5 is satisfied.

Proposition 2.1. *We have*

$$-\frac{\partial e^{v(x,t)}}{\partial t} + \rho e^{v(x,t)} \geq -C, \quad \forall t \geq 0, \quad \forall x \in M. \tag{2-2}$$

Proof. Set

$$R(x,t) = e^{-v(x,t)}(-\Delta v(x,t) + Q(x)).$$

We can rewrite equation (0-4), satisfied by v, in the form

$$\frac{\partial v(x,t)}{\partial t} = -(R(x,t) - \rho).$$

Hence

$$\frac{\partial R(x,t)}{\partial t} = R(x,t)(R(x,t) - \rho) + e^{-v(x,t)} \Delta R(x,t).$$

Define $R_{\min}(t) = \min_{x\in M} R(x,t)$. Using the maximum principle (as in (1-8), we may assume that $R_{\min}(t)$ is differentiable), we find

$$\frac{\partial R_{\min}(t)}{\partial t} \geq -\rho R_{\min}(t).$$

Integrate between 0 and t to obtain

$$R_{\min}(t) \geq e^{-\rho t} R_{\min}(0).$$

This implies that

$$-\frac{\partial e^{v(x,t)}}{\partial t} + \rho e^{v(x,t)} \geq -|R_{\min}(0)| e^{-\rho t + v(x,t)}. \tag{2-3}$$

Set $v_{\max}(t) = \max_{x \in M} v(x,t)$. By the maximum principle, we have

$$\frac{\partial}{\partial t} e^{v_{\max}(t)} \leq \rho\Big(\frac{1}{\rho}\|Q\|_{L^\infty(M)} + e^{v_{\max}(t)}\Big).$$

Integrating again between 0 and t, we get

$$e^{v_{\max}(t)-\rho t} \leq e^{v_{\max}(0)} + \frac{1}{\rho}\|Q\|_{L^\infty(M)} - \frac{1}{\rho}\|Q\|_{L^\infty(M)} e^{-\rho t} \leq C. \tag{2-4}$$

Combining (2-3) and (2-4), we finally conclude

$$-\frac{\partial e^{v(x,t)}}{\partial t} + \rho e^{v(x,t)} \geq -C|R_{\min}(0)| \geq -C. \qquad \square$$

We are now in position to bound $\|v(t)\|_{H^1(M)}$, $t \geq 0$, uniformly in time.

Proposition 2.2. *Let $\rho \in (8k\pi, 8(k+1)\pi)$, $k \in \mathbb{N}^*$ and $v(t) : M \to \mathbb{R}$ be the solution of* (0-4). *Suppose that*

$$J_\rho(v(t)) \geq -C, \quad \forall t \geq 0. \tag{2-5}$$

Then there exists a constant $\widetilde{C}$, depending on M, Q, ρ, α and $\|v_0\|_{C^{2+\alpha}(M)}$ but not on T, such that

$$\|v(t)\|_{H^1(M)} \leq \widetilde{C}, \quad \forall t \geq 0. \tag{2-6}$$

Proof. Thanks to (2-1) and (2-2), from Theorem 0.5 there exists a constant $C > 0$ such that

$$\|v(t_n)\|_{H^2(M)} \leq C,$$

where $(t_n)_n$ is the sequence defined in (2-1). By Sobolev's embedding theorem, it follows that $\|v(t_n)\|_{C^\alpha(M)} \leq C$ for all $\alpha \in (0,1)$. Since $\lim_{n\to+\infty} t_n = +\infty$, for all sufficiently large $t \geq 0$ there exists $n \in \mathbb{N}$ such that $t_n \leq t \leq t_{n+1}$. Moreover, since $|t_{n+1} - t_n| \leq 2$, we have $|t - t_n| \leq 2$. We claim that for all $p > 1$,

$$\int_M e^{pv(t)}\, dV \leq C, \quad \forall t \geq 0. \tag{2-7}$$

Since $v(t)$ satisfies (0-4), integrating by parts and using Young's inequality, we see that

$$\begin{aligned}\frac{\partial}{\partial t}\int_M e^{pv(t)}\,dV\\ &= -p(p-1)\int_M |\nabla v(t)|^2 e^{(p-1)v(t)}\,dV - p\int_M Qe^{(p-1)v(t)}\,dV + p\frac{\rho}{a}\int_M e^{pv(t)}\,dV\\ &\le C\int_M e^{(p-1)v(t)}\,dV + p\frac{\rho}{a}\int_M e^{pv(t)}\,dV\\ &\le C + C\int_M e^{pv(t)}\,dV.\end{aligned}$$

Setting $y(t) = \int_M e^{pv(t)}\,dV$ and integrating the previous inequality between t_n and t, it follows that

$$y(t) \le e^{C(t-t_n)}y(t_n) + C\big(e^{C(t-t_n)} - 1\big).$$

Since $\|v(t_n)\|_{C^\alpha(M)} \le C$, $\alpha \in (0, 1)$, and $|t - t_n| \le 2$, we have that (2-7) is satisfied.

Fix $t \ge 0$ and set

$$M_\varepsilon = \{x \in M : e^{v(x,t)} < \varepsilon\},$$

where $\varepsilon > 0$ is a real number which will be determined shortly. We have

$$\begin{aligned}1 = \int_M e^{v(t)}\,dV &= \int_{M_\varepsilon} e^{v(t)}\,dV + \int_{M\backslash M_\varepsilon} e^{v(t)}\,dV\\ &\le \varepsilon|M_\varepsilon| + |M\backslash M_\varepsilon|^{1-1/p}\biggl(\int_M e^{pv(t)}\,dV\biggr)^{1/p}.\end{aligned} \tag{2-8}$$

Thus, taking $\varepsilon = \dfrac{1}{2|M|}$, (2-7) implies

$$\tfrac{1}{2} \le C|M\backslash M_\varepsilon|^{1-1/p}.$$

Since $p > 1$, we get

$$|M\backslash M_\varepsilon| \ge \Big(\frac{1}{2C}\Big)^{p/(p-1)} > 0. \tag{2-9}$$

Set $A = M\backslash M_\varepsilon$, so that

$$\int_A v(t)\,dV \ge \ln\Big(\frac{1}{2|M|}\Big)|A|. \tag{2-10}$$

On the other hand, we have

$$\int_A v(t)\,dV \le \int_A e^{v(t)}\,dV \le 1.$$

From this inequality and (2-10), we deduce that there exists a constant C such that

$$\left|\int_A v(t)\,dV\right| \le C. \tag{2-11}$$

Arguing the same way as in Proposition 1.2, (2-9) and (2-11) imply that there exists a constant $\widetilde{C}$ not depending on t such that, for all $t \ge 0$,

$$\|v(t)\|_{H^1(M)} \le \widetilde{C}.$$ □

Proof of Theorem 0.2. First, we prove that

$$\int_M (\Delta v(t))^2\,dV \le C, \quad \forall t \ge 0 \tag{2-12}$$

following the arguments of Brendle [2003]. Set

$$V(t) = \frac{\partial v(t)}{\partial t}$$

and

$$y(t) = \int_M V^2(t) e^{v(t)}\,dV.$$

We claim that $\lim_{t\to+\infty} y(t) = 0$. By (2-6), we have for all $T \ge 0$

$$\int_0^T \int_M \Big(\frac{\partial v(t)}{\partial t}\Big)^2 e^{v(t)}\,dV dt \le J_\rho(v(0)) - J_\rho(v(T)) \le C, \tag{2-13}$$

where C is a constant not depending on T. Let ε be some positive real number. From (2-13), we deduce that there exists $t_0 \ge 0$ such that $y(t_0) \le \varepsilon$.

We want to prove that

$$y(t) \le 3\varepsilon, \quad \forall t \ge t_0.$$

Otherwise, define

$$t_1 = \inf\{t \ge t_0 : y(t) \ge 3\varepsilon\} < +\infty.$$

This implies that

$$y(t) \le 3\varepsilon, \quad \forall t_0 \le t \le t_1. \tag{2-14}$$

Since $\dfrac{\partial v(t)}{\partial t} = e^{-v(t)}(\Delta v(t) - Q) + \rho$, using (2-14) we arrive at

$$\int_M e^{-v(t)}(\Delta v(t) - Q)^2\,dV = y(t) + \rho^2 \le C_1, \quad \forall t_0 \le t \le t_1, \tag{2-15}$$

where C_1 denotes a constant depending on ε, and thus on t_1. From (2-7), we have for all $t \ge 0$

$$\int_M e^{3v(t)}\,dV \le C, \tag{2-16}$$

with C independent of t_1. Using Hölder's inequality, (2-15) and (2-16), we obtain for all $t_0 \le t \le t_1$

$$\int_M |\Delta v(t) - Q|^{3/2}\, dV \le \left(\int_M e^{-v(t)}(\Delta v(t) - Q)^2\, dV\right)^{3/4}\left(\int_M e^{3v(t)}\, dV\right)^{1/4} \le C_1.$$

Thus, $\int_M |\Delta v(t)|^{3/2}\, dV \le C_1$ for all $t_0 \le t \le t_1$. From Sobolev's embedding theorem, we get

$$|v(t)| \le C_1, \quad \forall t_0 \le t \le t_1. \tag{2-17}$$

On the other hand, we see that $V(t) = \partial v(t)/\partial t$ satisfies

$$\frac{\partial V(t)}{\partial t} = -V(t)e^{-v(t)}\Delta v(t) + e^{-v(t)}\Delta V(t) + QV(t)e^{-v(t)}. \tag{2-18}$$

Now, using (2-18), we have for all $t_0 \le t \le t_1$

$$\begin{aligned}\frac{\partial y(t)}{\partial t} &= \frac{\partial}{\partial t}\left(\int_M V^2(t)e^{v(t)}\, dV\right)\\ &= 2\int_M V(t)e^{v(t)}\big(e^{-v(t)}\Delta V(t) - V(t)e^{-v(t)}\Delta v(t) + QV(t)e^{-v(t)}\big)\, dV\\ &\qquad\qquad + \int_M V^3(t)e^{v(t)}\, dV\end{aligned}$$

Integrating by parts, we obtain

$$\frac{\partial y(t)}{\partial t} = -2\int_M |\nabla V(t)|^2\, dV - \int_M V^3(t)e^{v(t)}\, dV + 2\rho\int_M V^2(t)e^{v(t)}\, dV. \tag{2-19}$$

The Gagliardo–Nirenberg inequality now gives

$$\|V(t)\|_{L^3_{g_1}(M)} \le C\|V(t)\|^{2/3}_{L^2_{g_1}(M)}\|V(t)\|^{1/3}_{H^1_{g_1}(M)},$$

where the norms are taken with respect to the metric $g_1(t) = e^{v(t)}g$. From (2-17), notice that the first eigenvalue of the Laplacian $\tilde\lambda_1(t)$ with respect to the metric $g_1(t)$ satisfies, for all $t_0 \le t \le t_1$,

$$\tilde\lambda_1(t) \ge C_1. \tag{2-20}$$

Combining $\int_M Ve^v\, dV = 0$, Poincaré's inequality and (2-20), we have

$$\int_M e^v|V|^3\, dV \le C_1\left(\int_M V^2e^v\, dV\right)\left(\int_M |\nabla V|^2\, dV\right)^{1/2}. \tag{2-21}$$

Thus we obtain, from (2-19), (2-21) and Young's inequality,

$$\frac{\partial}{\partial t}\left(\int_M V^2e^v\, dV\right) \le C_1\left(\int_M V^2e^v\, dV\right)^2 + C\left(\int_M V^2e^v\, dV\right),$$

i.e.,

$$\frac{\partial}{\partial t} y(t) \le C_1 y^2(t) + C y(t).$$

Since $y(t_0) \le \varepsilon$ and $y(t_1) = 3\varepsilon$, we find

$$2\varepsilon \le y(t_1) - y(t_0) \le (C_1 + C) \int_{t_0}^{t_1} y(t)\, dt.$$

Choosing t_0 large enough, we have $(C_1 + C) \int_{t_0}^{+\infty} y(t)\, dt \le \varepsilon$, and thus we obtain a contradiction. We conclude that

$$y(t) \underset{t \to +\infty}{\longrightarrow} 0,$$

and thereby find $t_1 = +\infty$. This implies that all previous estimates hold for all $t \ge 0$. Thus, for all $t \ge 0$ we have $|v(t)| \le C$ and

$$\int_M e^{-v(t)} (\Delta v(t) - Q)^2\, dV \le C.$$

It follows that, for all $t \ge 0$, $\int_M (\Delta v(t))^2\, dV \le C$.

Thus, using (2-6), for all $t \ge 0$ we have $\|v(t)\|_{H^2(M)} \le C$. Therefore, there exist a function $v_\infty \in H^2(M)$ and a sequence $(t_n)_n$ with $\lim_{n \to +\infty} t_n = +\infty$ such that

$$v(t_n) \underset{n \to +\infty}{\longrightarrow} v_\infty \text{ weakly in } H^2(M)$$

and

$$v(t_n) \underset{n \to +\infty}{\longrightarrow} v_\infty \text{ in } C^\alpha(M), \alpha \in (0, 1).$$

It is easy to check that v_∞ is a solution to

$$-\Delta v_\infty + Q = \rho \frac{e^{v_\infty}}{\int_M e^{v_\infty}\, dV},$$

and, by bootstrap regularity arguments, we have $v_\infty \in C^\infty(M)$. To obtain that $\|v(t_n) - v_\infty\|_{H^2(M)} \underset{n \to +\infty}{\longrightarrow} 0$, notice that

$$\begin{aligned}
\int_M (\Delta v(t_n) - \Delta v_\infty)^2\, dV \\
&= \int_M \left(\frac{\rho}{a} \left(e^{v_\infty} - e^{v(t_n)} \right) + \frac{\partial e^{v(t_n)}}{\partial t} \right)^2 dV \\
&\le C \int_M \left(e^{v_\infty} - e^{v(t_n)} \right)^2 dV + C \int_M \left| \frac{\partial v}{\partial t}(t_n) \right|^2 e^{v(t_n)}\, dV \underset{n \to +\infty}{\longrightarrow} 0.
\end{aligned}$$

Since the flow is a gradient flow for the functional J_ρ, which is real analytic, from a general result of Simon [1983] we finally obtain that

$$\|v(t) - v_\infty\|_{H^2(M)} \underset{n \to +\infty}{\longrightarrow} 0. \qquad \square$$

Proof of Theorem 0.3. We prove the existence of an initial data $v_0 \in C^\infty(M)$ for the flow (0-4) such that the functional $J_\rho(v(t))$, $t \geq 0$, is uniformly bounded from below. From standard parabolic theory, it is easy to see that for $v_0 \in C^\infty(M)$, the solution v of (0-4) belongs to $C^\infty(M \times [0, +\infty))$.

Let X be the space of functions $C^\infty(M)$ endowed with the norm $\|\cdot\|_{C^{2+\alpha}(M)}$, and define

$$\Phi : X \times [0, +\infty) \longrightarrow C^\infty(M \times [0, +\infty))$$

by letting $\Phi(v, t)$ be a solution of

$$\begin{cases} \dfrac{\partial \Phi(v,t)}{\partial t} = e^{-\Phi(v,t)} \Delta \Phi(v,t) - e^{-\Phi(v,t)} Q + \dfrac{\rho}{\int_M e^{\Phi(v,t)}\, dV}, \\ \Phi(v, 0) = v. \end{cases}$$

Suppose that for all $v \in X$, we have

$$J_\rho(\Phi(v,t)) \underset{t \to +\infty}{\longrightarrow} -\infty. \tag{2-22}$$

Let $L > 0$. Following the same arguments as in [Malchiodi 2008], one can show that there exists $L_1 > 0$ such that $\{v \in X : J_\rho(v) \leq -L_1\}$ is not contractible. However, we prove that if (2-22) is satisfied then $\{v \in X : J_\rho(v) \leq -L\}$ is contractible. We proceed in two steps.

Step 1. Let $L > 0$ be fixed and

$$T_v = \inf\{t \geq 0 : J_\rho(\Phi(v,t)) \leq -L\},$$

then the function $T : C^{2+\alpha}(M) \to \mathbb{R}$, $v \mapsto T_v$ is continuous.

Proof of Step 1. From (2-22), we have

$$\{t \geq 0 : J_\rho(\Phi(v,t)) \leq -L\} \neq \varnothing,$$

and from the uniqueness of solutions of (0-4) having the same initial data, one can prove that $J_\rho(\Phi(v,t))$ is strictly decreasing on $[0, +\infty)$. Let $\bar{v} \in C^\infty(M)$ and $(v_n)_n \in C^\infty(M)$ be a sequence such that $\lim_{n\to+\infty} v_n = \bar{v}$ in $C^{2+\alpha}(M)$. We claim that $\lim_{n\to+\infty} T_{v,n} = T_{\bar{v}}$. To prove this, we consider two cases depending on the value of $J_\rho(\bar{v})$.

First case. Suppose that $J_\rho(\bar{v}) < -L$. Since the function $t \to J_\rho(\Phi(\bar{v},t))$ is decreasing, we have $J_\rho(\Phi(\bar{v},t)) < -L$ for all $t \geq 0$. We deduce that $T_{\bar{v}} = 0$. Since $\lim_{n\to+\infty} v_n = \bar{v}$ in $C^{2+\alpha}(M)$, it is easy to see that

$$J_\rho(v_n) \underset{n \to +\infty}{\longrightarrow} J_\rho(\bar{v}).$$

Thus, there exists $n_0 \in \mathbb{N}$ such that $J_\rho(v_n) \le -L$ for all $n \ge n_0$. So, we obtain that $T_{v_n} = 0 = T_{\bar v}$ for all $n \ge n_0$. This implies that

$$T_{v_n} \underset{n\to+\infty}{\longrightarrow} T_{\bar v}.$$

Second case. Suppose that $J_\rho(\bar v) \ge -L$. In this case, $T_{\bar v}$ verifies $J_\rho(\Phi(\bar v, T_{\bar v})) = -L$. Setting $T_n := T_{v_n}$ and supposing that T_n does not converge to $T_{\bar v}$, then, up to extracting a subsequence, there exists $\varepsilon_0 > 0$ such that $|T_n - T_{\bar v}| \ge \varepsilon_0$. So we have $T_n \ge \varepsilon_0 + T_{\bar v}$ or $T_n \le -\varepsilon_0 + T_{\bar v}$. Suppose, without loss of generality, that

$$T_n \ge \varepsilon_0 + T_{\bar v}. \tag{2-23}$$

Set $T = T_{\bar v} + \varepsilon_0 + 1$. Since $\lim_{n\to+\infty} v_n = \bar v$ in $C^{2+\alpha}(M)$ by Proposition 1.5, it is easy to see that

$$J_\rho(\Phi(v_n, t)) \underset{n\to+\infty}{\longrightarrow} J_\rho(\Phi(\bar v, t)), \tag{2-24}$$

for all t fixed in $[0, T]$. Since $t \to J_\rho(\Phi(\bar v, t))$ is strictly decreasing, we have

$$\alpha_1 = J_\rho(\Phi(\bar v, T_{\bar v})) - J_\rho(\Phi(\bar v, T_{\bar v} + \varepsilon_0)) > 0.$$

From (2-24), since $T_{\bar v} + \varepsilon_0 \in [0, T]$, we get

$$J_\rho(\Phi(v_n, T_{\bar v} + \varepsilon_0)) \underset{n\to+\infty}{\longrightarrow} J_\rho(\Phi(\bar v, T_{\bar v} + \varepsilon_0)) = -L - \alpha_1,$$

and from (2-23),

$$J_\rho(\Phi(v_n, T_n)) \le J_\rho(\Phi(v_n, T_{\bar v} + \varepsilon_0)).$$

This implies that, if n tends to $+\infty$, $-L \le -L - \alpha_1$. Thus we obtain a contradiction.

Step 2. If (2-22) holds, then the set $\{v \in X : J_\rho(v) \le -L\}$ is contractible.

Proof of Step 2. We construct a deformation retract from $\{v \in X\}$ into $\{v \in X : J_\rho(v) \le -L\}$. Since $\{v \in X\}$ is contractible, $\{v \in X : J_\rho(v) \le -L\}$ must also be contractible. We denote by h the one-to-one function defined by

$$h(t) : [0, 1) \to [0, +\infty), \quad t \mapsto \frac{t}{1-t},$$

and by $\eta(v, t) : X \times [0, 1] \to X$ the function defined by

$$\eta(v, t) = \begin{cases} \Phi(v, h(t)) & \text{if } h(t) \le T_v, \\ \Phi(v, T_v) & \text{if } h(t) \ge T_v. \end{cases}$$

First we prove that $\eta = \Phi \circ \Phi_1 : X \times [0, 1) \to X$ is continuous, in which $\Phi_1 : X \times [0, 1) \to X \times [0, +\infty)$ is the function defined by

$$\Phi_1(v,t)=\begin{cases}(v,h(t)) & \text{if } h(t)\le T_v,\\ (v,T_v) & \text{if } h(t)\ge T_v.\end{cases}$$

From Step 1, $\Phi_1 : X\times[0,1)\to X\times[0,+\infty)$ is a continuous function. Therefore, to prove that η is a continuous map from $X\times[0,1)\to X$, it is sufficient to prove that, for $T>0$ fixed, $\Phi : X\times[0,T]\to X$ is continuous.

Let $(v_n,t_n)\in C^\infty(M)\times[0,T]$ be such that $\lim_{n\to+\infty} v_n=v$ in $C^{2+\alpha}(M)$, where $v\in C^\infty(M)$ and $\lim_{n\to+\infty} t_n=t\in[0,T]$. Then we have

$$\begin{aligned}\text{(2-25)}\quad &\|\Phi(v_n,t_n)-\Phi(v,t)\|_{C^{2+\alpha}(M)}\\ &\qquad\le \|\Phi(v_n,t_n)-\Phi(v_n,t)\|_{C^{2+\alpha}(M)}+\|\Phi(v_n,t)-\Phi(v,t)\|_{C^{2+\alpha}(M)}.\end{aligned}$$

Since $\Phi(v_n,\cdot)\in C^\infty(M\times[0,T])$, Theorem 0.1 implies that for all $t\in[0,T]$,

$$\left\|\frac{\partial\Phi(v_n,t)}{\partial t}\right\|_{C^{2+\alpha}(M)}\le C_T,$$

where C_T denotes a constant not depending on n. We deduce that

$$\begin{aligned}\text{(2-26)}\quad &\|\Phi(v_n,t_n)-\Phi(v_n,t)\|_{C^{2+\alpha}(M)}\\ &\qquad=\left\|\int_{t_n}^{t}\frac{\partial\Phi(v_n,s)}{\partial s}\,ds\right\|_{C^{2+\alpha}(M)}\\ &\qquad\le |t_n-t|\max_{s\in[t_n,t]}\left\|\frac{\partial\Phi(v_n,s)}{\partial s}\right\|_{C^{2+\alpha}(M)}\underset{n\to+\infty}{\longrightarrow}0.\end{aligned}$$

On the other hand, using Proposition 1.5, we have for all $t\in[0,T]$

$$\text{(2-27)}\qquad \|\Phi(v_n,t)-\Phi(v,t)\|_{C^{2+\alpha}(M)}\le C_T\|v_n-v\|_{C^{2+\alpha}(M)}\underset{n\to+\infty}{\longrightarrow}0.$$

Combining (2-25) , (2-26) and (2-27), we find that

$$\|\Phi(v_n,t_n)-\Phi(v,t)\|_{C^{2+\alpha}(M)}\underset{n\to+\infty}{\longrightarrow}0.$$

Thus η is continuous from $X\times[0,1)\to X$. It remains to prove that it is continuous on $X\times[0,1]$. Let $(v_n,t_n)\in C^\infty(M)\times[0,1]$ be such that $\lim_{n\to+\infty} v_n=\bar v$ in $C^{2+\alpha}(M)$, where $\bar v\in C^\infty(M)$, and $\lim_{n\to+\infty} t_n=1$. From Step 1, we have

$$T_{v_n}=T_n\underset{n\to+\infty}{\longrightarrow}T_{\bar v}.$$

Since T_n is finite and $\lim_{n\to+\infty} t_n=1$, it follows that $\lim_{n\to+\infty} h(t_n)=+\infty$. So, for sufficiently large n, $h(t_n)\ge T_n$ and thus $\eta(v_n,t_n)=\Phi(v_n,T_n)$. We have, in the same way as (2-26) and (2-27), that

$$\begin{aligned}&\|\eta(v_n,t_n)-\eta(\bar v,1)\|_{C^{2+\alpha}(M)}\\ &\quad=\|\Phi(v_n,T_n)-\Phi(\bar v,T_{\bar v})\|_{C^{2+\alpha}(M)}\\ &\quad\le\|\Phi(v_n,T_n)-\Phi(\bar v,T_n)\|_{C^{2+\alpha}(M)}+\|\Phi(\bar v,T_n)-\Phi(\bar v,T_{\bar v})\|_{C^{2+\alpha}(M)}\underset{n\to+\infty}{\longrightarrow}0.\end{aligned}$$

Therefore η is continuous from $X \times [0, 1] \to X$.

Now it is easy to check that η is a deformation retract from X into the set $\{v \in X : J_\rho(v) \leq -L\}$. Hence this set is contractible. □

Nonconvergence of the flow: proof of Theorem 0.4. To prove Theorem 0.4, it is sufficient to prove that there exists a real number $C > 0$ depending on M, Q and ρ such that, for all $v_0 \in C^{2+\alpha}(M)$ satisfying $J_\rho(v_0) \leq -C$, the solution $v(t)$ of the flow (0-4), with $v(x, 0) = v_0(x)$ for all $x \in M$, satisfies

$$J_\rho(v(t)) \underset{t \to +\infty}{\longrightarrow} -\infty.$$

We recall (see [Li 1999]) that there exists a constant $C_0 \geq 0$ depending on M, Q and ρ such that

$$\|w\|_{C^{2+\alpha}(M)} \leq C_0 \tag{2-28}$$

for any solution $w \in C^{2+\alpha}(M)$, $\alpha \in (0, 1)$, of

$$-\Delta w + Q = \frac{\rho e^w}{\int_M e^w \, dV}.$$

Since $J_\rho(v(t))$ is decreasing, if $\lim_{t \to +\infty} J_\rho(v(t)) \neq -\infty$ then there exists $L \in \mathbb{R}$ such that

$$J_\rho(v(t)) \geq L, \quad \forall t \in [0, +\infty).$$

From Theorem 0.2, there is a function $v_\infty \in C^\infty(M)$ such that

$$\|v(t) - v_\infty\|_{H^2(M)} \underset{t \to +\infty}{\longrightarrow} 0$$

which is a solution of

$$-\Delta v_\infty + Q = \frac{\rho e^{v_\infty}}{\int_M e^{v_\infty} \, dV}. \tag{2-29}$$

It follows that

$$\|v_\infty\|_{C^{2+\alpha}(M)} \leq C_0,$$

where C_0 is the constant defined in (2-28). This implies that there exists a constant $\overline{C}$ depending on M, Q, ρ and C_0 such that

$$J(w_\infty) \geq -\overline{C}.$$

Since $J_\rho(v(t_1)) \leq J_\rho(v(t_2))$, for all $t_1 \geq t_2$, we have

$$J_\rho(v_0) \geq J_\rho(v_\infty) \geq -\overline{C}.$$

However, $J_\rho(v_0) \leq -C$ by hypothesis. Therefore, by choosing $C > \overline{C}$, we get a contradiction. □

References

[Baird et al. 2004] P. Baird, A. Fardoun, and R. Regbaoui, "The evolution of the scalar curvature of a surface to a prescribed function", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **3**:1 (2004), 17–38. MR 2005e:53100 Zbl 1170.58306

[Brendle 2003] S. Brendle, "Global existence and convergence for a higher order flow in conformal geometry", *Ann. of Math.* (2) **158**:1 (2003), 323–343. MR 2004e:53098 Zbl 1042.53016

[Brendle 2005] S. Brendle, "Convergence of the Yamabe flow for arbitrary initial energy", *J. Differential Geom.* **69**:2 (2005), 217–278. MR 2006e:53119 Zbl 1085.53028

[Brendle 2006] S. Brendle, "Convergence of the Q-curvature flow on S^4", *Adv. Math.* **205**:1 (2006), 1–32. MR 2007f:53075 Zbl 1104.53064

[Brezis and Merle 1991] H. Brezis and F. Merle, "Uniform estimates and blow-up behavior for solutions of $-\Delta u = V(x)e^u$ in two dimensions", *Comm. Partial Differential Equations* **16**:8-9 (1991), 1223–1253. MR 92m:35084 Zbl 0746.35006

[Brouttelande 2003] C. Brouttelande, "The best-constant problem for a family of Gagliardo–Nirenberg inequalities on a compact Riemannian manifold", *Proc. Edinb. Math. Soc.* (2) **46**:1 (2003), 117–146. MR 2004b:58025 Zbl 1031.58009

[Caffarelli and Yang 1995] L. A. Caffarelli and Y. S. Yang, "Vortex condensation in the Chern–Simons Higgs model: An existence theorem", *Comm. Math. Phys.* **168**:2 (1995), 321–336. MR 96b:81076 Zbl 0846.58063

[Caglioti et al. 1992] E. Caglioti, P.-L. Lions, C. Marchioro, and M. Pulvirenti, "A special class of stationary flows for two-dimensional Euler equations: A statistical mechanics description", *Comm. Math. Phys.* **143**:3 (1992), 501–525. MR 93i:82060 Zbl 0745.76001

[Castéras 2013] J.-b. Castéras, "Equivariant mean field flow", *J. Geom. Phys.* **74** (2013), 314–327. MR 3118588 Zbl 1282.35217

[Castéras 2015] J.-B. Castéras, "A mean field type flow part I: Compactness of solutions to a perturbed mean field type equation", *Calc. Var. Partial Differential Equations* **53**:1-2 (2015), 221–246. MR 3336318

[Chen and Lin 2003] C.-C. Chen and C.-S. Lin, "Topological degree for a mean field equation on Riemann surfaces", *Comm. Pure Appl. Math.* **56**:12 (2003), 1667–1727. MR 2004h:35065 Zbl 1032.58010

[Ding et al. 1999] W. Ding, J. Jost, J. Li, and G. Wang, "Existence results for mean field equations", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **16**:5 (1999), 653–666. MR 2000i:35061 Zbl 0937.35055

[Djadli 2008] Z. Djadli, "Existence result for the mean field problem on Riemann surfaces of all genuses", *Commun. Contemp. Math.* **10**:2 (2008), 205–220. MR 2009c:35115 Zbl 1151.53035

[Djadli and Malchiodi 2008] Z. Djadli and A. Malchiodi, "Existence of conformal metrics with constant Q-curvature", *Ann. of Math.* (2) **168**:3 (2008), 813–858. MR 2009h:53074 Zbl 1186.53050

[Friedman 1964] A. Friedman, *Partial differential equations of parabolic type*, Prentice-Hall, Englewood Cliffs, NJ, 1964. MR 31 #6062 Zbl 0144.34903

[Hamilton 1988] R. S. Hamilton, "The Ricci flow on surfaces", pp. 237–262 in *Mathematics and general relativity* (Santa Cruz, CA, 1986), edited by J. A. Isenberg, Contemp. Math. **71**, Amer. Math. Soc., Providence, RI, 1988. MR 89i:53029 Zbl 0663.53031

[Han 2003] J. Han, "Asymptotic limit for condensate solutions in the abelian Chern–Simons Higgs model", *Proc. Amer. Math. Soc.* **131**:6 (2003), 1839–1845. MR 2004b:58022 Zbl 1036.35034

[Hebey 1997] E. Hebey, *Introduction à l'analyse non linéaire sur les variétés*, Diderot Editeur, Paris, 1997. Zbl 0918.58001

[Lamm et al. 2009] T. Lamm, F. Robert, and M. Struwe, "The heat flow with a critical exponential nonlinearity", *J. Funct. Anal.* **257**:9 (2009), 2951–2998. MR 2011c:35284 Zbl 05618762

[Li 1999] Y. Y. Li, "Harnack type inequality: The method of moving planes", *Comm. Math. Phys.* **200**:2 (1999), 421–444. MR 2000c:58024 Zbl 0928.35057

[Li and Shafrir 1994] Y. Y. Li and I. Shafrir, "Blow-up analysis for solutions of $-\Delta u = Ve^u$ in dimension two", *Indiana Univ. Math. J.* **43**:4 (1994), 1255–1270. MR 96c:35048 Zbl 0842.35011

[Malchiodi 2008] A. Malchiodi, "Morse theory and a scalar field equation on compact surfaces", *Adv. Differential Equations* **13**:11-12 (2008), 1109–1129. MR 2009m:53095 Zbl 1175.53052

[Malchiodi and Struwe 2006] A. Malchiodi and M. Struwe, "Q-curvature flow on S^4", *J. Differential Geom.* **73**:1 (2006), 1–44. MR 2007c:53087 Zbl 1099.53034

[Moser 1970/71] J. Moser, "A sharp form of an inequality by N. Trudinger", *Indiana Univ. Math. J.* **20** (1970/71), 1077–1092. MR 46 #662 Zbl 0213.13001

[Schwetlick and Struwe 2003] H. Schwetlick and M. Struwe, "Convergence of the Yamabe flow for "large" energies", *J. Reine Angew. Math.* **562** (2003), 59–100. MR 2004h:53097 Zbl 1079.53100

[Simon 1983] L. Simon, "Asymptotics for a class of nonlinear evolution equations, with applications to geometric problems", *Ann. of Math.* (2) **118**:3 (1983), 525–571. MR 85b:58121 Zbl 0549.35071

[Struwe 2002] M. Struwe, "Curvature flows on surfaces", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **1**:2 (2002), 247–274. MR 2004f:53083 Zbl 1150.53025

[Struwe 2005] M. Struwe, "A flow approach to Nirenberg's problem", *Duke Math. J.* **128**:1 (2005), 19–64. MR 2005m:53125 Zbl 1087.53034

[Struwe and Tarantello 1998] M. Struwe and G. Tarantello, "On multivortex solutions in Chern–Simons gauge theory", *Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat.* (8) **1**:1 (1998), 109–121. MR 99c:58041 Zbl 0912.58046

[Tarantello 1996] G. Tarantello, "Multiple condensate solutions for the Chern–Simons–Higgs theory", *J. Math. Phys.* **37**:8 (1996), 3769–3796. MR 97f:58045 Zbl 0863.58081

[Trudinger 1967] N. S. Trudinger, "On imbeddings into Orlicz spaces and some applications", *J. Math. Mech.* **17** (1967), 473–483. MR 35 #7121 Zbl 0163.36402

[Yang 2001] Y. Yang, *Solitons in field theory and nonlinear analysis*, Springer, New York, 2001. MR 2002m:58001 Zbl 0982.35003

Received June 14, 2014. Revised December 2, 2014.

JEAN-BAPTISTE CASTÉRAS
UFRGS, INSTITUTO DE MATEMÁTICA
AV. BENTO GONCALVES 9500
91540-000 PORTO ALEGRE-RS
BRAZIL

jean-baptiste.casteras@univ-brest.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.347

ISOMETRIC EMBEDDING OF NEGATIVELY CURVED COMPLETE SURFACES IN LORENTZ–MINKOWSKI SPACE

BING-LONG CHEN AND LE YIN

The Hilbert–Efimov theorem states that any complete surface with curvature bounded above by a negative constant cannot be isometrically embedded in $\mathbb{R}^3$. We demonstrate that any simply connected smooth complete surface with curvature bounded above by a negative constant admits a smooth isometric embedding into the Lorentz–Minkowski space $\mathbb{R}^{2,1}$.

1. Introduction

Weyl [1916] posed the problem whether every abstract compact smooth simply connected 2-dimensional Riemannian manifold with positive curvature can be isometrically embedded in $\mathbb{R}^3$. Weyl's problem was investigated by Weyl, Lewy, Alexandrov, and others, and finally resolved (in the smooth category) by Nirenberg [1953] and Pogorelov [1952] independently. The generalization to the nonnegative curvature case was done by Guan and Li [1994] and Hong and Zuily [1995], though only $C^{1,1}$ embedding was obtained.

For noncompact convex surfaces, the problem was solved by Olovjanisnikov in the 1940s in the weak sense (see the survey article [Hong 2001]), and in smooth category by Pogorelov [1973]. The result has been generalized to the nonnegatively curved case in [Hong 1997]. For local isometric embeddings, there were important breakthroughs by C.-S. Lin, Q. Han and J.-X. Hong [Lin 1985; 1986; Han et al. 2003; Han 2005; 2006] (see also the survey articles [Hong 2001; Yau 2000] and the book [Han and Hong 2006]).

The story is completely different for surfaces with negative curvature. The famous Hilbert–Efimov theorem [Hilbert 1901; Efimov 1964] asserts that any complete surface with curvature bounded above by a negative constant cannot be realized in $\mathbb{R}^3$. If the complete surface is negatively but not strongly negatively curved, Hong [1993] found a sufficient and almost sharp condition for the existence of an isometric embedding in $\mathbb{R}^3$.

The authors were partially supported by grants NSFC11025107 and NSFC11101289.
MSC2010: primary 53C21; secondary 35J96.
Keywords: isometric embedding, Hilbert–Efimov theorem, Lorentz–Minkowski space.

On the other hand, the hyperbolic plane $\mathbb{H}^2$ admits a canonical smooth isometric embedding in the 3-dimensional Lorentz–Minkowski space $\mathbb{R}^{2,1}$ as a unit imaginary sphere $x_3^2 - (x_1^2 + x_2^2) = 1$. Here $\mathbb{R}^{2,1}$ denotes $\mathbb{R}^3$ equipped with the metric $ds^2 = dx_1^2 + dx_2^2 - dx_3^2$. Instead of the Euclidean space $\mathbb{R}^3$, it is proved that the Lorentz–Minkowski space $\mathbb{R}^{2,1}$ is the appropriate ambient space for the isometric embedding of strongly negatively curved surfaces.

We remark that the problem of isometric embedding of Riemannian manifolds into Lorentzian manifolds has been investigated by many authors. Schlenker [2001] found a Hilbert–Efimov type theorem in the anti-de Sitter space. A celebrated theorem of Schoen and Yau [1981] states that a 3-dimensional complete asymptotic flat Riemannian manifold (M^3, g, p) with dominant energy condition

$$R - |p_{ij}|^2 + (\mathrm{tr}_g\, p)^2 \geq 2|\nabla_j (p^{ij} - (\mathrm{tr}_g\, p) g^{ij})|_g \tag{1-1}$$

and zero ADM mass, where p_{ij} is a symmetric 2-tensor, admits an isometric embedding into $\mathbb{R}^{3,1}$ such that p_{ij} is the second fundamental form. Delanoë [1988] and Guan [2007] constructed smooth isometric embeddings of a negatively curved compact 2-disc $\bar{\mathcal{D}}$ with smooth boundary $\partial\mathcal{D}$ into Lorentz–Minkowski space $\mathbb{R}^{2,1}$.

The purpose of this paper is to find global isometric embeddings for complete negatively curved surfaces into $\mathbb{R}^{2,1}$. The main result is the following:

Theorem 1.1. *Let (M, g) be a smooth 2-dimensional simply connected complete Riemannian manifold with curvature K satisfying*

$$K \leq -C_1 \tag{1-2}$$

for some positive constant $C_1 > 0$. There exists a smooth isometric embedding $X : M \to \mathbb{R}^{2,1}$, and the spacelike submanifold $X(M)$ is a graph over $\mathbb{R}^2 \subset \mathbb{R}^{2,1}$, say $(x_1, x_2, 0) \mapsto (x_1, x_2, Z(x_1, x_2))$, satisfying

$$\sqrt{x_1^2 + x_2^2} \leq Z(x_1, x_2) \leq \sqrt{\frac{1}{C_1} + x_1^2 + x_2^2}. \tag{1-3}$$

Remark 1.2. It is likely that the solution of the isometric embedding problem is not unique (up to isometries of $\mathbb{R}^{2,1}$) if we drop the restriction (1-3). Actually, a remarkable fact (see [Guan et al. 2006]) is that there are many distinct isometric embeddings for the hyperbolic plane $\mathbb{H}^2$ into $\mathbb{R}^{2,1}$; some even have unbounded second fundamental forms and violate (1-3). In this sense, the natural isometric embedding provided by Theorem 1.1 is rather special. The construction and classification of all exotic embeddings deserve further study.

An interesting (but not direct) corollary of Theorem 1.1 is the following:

Theorem 1.3. *Let (M^2, g) be a compact 2-dimensional Riemannian manifold with negative Gaussian curvature. Then there is a smooth symmetric positive definite*

(0, 2)-tensor h_{ij}, which is invariant under the isometry group of (M^2, g), such that

$$
\begin{aligned}
R_{ijkl} &= -(h_{ik}h_{jl} - h_{il}h_{jk}), \\
\nabla_i h_{jk} &= \nabla_j h_{ik},
\end{aligned}
\tag{1-4}
$$

where R_{ijkl} is the curvature tensor of the manifold.

The proof of Theorem 1.3 relies on a uniqueness theorem of isometric embeddings. Unfortunately, the uniqueness of the isometric embeddings in Theorem 1.1 has not been proven. Technically, the proof of the uniqueness theorem involves the estimates of the second fundamental form. Note that by Remark 1.2, it is possible that some exotic embeddings may violate these estimates. We remark that the boundedness of the second fundamental form is not sufficient to guarantee the uniqueness of the isometric embedding (see [Guan et al. 2006, Theorem 2.3]).

The proof of the main theorem is reduced to solving certain equations of Monge–Ampère type

$$
\frac{\det(\nabla^2 u + g)}{\det(g)} = -K_g(|\nabla u|^2 + 2u)
\tag{1-5}
$$

on the whole manifold M. The corresponding Dirichlet problems may be solved on a sequence of exhausting domains Ω_l with some particular boundary values. The problem amounts to deriving certain uniform a priori estimates for these solutions u_l. The bulk of the present paper is devoted to these estimates.

In Section 2, we sketch a proof of the main theorem and derive the zeroth- and first-order estimates. In Section 3, we derive the second- and higher-order estimates and prove Theorem 1.1. In Section 4, to prove Theorem 1.3, we derive some estimates of the second fundamental form (Theorem 4.1) and obtain a uniqueness theorem (Theorem 4.4) for isometric embeddings. Finally, in the appendix we supply an alternative, straightforward argument for the second-order derivative estimate.

2. Zeroth- and first-order estimates

Sketch of proof. Suppose $X : M \to \mathbb{R}^{2,1}$ is an isometric embedding. Then $X(M)$ is a spacelike submanifold and the Gauss–Codazzi–Weingarten equations read as follows:

$$
\begin{aligned}
\nabla_i \nabla_j X &= h_{ij}\vec{n}, \\
\nabla_i \vec{n} &= h_{ij} g^{jk} X_k, \\
R_{ijkl} &= -(h_{ik}h_{jl} - h_{il}h_{jk}), \\
\nabla_i h_{jk} &= \nabla_j h_{ik},
\end{aligned}
\tag{2-1}
$$

where $\vec{n}$ is the normal vector, h_{ij} the second fundamental form, and R_{ijkl} the curvature tensor.

Let $u = -\frac{1}{2}\langle X, X\rangle$, where $\langle \cdot\,, \cdot\rangle$ is the Lorentz–Minkowski metric. By (2-1), we have

$$\begin{aligned} \nabla_i u &= -\langle X, X_i\rangle, \\ \nabla_i\nabla_j u &= -h_{ij}\langle \vec{n}, X\rangle - g_{ij}. \end{aligned} \tag{2-2}$$

Since

$$\begin{aligned} \langle X, X\rangle =& \sum_{i,j=1}^{2} g^{ij}\langle X, X_i\rangle\langle X, X_j\rangle - \langle X, \vec{n}\rangle^2 \\ =& |\nabla u|^2 - \langle X, \vec{n}\rangle^2, \end{aligned}$$

it follows that

$$\langle X, \vec{n}\rangle^2 = |\nabla u|^2 + 2u. \tag{2-3}$$

Combining (2-1), (2-2), and (2-3), we get

$$\frac{\det(\nabla^2 u + g)}{\det(g)} = -K_g(|\nabla u|^2 + 2u). \tag{2-4}$$

Note that the equation (2-4) satisfied by the function $-\frac{1}{2}\langle X, X\rangle$ is an intrinsic equation on the manifold (M, g).

Conversely, if we can find a bounded positive solution u of (2-4) on M, we will show that this yields an isometric embedding $X : (M, g) \to \mathbb{R}^{2,1}$ such that $-\frac{1}{2}\langle X, X\rangle = u$. To construct this isometric embedding, we need to introduce the polar coordinates in the open future timelike cone

$$\mathcal{I}^+ = \left\{(x_1, x_2, x_3) \in \mathbb{R}^{2,1} \;\middle|\; \sqrt{x_1^2 + x_2^2} < x_3\right\}.$$

In this polar coordinate system, the Lorentz–Minkowski metric takes the form

$$-dr^2 + r^2 ds_{\mathbb{H}}^2, \tag{2-5}$$

where $r = \sqrt{x_3^2 - x_1^2 - x_2^2}$ and $ds_{\mathbb{H}}^2$ is the hyperbolic metric ($K = -1$) of the unit imaginary sphere $r = 1$.

Proposition 2.1. *For a positive C^2 function u on M, define a new metric*

$$\bar{g} = \frac{g + \left(d\sqrt{2u}\right)^2}{2u}$$

on M. The Gaussian curvature $K_{\bar g}$ of $\bar g$ is given by

$$K_{\bar g} = -1 + \frac{\dfrac{\det(\nabla^2 u + g)}{\det(g)} + K_g(|\nabla u|^2 + 2u)}{\left(1 + \dfrac{|\nabla u|^2}{2u}\right)^2}. \tag{2-6}$$

Proof. The Gaussian curvature of the metric $g_1 \triangleq g + \left(d\sqrt{2u}\right)^2$ can be computed by the formula

$$K_{g_1} = \frac{1}{1+\left|\nabla\sqrt{2u}\right|^2}\left(K_g + \frac{\det\left(\nabla^2\sqrt{2u}\right)}{\det(g)\left(1+\left|\nabla\sqrt{2u}\right|^2\right)}\right).$$

(see [Guan 2007; Hong 2001]). From

$$g_1^{ij} = g^{ij} - \frac{u^i u^j}{2u + |\nabla u|^2} \quad \text{and} \quad {\Gamma_1}_{ij}^k - \Gamma_{ij}^k = \frac{u^k}{2u + |\nabla u|^2}\left(\nabla_{ij}^2 u - \frac{u_i u_j}{2u}\right)$$

we may calculate $\Delta_{g_1}\log u$ by the formula

$$\Delta_{g_1}\log u = g_1^{ij}\left(\nabla_{ij}^2 \log u - ({\Gamma_1}_{ij}^k - \Gamma_{ij}^k)(\log u)_k\right),$$

where $\nabla_{ij}^2 \log u$ is the Hessian of $\log u$ with respect to the metric g. By the curvature formula of conformal transformation $\bar g = g_1/2u$, a long but straightforward computation gives us

$$\frac{K_{\bar g}}{2u} = K_{g_1} + \frac{1}{2}\Delta_{g_1}\log u = -\frac{1}{2u} + \frac{1}{2u}\frac{\dfrac{\det(\nabla^2 u + g)}{\det(g)} + K_g(|\nabla u|^2 + 2u)}{\left(1 + \dfrac{|\nabla u|^2}{2u}\right)^2}. \qquad \square$$

Remark 2.2. If u is a bounded positive smooth solution to (2-4), then the metric $\bar g$ in Proposition 2.1 is complete and has constant curvature -1. Hence there exists an isometry $i : (M, \bar g) \to \mathbb{H} = \{r = 1\}$ and we can construct an embedding $I : (M, g) \to \mathcal{I}^+ \subset \mathbb{R}^{2,1}$ as $I(y) \triangleq \left(i(y), \sqrt{2u(y)}\right)$ in the polar coordinate system (2-5). It is clear that

$$I^*(-dr^2 + r^2 ds_{\mathbb{H}}^2) = -\left(d\sqrt{2u}\right)^2 + 2u i^* ds_{\mathbb{H}}^2 = -\left(d\sqrt{2u}\right)^2 + 2u\bar g = g, \tag{2-7}$$

which shows that the map I is the desired isometric embedding. The regularity of the embedding I follows from the regularity of u.

Hence the proof of Theorem 1.1 may be reduced to solving equation (2-4). The next theorem gives the required solution.

Theorem 2.3. *Under the assumptions of Theorem 1.1, equation* (2-4) *has a smooth bounded positive solution u such that* $0 < u \le 1/2C_1$.

The following strategy will be adapted to solve (2-4). We first solve it on a sequence of compact smooth exhausting domains $\Omega_1 \Subset \Omega_2 \Subset \cdots$. Let u_l be the solution on Ω_l. Fixing $x_0 \in M$, we show that for any nonnegative integer $k \geq 0$, there exists a constant $D_k > 0$ such that

(2-8) $$\sup_{\Omega_l \supset B(x_0,k+1)} |u_l|_{C^k(\bar{B}(x_0,k))} \leq D_k,$$

where the norm $C^k(\bar{B}(x_0, k))$ can be defined on some (indeed any) fixed finite coordinate covering of $\bar{B}(x_0, k)$. Once (2-8) has been obtained, we use the Arzelà–Ascoli theorem to extract a subsequence of u_l such that the limit is a smooth solution of equation (2-4).

Indeed, we choose simply $\Omega_l = B(x_0, l)$ and consider the Dirichlet problem

(2-9) $$\begin{cases} \dfrac{\det(\nabla^2 u + g)}{\det(g)} = -K_g(|\nabla u|^2 + 2u), \\ u|_{\partial B(x_0,l)} = \dfrac{1}{2C_2(l)}, \end{cases}$$

where $C_2(l) = \max_{x \in \bar{B}(x_0,l)}(-K_g(x))$.

Clearly, (2-9) has a subsolution $u_0 \equiv 1/2C_2(l)$, i.e.,

$$\frac{\det(\nabla^2 u_0 + g)}{\det(g)} \geq -K_g(|\nabla u_0|^2 + 2u_0).$$

By continuity methods, this implies that (2-9) admits a smooth solution u_l which satisfies $u_l \geq u_0$ and $\nabla^2 u_l + g > 0$ (see [Guan 1998]).

The main task of the subsequent sections is to derive a priori estimates for the solutions u_l so that (2-8) holds. For convenience, we drop the subscript l from u_l and Ω_l in the process of the computations.

Zeroth- and first-order estimates.

Proposition 2.4. *The solution u of the Dirichlet problem* (2-9) *satisfies*

(2-10) $$\frac{1}{2C_2(l)} \leq u \leq \frac{1}{2C_1}.$$

Proof. By applying the maximum principle to u, we have

$$\frac{1}{2C_2(l)} \leq u \leq \max\left\{\sup\nolimits_\Omega\left(-\frac{1}{2K_g}\right), \frac{1}{2C_2(l)}\right\} \leq \frac{1}{2C_1}. \qquad \square$$

Proposition 2.5 (first-order estimate). *The gradient of the solution u of* (2-9) *satisfies*

(2-11) $$|\nabla u| \leq \frac{2}{\sqrt{C_1}}.$$

Proof. We choose

$$\xi = \frac{1}{2C_2(l)} + \frac{2}{\sqrt{C_1}}(l - d(x_0, \cdot))$$

as a barrier function. Clearly ξ satisfies $\xi|_{\partial\Omega} = 1/2C_2(l)$ and $|\nabla\xi| \le 2/\sqrt{C_1}$. By the standard Hessian comparison theorem

$$\mathrm{Hess}(d)|_{(\nabla d)^\perp} \ge \sqrt{C_1}\coth\bigl(d\sqrt{C_1}\bigr),$$

we have

$$\Delta\xi = -\frac{2}{\sqrt{C_1}}\Delta d \le -2.$$

On the other hand, from $\nabla^2 u + g > 0$, we know that

$$\Delta u + 2 > 0.$$

Hence $\Delta(u - \xi) > 0$ on Ω. The maximum principle implies that $u \le \xi$. Therefore we have

$$|\nabla u|\big|_{\partial\Omega} \le |\nabla\xi| \le \frac{2}{\sqrt{C_1}}. \tag{2-12}$$

Now we consider the quantity $|\nabla u|^2 + 2u$. The maximum $\max_{\overline{\Omega}}(|\nabla u|^2 + 2u)$ is achieved either on the boundary or in the interior of the domain. In the former case, the maximum is bounded by $4/C_1 + 1/C_2(l)$ by (2-12). In the latter case, suppose the maximum is achieved at some point $\bar{x} \in \Omega$. Since

$$0 = \nabla_i(|\nabla u|^2 + 2u)(\bar{x}) = 2(u_{ij} + g_{ij})u_j(\bar{x})$$

and $u_{ij} + g_{ij} > 0$, it follows that $|\nabla u|(\bar{x}) = 0$, and therefore

$$\max_{\overline{\Omega}}(|\nabla u|^2 + 2u) \le \max_{\overline{\Omega}}(2u) \le \frac{1}{C_1}.$$

Combining both cases, we get

$$\sup_{\overline{\Omega}} |\nabla u| \le \max\left\{\frac{2}{\sqrt{C_1}}, \sqrt{\frac{1}{C_1} - \frac{1}{C_2(l)}}\right\} = \frac{2}{\sqrt{C_1}}. \qquad \square$$

Propositions 2.4 and 2.5 state that the function u and its gradient can be bounded from above by a constant independent of the domain Ω_l. Before estimating the lower bound of u, we need to construct cutoff functions around points where the values of u are not too large.

Lemma 2.6. *Fix $\tilde{x} \in M$, and suppose there exist a real number $r_0 > 0$ and a solution u of* (2-9) *defined on a domain $\Omega \supset B(\tilde{x}, r_0)$ satisfying*

$$u(\tilde{x}) < \frac{r_0}{2\sqrt{c_2}\coth(r_0\sqrt{c_2})}, \quad \text{where } c_2 = \max_{y \in \overline{B}(\tilde{x}, r_0)}(-K_g(y)). \tag{2-13}$$

Let $r_2 \triangleq 2\sqrt{c_2}\coth(r_0\sqrt{c_2})$ be the denominator in (2-13). *Then there are a domain $Q_{\tilde{x}} \subset B(\tilde{x}, r_0)$ containing $\tilde{x}$ and a function $\varphi^{\tilde{x}} \in C^2(\bar{Q}_{\tilde{x}})$ such that*

(i) $0 \le \varphi^{\tilde{x}} \le \dfrac{r_0}{r_2}$, $\quad \varphi^{\tilde{x}}\big|_{\partial Q_{\tilde{x}}} = 0$, *and*

$$\varphi^{\tilde{x}} \ge \frac{1}{2}\Big(\frac{r_0}{r_2} - u(\tilde{x})\Big) \quad \textit{on } B\bigg(\tilde{x}, \frac{\sqrt{C_1}}{6}\Big(\frac{r_0}{r_2} - u(\tilde{x})\Big)\bigg); \tag{2-14}$$

(ii) $|\nabla\varphi^{\tilde{x}}| \le \dfrac{3}{\sqrt{C_1}}$;

(iii) $\nabla^2\varphi^{\tilde{x}} \ge -(\nabla^2 u + g)$.

Proof. Set

$$\xi = u + \frac{d^2(\tilde{x}, \cdot)}{r_0 r_2}, \quad Q_{\tilde{x}} = \Big\{\xi < \frac{r_0}{r_2}\Big\}, \quad \varphi^{\tilde{x}} = \frac{r_0}{r_2} - \xi.$$

Then $\varphi^{\tilde{x}}$ satisfies (ii) by Proposition 2.5. By (ii) and assumption (2-13), we know (i) holds. To check that $\varphi^{\tilde{x}}$ satisfies (iii), we use the Hessian comparison theorem

$$\nabla^2 d^2(\tilde{x}, \cdot) \le 2d(\tilde{x}, \cdot)\sqrt{c_2}\coth\big(d(\tilde{x}, \cdot)\sqrt{c_2}\big)g$$

to conclude that

$$\nabla^2\xi \le \nabla^2 u + g. \qquad \square$$

Proposition 2.7 (lower bound for u). *For any $\tilde{x} \in M$, $r_0 > 0$, assume the solution u of* (2-9) *is defined on a domain $\Omega \supset B(\tilde{x}, r_0)$. Let c_2 and r_2 be defined as in Lemma 2.6. Then*

$$u(\tilde{x}) \ge \min\Big\{\frac{r_0}{2r_2}, \frac{C_1 r_0^2}{9r_2^2}, \frac{1}{32c_2}\Big\}. \tag{2-15}$$

Proof. Assume

$$u(\tilde{x}) < \frac{r_0}{2r_2}. \tag{2-16}$$

Clearly the condition (2-13) holds for this r_0. Consider the quantity u/ζ around $\tilde{x}$, where $\zeta = \varphi^{\tilde{x}}$ is the cutoff function in Lemma 2.6. Suppose the minimum of u/ζ is achieved at some point $\bar{x} \in \operatorname{supp}(\zeta)$. At the point $\bar{x}$, we have

$$\frac{\nabla u}{u} = \frac{\nabla\zeta}{\zeta} \quad \text{and} \quad 0 \le \nabla^2 \log\frac{u}{\zeta} = \frac{\nabla^2 u}{u} - \frac{\nabla^2\zeta}{\zeta}. \tag{2-17}$$

Diagonalize $u_{ij} = \lambda_i\delta_{ij}$ at $\bar{x}$ with an orthonormal basis. It follows from (2-17) that

$$\sum \frac{\nabla_{ii}\zeta}{1+\lambda_i} \le \frac{\zeta}{u}\Big(2 - \frac{2}{\sqrt{(1+\lambda_1)(1+\lambda_2)}}\Big) \tag{2-18}$$
$$= \frac{2\zeta}{u}\left(1 - \frac{1}{\sqrt{(-K_g)\big(|\nabla\zeta|^2 u^2/\zeta^2 + 2(u/\zeta)\zeta\big)}}\right).$$

Combining (2-18) and Lemma 2.6, we have

$$-2 \le \frac{2\zeta}{u}\left(1 - \frac{1}{\sqrt{(-K_g)\big(\frac{9}{C_1}\frac{u^2}{\zeta^2} + \frac{2r_0}{r_2}\frac{u}{\zeta}\big)}}\right). \tag{2-19}$$

If the denominator in (2-19) is at most $\frac{1}{2}$, we get

$$\frac{u}{\zeta} \ge 1. \tag{2-20}$$

On the other hand, if the denominator in (2-19) exceeds $\frac{1}{2}$, direct computation shows that

$$\frac{u}{\zeta} \ge \min\left\{\frac{2C_1 r_0}{9r_2}, \frac{r_2}{16 r_0 c_2}\right\}. \tag{2-21}$$

Combining (2-20) and (2-21), we have

$$u \ge \zeta \min\left\{1, \frac{2C_1 r_0}{9r_2}, \frac{r_2}{16 r_0 c_2}\right\}.$$

Recalling (2-14) we immediately obtain (2-15). □

Corollary 2.8. *For any $r_0 > 0$, there is a constant C depending only on r_0 and C_1 such that*

$$u(\tilde{x}) \ge \frac{C^{-1}}{\max_{\bar{B}(\tilde{x},r_0)}(-K_g(x))}, \tag{2-22}$$

for any solution u to (2-9) *defined on $\Omega \supset B(\tilde{x}, r_0)$.*

3. Second- and higher-order estimates

In this section, we give a purely local second-order derivative estimate. This estimate could be done by Heinz–Lewy "characteristic" theory for Monge–Ampère equations in dimension 2. The reader is referred to the lecture notes [Schulz 1990] for detailed exposition. To state the result in [Schulz 1990], we consider the Monge–Ampère equation for a function $z = z(x, y)$ on a domain $\mathcal{D} \subset \mathbb{R}^2$:

$$(z_{xx} + C)(z_{yy} + A) - (z_{xy} - B)^2 = K(x, y, z)D(x, y, z, z_x, z_y) > 0, \tag{3-1}$$

where A, B, C, D are functions of x, y, z, p, q, and $p = z_x, q = z_y$.

Assumption (i) $z \in C^{1,1}(\mathcal{D})$ and

$$|z_x| + |z_y| \le \mathcal{K}_1. \tag{3-2}$$

Assumption (ii) $A, B, C \in C^1(\mathcal{D} \times \mathbb{R}^3)$, $K \in C^\mu(\mathcal{D} \times \mathbb{R})$ for some $0 < \mu < 1$, $D \in C^1(\mathcal{D} \times \mathbb{R}^3)$, and

$$|A| + |B| + |C| + |D| \le \mathcal{A}_1, \tag{3-3}$$

$$K, D \ge 1/\mathcal{A}_2, \tag{3-4}$$

$$|\partial_{\mathcal{D}\times\mathbb{R}^3} A| + \cdots + |\partial_{\mathcal{D}\times\mathbb{R}^3} D| \le \mathcal{A}_3, \tag{3-5}$$

$$|K|_{C^\mu(\mathcal{D}\times\mathbb{R})} \le \mathcal{A}_4. \tag{3-6}$$

Assumption (iii) The functions

$$\begin{aligned} \phi_1(x, y) &= A_p, \\ \phi_2(x, y) &= A_q + 2B_p, \\ \phi_3(x, y) &= C_p + 2B_q, \\ \phi_4(x, y) &= C_q \end{aligned} \tag{3-7}$$

are Lipschitz continuous with

$$[\phi_1]^{\mathcal{D}}_{0,1} + \cdots + [\phi_4]^{\mathcal{D}}_{0,1} \le \mathcal{A}_5. \tag{3-8}$$

Theorem 3.1 [Schulz 1990, Theorem 9.4.1]. *Suppose $z \in C^{1,1}(\mathcal{D})$ is a solution of* (3-1) *such that the above Assumptions* (*i*), (*ii*), *and* (*iii*) *hold with the constants $\mathcal{K}_1, \mathcal{A}_1, \ldots, \mathcal{A}_5$. Then $z \in C^{2,\mu}_{\text{loc}}(\mathcal{D})$, and for any $\mathcal{D}' \Subset \mathcal{D}$ there is an interior estimate*

$$\|\partial^2 z\|_{C^\mu(\mathcal{D}')} \le C(\mu, \mathcal{K}_1, \mathcal{A}_1 \cdots \mathcal{A}_5, \operatorname{dist}(\mathcal{D}', \partial\mathcal{D})). \tag{3-9}$$

For any $\tilde{x} \in M$, to invoke the result in [Schulz 1990], we fix a local coordinate system $(x, y) \in \mathcal{D}$ in M around $\tilde{x}$. Take $z(x, y)$ to be a solution $u(x, y)$ of equation (2-9) defined on $\Omega \supset \mathcal{D}$. Then we find

$$\begin{aligned} A &= g_{22} - \Gamma^k_{22} p_k, \\ B &= -g_{12} + \Gamma^k_{12} p_k, \\ C &= g_{11} + \Gamma^k_{11} p_k, \\ D &= g^{kl} p_k p_l + 2z, \\ K(x, y, z) &= -K_g(x, y) \det(g_{ij}), \end{aligned} \tag{3-10}$$

where $p_1 = p$, $p_2 = q$.

Note that by Propositions 2.4, 2.5, and 2.7, we have estimated the upper bound of u, ∇u, and the lower bound of u in the coordinate system $\mathcal{D}$. This gives rise to a control of the constants $\mathcal{K}_1, \mathcal{A}_1, \dots, \mathcal{A}_5$ in terms of the geometry of $(\mathcal{D}, g)$. From Theorem 3.1, we have immediately

Proposition 3.2. *For any nonnegative integer $k \geq 0$, there exists a constant $D_k > 0$ such that*

$$\sup_{\Omega_l \supset B(x_0, k+1)} |u_l|_{C^{2,\mu}(\bar{B}(x_0, k+\frac{1}{2}))} \leq D_k, \tag{3-11}$$

where the norm $C^{2,\mu}(\bar{B}(x_0, k+\frac{1}{2}))$ can be defined on some (and any) fixed finite coordinate covering of $\bar{B}(x_0, k+\frac{1}{2})$.

We proceed to consider the third- and higher-order estimates (2-8). This may be done by the standard Schauder estimate for elliptic equations.

Proposition 3.3. *For any nonnegative integer $k \geq 0$, there exists a constant $D_k > 0$ such that*

$$\sup_{\Omega_l \supset B(x_0, k+1)} |u_l|_{C^{k}(\bar{B}(x_0, k))} \leq D_k, \tag{3-12}$$

where the norm $C^{k}(\bar{B}(x_0, k))$ can be defined on some (indeed any) fixed finite coordinate covering of $\bar{B}(x_0, k)$.

Proof. By (2-9), we see that $\nabla_i u$ satisfies an equation of the type

$$\hat{g}^{jm} v_{jm} = f(x, v, \nabla v), \tag{3-13}$$

where $\hat{g} = \nabla^2 u + g$. By the previous second-order estimate, we know (3-13) is uniformly elliptic on $B(x_0, k+\frac{1}{2})$ and the C^μ norm of $\hat{g}$ and f are uniformly bounded (independently of l). The result follows from the standard interior Schauder estimate and a bootstrap argument. □

Proof of Theorem 2.3. By Proposition 3.3 and the Arzelà–Ascoli theorem, we may extract a C^∞_{loc} convergent subsequence of u_l. The limit is the desired solution. □

Theorem 1.1 follows from Theorem 2.3 (see Remark 2.2).

4. Estimating the second fundamental form

In this section, we refine the result in Theorem 1.1. We prove that under an additional "smoothness" assumption on the Gaussian curvature, a particular embedding with controlled second fundamental form will be obtained. More precisely, we have

Theorem 4.1. *Let (M, g) be a smooth 2-dimensional simply connected complete Riemannian manifold whose Gaussian curvature satisfies*

$$-C_2 \leq K \leq -C_1 \tag{4-1}$$

and

$$\sup_{d(x,y)\le 1} \frac{|K(x)-K(y)|}{d(x,y)^{\mu}} \le C_{\mu} \tag{4-2}$$

for some positive constants $C_2 \ge C_1 > 0$, $1 > \mu > 0$, $C_\mu > 0$.

Then there exists a smooth isometric embedding $X : M \to \mathbb{R}^{2,1}$ *such that the space-like submanifold* $X(M)$ *is a graph over* $\mathbb{R}^2 \subset \mathbb{R}^{2,1} : (x_1, x_2, 0) \to (x_1, x_2, Z(x_1, x_2))$ *satisfying these conditions*:

(i) $$\sqrt{\frac{1}{C_2}+x_1^2+x_2^2} \le Z(x_1,x_2) \le \sqrt{\frac{1}{C_1}+x_1^2+x_2^2}.$$

(ii) $|A| \le C$, *where* A *is the second fundamental form of the submanifold* $X(M)$, *and the constant* C *only depends on* C_1, C_2, *and* C_μ.

The proof of Theorem 4.1 is based on Proposition 2.4 and Theorem 3.1 and the following result.

Proposition 4.2. *Under the assumptions* (4-1) *and* (4-2) *of Theorem 4.1, there exists* $R > 0$ *such that* M *admits a covering of isothermal coordinate charts* $\{(U_i, (u^1, u^2))\}$, *where with* $U_i = \{(u^1)^2 + (u^2)^2 < R^2\}$, *with these properties*:

(i) *For any* $y_0 \in M$, *there is* U_{i_0} *with* $y_0 \in \{(u^1)^2 + (u^2)^2 < R^2/4\} \subset U_{i_0}$.

(ii) *In each* U_i, *the metric* g *of* M *takes the form* $g = \psi\big((du^1)^2 + (du^2)^2\big)$, *with*

$$c^{-1} \le \psi \le c \quad \textit{and} \quad |\psi|_{C^{2,\mu}(U_i)} \le c_\mu, \tag{4-3}$$

for constants c *and* c_μ *independent of* i. *If additionally* (4-6) *is satisfied*, *we have*

$$|\psi|_{C^{l+1,\alpha}(U_i)} \le c_{l,\alpha} \quad \textit{for any } \alpha \in (0,1), \tag{4-4}$$

where the $c_{l,\alpha}$ *are constants independent of* i.

Proof of Theorem 4.1. By Proposition 2.4 and Theorem 1.1, we know there exists a smooth isometric embedding $X : M \to \mathbb{R}^{2,1}$ such that $u = -\frac{1}{2}\langle X, X\rangle$ satisfying $1/2C_2 \le u \le 1/2C_1$. Let R be the constant provided in Proposition 4.2. Let the coordinates (x, y) in equation (3-1) be the isothermal coordinates (u^1, u^2) in Proposition 4.2, $z(x, y) = u(x, y)$, and $\mathcal{D} = \{x^2 + y^2 < R^2/4\}$. In these coordinates, (3-10) becomes

$$\begin{aligned} A &= \psi - \Gamma^k_{22} p_k, \\ B &= \Gamma^k_{12} p_k, \\ C &= \psi + \Gamma^k_{11} p_k, \\ D &= \psi^{-1}(p_1^2 + p_2^2) + 2z, \\ K(x,y,z) &= -K_g(x,y)\psi^2. \end{aligned} \tag{4-5}$$

Estimate (4-3) and Proposition 2.5 imply that there is a constant C depending only on C_1, C_2, C_μ such that the constants in equations (3-2)–(3-8) can be bounded by C,

$$\mathcal{K}_1, \mathcal{A}_1, \dots, \mathcal{A}_5 \le C.$$

Theorem 3.1 implies $|\partial_{ij} u|_{C^\mu(B(0,R/4))} \le C$. Combining this with (4-3) in particular gives $|h_{ij}| \le C$. This proves (ii) in Theorem 4.1. $\square$

Remark 4.3. If the curvature covariant derivatives up to order l are assumed to be bounded in Theorem 4.1, i.e.,

$$\sum_{p=0}^{l} |\nabla^p K| \le \bar{C}_l \tag{4-6}$$

for some $l \ge 1$, then the covariant derivatives of the second fundamental form of $X(M)$ up to order $l-1$ are also bounded,

$$\sup_{x \in X(M)} \sum_{p=0}^{l-1} |\nabla^p A|(x) \le C \tag{4-7}$$

for some C depending only on $\bar{C}_l$.

Actually, if (4-6) is assumed, notice that (4-4) holds. Then (4-7) follows by the same argument as in Proposition 3.3.

An important application of Theorem 4.1 is to give a uniqueness theorem.

Theorem 4.4. *Under the assumptions of Theorem 4.1, let X be the isometric embedding constructed in Theorem 4.1.*

(i) *Let $\widetilde{X}$ be another isometric embedding of (M, g) into $\mathbb{R}^{2,1}$ such that $\widetilde{X}(M)$ is represented as a graph*

$$\sqrt{y_1^2 + y_2^2} \le \widetilde{Z}(y_1, y_2) \le \sqrt{\frac{1}{C} + y_1^2 + y_2^2} \tag{4-8}$$

in some Lorentz–Minkowski coordinate system $\{y_1, y_2, y_3\}$. Then there is an isometry $\iota \in \mathrm{Iso}(\mathbb{R}^{2,1})$ such that $\widetilde{X} = \iota \circ X$.

(ii) *There is an injective homomorphism $\rho : \mathrm{Iso}(M, g) \to \mathrm{Iso}(\mathbb{H}) \subset \mathrm{Iso}(\mathbb{R}^{2,1})$ such that*

$$X \circ \gamma = \rho(\gamma) \circ X \tag{4-9}$$

for any $\gamma \in \mathrm{Iso}(M, g)$, where $\mathrm{Iso}(M, g)$, $\mathrm{Iso}(\mathbb{H})$, and $\mathrm{Iso}(\mathbb{R}^{2,1})$ are the groups of isometries of M, the unit imaginary sphere in $\mathbb{R}^{2,1}$, and $\mathbb{R}^{2,1}$ respectively.

Proof. After an isometry $\tilde{\iota}$ of $\mathbb{R}^{2,1}$, $\tilde{\iota} \circ \widetilde{X}(M)$ can be pinched between the light cone and a hyperboloid associated to X, and we can define $\tilde{u} = -\frac{1}{2}\langle \tilde{\iota} \circ \widetilde{X}, \tilde{\iota} \circ \widetilde{X} \rangle$, which satisfies $0 < \tilde{u} \le C$.

Using the polar coordinates in Remark 2.2, we know that $\tilde{\iota} \circ \widetilde{X}$ is determined by $\tilde{u}$ and an isometry

$$\tilde{i} : \left(M, \frac{g + \left(d\sqrt{2\tilde{u}}\right)^2}{2\tilde{u}} \right) \to \mathbb{H}.$$

To show that $\tilde{\iota} \circ \widetilde{X}$ is congruent to X, it suffices to show that $u = \tilde{u}$. Indeed, once we have $u = \tilde{u}$, it follows that $\tilde{\iota} \circ \widetilde{X} = \sigma \circ X$, where $\sigma = \tilde{i} i^{-1} \in \mathrm{Iso}(\mathbb{H}) \subset \mathrm{Iso}(\mathbb{R}^{2,1})$. Then $\widetilde{X} = \iota \circ X$, where $\iota = \tilde{\iota}^{-1} \circ \sigma$.

We need some a priori estimates of $\tilde{u}$ up to second order. To this end, we use the powerful tool of the maximum principle in [Cheng and Yau 1980]. Since the curvature is assumed to be bounded, for any C^2 function F bounded from above, there is a sequence of $x_k \in M$ and $\varepsilon_k \to 0$ such that

$$\begin{aligned} \sup\nolimits_M F - F(x_k) &\le \varepsilon_k, \\ |\nabla F|(x_k) &\le \varepsilon_k, \\ \nabla^2 F(x_k) &\le \varepsilon_k g. \end{aligned} \tag{4-10}$$

Note that $\tilde{u}$ satisfies (2-4). Applying the above maximum principle to $\tilde{u}$ and $-\tilde{u}$, we immediately get

$$\frac{1}{2C_2} \le \tilde{u} \le \frac{1}{2C_1}.$$

We claim that the gradient of $\tilde{u}$ is also bounded, and more precisely, it satisfies

$$|\nabla \tilde{u}| \le \frac{1}{\sqrt{C_1}}.$$

Indeed, for any $\tilde{x} \in M$, let γ be a geodesic of unit speed such that $\gamma(0) = \tilde{x}$. We would like to control $\left|\frac{d}{dt}(\tilde{u} \circ \gamma)(0)\right|$. By the convexity of the function $\tilde{u} + \frac{1}{2}d^2(\tilde{x}, \ldots)$, we know

$$t\left|\frac{d}{dt}(\tilde{u} \circ \gamma)(0)\right| \le \max\{\tilde{u}(\gamma(t)) - \tilde{u}(\gamma(0)), \tilde{u}(\gamma(-t)) - \tilde{u}(\gamma(0))\} + \frac{t^2}{2}.$$

It follows that $\left|\frac{d}{dt}(\tilde{u} \circ \gamma)(0)\right| \le 1/\sqrt{C_1}$ by taking $t = 1/\sqrt{C_1}$. This implies $|\nabla \tilde{u}|(\tilde{x}) \le 1/\sqrt{C_1}$, and the claim is proved.

Combining the gradient estimate of $\tilde{u}$ with the proof of Proposition 3.2, we know that $|\nabla^2 \tilde{u}|$ is bounded.

Summarizing the above estimates, it follows that there is $C > 0$ such that

$$1/C \le u_t \le C, \quad |\nabla u_t| \le C, \quad \nabla^2 u_t + g \ge C^{-1} g, \tag{4-11}$$

where $u_t = u + t(\tilde{u} - u)$, $t \in [0, 1]$.

Note that u and $\tilde{u}$ satisfy the same equation (2-4). This implies

$$\int_0^1 (g+\nabla^2 u_t)^{ij}\, dt \nabla^2_{ij}(\tilde{u}-u) = \int_0^1 \frac{2\langle \nabla u_t, \nabla(\tilde{u}-u)\rangle + 2(\tilde{u}-u)}{|\nabla u_t|^2 + 2u_t}\, dt. \tag{4-12}$$

Let $F=\tilde{u}-u$ in (4-10). Combining (4-10) with (4-11) and (4-12), we have

$$C\varepsilon_k \geq \sup_M(\tilde{u}-u).$$

This gives $u \geq \tilde{u}$. Similarly, we have $u \leq \tilde{u}$. Hence $u=\tilde{u}$.

To prove (ii), one can show $u\circ\gamma = u$ for any $\gamma \in \operatorname{Iso}(M, g)$ by Cheng and Yau's maximum principle (4-10). This implies $\operatorname{Iso}(M, g) \subset \operatorname{Iso}(M, \bar{g})$, where

$$\bar{g} = \frac{g + \left(d\sqrt{2u}\right)^2}{2u}.$$

The desired injective homomorphism $\rho : \operatorname{Iso}(M, g) \to \operatorname{Iso}(\mathbb{H})$ is given by

$$\rho(\gamma) = i\circ\gamma\circ i^{-1}. \qquad \square$$

Proof of Theorem 1.3. Let $(\widetilde{M}, \tilde{g})$ be the universal cover of (M, g) with the induced metric. Then the curvature of $\tilde{g}$ is pinched between two negative constants, and all covariant derivatives of the curvature are bounded. By Theorem 4.4, there is an isometric embedding $X : (\widetilde{M}, \tilde{g}) \to \mathbb{R}^{2,1}$ and an injective homomorphism $\rho : \operatorname{Iso}(\widetilde{M}, \tilde{g}) \to \operatorname{Iso}(\mathbb{R}^{2,1})$ such that (4-9) holds. Let $\tilde{h}_{ij}$ be the second fundamental form of X. Note that the deck transformation group Γ of $\widetilde{M}$ over M is contained in $\operatorname{Iso}(\widetilde{M}, \tilde{g})$. Combining (4-9), we know $\tilde{h}_{ij}$ is invariant under Γ. This implies $\tilde{h}_{ij}$ descends to a tensor h_{ij} on M. This completes the proof of Theorem 1.3. $\square$

5. Appendix

The purpose of this appendix is to give an alternative method for the second-order estimate. The argument we present here is classical, straightforward, and may be generalized to higher dimensions (see [Guan and Li 1996]). The price to be paid is that this method requires some geometry of the background manifold. It works well on those points where the values of a solution u of (2-9) are not too large in comparison to the local geometry.

Proposition 5.1. *There exists $C>0$, depending only on C_1, satisfying the following property. Fixing $\tilde{x}\in M$, suppose there exist a real number $r_0>0$ and a solution u of (2-9) defined on a domain $\Omega_l \supset B(\tilde{x}, r_0)$ such that*

$$u(\tilde{x}) < \frac{r_0}{2\sqrt{c_2}\coth(r_0\sqrt{c_2})}, \quad \text{where } c_2 = \max_{y\in\bar{B}(\tilde{x}, r_0)}(-K_g(y)). \tag{5-1}$$

Let $r_2 \triangleq 2\sqrt{c_2}\coth(r_0\sqrt{c_2})$. *Then*

$$(5\text{-}2)\quad (g+\nabla^2 u)(x) \le \frac{e^{Cc_2'}}{r_0/r_2 - u(\tilde{x})}\left(1+\sqrt{c_4}\frac{r_0}{\sqrt{c_2}} + c_2'\left(1+\frac{r_0}{\sqrt{c_2}} + c_3\frac{r_0}{\sqrt{c_2}}\right)\right)$$

on $B\big(\tilde{x}, \frac{1}{6}\sqrt{C_1}\big(r_0/r_2 - u(\tilde{x})\big)\big)$, *where*

$$(5\text{-}3)\qquad \begin{aligned} c_2' &= \max_{x\in\bar{B}(\tilde{x},r_0+1)} (-K_g(x)),\\ c_3 &= \max_{x\in\bar{B}(\tilde{x},r_0)} |\nabla \log(-K_g(x))|,\\ c_4 &= \max_{x\in\bar{B}(\tilde{x},r_0)} |\nabla^2 \log(-K_g(x))|. \end{aligned}$$

Note that by Proposition 2.4, condition (5-1) can be justified at each $\tilde{x}$ (for suitable r_0) when the curvature K satisfies

$$(5\text{-}4)\qquad -C_2^2(d(x,x_0)+C_3)^2 \le K(x) \le -C_1$$

for some $x_0 \in M$ and positive constants $0 < C_2 < C_1 < C_3$.

Proof of Proposition 5.1. Consider an auxiliary function $STM \to \mathbb{R}$ on the unit tangent bundle of M, given by

$$(5\text{-}5)\qquad (x,\gamma) \mapsto \eta(x)(1+\nabla_{\gamma\gamma}u)e^{\frac{a}{2}(|\nabla u|^2+2u)(x)},$$

where $x\in M$, $\gamma \in T_xM$, $|\gamma|=1$, η is a cutoff function on M, and $a \ge 1$ is a constant to be specified later. Suppose the maximum

$$\max_{(x,\gamma)\in STM} \eta(1+\nabla_{\gamma\gamma}u)e^{\frac{a}{2}(|\nabla u|^2+2u)}$$

is achieved at $\bar{x} \in \operatorname{supp}(\eta)$ for some $\gamma \in T_{\bar{x}}M$ with $|\gamma|=1$. Diagonalize $u_{ij} = \lambda_i\delta_{ij}$ at $\bar{x}$ with the orthonormal eigenvectors e_i. Let $e_1=\gamma$. By parallel transport of each e_i along radial geodesics, we obtain a field of orthonormal frame $\{e_i\}$ near $\bar{x}$. The function

$$w = \eta(1+\nabla_{e_1,e_1}u)e^{\frac{a}{2}(|\nabla u|^2+2u)}$$

defined near $\bar{x}$ achieves its maximum at $\bar{x}$. In the following, we use C to denote various big constants depending only on C_1.

At the point $\bar{x}$, we have

$$(5\text{-}6)\qquad 0 = \nabla_i \log w = \frac{\nabla_{i11}u}{1+\lambda_1} + a(1+\lambda_i)u_i + \frac{\nabla_i \eta}{\eta}$$

and

$$(5\text{-}7)\quad 0 \ge \nabla_{ij}\log w = \frac{\nabla_{ij11}u}{1+\lambda_1} - \frac{\nabla_{i11}u\nabla_{j11}u}{(1+\lambda_1)^2} + a\big(u_k\nabla_{ijk}u + (\lambda_i+\lambda_i^2)\delta_{ij}\big) + \frac{\nabla_{ij}\eta}{\eta} - \frac{\nabla_i\eta\nabla_j\eta}{\eta^2}.$$

Let $f = f(x, z, p) \triangleq \log(-K) + \log(|\nabla u|^2 + 2u))$, where $z = u$, $p = \nabla u$. Differentiating equation (2-9), we get

$$(5\text{-}8)\qquad f_{11} = \frac{\nabla_{11ii}u}{1+\lambda_i} - \frac{(\nabla_{1ij}u)^2}{(1+\lambda_i)(1+\lambda_j)} \quad\text{and}\quad \nabla_k f = \frac{\nabla_{kii}u}{1+\lambda_i}.$$

Combining (5-7), (5-8), and the Ricci formula, we have

$$(5\text{-}9)\quad (1+\lambda_1)\Big(-\frac{\nabla_{ii}\eta}{(1+\lambda_i)\eta} + \frac{|\nabla_i\eta|^2}{(1+\lambda_i)\eta^2}\Big) + \frac{1}{1+\lambda_i}\Big(\frac{(\nabla_{i11}u)^2}{1+\lambda_1} - a(R_{ijip}u_pu_j + \lambda_i + \lambda_i^2)(1+\lambda_1) - (\nabla_i R_{i11p} + \nabla_1 R_{i1ip})u_p - 2R_{1i1i}(\lambda_1-\lambda_i)\Big) - \frac{(\nabla_{1ij}u)^2}{(1+\lambda_i)(1+\lambda_j)} \ge f_{11} + au_j\nabla_j f(1+\lambda_1).$$

By direct computations, we have (at $\bar{x}$)

$$(5\text{-}10)\quad f_{11} + au_k\nabla_k f(1+\lambda_1) \ge \frac{2}{|\nabla u|^2+2u}\Big\langle\nabla u, -a\nabla u - \frac{\nabla\eta}{\eta}\Big\rangle(1+\lambda_1) + \log(-K)_{11} + a\langle\nabla\log(-K), \nabla u\rangle(1+\lambda_1) + \frac{a(1+\lambda_1)|\nabla u|^2}{|\nabla u|^2+2u} - \frac{8(|\lambda_1|+1)^2}{|\nabla u|^2+2u} + 2\frac{R_{1j1l}u_ju_l}{|\nabla u|^2+2u}.$$

By (5-6), we have

$$(5\text{-}11)\quad \sum_i \frac{|\nabla_i\eta|^2}{(1+\lambda_i)\eta^2}(1+\lambda_1) = \frac{|\nabla_1\eta|^2}{\eta^2} + \sum_{i\ge 2}\bigg(\frac{|\nabla_{i11}u|^2}{(1+\lambda_1)(1+\lambda_i)} - 2a(1+\lambda_1)\nabla_i u\frac{\nabla_i\eta}{\eta} - a^2(1+\lambda_1)(1+\lambda_i)u_i^2\bigg).$$

Note that

$$\sum_{i,j}\frac{(\nabla_{1ij}u)^2}{(1+\lambda_i)(1+\lambda_j)}-\sum_i\frac{(\nabla_{i11}u)^2}{(1+\lambda_1)(1+\lambda_i)}-\sum_{i\geq 2}\frac{|\nabla_{i11}u|^2}{(1+\lambda_1)(1+\lambda_i)} \geq -4\frac{|\nabla u|}{|\nabla u|^2+2u}\frac{|\nabla\eta|}{\eta}(1+\lambda_1), \tag{5-12}$$

$$\sum_i -\frac{1}{1+\lambda_i}(R_{ijip}u_pu_j+\lambda_i+\lambda_i^2)(1+\lambda_1) \leq -\frac{2u}{|\nabla u|^2+2u}(1+\lambda_1)^2+2(1+\lambda_1), \tag{5-13}$$

$$\frac{1}{1+\lambda_i}\bigl(-(\nabla_i R_{i11p}+\nabla_1 R_{i1ip})u_p-2R_{1i1i}(\lambda_1-\lambda_i)\bigr) \leq \frac{2|\nabla u|}{|\nabla u|^2+2u}|\nabla\log(-K)|(1+\lambda_1)+\frac{2}{|\nabla u|^2+2u}(1+\lambda_1)^2+2K. \tag{5-14}$$

Multiplying both sides of (5-9) by η^2 and combining (5-10)–(5-14), we get

$$L_1(1+\lambda_1)^2\eta^2-L_2(1+\lambda_1)\eta-L_3\leq\eta(1+\lambda_1)\sum_{i\geq 1}\frac{-\nabla_{ii}\eta}{1+\lambda_i}, \tag{5-15}$$

where

$$\begin{aligned}
L_1 &= a\frac{2u}{|\nabla u|^2+2u}-\frac{10}{|\nabla u|^2+2u},\\
L_2 &= \Bigl(6\frac{|\nabla u|}{|\nabla u|^2+2u}+2a|\nabla u|\Bigr)|\nabla\eta|+2a\eta+\frac{|\nabla u|^2}{|\nabla u|^2+2u}a\eta\\
&\qquad +a|\nabla\log(-K)||\nabla u|\eta+\frac{2|\nabla u|}{|\nabla u|^2+2u}|\nabla\log(-K)|\eta,\\
L_3 &= |\nabla\eta|^2+\eta^2|\nabla^2\log(-K)|+2K\frac{|\nabla_1 u|^2+2u}{|\nabla u|^2+2u}\eta^2.
\end{aligned} \tag{5-16}$$

Note that by (5-1), Lemma 2.6 is applicable. Choose the cutoff function η in (5-5) to be $\varphi^{\tilde{x}}$ in Lemma 2.6, and consider the maximum of the quantity w on $Q_{\tilde{x}}$. From Lemma 2.6 (iii), we have

$$\eta(1+\lambda_1)\sum_{i\geq 1}\frac{-\nabla_{ii}\eta}{1+\lambda_i}\leq 2(1+\lambda_1)\eta. \tag{5-17}$$

Since $u(\bar{x})\geq C^{-1}{c_2'}^{-1}$ by Corollary 2.8, choosing $a=10Cc_2'$ in (5-16) and applying Lemma 2.6, we have

$$\begin{aligned} L_1 &\geq \frac{10}{|\nabla u|^2+2u} \geq 2C_1, \\ L_2 &\leq c_2'\Big(1+\frac{r_0}{\sqrt{c_2}}+c_3\frac{r_0}{\sqrt{c_2}}\Big), \\ L_3 &\leq C(1+c_4)\frac{r_0^2}{c_2}. \end{aligned} \tag{5-18}$$

From (5-15), (5-18), and (5-17), we have

$$\begin{aligned} (1+\lambda_1)\eta &\leq \max\left\{\sqrt{\frac{2L_3}{L_1}}, \frac{2(L_2+2)}{L_1}\right\} \\ &\leq C\left(1+\sqrt{c_4}\frac{r_0}{\sqrt{c_2}}+c_2'\left(1+\frac{r_0}{\sqrt{c_2}}+c_3\frac{r_0}{\sqrt{c_2}}\right)\right). \end{aligned} \tag{5-19}$$

Combining Lemma 2.6 (i) and (5-19), we have

$$(1+\lambda_1)(x) \leq \frac{e^{Cc_2'}}{r_0/r_2-u(\tilde{x})}\left(1+\sqrt{c_4}\frac{r_0}{\sqrt{c_2}}+c_2'\Big(1+\frac{r_0}{\sqrt{c_2}}+c_3\frac{r_0}{\sqrt{c_2}}\Big)\right)$$

on $B\big(\tilde{x}, \frac{1}{6}\sqrt{C_1}\big(r_0/r_2-u(\tilde{x})\big)\big)$. The proof of Proposition 5.1 is completed. □

Remark 5.2. Most computations in this section are just modifications of those in the classical theory of Monge–Ampère equations. A closer reference is [Guan and Li 1996], in which the Dirichlet problem of real Monge–Ampère equations on manifolds is systematically studied. The observation of this appendix is that these estimates can be localized under certain geometric conditions.

Acknowledgements

We are thankful to Professor J. X. Hong for beneficial comments and providing us with some references. We also thank Professor J. M. Schlenker for informing us of one interesting result in anti-de Sitter space. The first author is grateful to Professor S. T. Yau for many stimulating discussions.

References

[Cheng and Yau 1980] S. Y. Cheng and S. T. Yau, "On the existence of a complete Kähler metric on noncompact complex manifolds and the regularity of Fefferman's equation", *Comm. Pure Appl. Math.* **33**:4 (1980), 507–544. MR 82f:53074 Zbl 0506.53031

[Delanoë 1988] P. Delanoë, "Réalisations globalement régulières de disques strictement convexes dans les espaces d'Euclide et de Minkowski par la méthode de Weingarten", *Ann. Sci. École Norm. Sup.* (4) **21**:4 (1988), 637–652. MR 90f:53097 Zbl 0682.53062

[Efimov 1964] N. V. Efimov, "Generation of singularities on surfaces of negative curvature", *Mat. Sb. (N.S.)* **64 (106)** (1964), 286–320. In Russian. MR 29 #5203 Zbl 0126.37402

[Guan 1998] B. Guan, “The Dirichlet problem for Monge–Ampère equations in non-convex domains and spacelike hypersurfaces of constant Gauss curvature”, *Trans. Amer. Math. Soc.* **350**:12 (1998), 4955–4971. MR 99b:53055 Zbl 0919.35046

[Guan 2007] B. Guan, “Isometric embedding of negatively curved disks in the Minkowski space”, *Pure Appl. Math. Q.* **3**:3 (2007), 827–840. MR 2008k:53067 Zbl 1154.53030

[Guan and Li 1994] P. Guan and Y. Y. Li, “The Weyl problem with nonnegative Gauss curvature”, *J. Differential Geom.* **39**:2 (1994), 331–342. MR 95c:53051 Zbl 0796.53056

[Guan and Li 1996] B. Guan and Y. Y. Li, “Monge–Ampère equations on Riemannian manifolds”, *J. Differential Equations* **132**:1 (1996), 126–139. MR 97j:53040 Zbl 0866.58067

[Guan et al. 2006] B. Guan, H.-Y. Jian, and R. M. Schoen, “Entire spacelike hypersurfaces of prescribed Gauss curvature in Minkowski space”, *J. Reine Angew. Math.* **595** (2006), 167–188. MR 2007h:53088 Zbl 1097.53040

[Han 2005] Q. Han, “On the isometric embedding of surfaces with Gauss curvature changing sign cleanly”, *Comm. Pure Appl. Math.* **58**:2 (2005), 285–295. MR 2005g:53106 Zbl 1073.53005

[Han 2006] Q. Han, “Local isometric embedding of surfaces with Gauss curvature changing sign stably across a curve”, *Calc. Var. Partial Differential Equations* **25**:1 (2006), 79–103. MR 2006g:53090 Zbl 1101.53001

[Han and Hong 2006] Q. Han and J.-X. Hong, *Isometric embedding of Riemannian manifolds in Euclidean spaces*, Mathematical Surveys and Monographs **130**, Amer. Math. Soc., Providence, RI, 2006. MR 2008e:53055 Zbl 1113.53002

[Han et al. 2003] Q. Han, J.-X. Hong, and C.-S. Lin, “Local isometric embedding of surfaces with nonpositive Gaussian curvature”, *J. Differential Geom.* **63**:3 (2003), 475–520. MR 2004i:53081 Zbl 1070.53034

[Hilbert 1901] D. Hilbert, “Ueber Flächen von constanter Gaussscher Krümmung”, *Trans. Amer. Math. Soc.* **2**:1 (1901), 87–99. MR 1500557 JFM 32.0608.01

[Hong 1993] J. X. Hong, “Realization in $\mathbf{R}^3$ of complete Riemannian manifolds with negative curvature”, *Comm. Anal. Geom.* **1**:3-4 (1993), 487–514. MR 95d:53003 Zbl 0856.53004

[Hong 1997] J. Hong, “Isometric embedding in $\mathbf{R}^3$ complete noncompact nonnegatively curved surfaces”, *Manuscripta Math.* **94**:3 (1997), 271–286. MR 98m:53006 Zbl 0894.53050

[Hong 2001] J. Hong, “Recent developments of realization of surfaces in $\mathbf{R}^3$”, pp. 47–62 in *First International Congress of Chinese Mathematicians* (Beijing, 1998), edited by L. Yang and S. T. Yau, AMS/IP Stud. Adv. Math. **20**, Amer. Math. Soc., Providence, RI, 2001. MR 2002d:53081 Zbl 1056.53039

[Hong and Zuily 1995] J. Hong and C. Zuily, “Isometric embedding of the 2-sphere with nonnegative curvature in $\mathbf{R}^3$”, *Math. Z.* **219**:3 (1995), 323–334. MR 96e:53005 Zbl 0833.53049

[Lin 1985] C. S. Lin, “The local isometric embedding in $\mathbf{R}^3$ of 2-dimensional Riemannian manifolds with nonnegative curvature”, *J. Differential Geom.* **21**:2 (1985), 213–230. MR 87m:53073 Zbl 0584.53002

[Lin 1986] C. S. Lin, “The local isometric embedding in $\mathbf{R}^3$ of two-dimensional Riemannian manifolds with Gaussian curvature changing sign cleanly”, *Comm. Pure Appl. Math.* **39**:6 (1986), 867–887. MR 88e:53097 Zbl 0612.53013

[Nirenberg 1953] L. Nirenberg, “The Weyl and Minkowski problems in differential geometry in the large”, *Comm. Pure Appl. Math.* **6** (1953), 337–394. MR 15,347b Zbl 0051.12402

[Pogorelov 1952] A. V. Pogorelov, “Regularity of a convex surface with given Gaussian curvature”, *Mat. Sbornik N.S.* **31(73)** (1952), 88–103. In Russian. MR 14,679b Zbl 0048.40501

[Pogorelov 1973] A. V. Pogorelov, *Extrinsic geometry of convex surfaces*, Translations of Mathematical Monographs **35**, Amer. Math. Soc., Providence, RI, 1973. MR 49 #11439 Zbl 0311.53067

[Schlenker 2001] J.-M. Schlenker, "Surfaces à courbure extrinsèque négative dans l'espace hyperbolique", *Ann. Sci. École Norm. Sup.* (4) **34**:1 (2001), 79–130. MR 2003d:53096 Zbl 1029.53068

[Schoen and Yau 1981] R. Schoen and S. T. Yau, "Proof of the positive mass theorem, II", *Comm. Math. Phys.* **79**:2 (1981), 231–260. MR 83i:83045 Zbl 0494.53028

[Schulz 1990] F. Schulz, *Regularity theory for quasilinear elliptic systems and Monge–Ampère equations in two dimensions*, Lecture Notes in Mathematics **1445**, Springer, Berlin, 1990. MR 92d:35002 Zbl 0709.35038

[Weyl 1916] H. Weyl, "Über die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement", *Naturf. Ges.* **61** (1916), 40–72. JFM 46.1115.03

[Yau 2000] S. T. Yau, "Review of geometry and analysis", *Asian J. Math.* **4**:1 (2000), 235–278. MR 2002e:53002 Zbl 1031.53004

Received August 28, 2014.

BING-LONG CHEN
DEPARTMENT OF MATHEMATICS
SUN YAT-SEN UNIVERSITY
GUANGZHOU, 510275
CHINA
mcscbl@mail.sysu.edu.cn

LE YIN
COLLEGE OF MATHEMATICS AND COMPUTATIONAL SCIENCE
SHENZHEN UNIVERSITY
SHENZHEN, 518060
CHINA
lyin@szu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.369

THE COMPLEX MONGE–AMPÈRE EQUATION ON SOME COMPACT HERMITIAN MANIFOLDS

JIANCHUN CHU

We consider the complex Monge–Ampère equation on compact manifolds when the background metric is a Hermitian metric (in complex dimension 2) or a Hermitian metric satisfying an additional condition (in higher dimensions). We prove that the Laplacian estimate holds when F is in W^{1,q_0} for any $q_0 > 2n$. As an application, we show that, up to scaling, there exists a unique classical solution in W^{3,q_0} for the complex Monge–Ampère equation when F is in W^{1,q_0}.

1. Introduction

We consider the regularity problem of the complex Monge–Ampère equation on some compact Hermitian manifolds. Let (M, g) be a compact Hermitian manifold of complex dimension $n \geq 2$. For a real-valued function F on M, we consider the Monge–Ampère equation

$$\det(g_{i\bar{j}} + \phi_{i\bar{j}}) = e^F \det(g_{ij}),$$

with $(g_{i\bar{j}} + \phi_{i\bar{j}}) > 0$, for a real-valued function ϕ such that $\sup_M \phi = -1$. We write

$$\omega = \sqrt{-1} g_{i\bar{j}}\, dz^i \wedge d\bar{z}^j \quad \text{and} \quad \tilde{\omega} = \sqrt{-1} \tilde{g}_{i\bar{j}}\, dz^i \wedge d\bar{z}^j,$$

where $\tilde{g}_{i\bar{j}} = g_{i\bar{j}} + \phi_{i\bar{j}}$. Thus, the Monge–Ampère equation can be written as

$$\begin{cases} \tilde{\omega}^n = e^F \omega^n, \\ \tilde{\omega} = \omega + \sqrt{-1}\, \partial\bar{\partial}\phi > 0, \\ \sup_M \phi = -1. \end{cases} \tag{1-1}$$

For functions f, h and a holomorphic coordinate $z = (z^1, \ldots, z^n)$ we write

$$f_{i\bar{j}} = \frac{\partial^2 f}{\partial z^i \partial \bar{z}^j}, \qquad \Delta f = g^{i\bar{j}} f_{i\bar{j}}, \qquad \tilde{\Delta} f = \tilde{g}^{i\bar{j}} f_{i\bar{j}},$$
$$|\nabla f|^2 = g^{i\bar{j}} f_i f_{\bar{j}}, \quad |\tilde{\nabla} f|^2 = \tilde{g}^{i\bar{j}} f_i f_{\bar{j}}, \quad \langle \nabla f, \nabla h \rangle = g^{i\bar{j}} f_i h_{\bar{j}}.$$

MSC2010: 35J96, 53C55.
Keywords: complex Monge–Ampère equation, compact Hermitian manifold.

We use $\|f\|_{L^p(M,\omega)}$ and $\|\nabla^m f\|_{L^p(M,\omega)}$ to denote the corresponding norms with respect to (M,ω).

When ω is Kähler, the complex Monge–Ampère equation is very important. Calabi [1957] presented his famous conjecture and transformed that problem into (1-1). Yau [1978] proved the existence of the classical solution of (1-1) by using the continuity method and solved Calabi's conjecture.

The Dirichlet problem for the complex Monge–Ampère equation is also very important. Bedford and Taylor [1976; 1982] studied the weak solution. After their work, weak solutions of the complex Monge–Ampère equation have been studied extensively. There are many existence, uniqueness and regularity results of the complex Monge–Ampère equation under different conditions, and we refer the reader to [Błocki 2005; Demailly and Pali 2010; Dinew 2009; Eyssidieux et al. 2009; Guedj and Zeriahi 2007; Kołodziej 1998; 2008; Zhang 2006].

On the other hand, the classical solvability of the Dirichlet problem was established by Caffarelli, Kohn, Nirenberg and Spruck [1985] for strongly pseudoconvex domains in $\mathbb{C}^n$. The reader can also see [Krylov 1989; Krylov 1994]. For further information, we refer the reader to [Phong et al. 2012], which is a survey of some recent developments in the theory of the complex Monge–Ampère equation.

When ω is not Kähler, the existence of the solution of the complex Monge–Ampère equation has been studied under some assumptions on ω (see [Cherrier 1987; Guan and Li 2009; Hanani 1996; Tosatti and Weinkove 2010b]). For a general ω, Tosatti and Weinkove [2010a] obtained the key C^0-estimate. As an application, they showed that, up to scaling, the complex Monge–Ampère equation on a compact Hermitian manifold admits a smooth solution when the right hand side F is smooth.

Chen and He [2012] have proved that, on a compact Kähler manifold of complex dimension n, the Laplacian estimate and the gradient estimate hold and there exists a classical solution in W^{3,q_0} for the complex Monge–Ampère equation when the right-hand side F is in W^{1,q_0} for any $q_0 > 2n$.

In this paper, we generalize the work of Chen and He. We use a different method (we don't need the gradient estimate to get the Laplacian estimate) to consider the regularity problem of (1-1) on some compact Hermitian manifolds (including compact Kähler manifolds).

Definition 1.1. A compact Hermitian manifold (M,ω) of complex dimension n satisfies condition $(*)$ if, for any $\phi \in C^2(M)$ such that

$$\tilde{\omega} = \omega + \sqrt{-1}\,\partial\bar{\partial}\phi > 0, \quad \|\phi\|_{L^\infty(M,\omega)} \le \Lambda_1 \quad \text{and} \quad \Lambda_2^{-1}\omega^n \le \tilde{\omega}^n \le \Lambda_2\omega^n,$$

there exists a constant $C = C(\Lambda_1, \Lambda_2, M, \omega)$ such that

$$-C\omega^n \le \sqrt{-1}\partial\bar{\partial}\tilde{\omega}^{n-1} \le C\omega^n.$$

Remark 1.2. When $n = 2$, condition $(*)$ is trivial. Since

$$\partial\bar{\partial}\tilde{\omega} = \partial\bar{\partial}\omega,$$

all compact Hermitian manifolds of complex dimension 2 satisfy condition $(*)$.

Remark 1.3. When $n = 3$, if (M, ω) is a compact Hermitian manifold satisfying

$$\partial\bar{\partial}\omega = 0,$$

then we have

$$\partial\bar{\partial}\tilde{\omega}^2 = 2\partial\omega \wedge \bar{\partial}\omega,$$

which implies this Hermitian manifold (M, ω) satisfies condition $(*)$.

Remark 1.4. When $n \geq 4$, condition $(*)$ is not a very strong restricted condition. For example, if (M, ω) is a compact Hermitian manifold satisfying

$$\partial\bar{\partial}\omega = 0 \text{ and } \partial\bar{\partial}\omega^2 = 0, \tag{1-2}$$

then we can conclude that $\partial\bar{\partial}\omega^k = 0$ for all $1 \leq k \leq n-1$ (see, for example, [Fino and Tomassini 2011]), which implies that $\partial\bar{\partial}\tilde{\omega}^k = 0$ for all $1 \leq k \leq n-1$. Thus, such a Hermitian manifold (satisfying (1-2)) satisfies condition $(*)$. For example, the products of a complex curve with a Kähler metric and a complex surface with a non-Kähler Gauduchon metric satisfy (1-2). More examples are constructed in [Fino and Tomassini 2011].

Remark 1.5. All compact Kähler manifolds satisfy condition $(*)$.

Now, we state our Laplacian estimate as follows.

Theorem 1.6. *Let (M, ω) be a compact Hermitian manifold of complex dimension n. Assume that either*

(1) $n = 2$, *or*

(2) $n \geq 3$ *and (M, ω) satisfies condition $(*)$.*

For any $q_0 > 2n$, if ϕ is a smooth solution of (1-1), *then*

$$\|n + \Delta\phi\|_{L^\infty(M,\omega)} \leq C(\|F\|_{W^{1,q_0}(M,\omega)}, q_0, M, \omega).$$

Usually, we need the gradient estimate to derive the Laplacian estimate. However, the computation on Hermitian manifolds is more complicated due to the existence of torsion terms. As a result, the gradient estimate is very difficult to obtain. In order to solve this problem, we introduce a new method to obtain the Laplacian estimate directly. By using Moser's iteration [1960], L^p estimates (for example, see [Gilbarg and Trudinger 1977]) and some interpolation inequalities, we can obtain the Laplacian estimate without doing any calculations involving the gradient, which makes the argument simpler and clearer. Therefore, we believe that our ideas can be applied to other nonlinear equations on compact manifolds.

As an application of Theorem 1.6, we have the following theorem:

Theorem 1.7. *Assume that* (M, ω) *satisfies condition* (1) *or* (2) *of Theorem 1.6. Let F be a function in W^{1,q_0} for any $q_0 > 2n$. Then there exist a function $\phi \in W^{3,q_0}$ and a constant b such that*

$$\begin{cases} \tilde{\omega}^n = e^{F+b}\omega^n, \\ \tilde{\omega} = \omega + \sqrt{-1}\partial\bar{\partial}\phi > 0, \\ \sup_M \phi = -1. \end{cases}$$

2. Some preliminary computations

We need the following C^0-estimate from [Tosatti and Weinkove 2010a]:

Theorem 2.1. *For any compact Hermitian manifold (M, ω), if ϕ is a smooth solution of* (1-1), *then we have*

$$\|\phi\|_{L^\infty(M,\omega)} \le C,$$

where $C = C(\sup_M F, M, \omega)$.

We need the following lemma from [Tosatti and Weinkove 2015]:

Lemma 2.2. *Let (M, ω) be a compact Hermitian manifold of complex dimension n. If ϕ is a smooth solution of* (1-1), *then, for any $\epsilon > 0$, we have*

$$\tilde{\Delta}(\Delta\phi) + (\epsilon - 1)\frac{|\tilde{\nabla}(\Delta\phi)|^2}{(n + \Delta\phi)} \ge \Delta F - A(1 + 1/\epsilon)(n + \Delta\phi)(n - \tilde{\Delta}\phi), \tag{2-1}$$

where $A = A(M, \omega, \|F\|_{L^\infty(M,\omega)})$.

Proof. We need the following equation, which is [Tosatti and Weinkove 2015, (9.5)]:

$$\tilde{\Delta}(\log(\mathrm{tr}_g\,\tilde{g})) \ge \frac{2}{(\mathrm{tr}_g\,\tilde{g})^2}\,\mathrm{Re}(\tilde{g}^{k\bar{l}}T^i_{ik}(\mathrm{tr}_g\,\tilde{g})_{\bar{l}}) + \frac{\Delta F}{\mathrm{tr}_g\,\tilde{g}} - C_1\,\mathrm{tr}_{\tilde{g}}\,g - C_1,$$

where the tensor T is the torsion of (M, ω) and $C_1 = C_1(M, \omega, \|F\|_{L^\infty(M,\omega)})$. After some calculations, we have

$$\tilde{\Delta}(\Delta\phi) - \frac{|\tilde{\nabla}(\Delta\phi)|^2}{(n + \Delta\phi)} \ge \frac{2}{(n + \Delta\phi)}\,\mathrm{Re}(\tilde{g}^{k\bar{l}}T^i_{ik}(\Delta\phi)_{\bar{l}}) + \Delta F - C_2(n + \Delta\phi)(n - \tilde{\Delta}\phi),$$

where $C_2 = C_2(M, \omega, \|F\|_{L^\infty(M,\omega)})$; we have used that $\operatorname{tr}_{\tilde{g}} g = (n - \tilde{\Delta}\phi) \geq ne^{-F/n}$. By the Cauchy–Schwarz inequality, for any $\epsilon > 0$, we have that

$$\begin{aligned}\tilde{\Delta}(\Delta\phi) - \frac{|\tilde{\nabla}(\Delta\phi)|^2}{(n+\Delta\phi)}& \\ \geq -\epsilon\frac{|\tilde{\nabla}(\Delta\phi)|^2}{(n+\Delta\phi)} &- \frac{A}{\epsilon}(n+\Delta\phi)(n-\tilde{\Delta}\phi) + \Delta F - A(n+\Delta\phi)(n-\tilde{\Delta}\phi),\end{aligned}$$

where $A = A(M, \omega, \|F\|_{L^\infty(M,\omega)})$ and we have used that $(n + \Delta\phi) \geq ne^{F/n}$. □

Lemma 2.3. *Let (M, ω) be a compact Hermitian manifold of complex dimension n. If ϕ is a smooth solution of* (1-1), *then, for any $p \geq 1$, we have*

$$\begin{aligned}&\tilde{\Delta}(e^{f_p(\phi)}(n+\Delta\phi)^p) \\ &\qquad \geq C_1(p)(n+\Delta\phi)^{p+\frac{1}{n-1}} - C_2(p)(n+\Delta\phi)^p + pe^{f_p(\phi)}(n+\Delta\phi)^{p-1}\Delta F,\end{aligned}$$

where

$$\begin{aligned}f_p(\phi) &= e^{-A(p+3)\phi}, & A &= A(\|F\|_{L^\infty(M,\omega)}, M, \omega),\\ C_1(p) &= C_1(p, \|F\|_{L^\infty(M,\omega)}, M, \omega), & C_2(p) &= C_2(p, \|F\|_{L^\infty(M,\omega)}, M, \omega).\end{aligned}$$

Proof. By direct calculation, we have

$$\begin{aligned}\tilde{\Delta}(e^{f_p(\phi)}&(n+\Delta\phi)^p) \\ &= f_p' e^{f_p(\phi)}(\tilde{\Delta}\phi)(n+\Delta\phi)^p + (f_p'^2 + f_p'')e^{f_p(\phi)}|\tilde{\nabla}\phi|^2(n+\Delta\phi)^p \\ &\quad + pe^{f_p(\phi)}\tilde{\Delta}(\Delta\phi)(n+\Delta\phi)^{p-1} + p(p-1)e^{f_p(\phi)}|\tilde{\nabla}(\Delta\phi)|^2(n+\Delta\phi)^{p-2} \\ &\quad + 2pf_p' e^{f_p(\phi)}(n+\Delta\phi)^{p-1}\operatorname{Re}(\tilde{g}^{k\bar{l}}\phi_k(\Delta\phi)_{\bar{l}}).\end{aligned} \tag{2-2}$$

By the definition of $f_p(\phi)$, we have

$$\begin{cases} f_p'(\phi) = -A(p+3)e^{-A(p+3)\phi} < 0, \\ f_p''(\phi) = A^2(p+3)^2 e^{-A(p+3)\phi} > 0. \end{cases} \tag{2-3}$$

Thus, by the Cauchy–Schwarz inequality, we have

$$2\operatorname{Re}(\tilde{g}^{k\bar{l}}\phi_k(\Delta\phi)_{\bar{l}}) \leq \frac{(f_p'^2 + f_p'')(n+\Delta\phi)}{-pf_p'}|\tilde{\nabla}\phi|^2 + \frac{-pf_p'}{(f_p'^2 + f_p'')(n+\Delta\phi)}|\tilde{\nabla}(\Delta\phi)|^2,$$

which implies that

$$\begin{aligned}2pf_p' e^{f_p(\phi)}&(n+\Delta\phi)^{p-1}\operatorname{Re}(\tilde{g}^{k\bar{l}}\phi_k(\Delta\phi)_{\bar{l}}) \\ &\geq -(f_p'^2 + f_p'')e^{f_p(\phi)}|\tilde{\nabla}\phi|^2(n+\Delta\phi)^p \\ &\qquad - \frac{p^2 f_p'^2}{f_p'^2 + f_p''}e^{f_p(\phi)}(n+\Delta\phi)^{p-2}|\tilde{\nabla}(\Delta\phi)|^2.\end{aligned} \tag{2-4}$$

Combining (2-2) and (2-4), we have

$$\begin{aligned}
&\tilde{\Delta}(e^{f_p(\phi)}(n+\Delta\phi)^p)\\
&\qquad \geq f_p' e^{f_p(\phi)}(n+\Delta\phi)^p\tilde{\Delta}\phi + pe^{f_p(\phi)}\tilde{\Delta}(\Delta\phi)(n+\Delta\phi)^{p-1}\\
&\qquad\qquad + |\tilde{\nabla}(\Delta\phi)|^2(n+\Delta\phi)^{p-2}e^{f_p(\phi)}\left(p(p-1)-\frac{p^2 f_p'^2}{f_p'^2+f_p''}\right)\\
&\qquad \geq pe^{f_p(\phi)}(n+\Delta\phi)^{p-1}\left(\tilde{\Delta}(\Delta\phi)+\left(\frac{pf_p''}{(f_p')^2+f_p''}-1\right)\frac{|\tilde{\nabla}(\Delta\phi)|^2}{(n+\Delta\phi)}\right)\\
&\qquad\qquad + f_p' e^{f_p(\phi)}(n+\Delta\phi)^p\tilde{\Delta}\phi.
\end{aligned}$$

By Lemma 2.2 (take $\epsilon = pf_p''/((f_p')^2+f_p'')$), we obtain

$$\begin{aligned}
&\tilde{\Delta}(e^{f_p(\phi)}(n+\Delta\phi)^p)\\
&\qquad \geq f_p' e^{f_p(\phi)}(n+\Delta\phi)^p\tilde{\Delta}\phi + pe^{f_p(\phi)}(n+\Delta\phi)^{p-1}\Delta F\\
&\qquad\qquad - Ape^{f_p(\phi)}(n+\Delta\phi)^p(n-\tilde{\Delta}\phi)\left(1+\frac{(f_p')^2+f_p''}{pf_p''}\right)\\
&\qquad = nf_p' e^{f_p(\phi)}(n+\Delta\phi)^p + pe^{f_p(\phi)}(n+\Delta\phi)^{p-1}\Delta F\\
&\qquad\qquad + e^{f_p(\phi)}(n+\Delta\phi)^p(n-\tilde{\Delta}\phi)\left(-f_p' - Ap\left(1+\frac{(f_p')^2+f_p''}{pf_p''}\right)\right)\\
&\qquad \geq nf_p' e^{f_p(\phi)}(n+\Delta\phi)^p + pe^{f_p(\phi)}(n+\Delta\phi)^{p-1}\Delta F\\
&\qquad\qquad + Ae^{f_p(\phi)}(n+\Delta\phi)^p(n-\tilde{\Delta}\phi),
\end{aligned}\tag{2-5}$$

where we have used that $\sup_M \phi = -1$ and (2-3). It is clear that

$$\operatorname{tr}_g \tilde{g} \leq (\operatorname{tr}_{\tilde{g}} g)^{n-1}\frac{\det\tilde{g}}{\det g},$$

which implies that

$$(n+\Delta\phi) \leq (n-\tilde{\Delta}\phi)^{n-1}e^F. \tag{2-6}$$

Combining this with (2-5) and (2-6), the proof is complete. □

For convenience, we introduce some notation here: we set

$$u = e^{f_1(\phi)}(n+\Delta\phi). \tag{2-7}$$

Thus, by Young's inequality and Lemma 2.3, we have

$$\tilde{\Delta}u \geq e^{f_1(\phi)}\Delta F - \tilde{C}, \tag{2-8}$$

where $\tilde{C} = \tilde{C}(\|F\|_{L^\infty(M,\omega)}, M, \omega)$.

3. The Laplacian estimate

We remark that in this section our constants may differ from line to line.

Lemma 3.1. *Let (M, ω) be a compact Hermitian manifold. If ϕ is a smooth solution of* (1-1), *then, for any $f \in C^{\infty}(M)$, we have*

$$|\nabla f|^2 \le Cu|\tilde{\nabla} f|^2,$$

where u is defined in (2-7) *and $C = C(\|F\|_{L^{\infty}(M,\omega)}, M, \omega)$.*

Proof. By direct calculation, we have

$$|\nabla f|^2 \le (n+\Delta\phi)|\tilde{\nabla} f|^2.$$

Combining this with (2-7) and Theorem 2.1, the proof is complete. □

Lemma 3.2. *Under the assumptions of Theorem 1.6, for any $p \ge 0$, we have*

$$\begin{aligned}\int_M |\nabla(u^{\frac{p}{2}})|^2\omega^n \\ \le C(p^2+1)\int_M u^p(1+|\nabla F|^2)\omega^n + Cp\int_M u^p|\nabla\phi||\nabla F|\omega^n + C\int_M u^{p+1}\omega^n,\end{aligned}$$

where u is defined in (2-7) *and $C = C(\|F\|_{L^{\infty}(M,\omega)}, M, \omega)$.*

Proof. By Lemma 3.1 and direct calculation, we have

$$\begin{aligned}\int_M |\nabla(u^{\frac{p}{2}})|^2\omega^n &\le C_1\int_M u|\tilde{\nabla}(u^{\frac{p}{2}})|^2\tilde{\omega}^n \\ &= C_1 np\sqrt{-1}\int_M \partial u^p \wedge \bar{\partial} u \wedge \tilde{\omega}^{n-1} \\ &= -C_1 np\sqrt{-1}\int_M u^p\,\partial\bar{\partial} u \wedge \tilde{\omega}^{n-1} + \frac{C_1 np}{p+1}\sqrt{-1}\int_M \bar{\partial} u^{p+1} \wedge \partial\tilde{\omega}^{n-1} \\ &= -C_1 p\int_M u^p(\tilde{\Delta} u)\tilde{\omega}^n - \frac{C_1 np}{p+1}\sqrt{-1}\int_M u^{p+1}\,\partial\bar{\partial}\tilde{\omega}^{n-1},\end{aligned}$$

where $C_1 = C_1(\|F\|_{L^{\infty}(M,\omega)}, M, \omega)$. Since M satisfies condition $(*)$ (when $n = 2$, all Hermitian manifolds satisfy condition $(*)$), we have

$$-\frac{C_1 np}{p+1}\sqrt{-1}\int_M u^{p+1}\,\partial\bar{\partial}\tilde{\omega}^{n-1} \le C_2\int_M u^{p+1}\omega^n,$$

where $C_2 = C_2(\|F\|_{L^{\infty}(M,\omega)}, M, \omega)$ (Since $n = \dim_{\mathbb{C}} M$, we can absorb it into the constant C_2). By (2-8) and $\tilde{\omega}^n = e^F\omega^n$, we obtain

$$\begin{aligned}
-C_1 p\int_M u^p(\tilde{\Delta}u)\tilde{\omega}^n &\le C_3 p\int_M u^p(\tilde{C}-e^{f_1(\phi)}\Delta F)\tilde{\omega}^n\\
&\le C_3\tilde{C}p\int_M u^p\tilde{\omega}^n - C_3 p\int_M e^{f_1(\phi)}u^p(\Delta(e^F)-e^F|\nabla F|^2)\omega^n\\
&\le C_4 p\int_M u^p(1+|\nabla F|^2)\omega^n + C_3 p\int_M\langle\nabla(e^{f_1(\phi)}u^p),\nabla(e^F)\rangle\omega^n\\
&\qquad -\sqrt{-1}C_3 np\int_M e^{f_1(\phi)}u^p\bar{\partial}e^F\wedge\partial\omega^{n-1},
\end{aligned}$$

where $C_3 = C_3(\|F\|_{L^\infty(M,\omega)}, M, \omega)$, $C_4 = C_4(\|F\|_{L^\infty(M,\omega)}, M, \omega)$. It is clear that

$$\begin{aligned}
&C_3 p\int_M\langle\nabla(e^{f_1(\phi)}u^p),\nabla(e^F)\rangle\omega^n\\
&\quad = C_3 p\int_M u^p\langle\nabla(e^{f_1(\phi)}),\nabla(e^F)\rangle\omega^n + C_3 p\int_M e^{f_1(\phi)}\langle\nabla(u^p),\nabla(e^F)\rangle\omega^n\\
&\quad \le C_5 p\int_M u^p|\nabla F||\nabla\phi|\omega^n + \frac{1}{2}\int_M|\nabla u^{\frac{p}{2}}|^2\omega^n + C_5 p^2\int_M u^p|\nabla F|^2\omega^n,
\end{aligned}$$

where $C_5 = C_5(\|F\|_{L^\infty(M,\omega)}, M, \omega)$. Here we have used the Cauchy–Schwarz inequality. We notice that

$$-\sqrt{-1}C_3 np\int_M e^{f_1(\phi)}u^p\,\bar{\partial}e^F\wedge\partial\omega^{n-1} \le C_6 p\int_M u^p|\nabla F|\omega^n,$$

where $C_6 = C_6(\|F\|_{L^\infty(M,\omega)}, M, \omega)$ (Since $n = \dim_{\mathbb{C}} M$, we can absorb it into the constant C_6). Combining the above inequalities, we complete the proof. □

Theorem 3.3. *Under the assumptions of Theorem 1.6, we have*

$$\|u\|_{L^\infty(M,\omega)} \le C\big(\|u\|_{L^{\frac{q_0}{2}}(M,\omega)}, \|F\|_{W^{1,q_0}(M,\omega)}, q_0, M, \omega\big).$$

Proof. Without loss of generality, we can assume that $q_0 < \infty$. We use the iteration method (see [Moser 1960]). By the Sobolev inequality (Corollary A.2) and Lemma 3.2, for $p \ge 1$ we have

$$\begin{aligned}
&\left(\int_M u^{p\beta}\omega^n\right)^{\frac{1}{\beta}}\\
&\quad\le C_1\int_M u^p\omega^n + C_1\int_M|\nabla(u^{\frac{p}{2}})|^2\omega^n\\
&\quad\le C_1\int_M u^p\omega^n + C_1 p^2\int_M u^p(1+|\nabla F|^2)\omega^n\\
&\qquad + C_1 p\int_M u^p|\nabla\phi||\nabla F|\omega^n + C_1\int_M u^{p+1}\omega^n\\
&\quad\le C_1 p^2\int_M u^{p+1}\omega^n + C_1 p^2\int_M u^p|\nabla F|^2\omega^n + C_1 p^2\int_M u^p|\nabla\phi||\nabla F|\omega^n,
\end{aligned}$$

where $\beta = n/(n-1)$ and $C_1 = C_1(\|F\|_{L^\infty(M,\omega)}, M, \omega)$. Here we have used Young's inequality and the inequality $p \le p^2$. By the Hölder inequality, we have

$$\int_M u^p |\nabla F|^2 \omega^n \le \left(\int_M u^{pr_0}\omega^n\right)^{\frac{1}{r_0}} \left(\int_M |\nabla F|^{q_0}\omega^n\right)^{\frac{2}{q_0}}$$

and

$$\int_M u^p |\nabla\phi||\nabla F|\omega^n \le \left(\int_M u^{pr_0}\omega^n\right)^{\frac{1}{r_0}} \left(\int_M |\nabla\phi|^{q_0}\omega^n\right)^{\frac{1}{q_0}} \left(\int_M |\nabla F|^{q_0}\omega^n\right)^{\frac{1}{q_0}},$$

where $1/r_0 + 2/q_0 = 1$. Combining the above inequalities, when $pr_0 \ge p+1$ (that is, $p \ge (q_0-2)/2$), we obtain

$$\begin{aligned}
\|u\|_{L^{p\beta}(M,\omega)} &\le (C_2 p^2(\|\nabla\phi\|_{L^{q_0}(M,\omega)}+1))^{\frac{1}{p}} \big(\|u\|^{\frac{p+1}{p}}_{L^{p+1}(M,\omega)} + \|u\|_{L^{pr_0}(M,\omega)}\big)\\
&\le (C_2 p^2(\|\nabla\phi\|_{L^{q_0}(M,\omega)}+1))^{\frac{1}{p}} \|u\|^{\frac{p+1}{p}}_{L^{pr_0}(M,\omega)},
\end{aligned}$$

where $C_2 = C_2(\|F\|_{W^{1,q_0}(M,\omega)}, q_0, M, \omega)$. By Lemma A.6, we have

$$\begin{aligned}
\|\nabla\phi\|_{L^{q_0}(M,\omega)} &\le C_3\|u\|_{L^{\frac{2nq_0}{2n+q_0}}(M,\omega)} + C_3\\
&\le C_3\|u\|_{L^{\frac{q_0}{2}}(M,\omega)} + C_3,
\end{aligned}$$

where $C_3 = C_3(q_0, \|F\|_\infty, M, \omega)$. Thus, for any $k \ge 0$, we have

$$\|u\|_{L^{p_k\beta}(M,\omega)} \le a_k \|u\|^{b_k}_{L^{p_k r_0}(M,\omega)}, \tag{3-1}$$

where

$$\begin{aligned}
a_k &= \big(C_4 p_k^2\big(\|u\|_{L^{\frac{q_0}{2}}(M,\omega)}+1\big)\big)^{\frac{1}{p_k}}, &\quad C_4 &= C_4(\|F\|_{W1,q_0(M,\omega)}, q_0, M, \omega),\\
b_k &= \frac{p_k+1}{p_k}, & p_k &= \frac{q_0-2}{2}\left(\frac{\beta}{r_0}\right)^k.
\end{aligned}$$

Here we point out that $q_0 > 2n$ implies that $\beta/r_0 > 1$. By (3-1), we have

$$\|u\|_{L^{p_k\beta}(M,\omega)} \le a_k a_{k-1}^{b_k} \cdots a_0^{b_k\cdots b_1} \|u\|^{b_k\cdots b_0}_{L^{p_0 r_0}(M,\omega)}. \tag{3-2}$$

Without loss of generality, we can assume that $a_k \ge 1$ for $k \ge 0$. We observe that $\prod_{i=0}^\infty b_k$ and $\prod_{i=0}^\infty a_k$ are convergent. In (3-2), letting $k \to \infty$, we obtain

$$\|u\|_{L^\infty(M,\omega)} \le C\big(\|u\|_{L^{\frac{q_0}{2}}(M,\omega)}, \|F\|_{W^{1,q_0}(M,\omega)}, q_0, M, \omega\big). \qquad \square$$

Lemma 3.4. *Under the assumptions of Theorem 1.6, for any $p \ge 1$, we have*

$$\int_M u^{p+\frac{1}{n-1}}\omega^n \le C(p)\int_M u^{p-1}|\nabla\phi||\nabla F|\omega^n + C(p)\int_M u^{p-1}|\nabla F|^2\omega^n + C(p),$$

where $C(p) = C(p, \|F\|_{L^\infty(M,\omega)}, M, \omega)$.

Proof. Starting with Lemma 2.3 and then integrating over $(M, \tilde{\omega})$, for any $p \geq 1$ we obtain

$$\begin{aligned}\int_M & \tilde{\Delta}(e^{f_p(\phi)}(n+\Delta\phi)^p)\tilde{\omega}^n \\ &\geq C_1(p)\int_M u^{p+\frac{1}{n-1}}\tilde{\omega}^n - C_2(p)\int_M u^p\tilde{\omega}^n + p\int_M e^{f_p(\phi)}(n+\Delta\phi)^{p-1}\Delta F e^F\omega^n,\end{aligned}$$

where $C_1(p) = C_1(p, \|F\|_{L^\infty(M,\omega)}, M, \omega)$ and $C_2(p) = C_2(p, \|F\|_{L^\infty(M,\omega)}, M, \omega)$. Here we have used (2-7) and Theorem 2.1. Since M satisfies condition $(*)$, we have

$$\begin{aligned}\int_M \tilde{\Delta}(e^{f_p(\phi)}(n+\Delta\phi)^p)\tilde{\omega}^n &= n\sqrt{-1}\int_M \partial\bar{\partial}(e^{f_p(\phi)}(n+\Delta\phi)^p)\wedge\tilde{\omega}^{n-1} \\ &= n\sqrt{-1}\int_M e^{f_p(\phi)}(n+\Delta\phi)^p\,\partial\bar{\partial}\tilde{\omega}^{n-1} \\ &\leq C_3(p)\int_M u^p\omega^n,\end{aligned}$$

where $C_3(p) = C_3(p, \|F\|_{L^\infty(M,\omega)}, M, \omega)$ (Since $n = \dim_{\mathbb{C}} M$, we can absorb it into the constant C_3). Combining the above inequalities, we compute that

$$\begin{aligned}\int_M & u^{p+\frac{1}{n-1}}\omega^n \\ &\leq C_4(p)\int_M e^{f_p(\phi)}(n+\Delta\phi)^{p-1}(|\nabla F|^2 e^F - \Delta(e^F))\omega^n + C_5(p)\int_M u^p\omega^n \\ &\leq C_5(p)\int_M u^{p-1}|\nabla F|^2\omega^n + C_4(p)\int_M \langle\nabla(e^{f_p(\phi)}(n+\Delta\phi)^{p-1}), \nabla e^F\rangle\omega^n \\ &\qquad - C_4(p)n\sqrt{-1}\int_M e^{f_p}(n+\Delta\phi)^{p-1}\,\bar{\partial}e^F\wedge\partial\omega^{n-1} + C_5(p)\int_M u^p\omega^n \\ &\leq C_5(p)\int_M u^p\omega^n + C_5(p)\int_M u^{p-1}|\nabla F|^2\omega^n + C_5(p)\int_M u^{p-1}|\nabla F|\omega^n \\ &\qquad + C_5(p)\int_M |\nabla(u^{p-1})||\nabla F|\omega^n + C_5(p)\int_M u^{p-1}|\nabla\phi||\nabla F|\omega^n,\end{aligned}$$

where $C_4(p) = C_4(p, \|F\|_{L^\infty(M,\omega)}, M, \omega)$ and $C_5(p) = C_5(p, \|F\|_{L^\infty(M,\omega)}, M, \omega)$ (Since $n = \dim_{\mathbb{C}} M$, we can absorb it into the constant C_5). By the Cauchy–Schwarz inequality, we have

$$\begin{aligned}C_5(p)\int_M |\nabla(u^{p-1})||\nabla F|\omega^n &= C_5(p)\int_M |\nabla(u^{\frac{p-1}{2}})|u^{\frac{p-1}{2}}|\nabla F|\omega^n \\ &\leq C_5(p)\int_M |\nabla(u^{\frac{p-1}{2}})|^2\omega^n + C_5(p)\int_M u^{p-1}|\nabla F|^2\omega^n.\end{aligned}$$

Combining this with the above inequalities and Lemma 3.2, we get

$$\begin{aligned}\int_M & u^{p+\frac{1}{n-1}}\omega^n \\ &\le C_6(p)\int_M u^p\omega^n + C_6(p)\int_M u^{p-1}|\nabla\phi||\nabla F|\omega^n + C_6(p)\int_M u^{p-1}|\nabla F|^2\omega^n,\end{aligned}$$

where $C_6(p) = C_6(p, \|F\|_{L^\infty(M,\omega)}, M, \omega)$. Using Young's inequality, we complete the proof. □

Now, we are in a position to prove Theorem 1.6.

Proof of Theorem 1.6. Without loss of generality, we assume that $q_0 < \infty$. By Lemma 3.4 and $F \in W^{1,q_0}$, for any $p \ge 1$, we have

$$\begin{aligned}\int_M u^{p+\frac{1}{n-1}}\omega^n &\le C_1(p)\int_M u^{p-1}|\nabla\phi||\nabla F|\omega^n + C_1(p)\int_M u^{p-1}|\nabla F|^2\omega^n + C_1(p) \\ &\le C_1(p)\int_M u^{p-1}|\nabla\phi|^2\omega^n + C_2(p)\int_M u^{(p-1)\frac{q_0}{q_0-2}}\omega^n + C_2(p),\end{aligned}$$

where $C_1(p) = C_1(p, \|F\|_{L^\infty(M,\omega)}, M, \omega)$, $C_2(p) = C_2(p, \|F\|_{W^{1,q_0}(M,\omega)}, q_0, M, \omega)$ and we have used the Hölder inequality in the last line. When $p \ge 1$ satisfies that

$$p + \frac{1}{n-1} > (p-1)\frac{q_0}{q_0-2}, \quad \text{or equivalently} \quad p < \frac{q_0-2}{2n-2} + \frac{q_0}{2},$$

we can use Young's inequality to get the inequality

$$\int_M u^{p+\frac{1}{n-1}}\omega^n \le C_3(p)\int_M u^{p-1}|\nabla\phi|^2\omega^n + C_3(p),$$

where $C_3(p) = C_3(p, \|F\|_{W^{1,q_0}(M,\omega)}, q_0, M, \omega)$. Now, we take $p = q_0/2 - 1/(n-1)$, we obtain

$$\begin{aligned}\int_M u^{\frac{q_0}{2}}\omega^n &\le C_4\int_M u^{\frac{q_0}{2}-\beta}|\nabla\phi|^2\omega^n + C_4 \\ &\le \frac{1}{2}\int_M u^{(\frac{q_0}{2}-\beta)\frac{q_0}{q_0-2\beta}}\omega^n + C_4\int_M |\nabla\phi|^{\frac{q_0}{\beta}}\omega^n + C_4,\end{aligned}$$

where $C_4 = C_4(\|F\|_{W^{1,q_0}(M,\omega)}, q_0, M, \omega)$ and $\beta = n/(n-1)$. It then follows that

$$\|u\|_{L^{\frac{q_0}{2}}(M,\omega)} \le C_4\|\nabla\phi\|^{\frac{2}{\beta}}_{L^{\frac{q_0}{\beta}}(M,\omega)} + C_4. \tag{3-3}$$

By Lemma A.7, we have

$$\|\nabla\phi\|_{L^{\frac{q_0}{\beta}}(M,\omega)} \le C_5\|u\|^{\frac{1}{2}}_{L^{\frac{q_0}{2\beta}}(M,\omega)} + C_5, \tag{3-4}$$

where $C_5 = C_5(q_0, \|F\|_{L^\infty(M,\omega)}, M, \omega)$. Combining (3-3), (3-4) and $\beta > 1$, we get

$$\|u\|_{L^{\frac{q_0}{2}}(M,\omega)} \leq C_6(\|F\|_{W^{1,q_0}(M,\omega)}, q_0, M, \omega).$$

By Theorem 3.3, we complete the proof. □

4. The Hölder estimate of second order, and solving the equation

We note that when F is in W^{1,q_0}, for any $q_0 > 2n$, Sobolev embedding implies that $F \in C^{\alpha_0}$, where $\alpha_0 = 1 - 2n/q_0$. By Theorem 1.1 of [Tosatti et al. 2014], we have the following theorem:

Theorem 4.1. *Let (M, ω) be a compact Hermitian manifold. If ϕ is a smooth solution of* (1-1) *and $F \in C^{\alpha_0}$, then there exists a constant $\alpha \in (0, 1)$ such that*

$$\|\phi\|_{C^{2,\alpha}(M,\omega)} \leq C,$$

where α and C depend only on $\|\phi\|_{L^\infty(M,\omega)}$, $\|\Delta\phi\|_{L^\infty(M,\omega)}$, α_0, $\|F\|_{C^{\alpha_0}(M,\omega)}$, q_0, M and ω.

Now we are in a position to prove Theorem 1.7.

Proof of Theorem 1.7. Our argument here is similar to the argument in [Chen and He 2012]. Let $F \in W^{1,q_0}$ on M such that $\|F\|_{W^{1,q_0}(M,\omega)} \leq \Lambda$ for some positive constant Λ. Let $\{F_k\}$ be a sequence of smooth functions such that $F_k \to F$ in W^{1,q_0}. In particular, we can assume that $\|F_k\|_{W^{1,q_0}(M,\omega)} \leq \Lambda + 1$ for any k. By [Tosatti and Weinkove 2010a], there is a unique smooth solution ϕ_k and constant b_k such that

$$\det(g_{i\bar{j}} + (\phi_k)_{i\bar{j}}) = e^{F+b_k} \det(g_{i\bar{j}}),$$

and such that $(g_{i\bar{j}} + (\phi_k)_{i\bar{j}}) > 0$ with normalized condition $\sup_M \phi_k = -1$. By the maximum principle, we have

$$|b_k| \leq C_1(\|F_k\|_{L^\infty(M,\omega)}, M, \omega). \tag{4-1}$$

By Theorem 1.6, Theorem 2.1 and Theorem 4.1, there exists a constant $\alpha \in (0, 1)$ such that

$$\|\phi_k\|_{C^{2,\alpha}(M,\omega)} \leq C_2(\|F_k\|_{W^{1,q_0}(M,\omega)}, q_0, M, \omega).$$

To get a W^{3,q_0}-estimate, we can localize the estimate as follows. Let ∂ denote an arbitrary first-order differential operator in a domain $\Omega \subset M$. Since we have a $C^{2,\alpha}$-estimate, we compute that

$$\tilde{\Delta}_{g_k}(\partial\phi_k) = \partial(F_k + \log(\det(g_{i\bar{j}}))) - (g_k)^{i\bar{j}}\partial g_{i\bar{j}}$$

in Ω, where $(g_k)_{i\bar{j}} = g_{i\bar{j}} + (\phi_k)_{i\bar{j}}$. Since $\tilde{\Delta}_{g_k}$ is a uniform elliptic operator, by L^p estimates (for example, see [Gilbarg and Trudinger 1977]), for any $\Omega' \subset \Omega$ we have

$$\|\partial\phi_k\|_{W^{2,q_0}(\Omega',\omega)} \leq C_3(\Omega, \Omega', q_0, \Lambda, \omega),$$

which implies

$$\|\phi_k\|_{W^{3,q_0}(M,\omega)} \leq C_4(\|F\|_{W^{1,q_0}(M,\omega)}, q_0, \Lambda, M, \omega). \tag{4-2}$$

By (4-1) and (4-2), we know that there is a subsequence $\{(\phi_{k_l}, b_{k_l})\}$ of $\{(\phi_k, b_k)\}$ such that $\{b_{k_l}\}$ converges to b and $\{\phi_{k_l}\}$ weakly converges to $\phi \in W^{3,q_0}$ such that $(g_{i\bar{j}} + \phi_{i\bar{j}}) > 0$, which defines a W^{1,q_0} Hermitian metric. Since the Sobolev embedding $W^{3,q_0} \hookrightarrow C^2$ is compact, the subsequence $\{\phi_{k_l}\}$ converges to ϕ in C^2. Hence ϕ with constant b is a classical solution of the complex Monge–Ampère equation. The uniqueness follows from Remark 5.1 in [Tosatti and Weinkove 2010b]. □

Appendix

Let $g_{\mathbb{R}}$ denote the Riemannian metric induced by g; thus $(M, g_{\mathbb{R}})$ is a Riemannian manifold of real dimension $2n$. In this appendix, we deduce some interpolation inequalities on the Hermitian manifold (M, ω) by using some fundamental inequalities on the Riemannian manifold $(M, g_{\mathbb{R}})$.

Let us recall the definition of $g_{\mathbb{R}}$ first. For any local holomorphic coordinates $(z^1, \dots, z^n)$ with $z^i = x^i + \sqrt{-1}y^i$, $(x^1, \dots, x^n, y^1, \dots, y^n)$ forms a smooth local coordinate system. We define

$$g_{\mathbb{R}}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = g_{\mathbb{R}}\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}\right) = 2\,\mathrm{Re}(g_{i\bar{j}}),$$

while

$$g_{\mathbb{R}}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) = 2\,\mathrm{Im}(g_{i\bar{j}}).$$

For the Riemannian metric $g_{\mathbb{R}}$, let $\nabla_{\mathbb{R}}$ and $dV_{\mathbb{R}}$ denote the Levi-Civita connection and the volume form, respectively. By direct calculation, we have

$$dV_{\mathbb{R}} = \frac{1}{n!}\omega^n. \tag{A-1}$$

For convenience, we introduce some notation. For any function $f \in C^\infty(M)$, let $\nabla^m_{\mathbb{R}} f$ and $\Delta_{\mathbb{R}} f$ denote the m-th covariant derivative and the Laplacian of f with respect to $g_{\mathbb{R}}$. Let $\|f\|_{L^p(M,g_{\mathbb{R}})}$ and $\|\nabla^m_{\mathbb{R}} f\|_{L^p(M,g_{\mathbb{R}})}$ denote the corresponding norms with respect to $(M, g_{\mathbb{R}})$.

Thus, by (A-1) and some calculation, we have the following lemma:

Lemma A.1. *For any $f \in C^\infty(M)$, we have*

$$\|f\|_{L^p(M,g_{\mathbb{R}})} = C_1(p)\|f\|_{L^p(M,\omega)} \quad \textit{and} \quad \|\nabla_{\mathbb{R}} f\|_{L^p(M,g_{\mathbb{R}})} = C_2(p)\|\nabla f\|_{L^p(M,\omega)},$$

where $C_1(p) = C_1(p,n)$ and $C_2(p) = C_2(p,n)$.

Corollary A.2. *For any $f \in C^\infty(M)$, we have the Sobolev inequality*

$$\left(\int_M f^{2\beta}\omega^n\right)^{\frac{1}{\beta}} \le C\int_M f^2\omega^n + C\int_M |\nabla f|^2\omega^n,$$

where $\beta = n/(n-1)$ and $C = C(M,\omega)$.

Proof. By the Sobolev embedding $W^{1,2}(M, g_{\mathbb{R}}) \hookrightarrow L^{2\beta}(M, g_{\mathbb{R}})$, we have

$$\left(\int_M f^{2\beta}\, dV_{\mathbb{R}}\right)^{\frac{1}{\beta}} \le C_s\int_M f^2\, dV_{\mathbb{R}} + C_s\int_M |\nabla_{\mathbb{R}} f|^2\, dV_{\mathbb{R}},$$

where $C_s = C_s(M, g_{\mathbb{R}})$. Thus, combining this with Lemma A.1, we complete the proof. □

Since $(M, g_{\mathbb{R}})$ is a Riemannian manifold of real dimension $2n$, we have the following interpolation inequality (for example, see [Aubin 1998]):

Theorem A.3. *Let q, r be real numbers such that $1 \le q, r \le +\infty$ and j, m integers such that $0 \le j < m$. Then there exists a constant*

$$C = C(M, g_{\mathbb{R}}, m, j, q, r, \alpha)$$

such that, for all $f \in C^\infty(M)$ with $\int_M f\, dV_{\mathbb{R}} = 0$, we have

$$\|\nabla^j_{\mathbb{R}} f\|_{L^p(M,g_{\mathbb{R}})} \le C\|\nabla^m f\|^{\alpha}_{L^r(M,g_{\mathbb{R}})}\|f\|^{1-\alpha}_{L^q(M,g_{\mathbb{R}})}, \tag{A-2}$$

where

$$\frac{1}{p} = \frac{j}{2n} + \alpha\left(\frac{1}{r} - \frac{m}{2n}\right) + (1-\alpha)\frac{1}{q}$$

for all α in the interval $j/m \le \alpha \le 1$, for which p is nonnegative. If $r = 2n/(m-j) \ne 1$, then (A-2) *is not valid for $\alpha = 1$.*

Corollary A.4. *Let $f \in C^\infty(M)$; for any $\epsilon > 0$ and $1 \le p < \infty$, we have*

$$\|\nabla_{\mathbb{R}} f\|_{L^p(M,g_{\mathbb{R}})} \le \epsilon\|\nabla^2_{\mathbb{R}} f\|_{L^p(M,g_{\mathbb{R}})} + C(\epsilon, p)\|f\|_{L^p(M,g_{\mathbb{R}})},$$

where $C(\epsilon, p) = C(\epsilon, p, M, \omega)$.

Proof. Set $\tilde f = f - 1/\mathrm{Vol}(M, g_{\mathbb{R}})\int_M f\, dV_{\mathbb{R}}$; then $\int_M \tilde f\, dV_{\mathbb{R}} = 0$. By Theorem A.3 we have

$$\|\nabla_{\mathbb{R}} \tilde f\|_{L^p(M,g_{\mathbb{R}})} \le C_1(p)\|\nabla^2_{\mathbb{R}} \tilde f\|^{\frac{1}{2}}_{L^p(M,g_{\mathbb{R}})}\|\tilde f\|^{\frac{1}{2}}_{L^p(M,g_{\mathbb{R}})},$$

where $C_1(p) = C_1(p, M, g_{\mathbb{R}})$. Thus, by the Cauchy–Schwarz inequality, for any $\epsilon > 0$ we obtain

$$\|\nabla_{\mathbb{R}} \tilde{f}\|_{L^p(M,g_{\mathbb{R}})} \le \epsilon \|\nabla^2_{\mathbb{R}} \tilde{f}\|_{L^p(M,g_{\mathbb{R}})} + C_2(\epsilon, p)\|\tilde{f}\|_{L^p(M,g_{\mathbb{R}})},$$

where $C_2(\epsilon, p) = C_2(\epsilon, p, M, g_{\mathbb{R}})$. By the definition of $\tilde{f}$, the proof is complete. □

Lemma A.5. *Let* (M, ω) *be a compact Hermitian manifold of complex dimension* n. *If* ϕ *is a smooth solution of* (1-1), *then, for any* $1 < p < \infty$, *we have*

$$\|\Delta_{\mathbb{R}}\phi\|_{L^p(M,\omega)} \le C_1(p)\|\Delta\phi\|_{L^p(M,\omega)} + C_2(p),$$

where $C_1 = C_1(p, n)$ *and* $C_2(p) = C_2(p, \|F\|_{L^\infty(M,\omega)}, M, \omega)$.

Proof. After some calculations, we have

$$\|\Delta_{\mathbb{R}}\phi\|_{L^p(M,g_{\mathbb{R}})} \le 2\|\Delta\phi\|_{L^p(M,g_{\mathbb{R}})} + C_3(p)\|\nabla_{\mathbb{R}}\phi\|_{L^p(M,g_{\mathbb{R}})}, \tag{A-3}$$

where $C_3 = C_3(p, M, \omega)$. For (A-3), one can find more details in [Tosatti 2007] (Lemma 3.2 there shows the exact relation between $\Delta_{\mathbb{R}}$ and 2Δ). By Corollary A.4 we obtain

$$C_3(p)\|\nabla_{\mathbb{R}}\phi\|_{L^p(M,g_{\mathbb{R}})} \le \tfrac{1}{2}\|\Delta_{\mathbb{R}}\phi\|_{L^p(M,g_{\mathbb{R}})} + C_4(p)\|\phi\|_{L^p(M,g_{\mathbb{R}})}, \tag{A-4}$$

where $C_4 = C_4(p, M, \omega)$. Combining this with (A-3) and (A-4), we obtain

$$\|\Delta_{\mathbb{R}}\phi\|_{L^p(M,g_{\mathbb{R}})} \le 4\|\Delta\phi\|_{L^p(M,g_{\mathbb{R}})} + C_5(p)\|\phi\|_{L^p(M,g_{\mathbb{R}})},$$

where $C_5 = C_5(p, M, \omega)$. By Theorem 2.1 and Lemma A.1, the proof is complete. □

Lemma A.6. *Under the assumptions of Theorem 1.6, for any* $1 < p < 2n$ *we have*

$$\|\nabla\phi\|_{L^{\frac{2np}{2n-p}}(M,\omega)} \le C(p)\|u\|_{L^p(M,\omega)} + C(p),$$

where u *is defined in* (2-7) *and* $C(p) = C(p, \|F\|_{L^\infty(M,\omega)}, M, \omega)$.

Proof. By the Sobolev embedding $W^{2,p}(M, g_{\mathbb{R}}) \hookrightarrow W^{1,\frac{2np}{2n-p}}(M, g_{\mathbb{R}})$, we have

$$\|\nabla_{\mathbb{R}}\phi\|_{L^{\frac{2np}{2n-p}}(M,g_{\mathbb{R}})} \le C_1(p)\|\nabla^2_{\mathbb{R}}\phi\|_{L^p(M,g_{\mathbb{R}})} + C_1(p)\|\nabla_{\mathbb{R}}\phi\|_{L^p(M,g_{\mathbb{R}})} + C_1(p)\|\phi\|_{L^p(M,g_{\mathbb{R}})},$$

where $C_1(p) = C_1(p, M, g_{\mathbb{R}})$. Combining this with Corollary A.4, we have

$$\|\nabla\phi\|_{L^{\frac{2np}{2n-p}}(M,g_{\mathbb{R}})} \le C_2(p)\|\nabla^2_{\mathbb{R}}\phi\|_{L^p(M,g_{\mathbb{R}})} + C_2(p)\|\phi\|_{L^p(M,g_{\mathbb{R}})},$$

where $C_2(p) = C_2(p, M, g_{\mathbb{R}})$. By Theorem 2.1 and L^p estimates, we have

$$\|\nabla\phi\|_{L^{\frac{2np}{2n-p}}(M,g_{\mathbb{R}})} \le C_3(p)\|\Delta_{\mathbb{R}}\phi\|_{L^p(M,g_{\mathbb{R}})} + C_3(p),$$

where $C_3(p) = C_3(p, \|F\|_{L^\infty(M,\omega)}, M, g_{\mathbb{R}})$. By Lemma A.1 and Lemma A.5, we have

$$\|\nabla\phi\|_{L^{\frac{2np}{2n-p}}(M,\omega)} \leq C_4(p)\|\Delta\phi\|_{L^p(M,\omega)} + C_4(p),$$

where $C_4(p) = C_4(p, \|F\|_{L^\infty(M,\omega)}, M, g_{\mathbb{R}})$. By (2-7) and Theorem 2.1, the proof is complete. □

Lemma A.7. *Let p, r be real numbers such that $1 < p, r < \infty$. Under the assumptions of Theorem 1.6, we have*

$$\|\nabla\phi\|_{L^p(M,\omega)} \leq C(p,r)\|u\|_{L^r}^{\alpha} + C(p,r),$$

where $C(p,r) = C(p, r, \|F\|_{L^\infty(M,\omega)}, M, \omega)$ and

$$\frac{1}{p} = \frac{1}{2n} + \alpha\Big(\frac{1}{r} - \frac{1}{n}\Big)$$

for α in the interval $\frac{1}{2} \leq \alpha < 1$.

Proof. Set $\tilde{\phi} = \phi - 1/\mathrm{Vol}(M, g_R)\int_M \phi\, dV_{\mathbb{R}}$; then $\int_M \tilde{\phi}\, dV_{\mathbb{R}} = 0$. By Theorem 2.1, Lemma A.1 and Theorem A.3, we have

$$\|\nabla_{\mathbb{R}}\tilde{\phi}\|_{L^p(M,g_{\mathbb{R}})} \leq C_1(p,r)\|\nabla^2_{\mathbb{R}}\tilde{\phi}\|^{\alpha}_{L^r(M,g_{\mathbb{R}})},$$

which implies that

$$\|\nabla_{\mathbb{R}}\phi\|_{L^p(M,g_{\mathbb{R}})} \leq C_1(p,r)\|\nabla^2_{\mathbb{R}}\phi\|^{\alpha}_{L^r(M,g_{\mathbb{R}})},$$

where $C_1(p,r) = C_1(p, r, \|F\|_{L^\infty(M,\omega)}, M, \omega)$ and

$$\alpha = \frac{(2n-p)r}{(2n-2r)p}.$$

Combining Lemma A.1, Lemma A.5, (2-7) and L^p estimates, the proof is complete. □

Acknowledgements

The author would like to thank his adviser Gang Tian for leading him to study the complex Monge–Ampère equation, constant encouragement and several useful comments on an earlier version of this paper. The author would also like to thank Valentino Tosatti for his helpful comments and suggestions, especially for pointing out that Lemma 2.2 holds when the background metric is not balanced when $n \geq 3$, which helped the author remove the balanced condition assumption for $n \geq 3$ in an earlier version of this paper. The author would also like to thank Wenshuai Jiang and Feng Wang for many helpful conversations.

References

[Aubin 1998] T. Aubin, *Some nonlinear problems in Riemannian geometry*, Springer, Berlin, 1998. MR 99i:58001 Zbl 0896.53003

[Bedford and Taylor 1976] E. Bedford and B. A. Taylor, "The Dirichlet problem for a complex Monge–Ampère equation", *Invent. Math.* **37**:1 (1976), 1–44. MR 56 #3351 Zbl 0315.31007

[Bedford and Taylor 1982] E. Bedford and B. A. Taylor, "A new capacity for plurisubharmonic functions", *Acta Math.* **149**:1 (1982), 1–40. MR 84d:32024 Zbl 0547.32012

[Błocki 2005] Z. Błocki, "On uniform estimate in Calabi–Yau theorem", *Sci. China Ser. A* **48**:1S (2005), 244–247. MR 2006c:32050 Zbl 1128.32025

[Caffarelli et al. 1985] L. A. Caffarelli, J. J. Kohn, L. Nirenberg, and J. Spruck, "The Dirichlet problem for nonlinear second-order elliptic equations, II: Complex Monge–Ampère, and uniformly elliptic, equations", *Comm. Pure Appl. Math.* **38**:2 (1985), 209–252. MR 87f:35097 Zbl 0598.35048

[Calabi 1957] E. Calabi, "On Kähler manifolds with vanishing canonical class", pp. 78–89 in *Algebraic geometry and topology: a symposium in honor of S. Lefschetz* (Princeton, NJ, 1954), edited by R. H. Fox et al., Princeton Mathematical Series **12**, Princeton University Press, 1957. MR 19,62b Zbl 0080.15002

[Chen and He 2012] X. X. Chen and W. Y. He, "The complex Monge–Ampère equation on compact Kähler manifolds", *Math. Ann.* **354**:4 (2012), 1583–1600. MR 2993005 Zbl 1253.35070

[Cherrier 1987] P. Cherrier, "Équations de Monge–Ampère sur les variétés Hermitiennes compactes", *Bull. Sci. Math.* (2) **111**:4 (1987), 343–385. MR 89d:58131 Zbl 0629.58028

[Demailly and Pali 2010] J.-P. Demailly and N. Pali, "Degenerate complex Monge–Ampère equations over compact Kähler manifolds", *Int. J. Math.* **21**:3 (2010), 357–405. MR 2012e:32039 Zbl 1191.53029

[Dinew 2009] S. Dinew, "Uniqueness in $\mathcal{E}(X,\omega)$", *J. Funct. Anal.* **256**:7 (2009), 2113–2122. MR 2010e:32037 Zbl 1171.32024

[Eyssidieux et al. 2009] P. Eyssidieux, V. Guedj, and A. Zeriahi, "Singular Kähler–Einstein metrics", *J. Amer. Math. Soc.* **22**:3 (2009), 607–639. MR 2010k:32031 Zbl 1215.32017

[Fino and Tomassini 2011] A. Fino and A. Tomassini, "On astheno-Kähler metrics", *J. Lond. Math. Soc.* (2) **83**:2 (2011), 290–308. MR 2012d:53230 Zbl 1215.53066

[Gilbarg and Trudinger 1977] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Grundlehren der mathematischen Wissenschaften **224**, Springer, Berlin, 1977. Second ed. in 1983. Rev. 3rd printing in 2001. MR 57 #13109 Zbl 0361.35003

[Guan and Li 2009] B. Guan and Q. Li, "Complex Monge–Ampere equations on Hermitian manifolds", preprint, 2009. arXiv 0906.3548v1

[Guedj and Zeriahi 2007] V. Guedj and A. Zeriahi, "The weighted Monge–Ampère energy of quasiplurisubharmonic functions", *J. Funct. Anal.* **250**:2 (2007), 442–482. MR 2008h:32056 Zbl 1143.32022

[Hanani 1996] A. Hanani, "Équations du type de Monge–Ampère sur les variétés Hermitiennes compactes", *J. Funct. Anal.* **137**:1 (1996), 49–75. MR 97c:32018 Zbl 0847.53045

[Kołodziej 1998] S. Kołodziej, "The complex Monge–Ampère equation", *Acta Math.* **180**:1 (1998), 69–117. MR 99h:32017 Zbl 0913.35043

[Kołodziej 2008] S. Kołodziej, "Hölder continuity of solutions to the complex Monge–Ampère equation with the right-hand side in L^p: the case of compact Kähler manifolds", *Math. Ann.* **342**:2 (2008), 379–386. MR 2009g:32079 Zbl 1149.32018

[Krylov 1989] N. V. Krylov, “Гладкость функции выигрыша для управляемого диффузионного процесса в области”, *Izv. Akad. Nauk SSSR Ser. Mat.* **53**:1 (1989), 66–96. Translated as “Smoothness of the value function for a controllable diffusion process in a domain” in *Math. USSR-Izv.* **34**:1 (1990), 65–95. MR 90f:93040 Zbl 0701.93054

[Krylov 1994] N. V. Krylov, “On analogues of the simplest Monge–Ampère equation”, *C. R. Acad. Sci. Paris Sér. I Math.* **318**:4 (1994), 321–325. MR 95b:35060 Zbl 0805.35020

[Moser 1960] J. Moser, “A new proof of De Giorgi’s theorem concerning the regularity problem for elliptic differential equations”, *Comm. Pure Appl. Math.* **13** (1960), 457–468. MR 30 #332 Zbl 0111.09301

[Phong et al. 2012] D. H. Phong, J. Song, and J. Sturm, “Complex Monge–Ampère equations”, pp. 327–410 in *In memory of C. C. Hsiung: lectures given at the JDG Symposium on Geometry and Topology* (Bethlehem, PA, 2010), edited by H.-D. Cao and S.-T. Yau, Surveys in Differential Geometry **17**, International Press, Somerville, MA, 2012. MR 3076065 Zbl 06322503 arXiv 1209.2203v2

[Tosatti 2007] V. Tosatti, “A general Schwarz lemma for almost-Hermitian manifolds”, *Comm. Anal. Geom.* **15**:5 (2007), 1063–1086. MR 2009b:53048 Zbl 1145.53019

[Tosatti and Weinkove 2010a] V. Tosatti and B. Weinkove, “The complex Monge–Ampère equation on compact Hermitian manifolds”, *J. Amer. Math. Soc.* **23**:4 (2010), 1187–1195. MR 2012c:32055 Zbl 1208.53075

[Tosatti and Weinkove 2010b] V. Tosatti and B. Weinkove, “Estimates for the complex Monge–Ampère equation on Hermitian and balanced manifolds”, *Asian J. Math.* **14**:1 (2010), 19–40. MR 2011h:32043 Zbl 1208.32034

[Tosatti and Weinkove 2015] V. Tosatti and B. Weinkove, “On the evolution of a Hermitian metric by its Chern–Ricci form”, *J. Differential Geom.* **99**:1 (2015), 125–163. MR 3299824 Zbl 06399561 arXiv 1201.0312v2

[Tosatti et al. 2014] V. Tosatti, Y. Wang, B. Weinkove, and X. K. Yang, “$C^{2,\alpha}$ estimates for nonlinear elliptic equations in complex and almost complex geometry”, preprint, 2014. arXiv 1402.0554v1

[Yau 1978] S.-T. Yau, “On the Ricci curvature of a compact Kähler manifold and the complex Monge–Ampère equation. I”, *Comm. Pure Appl. Math.* **31**:3 (1978), 339–411. MR 81d:53045 Zbl 0369.53059

[Zhang 2006] Z. Zhang, “On degenerate Monge–Ampère equations over closed Kähler manifolds”, *Int. Math. Res. Not.* **2006** (2006), Art. ID 63640. MR 2007b:32058 Zbl 1112.32021 arXiv math/0603465v2

Received May 15, 2014. Revised October 22, 2014.

JIANCHUN CHU
SCHOOL OF MATHEMATICAL SCIENCES
PEKING UNIVERSITY
YIHEYUAN ROAD 5
BEIJING, 100871
CHINA
chujianchun@pku.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.387

TOPOLOGICAL AND PHYSICAL LINK THEORY ARE DISTINCT

ALEXANDER COWARD AND JOEL HASS

Physical knots and links are one-dimensional submanifolds of $\mathbb{R}^3$ with fixed length and thickness. We show that isotopy classes in this category can differ from those of classical knot and link theory. In particular we exhibit a Gordian split link, a two-component link that is split in the classical theory but cannot be split with a physical isotopy.

1. Introduction

The theory of knots and links studies one-dimensional submanifolds of $\mathbb{R}^3$. These are often described as loops of string, or rope, with their ends glued together. Real ropes however are not one-dimensional, but have a positive thickness and a finite length. Indeed, most physical applications of knot theory are related more closely to the theory of knots of fixed thickness and length than to classical knot theory. For example, biologists are interested in knotted curves of fixed thickness and length when studying properties of DNA [Cantarella et al. 1998] and protein molecules [Liang and Mislow 1994]. In these applications the thickness of the curve modeling the molecule plays an essential role in determining the possible configurations.

In this paper we show that the equivalence class of a link in $\mathbb{R}^3$ under an isotopy that preserves thickness and length can be distinct from the classical equivalence class under isotopy. We thus show for the first time that the theory of physically realistic curves of fixed thickness and length in $\mathbb{R}^3$ is distinct from the classical theory of knots and links.

The two most fundamental problems concerning physical knots and links are to show the existence of a *Gordian unknot* and a *Gordian split link*. A Gordian unknot is a loop of fixed thickness and length whose core is unknotted, but which cannot be deformed to a round circle by an isotopy fixing its length and thickness. A Gordian split link is a pair of loops of fixed thickness whose core curves can be split, or isotoped so that its two components are separated by a plane, but cannot be split by an isotopy fixing each component's length and thickness. In this paper we establish the existence of such a link.

MSC2010: 53A04, 57M25.
Keywords: physical knot theory, Gordian, split link.

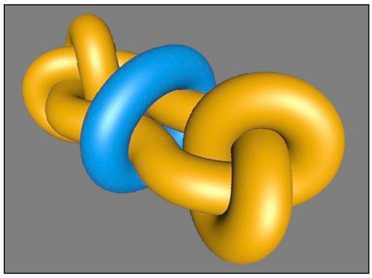

Figure 1. A Gordian split link.

Theorem 1.1. *A Gordian split link exists.*

The proof of Theorem 1.1 is by explicit construction of a link, illustrated in Figure 1, that can be topologically but not physically split.

There has been extensive investigation into the properties of shortest representatives of physical knots, called *tight* [Cantarella et al. 1998] or *ideal* knots [Katritch et al. 1997], and into the possible existence of Gordian unknots [Buck and Simon 1997; Cantarella et al. 1998; Diao et al. 1999; Gonzalez and Maddocks 1999; Katritch et al. 1997]. A candidate Gordian unknot was suggested by Freedman, He and Wang [Freedman et al. 1994], who studied energies associated to curves in $\mathbb{R}^3$ and associated gradient flows [He 2002]. This curve was studied numerically by Pierański [1998], who developed a computer program called SONO (Shrink On No Overlaps) to numerically shorten a curve of fixed thickness while avoiding overlaps. The program unexpectedly succeeded in unraveling the Freedman–He–Wang example. However there are more complicated examples that do fail to unravel under SONO and hence give numerical evidence for the existence of Gordian unknots. Extensive tables of physical knots of minimal length in various isotopy classes have been experimentally derived [Ashton et al. 2011]. A proof of the existence of Gordian unknots or Gordian split links based on a rigorous analysis of such algorithms is plausible but has not yet been found. In related work, Nabutovsky [1995] showed that n-dimensional spheres of fixed thickness in $\mathbb{R}^{n+1}$ can be knotted for dimensions $n \geq 5$. Cantarella and Johnston [1998] showed that the theory of polygonal knots of fixed edge lengths is distinct from classical knot theory.

We now give precise definitions. We say that a knot or link L in $\mathbb{R}^3$ is *r-thick* if it is differentiable and its open radius-r normal disk bundle is embedded. This means that the collection of flat, radius-r two-disks intersecting L perpendicularly at their centers have mutually disjoint interiors. An isotopy of a knot or link maintaining r-thickness throughout is called an *r-thick isotopy*. By rescaling we may take $r = 1$ and take *thick* to mean 1-thick. A *physical isotopy* is a thick isotopy of a knot or link that preserves the length of each component. This reflects the real world properties of links composed of nonstretchable rope of fixed radius.

Theorem 1.1 is proved by an explicit construction of a two-component thick link L that is split but admits no physical isotopy splitting its components. To construct this link we begin by placing two points A and B at $(1, 0, 0)$ and $(-1, 0, 0)$. Let AB denote the straight line between these two points. The first component L_1 of L is any thick curve encircling AB in the xz-plane, disjoint from the open, radius-2 neighborhood of AB. The length of L_1 is at least $4\pi + 4 \approx 16.566$, and this length can be realized by taking L_1 to be the boundary of the radius-2 neighborhood of AB in the xz-plane. To construct the other component, join the two points A and B by an arc α satisfying the following three conditions:

(1) The union of α with AB forms a nontrivial knot contained in the half-space $y \geq 0$.

(2) The arc α meets the xz-plane only at its endpoints and is perpendicular to the xz-plane at these points.

(3) The union of L_1, α and the reflection of α across the xz-plane forms a thick link.

The union of α and its reflection in the xz-plane is the second component L_2 of the thick link L. Figure 1 shows an example of such a link.

Theorem 1.1 follows immediately from the following result, which gives an explicit lower bound on the length required for L_1, the unknotted component of L, if L can be split by a physical isotopy.

Theorem 1.2. *If there is a physical isotopy of $L = L_1 \sqcup L_2$ that splits its two components, then the length of L_1 must be at least $4\pi + 6 \approx 18.566$.*

Since the link L can be constructed with the length of the unknotted component L_1 equal to $4\pi + 4$, this result implies Theorem 1.1.

The paper is arranged as follows. In Section 2 we give a lower bound on the boundary length of a disk of nonpositive curvature containing three disjoint disks of radius 1. In Section 3 we show that if a family of disks spanning L_1 gives a homotopy from a disk in the xz-plane to a disk disjoint from L_2 and each disk in the homotopy intersects a neighborhood of L_2 in at most two components containing points of L_2, then L_2 is unknotted. In Section 4 we bring these results together to prove Theorem 1.2. We conclude with a short list of open problems.

2. An isoperimetric inequality

To prove Theorem 1.2, we show that if the unknotted component L_1 has length less than $4\pi + 6$, then there are severe restrictions on how the other component can pass through a natural spanning disc for L_1. This spanning disc, to be defined in Section 4, is a cone with cone angle at least 2π, and hence a CAT(0) space. We are therefore led to finding a lower bound on the length of a curve in a CAT(0) surface

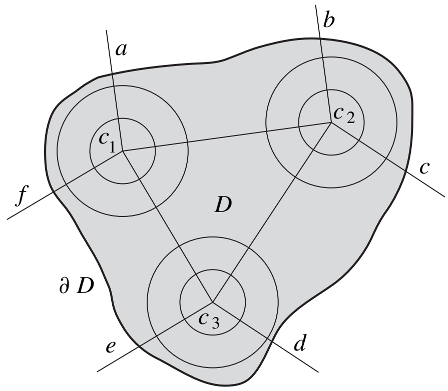

Figure 2. The boundary of this disk has length at least $6+4\pi$.

that bounds a disk enclosing three or more nonoverlapping subdisks of radius 1, as in Figure 2. This next result is based on an argument for flat metrics given in [Cantarella et al. 2002].

Proposition 2.1. *Let P be a complete* CAT(0) *surface and let D_1, D_2, D_3 be three subdisks of P with disjoint interiors and radius* 1. *Let $D \subset P$ be a disk containing D_1, D_2, D_3 such that ∂D has distance at least* 1 *to any of D_1, D_2, D_3. Then the length of ∂D is at least* $4\pi + 6$.

Proof. Let T be the triangle with vertices at the center points c_1, c_2, c_3 of the disks D_1, D_2, D_3. Each edge of T has length at least 2. Let a, b, c, d, e, f denote perpendicular rays from the sides of T at its vertices, as in Figure 2. The curve ∂D intersects each of a, b, c, d, e, f in at least one point. We pick one such point for each, ordered cyclically around ∂D, and refer to the intervening arcs of ∂D as the parts of ∂D between them. Since the edge of T between c_1 and c_2 is perpendicular to a and to b, it realizes the minimal distance of any path between them. Thus the length of the part of ∂D between a and b is at least 2. We can argue similarly for the length of ∂D between c and d, and between e and f. Thus these three parts of ∂D have total length at least 6.

The sum of the interior angles of T is at most π, so the sum of the three angles between f and a, between b and c, and between d and e is at least 2π. Radial projection projects the remaining parts of ∂D onto three circular arcs with total angle at least 2π. In a CAT(0) space, radius-decreasing radial projection onto a circle of constant radius is length-decreasing. Since a circle of radius 2 has length at least 4π, it follows that the length of ∂D is at least $6+4\pi$. □

Remark. A similar argument shows that a curve enclosing two disks has length at least $4\pi+4$ and that the length of a curve enclosing $n > 3$ disks is at least $4\pi+2n$ if the centers of the disks form the vertices of a convex polygon.

3. Sweepouts of solid tori

The following proposition gives a generalization of the fact that a 1-bridge knot is unknotted. It considers a generic 1-parameter family of disks, possibly singular, whose interiors sweep across a region containing a solid torus and concludes that if each disk meets the solid torus in at most two components that cross its core, then the solid torus is unknotted. To simplify the argument we restrict to a setting where T and c have a reflectional symmetry. Roughly speaking, this allows us to argue that if a family of arcs forms a partial spanning disk that fills in half of a curve, then the entire curve bounds a disk formed from reflection of this partial spanning disk and is thus unknotted.

Proposition 3.1. *Let T be a solid torus in $\mathbb{R}^3$ with core c, such that both T and c are symmetric under reflection r in the xz-plane. Suppose there is a homotopy of the disc $g_t : D \to \mathbb{R}^3$, $t \in [-1, 1]$, with the following properties*:

(1) *The curve $g_t(\partial D)$ is disjoint from T for all $t \in [-1, 1]$.*

(2) *The family of disks $g_t(D)$ is symmetric under reflection r in the xz-plane*; *i.e.*, *$g_0(D)$ is contained in the xz-plane and $g_{-t} = r \circ g_t$.*

(3) *The preimage $g_0^{-1}(T)$ has two components*, *each containing a single point of $g_0^{-1}(c)$.*

(4) *The disk $g_1(D)$ is disjoint from c.*

(5) *For all $t \in [-1, 1]$ the preimage $g_t^{-1}(T) \subset D$ has at most two components that contain a point of $g_t^{-1}(c)$.*

(6) *The map g_t is generic with respect to the pair (T, c).*

Then c is unknotted.

Assumption (6) means that g_t is transverse to c with the exception of a finite number of times t at which a birth or death of a pair of points of $g_t^{-1}(c)$ occurs and that g_t is transverse to ∂T at these times. Additionally, g_t is transverse to ∂T except for a finite number of times at which $g_t^{-1}(\partial T)$ consists of finitely many simple closed curves and a single component that is either a bouquet of finitely many circles (at a general saddle-type singularity) or a point (at a birth or death singularity).

Proof of Proposition 3.1. To show that c is unknotted, we will form a spanning disk E for c that is traced by a continuous family of arcs in $\mathbb{R}^3$, each arc having endpoints on c and interior disjoint from c. These arcs are of two types. The first type will lie on $g_t(D)$ and vary continuously with t except at finitely many times t when it jumps from one arc on $g_t(D)$ to another; the second type will interpolate continuously between the arcs just before and just after these jumps.

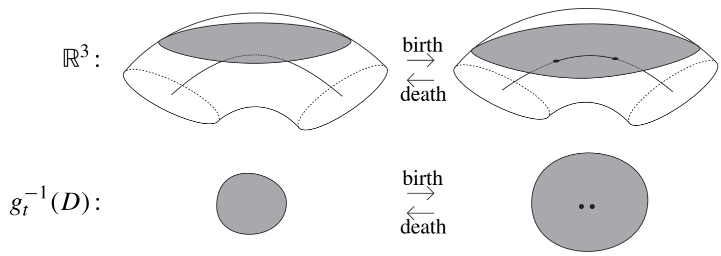

Figure 3. A birth or death adds or removes two points of $g_t^{-1}(c)$ from a component of $g_t^{-1}(T)$.

For a map $g : D \to \mathbb{R}^3$ we call a point of $g^{-1}(c) \subset D$ a *dot* and a component of $g^{-1}(T)$ that contains at least one point of $g^{-1}(c)$ a *dotted component*. Thus $g_0^{-1}(T)$ contains two dotted components, each with a single dot. A birth or death changes the number of points of $g_t^{-1}(c)$ in a component of $g_t^{-1}(T)$ by two, oppositely oriented, as illustrated in Figure 3.

At time $t = 0$, the preimage $g_0^{-1}(c) \subset D$ consists of a pair of points, one in each dotted component. Let α_0 be an arc joining these two points in D, with interior disjoint from $g_0^{-1}(c) \subset D$. As t increases, we take α_t to vary continuously through arcs in D, joining dots in distinct dotted components with interiors disjoint from the dots. There is no obstruction to this while the collection of dots in D is changing by an isotopy. As long as the number of dotted components does not drop, there are two possible obstructions to the extension of α_t as t increases:

(1) Part of the arc α_t may run between two dots that come together and disappear in a death singularity.

(2) One of the endpoints of α_t may disappear in a death singularity.

In contrast, birth singularities do not pose a problem for the extension of the family of arcs past the time at which they occur.

Let t_1 be the time of the first death singularity. To avoid the two problems above we pick a small $\varepsilon > 0$ and at time $t_1' = t_1 - \varepsilon$ we jump from $\alpha_{t_1'}^- := \alpha_{t_1'}$ to a different arc $\alpha_{t_1'}^+$ that also joins points of $g_{t_1'}^{-1}(c)$ in distinct dotted components but that avoids a neighborhood of the death singularity. We will show how to construct $\alpha_{t_1'}^+$ so that this jump can be filled in appropriately for the construction of the disk E. We will then extend the family of arcs α_t for $t > t_1'$ by starting with $\alpha_{t_1'}^+$ and continuing past t_1 until just before the next death singularity occurs at some time $t_2 > t_1$. In the first case we show that the arc α_t can be replaced with one that avoids a neighborhood of the death singularity. In the second case we show that α_t can be replaced with an arc having an endpoint that avoids the death singularity.

For a continuous map $g : D \to \mathbb{R}^3$, we say that two arcs in the disk D joining points of $g^{-1}(c)$ in distinct dotted components of $g^{-1}(T)$ are *g-equivalent* if their

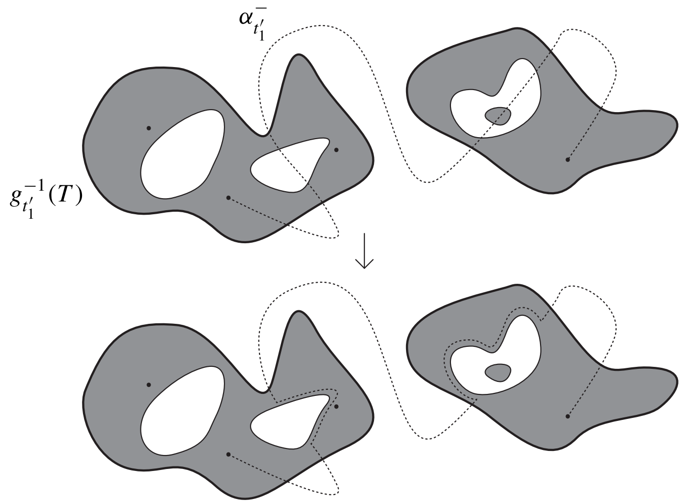

Figure 4. Removing intersections of $\alpha^-_{t'_1}$ with secondary boundary components of $g^{-1}_{t'_1}(T)$.

images under g are homotopic through arcs in $\mathbb{R}^3$ whose interiors are disjoint from c and whose endpoints lie on c. Our goal is then to construct the arc $\alpha^+_{t'_1}$ so that it is $g_{t'_1}$-equivalent to $\alpha^-_{t'_1}$.

When g_t is transverse to ∂T the dotted components of $g_t^{-1}(T)$ form planar subsurfaces of D, each a disk with holes. A boundary curve of a dotted component that has dots on both of its sides in D is called *primary*, and boundary components with all dots on one side are called *secondary*.

If $\alpha^-_{t'_1}$ leaves a dotted component X through a secondary boundary component b, it must reenter X through b, since b is separating in D. Let β be a subarc of $\alpha^-_{t'_1}$ running between two successive intersections of $\alpha^-_{t'_1}$ with b and with interior outside of X. Then β runs through either a dotless disc or a dotless annulus in D. We can then homotope β into b rel endpoints without crossing any dots. Push β a little further into the interior of X. Repeating this process we can homotope $\alpha^-_{t'_1}$ rel endpoints without crossing dots and so that it crosses only primary boundary components. See Figure 4. We abuse notation somewhat and continue to refer to this arc as $\alpha^-_{t'_1}$.

Our next goal is to arrange for $\alpha^-_{t'_1}$ to pass through each primary boundary component exactly once. We will achieve this with the following lemma.

Lemma 3.2. *Suppose T is an embedded solid torus in $\mathbb{R}^3$ with core c, and suppose $g : D \to \mathbb{R}^3$ is a continuous map of a disc into $\mathbb{R}^3$ for which $g^{-1}(T)$ has two dotted components, each with image having algebraic intersection number ± 1 with c. Let α be an arc in D, joining dots in distinct dotted components of $g^{-1}(T)$, and with*

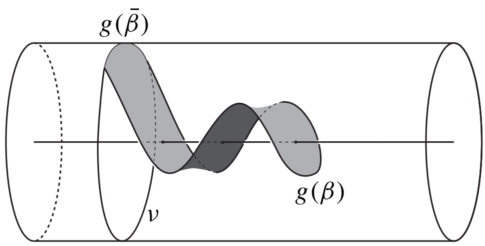

Figure 5. Two ways to homotope $g(\beta)$ onto ∂T.

interior disjoint from $g^{-1}(c)$. Let $\beta \subset g^{-1}(T)$ be a subarc of α that starts and ends on the same primary boundary component b of $g^{-1}(T)$. Then there is an arc $\beta' \subset b$ with the same endpoints as β and with the property that replacing β with β' in α yields an arc α' that is g-equivalent to α.

Proof. There is a homotopy of β in D rel endpoints to an arc $\bar{\beta}$ contained in b, possibly crossing dots. So $g(\beta)$ is homotopic rel endpoints in $\mathbb{R}^3$ to an arc $g(\bar{\beta})$ in $\partial T \cap g(D)$. This homotopy may pass outside T, as β slides over holes of the dotted component. However the boundaries of these holes are secondary, since there are precisely two dotted components, and therefore bound disks in $\mathbb{R}^3$ disjoint from c. It follows that they have image under g that is homotopically trivial on ∂T or they have images on ∂T that are nontrivial and bound disks in the exterior of T. In the latter case T is unknotted, and we are done. Thus we can assume that $g(\beta)$ and $g(\bar{\beta})$ are homotopic rel endpoints in T. The arc $g(\beta)$ is also homotopic rel endpoints in $T - c$, by radial projection away from c, to an arc ν on ∂T. See Figure 5, which for clarity shows only part of $g(D)$.

Now, $g(\bar{\beta})$ and ν are homotopic rel endpoints in T and so in ∂T they differ by a multiple of a meridian. Note that the curve $g(b)$ is a meridian, since it bounds a disk in T meeting c algebraically once. So $g(\beta)$ can be homotoped rel endpoints in the complement of c in T to ν, and then in turn to a curve formed by concatenating $g(\bar{\beta})$ with a multiple of $g(b)$. Take β' to be $\bar{\beta}$ followed by this multiple of b. $\square$

Now suppose that β is a subarc of $\alpha_{t_1'}^-$ that enters and leaves a dotted component. Using Lemma 3.2 we can replace it with a $g_{t_1'}$-equivalent arc β' that lies entirely on $g_{t_1'}^{-1}(\partial T)$, and then perturb β' slightly so that it is disjoint from $g_{t_1'}^{-1}(T)$, as illustrated in Figure 6. In this way we replace $\alpha_{t_1'}^-$ with a $g_{t_1'}$-equivalent arc that has fewer intersections with dotted components, and by repeating we may remove all subarcs of $\alpha_{t_1'}^-$ that enter and leave a dotted component. We continue to refer to the resulting arc as $\alpha_{t_1'}^-$. Note that $\alpha_{t_1'}^-$ may now intersect itself.

We have found an arc $g_{t_1'}$-equivalent to the original arc $\alpha_{t_1'}^-$ that starts at a dot in one dotted component, exits that dotted component, then enters the second dotted component and finally ends at a dot. The following lemma allows us to

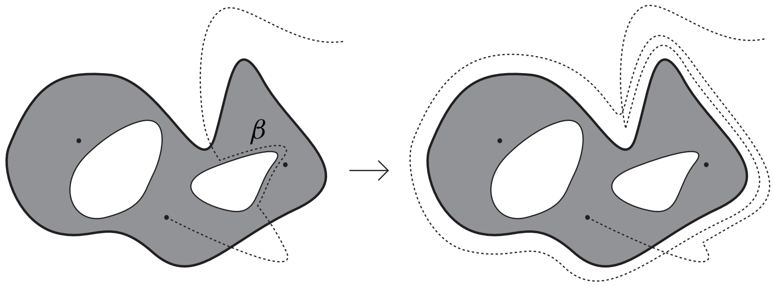

Figure 6. Removing an arc in D that starts and ends on the same primary boundary component of $g^{-1}(T)$.

find a $g_{t_1'}$-equivalent arc that replaces a subarc running from a primary boundary component to a dot, with any other arc running from the same point to a dot and not leaving $g_{t_1'}^{-1}(T)$.

We define an arc in a solid torus T with core c to be *core-to-boundary* if it has one endpoint on ∂T, the other endpoint on c, and interior disjoint from c.

Lemma 3.3. *Let T be a solid torus with core c. Let γ and γ' be core-to-boundary arcs in T with $\gamma \cap \partial T = \gamma' \cap \partial T$. Then γ and γ' can be joined by a homotopy of core-to-boundary arcs, keeping the endpoint on ∂T fixed.*

Proof. We can lift γ and γ' to the universal cover of T, which is homeomorphic to (disk) $\times \mathbb{R}$, so that their common endpoint on ∂T lifts to the same point x while c lifts to $\{0\} \times \mathbb{R}$. Each lift can be homotoped, keeping x fixed and moving points only along the $\mathbb{R}$-factor, to the slice (disk) $\times$ {point} containing x. Further, the resulting arcs are homotopic rel endpoints in (disk) $\times$ {point} via arcs that miss $\{0\} \times$ {point} in their interior. Therefore the lifts of γ and γ' are homotopic through arcs joining x to $\{0\} \times \mathbb{R}$ and with interiors disjoint from $\{0\} \times \mathbb{R}$. The projection of this homotopy to T gives a homotopy joining γ and γ' through core-to-boundary arcs in T. □

Now suppose that a death singularity takes place in a dotted component containing a segment of $\alpha^-_{t_1'}$. Let γ_1 be the segment of $\alpha^-_{t_1'}$, running from a dot to the primary boundary component. Choose an arc γ_1' in the same dotted component that runs from a dot to the point where γ_1 exits the dotted component, so that γ_1' is disjoint from a neighborhood containing the two dying dots. This is possible because the number of dots in each dotted component is odd. By Lemma 3.3, $\alpha^-_{t_1'}$ is $g_{t_1'}$-equivalent to the arc formed by replacing γ_1 by γ_1'. We take $\alpha^+_{t_1'}$ to be the arc, $g_{t_1'}$-equivalent to $\alpha^-_{t_1'}$, which is obtained from $\alpha^-_{t_1'}$ after making all these changes. See Figure 7. See also Figure 8 for an example, illustrated in $\mathbb{R}^3$, of how Lemma 3.3 may be applied.

We have constructed a family of arcs α_t in D that varies continuously until time $t_1' = t_1 - \varepsilon$. It then jumps from $\alpha^-_{t_1'}$ to the $g_{t_1'}$-equivalent arc $\alpha^+_{t_1'}$. The arc $\alpha^+_{t_1'}$ was

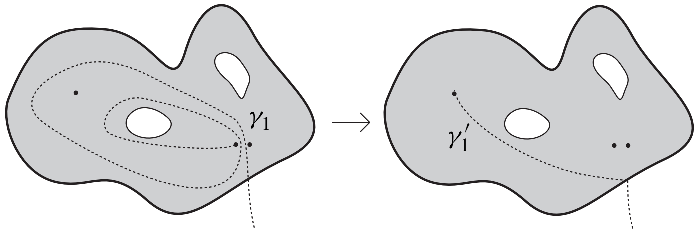

Figure 7. Changing $\alpha_{t_1'}^-$ so that it avoids a death singularity.

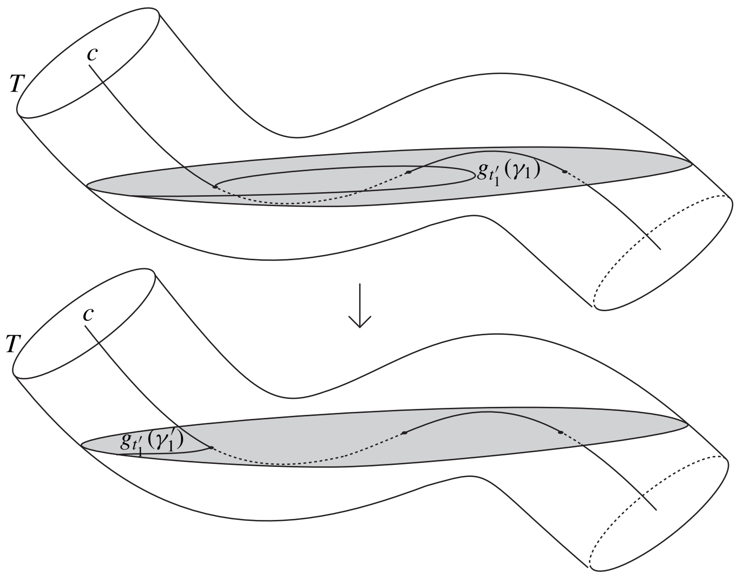

Figure 8. Illustrating Lemma 3.3, where an arc can be homotoped to a new arc through arcs in the complement of c.

chosen to avoid a neighborhood of the death singularity at time t_1, so we can extend the family of arcs α_t past time $t = t_1$ until just before the next death singularity at time $t = t_2$. We then repeat this process.

Eventually, at time $t = t_l \in (0, 1)$ the two dotted components must merge along ${g_{t_l}}^{-1}(\partial T)$. At this time ${g_{t_l}}^{-1}(\partial T)$ consists of a collection of finitely many simple closed curves and one bouquet of finitely many circles embedded in D. The arc α_{t_l} begins and ends at a dot, and may run in and out of the single dotted component.

We now look at the complementary components in D of the single dotted component of ${g_{t_l}}^{-1}(T)$. Each of these is either a disk or an annulus with ∂D as one boundary component. Any subarc β of α_{t_l} that runs out of a dotted component into a complementary component X eventually leaves X and reenters the dotted component. Since X is either a disk or an annulus having ∂D as one of its boundary components, we can homotope β rel endpoints off X and into the dotted component without passing through any dots, since all dots lie within the single dotted component. In this way we can homotope α_{t_l} so that it lies entirely within the dotted component of ${g_{t_l}}^{-1}(T)$. This means that $g_{t_l}(\alpha_{t_l})$ now lies entirely within T. It is then straightforward to shrink $g_{t_l}(\alpha_{t_l})$ within T, keeping its interior disjoint from c and its endpoints on c, until it collapses to a point on c.

We now form a spanning disk E for c. Let $\beta_t = g_t \circ \alpha_t,\ t \in [0, t_l]$. Then β_t is a family of arcs in $\mathbb{R}^3$ whose endpoints lie on c and whose interiors are disjoint from c. These arcs vary continuously except at finitely many times $t'_1, t'_2, \dots, t'_n$ just before death singularities. At these times the limiting arcs $\alpha^-_{t'_i}$ and $\alpha^+_{t'_i}$, as t approaches t'_i from below and above, are $g_{t'_i}$-equivalent. Finally, α_{t_l} is homotopic to a point on c via arcs that start and end on c but have interiors disjoint from c. Therefore there is a family of arcs, with endpoints on c sweeping out a disk with interior in the complement of c, that represents a homotopy of β_0 to a point in c. Let $\bar{E}$ denote the union of these arcs in $\mathbb{R}^3$ and take E to be the disk obtained by taking the union of $\bar{E}$ with its reflection in the xz-plane. Note that the interior of E does not intersect c and that $\partial E \subseteq c$.

Let a be one of the two points of intersection of c with $g_0(D)$. Then ∂E intersects a in an odd number of points, one coming from β_0 and an additional even number coming from equal numbers of intersections of a with $\partial \bar{E}$ and its reflection. So ∂E is nontrivial in $\pi_1(c)$. By the Loop Theorem [Papakyriakopoulos 1957], c is the boundary of an embedded disc and therefore unknotted. □

4. Proof of Theorem 1.2

We now prove Theorem 1.2, showing that if L can be split via a physical isotopy then the length of L_1 must be at least $4\pi + 6$. Note that in this isotopy both components may move.

Assume there is a physical isotopy I_s, $s \in [0, 1]$, of $\mathbb{R}^3$ with I_0 the identity, I_1 taking L_1 and L_2 to opposite sides of a plane, and with the length of the unknotted component L_1 being less than $4\pi + 6$. We will derive a contradiction.

During the course of the isotopy I_s it is possible that the radius-one solid torus neighborhoods of the two link components bump against themselves or each other. We describe a slight modification of the isotopy that keeps the two components embedded and disjoint. Throughout the isotopy, tubular neighborhoods of any radius $r < 1$ give embedded disjoint solid torus neighborhoods of each of L_1 and L_2. Take $r = 1 - \varepsilon'$ to be slightly less than 1 and then rescale the entire isotopy I_s by $1/(1 - \varepsilon')$. This restores the radius to 1 at the cost of slightly lengthening L_1 and L_2. With ε' small, the length of L_1 remains below $4\pi + 6$. The rescaled physical isotopy is then $(1 + \varepsilon)$-thick, with $\varepsilon = \varepsilon'/(1 - \varepsilon')$.

For a curve c in $\mathbb{R}^3$, let $T(c)$ denote the radius-1 tubular neighborhood of c. Without loss of generality we can assume that the isotopy I_s preserves $T(L_1)$ and $T(L_2)$, so that $I_s(T(L_i)) = T(I_s(L_i))$ for $i = 1, 2$. We then let T_s denote the solid torus $T(I_s(L_2))$. For each $s \in [0, 1]$, let x_s be the center of mass of the embedded curve $I_s(L_1)$ and let $f_s : D \to \mathbb{R}^3$ parametrize the disk forming the cone over $I_s(L_1)$ with cone point x_s. Since the cone point is inside the convex hull, its cone angle is always at least 2π [Cantarella et al. 2002; Gage 1980; Gromov 1983]. The maps $f_s : D \to \mathbb{R}^3$ induce a family of metrics on the disk D, parametrized by s, in which each disc is flat except at the cone point and is therefore a subdisk of a complete CAT(0) surface obtained by extending the rays from the cone point to infinity.

Now perturb f_s, $s \in [0, 1]$, so that the family of maps f_s is generic, but leaving f_s unchanged for s in a small neighborhood of 0 and unchanged on ∂D for all s. By generic, we mean that

(1) f_s is transverse to c except for a finite number of times s at which a birth or death of a pair of points of $f_s^{-1}(c)$ occurs;

(2) f_s is transverse to ∂T at these times; and

(3) f_s is transverse to ∂T except for a finite number of times when $f_s^{-1}(\partial T)$ consists of finitely many simple closed curves and a single component that is a bouquet of finitely many circles (in the case of a saddle-type singularity) or a single point (in the case of a birth or death singularity).

Genericity can be achieved by approximating the appropriate parts of f_s by PL maps and using general position. The perturbation can be made arbitrarily $\mathscr{C}^0$-small, and we let f'_s denote the result of perturbing f_s in this manner.

Each component of $f_s'^{-1}(T_s)$ in D that contains a point of $f_s'^{-1}(I_s(L_2))$ contains a disc of radius 1 in D enclosing that point, measured in the induced metric. The distance in D of ∂D from each of these components is at least 1. Suppose for a contradiction that there are three components of $f_s'^{-1}(T_s)$ containing a point of $f_s'^{-1}(I_s(L_2))$ for some s and furthermore suppose that this is true no matter how small we made the perturbation of f_s that gave f'_s. Then $f_s^{-1}(T_s)$ contains three disks of radius 1 with disjoint interiors, and with ∂D having distance at least 1 from each disk. This contradicts Proposition 2.1, since L_1 has length less than $4\pi + 6$. Hence, by taking the perturbation to obtain f'_s to be sufficiently small, we can arrange that for all s there are at most two components of $f_s'^{-1}(T_s)$ containing a point of $f_s'^{-1}(I_s(L_2))$.

We now define a family of disks $h_s : D \to \mathbb{R}^3$, $0 \le s \le 1$, by setting $h_s = I_s^{-1} \circ f'_s$. Extend h_s to $-1 \le s \le 0$ by reflecting through the xz-plane, setting $h_s = r \circ h_{-s}$ for $s < 0$. Each h_s maps ∂D to L_1 and for $s \in [-1, 1]$ the preimage $h_s^{-1}(T(L_2))$ has at most two components containing a point of $h_s^{-1}(L_2)$. Moreover $h_0(D)$ lies in the xz-plane, and $h_1(D)$ and $h_{-1}(D)$ are disjoint from L_2. The disks $h_s(D)$ now satisfy all the conditions of Proposition 3.1, implying that L_2 is unknotted. This contradiction proves Theorem 1.2. □

5. Some open problems

The following related problems remain open.

(1) Does there exist a Gordian unknot?

(2) Can the methods of Theorem 1.2 be extended to produce a Gordian split link with two unknotted components?

(3) In Theorem 1.2 the length of each component is fixed. One can formulate a problem where the sum of the component lengths is fixed but the individual components are allowed to stretch. Is there a Gordian split link in that setting?

Acknowledgements

We thank the anonymous referee for helpful comments that led to the improvement of this paper.

References

[Ashton et al. 2011] T. Ashton, J. Cantarella, M. Piatek, and E. J. Rawdon, "Knot tightening by constrained gradient descent", *Exp. Math.* **20**:1 (2011), 57–90. MR 2802724 Zbl 1261.49007

[Buck and Simon 1997] G. Buck and J. Simon, "Energy and length of knots", pp. 219–234 in *Lectures at KNOTS '96 (Tokyo)*, Ser. Knots Everything **15**, World Sci. Publ., 1997. MR 98f:58054 Zbl 0931.57007

[Cantarella and Johnston 1998] J. Cantarella and H. Johnston, "Nontrivial embeddings of polygonal intervals and unknots in 3-space", *J. Knot Theory Ramifications* **7**:8 (1998), 1027–1039. MR 99m:57002

[Cantarella et al. 1998] J. Cantarella, R. Kusner, and J. Sullivan, "Tight knot values deviate from linear relations", *Nature* **392**:6673 (1998), 237–238.

[Cantarella et al. 2002] J. Cantarella, R. B. Kusner, and J. M. Sullivan, "On the minimum ropelength of knots and links", *Invent. Math.* **150**:2 (2002), 257–286. MR 2003h:58014 Zbl 1036.57001

[Diao et al. 1999] Y. Diao, C. Ernst, and E. J. Janse van Rensburg, "Thicknesses of knots", *Math. Proc. Cambridge Philos. Soc.* **126**:2 (1999), 293–310. MR 2000e:57008

[Freedman et al. 1994] M. H. Freedman, Z.-X. He, and Z. Wang, "Möbius energy of knots and unknots", *Ann. of Math.* (2) **139**:1 (1994), 1–50. MR 94j:58038 Zbl 0817.57011

[Gage 1980] M. E. Gage, "A proof of Gehring's linked spheres conjecture", *Duke Math. J.* **47**:3 (1980), 615–620. MR 82b:52017 Zbl 0463.57007

[Gonzalez and Maddocks 1999] O. Gonzalez and J. H. Maddocks, "Global curvature, thickness, and the ideal shapes of knots", *Proc. Natl. Acad. Sci. USA* **96**:9 (1999), 4769–4773. MR 2000c:57008 Zbl 1057.57500

[Gromov 1983] M. Gromov, "Filling Riemannian manifolds", *J. Differential Geom.* **18**:1 (1983), 1–147. MR 85h:53029 Zbl 0515.53037

[He 2002] Z.-X. He, "On the minimizers of the Möbius cross energy of links", *Experiment. Math.* **11**:2 (2002), 244–248. MR 2003k:58016 Zbl 1116.58300

[Katritch et al. 1997] V. Katritch, W. K. Olson, P. Pierański, J. Dubochet, and A. Stasiak, "Properties of ideal composite knots", *Nature* **388**:6638 (1997), 148–151.

[Liang and Mislow 1994] C. Liang and K. Mislow, "Knots in proteins", *J. Am. Chem. Soc.* **24** (1994), 11189–11190.

[Nabutovsky 1995] A. Nabutovsky, "Non-recursive functions, knots "with thick ropes", and self-clenching "thick" hyperspheres", *Comm. Pure Appl. Math.* **48**:4 (1995), 381–428. MR 96d:58025

[Papakyriakopoulos 1957] C. D. Papakyriakopoulos, "On Dehn's lemma and the asphericity of knots", *Ann. of Math.* (2) **66** (1957), 1–26. MR 19,761a Zbl 0078.16402

[Pierański 1998] P. Pierański, "In search of ideal knots", pp. 20–41 in *Ideal knots*, Ser. Knots Everything **19**, World Sci. Publ., 1998. MR 1702021 Zbl 0943.57005

Received August 6, 2012. Revised September 10, 2013.

ALEXANDER COWARD
DEPARTMENT OF MATHEMATICS
UC BERKELEY
970 EVANS HALL #3840
BERKELEY, CA 94720
UNITED STATES
coward@math.berkeley.edu

JOEL HASS
MATHEMATICS DEPARTMENT
UNIVERSITY OF CALIFORNIA
DAVIS, CA 95616
UNITED STATES
hass@math.ucdavis.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.401

THE MEASURES OF ASYMMETRY FOR COPRODUCTS OF CONVEX BODIES

QI GUO, JINFENG GUO AND XUNLI SU

In previous work, we introduced a family of p-measures of asymmetry for convex bodies, which have the well-known Minkowski measure of asymmetry as a particular case. We now reveal more properties of 1-measure and ∞-measure and give some calculating formulas of p-measures, in particular, for the so-called coproducts of convex bodies.

The measures of asymmetry for convex bodies, which in principle can be traced back to an early paper by Minkowski [1897], have been studied for a long time [Asplund et al. 1962; Besicovitch 1948; Chakerian and Stein 1964; Eggleston 1952; Klee 1953; Rogers and Shephard 1958; Stein 1956]. In particular, after B. Grünbaum formulated in his well-known paper [1963] a general definition of measures of (central) asymmetry (or symmetry), many mathematicians have contributed their efforts to this topic: studying the properties/applications of those known measures of asymmetry [Böröczky 2010; Dziechcińska-Halamoda and Szwiec 1985; Ekström 2000; Gluskin and Litvak 2008; Groemer 2000; Groemer and Wallen 2001; Guo 2005; Guo and Kaijser 1999; 2003; 2002; Hug and Schneider 2007; Kaiser 1996; Petitjean 2003; Schneider 2009; Soltan 2005; Mizushima 2000; Toth 2009; 2008], looking for new ones or studying other types of measure of asymmetry [Tuzikov et al. 2000; Tuzikov et al. 1997; Zouaki 2003]. Several such measures, most of which are related to extremal problems, are proposed and investigated.

In [Guo 2012], we found a family of measures of asymmetry $\mathrm{as}_p(\cdot)$ for convex bodies, called the p-measures of asymmetry ($1 \le p \le \infty$) (see definition below), which have the well-known Minkowski measure as a particular case. It turns out that p-measures do share some nice properties with the Minkowski measure and might be useful for further research.

As shown in [Guo 2012], for any convex body C and $1 \le p < \infty$, we have $\mathrm{as}_p(C) \le \mathrm{as}_\infty(C)$ in general, and equality holds if C is a symmetric convex body or a simplex. Equality also holds for some nontrivial (i.e., neither symmetric nor a simplex) convex bodies (see examples in Remark 2.3 below). Since, in some sense,

---

Project supported by the National Natural Science Foundation of China (No. 11271282; 11371013).
MSC2010: 52A20, 52A38, 52A39, 52A40.
Keywords: Convex bodies, p-measure of asymmetry, Mixed volume, Mixed area measure.

$\mathrm{as}_\infty(C)$ is a maximal value and $\mathrm{as}_1(C)$ is a mean of a certain function related to C, defined on the unit sphere, it is interesting to consider the following question: Under what conditions is it true that $\mathrm{as}_1(C) = \mathrm{as}_\infty(C)$ (and therefore that all $\mathrm{as}_p(C)$ coincide)?

In this article, we reveal more properties of 1-measure and ∞-measure and give some calculating formulas of p-measures, in particular, for coproducts of convex bodies (see definition below). We will also formulate some questions related to the question above. From now on, we will simply write asymmetry instead of central asymmetry.

1. Preliminary

Let $\mathbb{R}^n$ denote the usual n-dimensional Euclidean space and $\langle\cdot,\cdot\rangle$ the canonical inner product on $\mathbb{R}^n$. Denote by $\mathcal{K}^n$ the class of all convex bodies (compact convex sets with nonempty interior) in $\mathbb{R}^n$, by $\mathrm{Aff}(\mathbb{R}^n)$ the family of all affine maps from $\mathbb{R}^n$ to $\mathbb{R}^n$, and by $\mathrm{aff}(\mathbb{R}^n)$ the family of all affine functionals on $\mathbb{R}^n$, which forms an $(n+1)$-dimensional linear space under the ordinary addition and scalar multiplication of functions.

We adopt the following notation and terms from [Schneider 1993].

For $C_1, \dots, C_n \in \mathcal{K}^n$, denote by $V(C_1, \dots, C_n)$ the mixed volume of $C_1, \dots, C_n$ and let $V(C[k])$ be an abbreviated notation for

$$V(C, \overset{(k)}{\dots}, C, -C, \overset{(n-k)}{\dots}, -C), \quad 0 \le k \le n.$$

Similarly, denote by $S(C_1, \dots, C_{n-1}, \cdot)$ the mixed area measure (of $C_1, \dots, C_{n-1}$) on $\mathbb{S}^{n-1}$, the $(n-1)$-dimensional unit sphere. It is stated in [Schneider 1993] that $V(C[0]) = V([n]) = V_n(C)$, where $V_n(\cdot)$ denotes the n-dimensional volume, and $S(C, \overset{(n-1)}{\dots}, C, \cdot) = S_{n-1}(C, \cdot)$, the surface area measure of C on $\mathbb{S}^{n-1}$.

For $\alpha \in \mathbb{R}$ and $u \in \mathbb{S}^{n-1}$, set $H_{u,\alpha} = \{x \in \mathbb{R}^n \mid \langle x, u\rangle = \alpha\}$ and notice that $H_{u,\alpha}$ is a hyperplane.

For $C \in \mathcal{K}^n$ and $x \in \mathbb{R}^n$, we define the support function of C based at x by

$$h_x(C, u) := \sup_{x \in C} \langle y - x, u\rangle, \quad \text{for all } u \in \mathbb{S}^{n-1}.$$

Denote $F(C, u) := C \cap H_{u, h_x(C,u)}$, which is independent of x and called the support set (of C) in the direction u.

It is shown in Theorem 5.1.6 of [Schneider 1993] that, for each $x \in \mathbb{R}^n$,

$$(*) \qquad \begin{aligned} V(C[n-1]) &= \frac{1}{n} \int_{\mathbb{S}^{n-1}} h_x(C, -u)\, dS_{n-1}(C, u), \\ V_n(C) &= \frac{1}{n} \int_{\mathbb{S}^{n-1}} h_x(C, u)\, dS_{n-1}(C, u). \end{aligned}$$

Given $C \in \mathcal{K}^n$, for $x \in \operatorname{int}(C)$, we write

$$\mu_p(C,x) := \begin{cases} \left(\int_{\mathbb{S}^{n-1}} \alpha_x(C,u)^p \, dm_x(C,u)\right)^{1/p} & \text{if } 1 \le p < \infty, \\ \sup_{u \in \mathbb{S}^{n-1}} \alpha_x(C,u) & \text{if } p = \infty, \end{cases}$$

where $\alpha_x(C,u) := h_x(C,-u)/h_x(C,u)$ and, for measurable $\omega \subset \mathbb{S}^{n-1}$,

$$m_x(C,\omega) := \frac{\int_\omega h_x(C,u)\, dS_{n-1}(C,u)}{nV_n(C)} = \frac{\int_\omega h_x(C,u)\, dS_{n-1}(C,u)}{\int_{\mathbb{S}^{n-1}} h_x(C,u)\, dS_{n-1}(C,u)},$$

which is a probability measure on $\mathbb{S}^{n-1}$.

Remark 1.1. If C is a polytope with (all) facets $F(C,u_i)$, $i = 1, 2, \ldots, m$, where u_i are outer normal vectors, then the measures $S_{n-1}(C,\cdot)$ and $m_x(C,\cdot)$ are linear combinations of m Dirac measures δ_{u_i}, $i = 1, 2, \ldots, m$, and so the integrals appearing above are just finite sums.

Definition [Guo 2012]. For $C \in \mathcal{K}^n$, we define its p-measure of asymmetry $(1 \le p \le \infty)$ $\operatorname{as}_p(C)$ by

$$\operatorname{as}_p(C) := \inf_{x \in \operatorname{int}(C)} \mu_p(C,x).$$

A point $x \in \operatorname{int}(C)$ satisfying $\mu_p(C,x) = \operatorname{as}_p(C)$ is called a p-critical point of C. The set of all p-critical points is called the p-critical set (of C), denoted by $\mathcal{C}_p(C)$.

Remark 1.2. (i) The measure

$$\operatorname{as}_\infty(C) = \inf_{x \in \operatorname{int}(C)} \sup_{u \in \mathbb{S}^{n-1}} \frac{h_x(C,-u)}{h_x(C,u)}$$

is nothing else but the Minkowski measure of asymmetry (of C). The measure $\operatorname{as}_1(C)$ is the (minimal) mean of $h_x(C,-u)/h_x(C,u)$ (against $m_x(C,\cdot)$) among all $x \in \operatorname{int}(C)$, which is in fact independent of x (i.e., $\mu_1(C,x) = \operatorname{as}_1(C)$ for all $x \in \operatorname{int}(C)$; see (*)).

(ii) It is shown in [Guo 2012] that if defining, for any $\varepsilon \ge 0$,

$$\phi(\varepsilon) := \left(\frac{V_n(C - \varepsilon C)}{V_n(C)}\right)^{1/n},$$

then $\operatorname{as}_1(C) = \phi'_+(0)$ (in fact, this is the definition of 1-measure in [Guo 2012]).

(iii) By definition 1 and (*) above, we see that $\operatorname{as}_1(C) = V(C[n-1])/V_n(C)$.

(iv) It is proved in [Guo 2012] that, for $1 < p < \infty$, $\mathcal{C}_p$ is a singleton.

One of the main results in [Guo 2012] is the following theorem.

Theorem 1.3. *For any* $1 \le p, q \le \infty$, *the following statements are true*:

(i) $\mathrm{as}_p(\cdot)$ *is affinely invariant*, *i.e.*, $\mathrm{as}_p(C) = \mathrm{as}_p(\mathrm{T}(C))$ *for any* $C \in \mathcal{K}^n$ *and any invertible* $\mathrm{T} \in \mathrm{Aff}(\mathbb{R}^n)$.

(ii) $\mathrm{as}_p(C) \le \mathrm{as}_q(C)$, *for any* $C \in \mathcal{K}^n$ *and* $1 \le p < q \le \infty$.

(iii) $1 \le \mathrm{as}_p(C) \le n$, $\mathrm{as}_p(C) = 1$ *if and only if* C *is symmetric*, *and* $\mathrm{as}_p(C) = n$ *if and only if* C *is a simplex.*

2. The 1-measure of asymmetry for coproducts of convex bodies

In [Guo 2012] we showed that p-measures do share some nice properties with Minkowski's measure. Here we will present more.

We first recall a conclusion in [Guo and Kaijser 2002]: for any $(n-1)$-dimensional convex set $C \subset \mathbb{R}^n$, $\mathrm{as}_\infty(\widehat{C}_z) = \mathrm{as}_\infty(C) + 1$, where $\widehat{C}_z := \mathrm{conv}(C, z)$ is the convex hull of $C \cup \{z\}$ (called the cone with vertex z and base C) and z is not in the affine hull of C (where $\mathrm{as}_\infty(C)$ is computed in the $(n-1)$-dimensional space). Furthermore, all ∞-critical points x^* of $\widehat{C}_z$ are of the form

$$x^* = \frac{1}{2 + \mathrm{as}_\infty(C)} z + \left(\frac{1 + \mathrm{as}_\infty(C)}{2 + \mathrm{as}_\infty(C)} \right) x',$$

where x' is an ∞-critical point of C.

We show that a similar conclusion holds for 1-measure but not for 2-measure. Further, we extend the result to the so-called coproducts of subsets which are a generalization of cones (see definition below).

Let us start with cones.

Theorem 2.1. *Let* C, z *be as above. Then*

(i) $\mathrm{as}_1(\widehat{C}_z) = \mathrm{as}_1(C) + 1$.

(ii) $\mathrm{as}_2(\widehat{C}_z)^2 = \mathrm{as}_2(C)^2 + 2\sqrt{\mathrm{as}_2(C)^2 + 2\mathrm{as}_1(C) + 1} - 1$. *Consequently we have* $\mathrm{as}_2(\widehat{C}_z) \le \mathrm{as}_2(C) + 1$, *where equality holds if and only if* $\mathrm{as}_2(C) = \mathrm{as}_1(C)$.

To prove Theorem 2.1, the following lemma is needed.

Lemma 2.2. *Let* C, z *be the same as in Theorem 2.1. For* x *in* $\mathrm{ri}(C)$, *the relative interior of* C, *if* $z_\lambda = \lambda z + (1-\lambda)x$ $(0 < \lambda < 1)$, *then*, *for any* $1 \le p < \infty$,

$$\mu_p(\widehat{C}_z, z_\lambda)^p = \lambda \left(\frac{1-\lambda}{\lambda} \right)^p + \frac{1}{(1-\lambda)^{p-1}} \int_{\mathbb{S}^{n-2}} (\lambda + \alpha_x(C, u))^p \, dm_x(C, u).$$

Proof. Since the family of $(n-1)$-dimensional polytopes is dense in $\mathcal{K}^{n-1}$, and $C_i \to C$ implies $\widehat{C}_i \to \widehat{C}$ (with respect to the Hausdorff metric), and $S_{n-1}(C, \cdot)$ is weakly continuous, we may assume, without loss of generality, that

$$C = \mathrm{conv}(v_1, \ldots, v_l)$$

is an $(n-1)$-dimensional polytope, where v_i are (all) vertices of C.

Thus

$$\widehat{C} = \operatorname{conv}(v_1, \dots, v_l, z)$$

is the n-dimensional polytope with vertices $v_1, \dots, v_l$ and z. Furthermore, if all facets of C are F_i $(1 \le i \le m)$, then all facets of $\widehat{C}$ are $\widehat{F_i} = \operatorname{conv}(F_i, z)$ $(1 \le i \le m)$ and C. We denote by $\tilde{u}_0 \in \mathbb{S}^{n-1}$ the outer normal vector of C (as a facet of $\widehat{C}$), by $\tilde{u}_i \in \mathbb{S}^{n-1}$ the outer normal vector of $\widehat{F_i}$, and by $u_i \in \mathbb{S}^{n-2} \equiv H^* \cap \mathbb{S}^{n-1}$, where H^* denotes the $(n-1)$-dimensional subspace parallel to the affine hull H of C, the outer normal vector of F_i (as a facet of C).

Use the fact that $\langle z - x, \tilde{u}_i \rangle = h_x(\widehat{C}, \tilde{u}_i)$ to observe that

$$\begin{aligned} h_{z_\lambda}(\widehat{C}, \tilde{u}_i) &= \sup_{y \in \widehat{C}} \langle y - z_\lambda, \tilde{u}_i \rangle = \sup_{y \in \widehat{C}} \langle (y + x - z_\lambda) - x, \tilde{u}_i \rangle \\ &= \sup_{y' \in \widehat{C} + x - z_\lambda} \langle y' - x, \tilde{u}_i \rangle = h_x(\widehat{C} + x - z_\lambda, \tilde{u}_i) \\ &= h_x(\widehat{C} + \lambda(x - z), \tilde{u}_i) = h_x(\widehat{C}, \tilde{u}_i) - \lambda \langle z - x, \tilde{u}_i \rangle \\ &= (1-\lambda) h_x(\widehat{C}, \tilde{u}_i). \end{aligned}$$

This, together with the fact that $h_x(\widehat{C}, \tilde{u}_i) + h_x(\widehat{C}, -\tilde{u}_i)$ is just the width of $\widehat{C}$ along $\tilde{u}_i$ and does not depend on the choice of x, in turn leads to

$$h_{z_\lambda}(\widehat{C}, -\tilde{u}_i) = \lambda h_x(\widehat{C}, \tilde{u}_i) + h_x(\widehat{C}, -\tilde{u}_i).$$

Finally we get

$$\begin{aligned} \alpha_{z_\lambda}(\widehat{C}, \tilde{u}_i) &= \frac{h_{z_\lambda}(\widehat{C}, -\tilde{u}_i)}{h_{z_\lambda}(\widehat{C}, \tilde{u}_i)} = \frac{\lambda}{1-\lambda} + \frac{1}{1-\lambda} \frac{h_x(\widehat{C}, -\tilde{u}_i)}{h_x(\widehat{C}, \tilde{u}_i)} \\ &= \frac{\lambda}{1-\lambda} + \frac{1}{1-\lambda} \frac{h_x(C, -u_i)}{h_x(C, u_i)} = \frac{\lambda}{1-\lambda} + \frac{1}{1-\lambda} \alpha_x(C, u_i). \end{aligned} \tag{2-1}$$

Furthermore,

$$\begin{aligned} h_{z_\lambda}(\widehat{C}, \tilde{u}_i) V_{n-1}(\widehat{F_i}) &= (1-\lambda) h_x(\widehat{C}, \tilde{u}_i) V_{n-1}(\widehat{F_i}) = (1-\lambda) n V_n(\widehat{C_i}) \\ &= (1-\lambda)(h_{z_\lambda}(\widehat{C}, u_0) + h_{z_\lambda}(\widehat{C}, -u_0)) V_{n-1}(C_i) \\ &= (1-\lambda)(h_{z_\lambda}(\widehat{C}, u_0) + h_{z_\lambda}(\widehat{C}, -u_0)) \frac{h_x(C, u_i)}{n-1} V_{n-2}(F_i), \end{aligned} \tag{2-2}$$

where $\widehat{C_i}$ and C_i denote, respectively, the n-dimensional body $\operatorname{conv}(z, x, F_i)$ and the $(n-1)$-dimensional body $\operatorname{conv}(x, F_i)$.

Now, by (2-1), (2-2) and the fact that

$$n V_n(\widehat{C}) = (h_{z_\lambda}(\widehat{C}, u_0) + h_{z_\lambda}(\widehat{C}, -u_0)) V_{n-1}(C),$$

it follows that

$$
\begin{aligned}
&\mu_p(\widehat{C}, z_\lambda)^p \\
&\quad = \int_{\mathbb{S}^{n-1}} \alpha_{z_\lambda}(\widehat{C}, u)^p \, dm_{z_\lambda}(\widehat{C}, u) \\
&\quad = \frac{\alpha_{z_\lambda}(\widehat{C}, u_0)^p h_{z_\lambda}(\widehat{C}, u_0) V_{n-1}(C) + \sum_{i=1}^m \alpha_{z_\lambda}(\widehat{C}, \tilde{u}_i)^p h_{z_\lambda}(\widehat{C}, \tilde{u}_i) V_{n-1}(\widehat{F}_i)}{n V_n(\widehat{C})} \\
&\quad = \frac{\alpha_{z_\lambda}(\widehat{C}, u_0)^p h_{z_\lambda}(\widehat{C}, u_0)}{h_{z_\lambda}(\widehat{C}, u_0) + h_{z_\lambda}(\widehat{C}, -u_0)} + \frac{\frac{1-\lambda}{n-1} \sum_{i=1}^m \alpha_{z_\lambda}(\widehat{C}, \tilde{u}_i)^p h_x(C, u_i) V_{n-2}(F_i)}{V_{n-1}(C)} \\
&\quad = \lambda\Big(\frac{1-\lambda}{\lambda}\Big)^p + \frac{1}{(1-\lambda)^{p-1}} \int_{\mathbb{S}^{n-2}} (\lambda + \alpha_x(C, u))^p \, dm_x(C, u),
\end{aligned}
$$

where we used the equalities $h_{z_\lambda}(\widehat{C}, u_0)/(h_{z_\lambda}(\widehat{C}, u_0) + h_{z_\lambda}(\widehat{C}, -u_0)) = \lambda$ and $\alpha_{z_\lambda}(\widehat{C}, u_0) = (1-\lambda)/\lambda$. □

Proof of Theorem 2.1. (i) In Lemma 2.2, taking $p = 1$, we have, for any $x \in \mathrm{ri}(C)$, $0 < \lambda < 1$,

$$
\begin{aligned}
\mathrm{as}_1(\widehat{C}_z) &= 1 - \lambda + \int_{\mathbb{S}^{n-2}} (\lambda + \alpha_x(C, u)) \, dm_x(C, u) \\
&= 1 + \int_{\mathbb{S}^{n-2}} \alpha_x(C, u) \, dm_x(C, u) = 1 + \mathrm{as}_1(C).
\end{aligned}
$$

(ii) In Lemma 2.2, taking $p = 2$ and noticing $\mu_1(C, x) = \mathrm{as}_1(C)$, we have

$$
\begin{aligned}
\mu_2(\widehat{C}, z_\lambda)^2 &= \frac{(1-\lambda)^2}{\lambda} + \frac{1}{1-\lambda} \int_{\mathbb{S}^{n-2}} (\lambda + \alpha_x(C, u))^2 \, dm_x(C, u) \\
&= \frac{(1-\lambda)^2}{\lambda} + \frac{\lambda^2}{1-\lambda} + \frac{2\lambda}{1-\lambda} \mathrm{as}_1(C) + \frac{1}{1-\lambda} \mu_2(C, x)^2 =: A(\lambda).
\end{aligned}
$$

Letting

$$
A'(\lambda) = \frac{(\mu_2(C, x)^2 + 2\mathrm{as}_1(C))\lambda^2 + 2\lambda - 1}{\lambda^2(1-\lambda)^2} = 0,
$$

we get $\lambda_0 = \big(\sqrt{\mu_2(C, x)^2 + 2\mathrm{as}_1(C) + 1} + 1\big)^{-1}$. Thus, with an elementary computation, it follows that

$$
\begin{aligned}
\min_{0<\lambda<1} \mu_2(\widehat{C}, z_\lambda)^2 &= A(\lambda_0) = \mu_2(\widehat{C}, z_{\lambda_0})^2 \\
&= \mu_2(C, x)^2 + 2\sqrt{\mu_2(C, x)^2 + 2\mathrm{as}_1(C) + 1} - 1.
\end{aligned}
$$

Therefore

$$
\mathrm{as}_2(\widehat{C}_z)^2 = \min_{x \in \mathrm{ri}(C)} \min_{0<\lambda<1} \mu_2(\widehat{C}, z_\lambda)^2 = \mathrm{as}_2(C)^2 + 2\sqrt{\mathrm{as}_2(C)^2 + 2\mathrm{as}_1(C) + 1} - 1.
$$

Since $\mathrm{as}_1(C) \le \mathrm{as}_2(C)$ by Theorem 1.3, we get

$$\mathrm{as}_2(\widehat{C}_z)^2 \le \mathrm{as}_2(C)^2 + 2\sqrt{\mathrm{as}_2(C)^2 + 2\mathrm{as}_2(C) + 1} - 1 = (\mathrm{as}_2(C) + 1)^2$$

which implies that $\mathrm{as}_2(\widehat{C}_z) \le \mathrm{as}_2(C) + 1$ and that equality holds if and only if $\mathrm{as}_2(C) = \mathrm{as}_1(C)$. □

Remark 2.3. Theorem 2.1 indicates that there are nontrivial $C \in \mathcal{K}^n$ $(n \ge 4)$ such that $\mathrm{as}_1(C) = \mathrm{as}_\infty(C)$: in $\mathbb{R}^n$, taking a symmetric $D \in \mathcal{K}^{n-2}$ and forming $\widehat{D}_y \in \mathcal{K}^{n-1}$, for each $C := \mathrm{conv}(z, \widehat{D}_y) \in \mathcal{K}^n$, we have, by Theorems 1.3 and 2.1 and Theorem 2 in [Guo and Kaijser 2002], $\mathrm{as}_1(C) = \mathrm{as}_2(C) = \mathrm{as}_\infty(C) = 3$, while clearly C is neither a symmetric convex body nor a simplex.

Now we introduce the so-called coproduct of subsets in different spaces and then generalize Theorem 2.1.

Definition. Given $C \subset \mathbb{R}^m$ and $D \subset \mathbb{R}^n$ $(m, n \ge 0)$, we define the coproduct body $C \amalg D \subset \mathbb{R}^{m+n+1}$ as

$$C \amalg D := \bigcup_{0 \le \lambda \le 1} (1-\lambda)C \times \lambda D \times \{\lambda\} = \{((1-\lambda)x, \lambda y, \lambda) \mid x \in C,\ y \in D,\ 0 \le \lambda \le 1\}.$$

Remark 2.4. (i) If both C and D are convex, then $C \amalg D = \mathrm{conv}(C \cup \tilde{D})$, where $\tilde{D} = \{0\} \times D \times \{1\} = \{(0, y, 1) \mid y \in D\}$ (in particular, $C \amalg D$ is convex). For example, $[-1, 1] \amalg [-1, 1] = \mathrm{conv}\{(1, 0, 0), (-1, 0, 0), (0, 1, 1), (0, -1, 1)\}$, a 3-dimensional simplex. In general, $[a, b] \amalg [c, d]$ is a 3-dimensional simplex.

(ii) If $C = \{v\}$ is a singleton and D is convex, then $C \amalg D$ reduces to the cone with vertex v and base D.

The next proposition, which may be checked easily, shows that, in a sense, the coproduct operation is the dual of product operation. For $C \in \mathcal{K}^n$, denote

$$C^A = \{f \in \mathrm{aff}(\mathbb{R}^n) \mid f(C) \subset [-1, 1]\}.$$

Proposition 2.5. *For any $C \in \mathcal{K}^m$ and $D \in \mathcal{K}^n$ under the correspondence*

$$(C \amalg D)^A \ni f \longleftrightarrow (f_{|C}, f_{|D}) \in C^A \times D^A,$$

where $(f_{|C}, f_{|D})\big(((1-\lambda)x, \lambda y, \lambda)\big) := (1-\lambda) f_{|C}(x) + \lambda f_{|D}(y)$, *we have*

$$(C \amalg D)^A = C^A \times D^A.$$

Now we can generalize (i) in Theorem 2.1 to the coproduct bodies.

Theorem 2.6. *For any $C \in \mathcal{K}^m$ and $D \in \mathcal{K}^n$ $(m, n \ge 0)$,*

$$\mathrm{as}_1(C \amalg D) = \mathrm{as}_1(C) + \mathrm{as}_1(D) + 1,$$

where we take the convention that $\mathrm{as}_1(C)$ *(or* $\mathrm{as}_1(D)$*)* $= 0$ *if* C *(or* D*)* $\in \mathcal{K}^0$.

In order to prove Theorem 2.6, more lemmas are needed. For any $\lambda \in \mathbb{R}$, $\varepsilon \geq 0$ and $A \subset \mathbb{R}^{m+n+1}$, denote

$$\mathbb{H}_\lambda := \mathbb{R}^m \times \mathbb{R}^n \times \{\lambda\}, \quad A_\lambda := A \cap \mathbb{H}_\lambda, \quad A_{\lambda,\varepsilon} := [A - \varepsilon A]_\lambda.$$

Lemma 2.7. *For any $C \in \mathcal{K}^m$ and $D \in \mathcal{K}^n$ $(m, n \geq 1)$,*

$$V_{m+n+1}(C \amalg D) = B(m+1, n+1) V_m(C) V_n(D),$$

where $B(\cdot, \cdot)$ is the Beta function.

Proof. Since $[C \amalg D]_\lambda = (1-\lambda)C \times \lambda D \times \{\lambda\}$ for $0 \leq \lambda \leq 1$, we have

$$V_{m+n}([C \amalg D]_\lambda) = V_{m+n}((1-\lambda)C \times \lambda D) = (1-\lambda)^m \lambda^n V_m(C) V_n(D).$$

Hence

$$V_{m+n+1}(C \amalg D) = \int_0^1 (1-\lambda)^m \lambda^n V_m(C) V_n(D)\, d\lambda = B(m+1, n+1) V_m(C) V_n(D).$$

□

Lemma 2.8. *If $o \in C$ and $o \in D$, then for any $0 \leq \varepsilon < 1$ and $0 \leq \lambda \leq 1 - \varepsilon$,*

$$V_{m+n}([C \amalg D]_{\lambda,\varepsilon}) = V_{m+n}\big(((1-\lambda)C - \varepsilon C)) \times ((\lambda+\varepsilon)D - \varepsilon D)\big) - \varepsilon^2 P^*(\lambda, \varepsilon),$$

where $P^*(\lambda, \varepsilon)$ is a polynomial of λ and ε.

Proof. Since

$$\begin{aligned} C \amalg D - \varepsilon(C \amalg D) \\ &= \bigcup_{0 \leq \mu, \nu \leq 1} ((1-\mu)C \times \mu D \times \{\mu\} - \varepsilon(1-\nu)C \times \varepsilon\nu D \times \{\varepsilon\nu\}) \\ &= \bigcup_{0 \leq \mu, \nu \leq 1} ((1-\mu)C - \varepsilon(1-\nu)C) \times (\mu D - \varepsilon\nu D) \times \{\mu - \varepsilon\nu\}, \end{aligned}$$

we have

$$[C \amalg D]_{\lambda,\varepsilon} = \bigcup_{\mu - \varepsilon\nu = \lambda} ((1-\mu)C - \varepsilon(1-\nu)C) \times (\mu D - \varepsilon\nu D) \times \{\lambda\}.$$

Thus,

$$[C \amalg D]_{\lambda,\varepsilon} \supset ((1-(\lambda+\varepsilon))C) \times ((\lambda+\varepsilon)D - \varepsilon D) \times \{\lambda\} =: E_1 \tag{2-3}$$

(the set when $\nu = 1$ and so $\mu = \lambda + \varepsilon$) and

$$[C \amalg D]_{\lambda,\varepsilon} \supset ((1-\lambda)C - \varepsilon C) \times (\lambda D) \times \{\lambda\} =: E_2 \tag{2-4}$$

(the set when $\nu = 0$ and so $\mu = \lambda$). We also have

$$[C \amalg D]_{\lambda,\varepsilon} \subset ((1-\lambda)C - \varepsilon C) \times ((\lambda+\varepsilon)D - \varepsilon D) \times \{\lambda\} =: E_3, \tag{2-5}$$

since, for $0 \le \mu,\ \nu \le 1$ with $\mu - \varepsilon\nu = \lambda$ (notice that $\lambda \le \mu \le \lambda + \varepsilon$ and that $o \in C$, $o \in D$),

$$\begin{aligned}(1-\mu)C - \varepsilon(1-\nu)C &\subset (1-\lambda)C - \varepsilon C,\\ \mu D - \varepsilon\nu D \subset \mu D - \varepsilon D &\subset (\lambda+\varepsilon)D - \varepsilon D.\end{aligned} \tag{2-6}$$

Now, setting

$$P(\lambda, \varepsilon) := V_{m+n}(E_3) - V_{m+n}([C \amalg D]_{\lambda,\varepsilon}),$$

which is a polynomial of λ and ε, we have by (2-3), (2-4), (2-5) and the fact that $(1-\lambda-\varepsilon)C \subset (1-\lambda)C - \varepsilon C$, $\lambda D \subset (\lambda+\varepsilon)D - \varepsilon D$,

$$\begin{aligned}0 &\le P(\lambda, \varepsilon)\\ &\le V_{m+n}(E_3) - V_{m+n}(E_1 \cup E_2)\\ &= V_{m+n}(E_3) - V_{m+n}(E_1) - V_{m+n}(E_2) + V_{m+n}(E_1 \cap E_2).\end{aligned}$$

By the polynomial expansion of the Minkowski sum (see Theorem 5.1.6 in [Schneider 1993]),

$$\begin{aligned}V_{m+n}(E_3) &= V_m((1-\lambda)C - \varepsilon C)V_n((\lambda+\varepsilon)D - \varepsilon D)\\ &= ((1-\lambda)^m V_m(C) + m\varepsilon(1-\lambda)^{m-1}V(C[m-1]) + \varepsilon^2 P_1'(\lambda, \varepsilon))\\ &\quad \times((\lambda+\varepsilon)^n V_n(D) + n\varepsilon(\lambda+\varepsilon)^{n-1}V(D[n-1]) + \varepsilon^2 P_1''(\lambda, \varepsilon))\\ &= (1-\lambda)^m(\lambda+\varepsilon)^n V_m(C)V_n(D)\\ &\quad + n\varepsilon(1-\lambda)^m(\lambda+\varepsilon)^{n-1}V_m(C)V(D[n-1])\\ &\quad\quad + m\varepsilon(1-\lambda)^{m-1}(\lambda+\varepsilon)^n V(C[m-1])V_n(D) + \varepsilon^2 P_1(\lambda, \varepsilon),\end{aligned} \tag{2-7}$$

$$\begin{aligned}V_{m+n}(E_1) &= V_m((1-\lambda-\varepsilon)C)V_n((\lambda+\varepsilon)D - \varepsilon D)\\ &= (1-\lambda-\varepsilon)^m V_m(C)\\ &\quad \times((\lambda+\varepsilon)^n V_n(D) + n\varepsilon(\lambda+\varepsilon)^{n-1}V(D[n-1]) + \varepsilon^2 P_2'(\lambda, \varepsilon))\\ &= (1-\lambda-\varepsilon)^m(\lambda+\varepsilon)^n V_m(C)V_n(D)\\ &\quad + n\varepsilon(1-\lambda-\varepsilon)^m(\lambda+\varepsilon)^{n-1}V_m(C)V(D[n-1]) + \varepsilon^2 P_2(\lambda, \varepsilon),\end{aligned}$$

$$\begin{aligned}V_{m+n}(E_2) &= V_m((1-\lambda)C - \varepsilon C)V_n(\lambda D)\\ &= ((1-\lambda)^m V_m(C) + m\varepsilon(1-\lambda)^{m-1}V(C[m-1]) + \varepsilon^2 P_3'(\lambda, \varepsilon))\lambda^n V_n(D)\\ &= (1-\lambda)^m\lambda^n V_m(C)V_n(D)\\ &\quad + m\varepsilon(1-\lambda)^{m-1}\lambda^n V(C[m-1])V_n(D) + \varepsilon^2 P_3(\lambda, \varepsilon),\end{aligned}$$

$$V_{m+n}(E_1 \cap E_2) = V_m((1-\lambda-\varepsilon)C)V_n(\lambda D) = (1-\lambda-\varepsilon)^m\lambda^n V_m(C)V_n(D),$$

where P_i', P_i'', P_i are polynomials of λ and ε, and

$$\begin{aligned}(1-\lambda)^m(\lambda+\varepsilon)^n &- (1-\lambda-\varepsilon)^m(\lambda+\varepsilon)^n - (1-\lambda)^m\lambda^n + (1-\lambda-\varepsilon)^m\lambda^n \\ &= ((1-\lambda)^m - (1-\lambda-\varepsilon)^m)(\lambda+\varepsilon)^n - ((1-\lambda)^m - (1-\lambda-\varepsilon)^m)\lambda^n \\ &= ((1-\lambda)^m - (1-\lambda-\varepsilon)^m)((\lambda+\varepsilon)^n - \lambda^n) \\ &= \varepsilon^2 Q_1(\lambda, \varepsilon),\end{aligned}$$

$$\begin{aligned}n\varepsilon(1-\lambda)^m(\lambda+\varepsilon)^{n-1} - n\varepsilon(1-\lambda-\varepsilon)^m(\lambda+\varepsilon)^{n-1} &= n\varepsilon((1-\lambda)^m - (1-\lambda-\varepsilon)^m)(\lambda+\varepsilon)^{n-1} \\ &= \varepsilon^2 Q_2(\lambda, \varepsilon),\end{aligned}$$

$$m\varepsilon(1-\lambda)^{m-1}((\lambda+\varepsilon)^n - \lambda^n) = \varepsilon^2 Q_3(\lambda, \varepsilon),$$

where Q_i are polynomials of λ and ε. Thus

$$V_{m+n}(E_3) - V_{m+n}(E_1) - V_{m+n}(E_2) + V_{m+n}(E_1 \cap E_2) = \varepsilon^2 Q(\lambda, \varepsilon)$$

for some polynomial $Q(\lambda, \varepsilon)$, and in turn $P(\lambda, \varepsilon) = \varepsilon^2 P^*(\lambda, \varepsilon)$ for some polynomial $P^*(\lambda, \varepsilon)$. □

Lemma 2.9. *For any $C \in \mathcal{K}^m$ and $D \in \mathcal{K}^n$ with $o \in C$, $o \in D$ $(m, n \geq 1)$,*

$$\begin{aligned}\frac{d}{d\varepsilon}\left(\int_0^{1-\varepsilon} V_{m+n}([C \amalg D]_{\lambda,\varepsilon})\, d\lambda\right)_{|\varepsilon=0} &= (m+n+1)B(m+1, n+1) \\ &\quad \times (V_m(C)V_n(D) + V(C[m-1])V_n(D) + V_m(C)V(D[n-1])).\end{aligned}$$

Proof. By Lemma 2.8 and (2-7), we have

$$\begin{aligned}V_{m+n}&([C \amalg D]_{\lambda,\varepsilon}) \\ &= V_{m+n}(E_3) - \varepsilon^2 P^*(\lambda, \varepsilon) \\ &= (1-\lambda)^m(\lambda+\varepsilon)^n V_m(C)V_n(D) + m\varepsilon(1-\lambda)^{m-1}(\lambda+\varepsilon)^n V(C[m-1])V_n(D) \\ &\qquad + n\varepsilon(\lambda+\varepsilon)^{n-1}(1-\lambda)^m V_m(C)V(D[n-1]) + \varepsilon^2 P_1(\lambda, \varepsilon) - \varepsilon^2 P^*(\lambda, \varepsilon).\end{aligned}$$

Thus, since

$$\frac{d}{d\varepsilon}\left(\int_0^{1-\varepsilon} (1-\lambda)^m(\lambda+\varepsilon)^n d\lambda\right)_{|\varepsilon=0} = nB(m+1, n),$$

$$\frac{d}{d\varepsilon}\left(\varepsilon\int_0^{1-\varepsilon} (1-\lambda)^{m-1}(\lambda+\varepsilon)^n d\lambda\right)_{|\varepsilon=0} = B(m, n+1),$$

$$\frac{d}{d\varepsilon}\left(\varepsilon\int_0^{1-\varepsilon}(1-\lambda)^m(\lambda+\varepsilon)^{n-1}d\lambda\right)_{|\varepsilon=0}=B(m+1,n),$$
$$\frac{d}{d\varepsilon}\left(\varepsilon^2\int_0^{1-\varepsilon}(P_1(\lambda,\varepsilon)-P^*(\lambda,\varepsilon))d\lambda\right)_{|\varepsilon=0}=0,$$

and because $mB(m,n+1)=nB(m+1,n)=(m+n+1)B(m+1,n+1)$, we get

$$\begin{aligned}\frac{d}{d\varepsilon}&\left(\int_0^{1-\varepsilon}V_{m+n}([C\amalg D]_{\lambda,\varepsilon})\,d\lambda\right)_{|\varepsilon=0}\\&=nB(m+1,n)V_m(C)V_n(D)+mB(m,n+1)V(C[m-1])V_n(D)\\&\qquad\qquad+nB(m+1,n)V_m(C)V(D[n-1])\\&=(m+n+1)B(m+1,n+1)\\&\qquad\times(V_m(C)V_n(D)+V(C[m-1])V_n(D)+V_m(C)V(D[n-1])).\qquad\square\end{aligned}$$

The following simple fact will be needed in the proof of Theorem 2.6.

Fact 2.10. *Suppose* $0\le u(t)\le v(t)$ *and* $u(0)=v(0)=0$. *If*

$$\frac{dv(t)}{dt^+}_{|t=0}=0,$$

then

$$0\le\lim_{t\to0^+}\frac{u(t)-u(0)}{t-0}\le\lim_{t\to0^+}\frac{v(t)-v(0)}{t-0}=0,$$

i.e.,

$$\frac{du(t)}{dt^+}_{|t=0}=0\quad\left(or\ \frac{du(t)}{dt}_{|t=0}=0\ \textit{if it exists}\right),$$

where d/dt^+ denotes the right derivative.

Proof of Theorem 2.6. If $m=n=0$, then $\mathrm{as}_1(C)=\mathrm{as}_1(D)=0$ and $C\amalg D$ is just the segment with ends o and $(0,0,1)$. Hence $\mathrm{as}_1(C\amalg D)=1=\mathrm{as}_1(C)+\mathrm{as}_1(D)+1$.

If $m=0, n\ge1$ (or $m\ge1, n=0$), it reduces to (i) in Theorem 2.1 (see (ii) in Remark 2.4).

Now we assume $m,n\ge1$ and $o\in C, o\in D$ (since $\mathrm{as}_1(\cdot)$ is affine invariant). Notice that $C\amalg D-\varepsilon(C\amalg D)$ is located in between $\mathbb{H}_{-\varepsilon}$ and $\mathbb{H}_1$ since $C\amalg D$ is located in between $\mathbb{H}_0$ and $\mathbb{H}_1$.

By the polynomial expansion of the Minkowski sum, we have that

$$\begin{aligned}\text{(2-8)}\quad(m+n+1)&V((C\amalg D)[m+n])\\&=\frac{d}{d\varepsilon}V_{m+n+1}(C\amalg D-\varepsilon C\amalg D)_{|\varepsilon=0}\\&=\frac{d}{d\varepsilon}\left[\left(\int_{-\varepsilon}^0+\int_0^{1-\varepsilon}+\int_{1-\varepsilon}^1\right)V_{m+n}([C\amalg D]_{\lambda,\varepsilon})\,d\lambda\right]_{|\varepsilon=0}.\end{aligned}$$

In order to compute

$$\frac{d}{d\varepsilon}\biggl(\int_{-\varepsilon}^{0} V_{m+n}([C \amalg D]_{\lambda,\varepsilon})\,d\lambda\biggr)_{|\varepsilon=0},$$

we observe that if $-\varepsilon \le \lambda \le 0$ and $\mu - \varepsilon\nu = \lambda$, then

$$((1-\mu)C - \varepsilon(1-\nu)C) \times (\mu D - \varepsilon\nu D) \subset (C - \varepsilon C) \times ((\lambda+\varepsilon)(D-D) + \lambda D),$$

since $(1-\mu)C - \varepsilon(1-\nu)C \subset C - \varepsilon C$ and

$$\mu D - \varepsilon\nu D = \mu D - (\mu-\lambda)D = \mu(D-D) - \lambda D \subset (\lambda+\varepsilon)(D-D) + \lambda D$$

(notice that $-\lambda \ge 0$, $o \in D - D$ and that $\mu - \varepsilon\nu = \lambda$ implies $\mu \le \lambda + \varepsilon$). So

(2-9)
$$0 \le \int_{-\varepsilon}^{0} V_{m+n}([C \amalg D]_{\lambda,\varepsilon})\,d\lambda \le \int_{-\varepsilon}^{0} V_m(C - \varepsilon C) V_n((\lambda+\varepsilon)(D-D) + \lambda D)\,d\lambda.$$

Denote $f(\lambda,\varepsilon) := V_m(C - \varepsilon C)V_n((\lambda+\varepsilon)(D-D)+\lambda D)$, which is a polynomial of λ and ε by the polynomial expansion of the Minkowski sum, and $f(0,0) = 0$. Thus

$$\frac{d}{d\varepsilon}\biggl(\int_{-\varepsilon}^{0} f(\lambda,\varepsilon)\,d\lambda\biggr)_{|\varepsilon=0} = f(0,0) = 0,$$

which, together with (2-9) and Fact 2.10, leads to

$$\frac{d}{d\varepsilon}\biggl(\int_{-\varepsilon}^{0} V_{m+n}([C \amalg D]_{\lambda,\varepsilon})\,d\lambda\biggr)_{|\varepsilon=0} = 0. \tag{2-10}$$

Similarly, we have

$$\frac{d}{d\varepsilon}\biggl(\int_{1-\varepsilon}^{1} V_{m+n}([C \amalg D]_{\lambda,\varepsilon})\,d\lambda\biggr)_{|\varepsilon=0} = 0. \tag{2-11}$$

Now, (2-8), (2-10), (2-11) and Lemma 2.9 show that

$$\begin{aligned}
&(m+n+1)V((C \amalg D)[m+n])\\
&\qquad = \frac{d}{d\varepsilon}\biggl(\int_{0}^{1-\varepsilon} V_{m+n}([C \amalg D]_{\lambda,\varepsilon})\,d\lambda\biggr)_{|\varepsilon=0}\\
&\qquad = (m+n+1)B(m+1,n+1)\\
&\qquad\qquad \times (V_m(C)V_n(D) + V(C[m-1])V_n(D) + V_m(C)V(D[n-1])),
\end{aligned}$$

which, together with (iii) in Remark 1.2 and Lemma 2.7, leads to

$$\begin{aligned}\mathrm{as}_1(C \amalg D) &= \frac{V((C \amalg D)[m+n])}{V_{m+n+1}(C \amalg D)} \\ &= \frac{V_m(C)V_n(D) + V(C[m-1])V_n(D) + V_m(C)V(D[n-1])}{V_m(C)V_n(D)} \\ &= 1 + \mathrm{as}_1(C) + \mathrm{as}_1(D). \qquad \square\end{aligned}$$

3. The Minkowski measure of coproducts of convex bodies

In this section, we will show that Theorem 2.6 also holds for the well-known Minkowski measure as_∞.

First, given $C \in \mathcal{K}^n$, for any fixed $x \in \mathrm{int}(C)$, define

$$\gamma(C, x) := \sup\{f(x) \mid f \in C^a\},$$

where $C^a := \{f \in \mathrm{aff}(R^n) \mid f(C) = [-1, 1]\}$. It is easy to check (see [Guo 2005]) that $\mu_\infty(C, x) = (1 + \gamma(C, x))/(1 - \gamma(C, x))$. Defining a measure of asymmetry $\mathrm{As}(C)$ of C by

$$\mathrm{As}(C) = \inf_{x \in \mathrm{int}(C)} \gamma(C, x),$$

we have $0 \le \mathrm{As}(C) \le (n-1)/(n+1)$ for $C \in \mathcal{K}^n$ and

$$\mathrm{As}(C) = \frac{\mathrm{as}_\infty(C) - 1}{\mathrm{as}_\infty(C) + 1} \quad \text{or} \quad \mathrm{as}_\infty(C) = \frac{1 + \mathrm{As}(C)}{1 - \mathrm{As}(C)}.$$

Then x is an ∞-critical point if and only if it is an As-critical point, and it is reasonable to study the Minkowski measure $\mathrm{as}_\infty(C)$ in terms of $\gamma(C, x)$ and $\mathrm{As}(C)$.

Definition. For $C \in \mathcal{K}^m$ and $D \in \mathcal{K}^n$, we define the affine direct sum of C^a and D^a, $C^a \boxplus D^a \subset \mathrm{aff}(\mathbb{R}^m) \times \mathrm{aff}(\mathbb{R}^n)$, by

$$C^a \boxplus D^a := (\{1_C\} \times D^a) \cup (C^a \times \{1_D\}),$$

where 1_C and 1_D denote the constant function 1 respectively on $\mathbb{R}^m$ and $\mathbb{R}^n$.

Under the same correspondence as in Proposition 2.5, $C^a \boxplus D^a$ can be identified with a subset of $(C \amalg D)^a$, and it is easy to check that

$$C^a \boxplus D^a \subsetneq (C \amalg D)^a \subsetneq (C \amalg D)^A.$$

Lemma 3.1. *Given $C \in \mathcal{K}^m$ and $D \in \mathcal{K}^n$, for any fixed $z = ((1-\lambda)x, \lambda y, \lambda)$ in* $\mathrm{int}(C \amalg D)$ *(i.e., $x \in \mathrm{ri}(C)$, $y \in \mathrm{ri}(D)$ and $0 < \lambda < 1$),*

$$\gamma_z := \gamma(C \amalg D, z) = \sup_{(f,g) \in (C \amalg D)^a} (f, g)(z) = \sup_{(f,g) \in C^a \boxplus D^a} (f, g)(z),$$

where $f \in \mathrm{aff}(\mathbb{R}^m)$, $g \in \mathrm{aff}(\mathbb{R}^n)$ and $(f, g)(z) := (1-\lambda)f(x) + \lambda g(y)$.

Proof. By a standard compactness argument, there is $(f_0, g_0) \in (C \amalg D)^a$ such that

$$\gamma_z = (f_0, g_0)(z) = (1-\lambda) f_0(x) + \lambda g_0(y). \tag{3-1}$$

Now we will show that $f_0 = 1_C$ and $g_0 \in D^a$ or $g_0 = 1_C$ and $f_0 \in C^a$.

To see this, we observe first that $f_0(C) \subset [-1, 1]$, $g_0(D) \subset [-1, 1]$ and

$$(f_0, g_0)(C \amalg D) = \operatorname{conv}(f_0(C) \cup g_0(D)), \tag{3-2}$$

which can be easily checked by the definition of $C \amalg D$ (in fact, this holds for any $(f, g) \in (C \amalg D)^a$).

Then we claim that $1 \in f_0(C)$ and $1 \in g_0(D)$. Suppose it is not true that, say, $1 \notin f_0(C)$. Then (3-2) implies that $1 \in g_0(D)$ since $(f_0, g_0)(C \amalg D) = [-1, 1]$, and either -1 is in $f_0(C)$ or $g_0(D)$. However, we will see that in either case there is a contradiction.

If $-1 \in g_0(D)$, then $g_0 \in D^a$. Thus $(1_C, g_0) \in (C \amalg D)^a$ and we have the inequality $(1_C, g_0)(z) > (f_0, g_0)(z)$, which contradicts (3-1).

If $-1 \in f_0(C)$, then we can find $f_1 \in C^a$ such that $\{f_1 = -1\} = \{f_0 = -1\}$, which implies that $f_1(x) > f_0(x)$ (since $1 \notin f_0(C)$). Thus $(f_1, g_0) \in (C \amalg D)^a$ and $(f_1, g_0)(z) > (f_0, g_0)(z)$ which contradicts (3-1) too. Hence we have confirmed our claim.

Now, with a similar argument, we can show that -1 is in $f_0(C)$ or $g_0(D)$.

If $-1 \in g_0(D)$, then $g_0 \in D^a$, and we must have $f_0 = 1_C$ since $(1_C, g_0) \in (C \amalg D)^a$ and $(1_C, g_0)(z) > (f, g_0)(z)$ for all $f \neq 1_C$. Thus $(f_0, g_0) = (1_C, g_0) \in C^a \boxplus D^a$.

Similarly, if $-1 \in f_0(C)$, then $g_0 = 1_D$ and so $(f_0, g_0) = (f_0, 1_D) \in C^a \boxplus D^a$. □

Now we can prove the following generalization of Theorem 2 in [Guo and Kaijser 2002].

Theorem 3.2. *For any $C \in \mathcal{K}^m$ and $D \in \mathcal{K}^n$ $(m, n \geq 0)$,*

$$\operatorname{as}_\infty(C \amalg D) = \operatorname{as}_\infty(C) + \operatorname{as}_\infty(D) + 1,$$

where we take the convention that $\operatorname{as}_\infty(C) = 0$ *for $C \in \mathcal{K}^0$. Moreover, all ∞-critical points z^* of $C \amalg D$ have the form*

$$z^* = \frac{1-\gamma_y}{2-\gamma_x-\gamma_y} x^* + \frac{1-\gamma_x}{2-\gamma_x-\gamma_y} y^*,$$

where $x^* = (x, 0, 0)$ with x being an ∞-critical point of C, and $y^* = (0, y, 1)$ with y being an ∞-critical point of D, and $\gamma_x := \gamma(C, x)$, $\gamma_y := \gamma(D, y)$.

Proof. If $m = n = 0$, the same argument as in the proof of Theorem 2.6 can be applied.

If $m = 0$, $n \geq 1$ or $m \geq 1$, $n = 0$, it reduces to Theorem 2 in [Guo and Kaijser 2002].

Now assume $m \geq 1,\ n \geq 1$. We first prove a general result: for any $\bar{x} := (x, 0, 0)$ with $x \in \mathrm{ri}(C)$ and $\bar{y} := (0, y, 1)$ with $y \in \mathrm{ri}(D)$,

$$\min_{z \in (\bar{x}, \bar{y})} \gamma(C \amalg D, z) = \gamma(C \amalg D, z_0) = \frac{1 - \gamma_x \gamma_y}{2 - \gamma_x - \gamma_y}, \tag{3-3}$$

where $(\bar{x}, \bar{y})$ is the open interval with $\bar{x}$, $\bar{y}$ as ends and

$$z_0 = \frac{1 - \gamma_y}{2 - \gamma_x - \gamma_y}\bar{x} + \frac{1 - \gamma_x}{2 - \gamma_x - \gamma_y}\bar{y}.$$

In fact, for any $(1, g), (f, 1) \in C^a \boxplus D^a$,

$$\begin{aligned}(1, g)(z_0) &= \frac{1 - \gamma_y}{2 - \gamma_x - \gamma_y} + \frac{1 - \gamma_x}{2 - \gamma_x - \gamma_y} g(y)\\ &\leq \frac{1 - \gamma_y}{2 - \gamma_x - \gamma_y} + \frac{1 - \gamma_x}{2 - \gamma_x - \gamma_y}\gamma_y = \frac{1 - \gamma_x \gamma_y}{2 - \gamma_x - \gamma_y},\end{aligned}$$

$$\begin{aligned}(f, 1)(z_0) &= \frac{1 - \gamma_y}{2 - \gamma_x - \gamma_y} f(x) + \frac{1 - \gamma_x}{2 - \gamma_x - \gamma_y}\\ &\leq \frac{1 - \gamma_y}{2 - \gamma_x - \gamma_y}\gamma_x + \frac{1 - \gamma_x}{2 - \gamma_x - \gamma_y} = \frac{1 - \gamma_x \gamma_y}{2 - \gamma_x - \gamma_y},\end{aligned}$$

with equality in the first formula if $g \in D^a$ such that $g(y) = \gamma_y$ and equality in the second formula if $f \in C^a$ such that $f(x) = \gamma_x$. So by Lemma 3.1, we have $\gamma_{z_0} = (1 - \gamma_x \gamma_y)/(2 - \gamma_x - \gamma_y)$.

Now, for $z = \lambda\bar{x} + (1 - \lambda)\bar{y} \in [\bar{x}, \bar{y}]$, if $\lambda > (1 - \gamma_y)/(2 - \gamma_x - \gamma_y)$, we choose $g_0 \in D^a$ such that $g_0(y) = \gamma_y$. Then

$$\begin{aligned}(1, g_0)(z) &= \lambda + (1 - \lambda)\gamma_y = (1 - \gamma_y)\lambda + \gamma_y\\ &> (1 - \gamma_y)\frac{1 - \gamma_y}{2 - \gamma_x - \gamma_y} + \gamma_y = \frac{1 - \gamma_x \gamma_y}{2 - \gamma_x - \gamma_y} = \gamma_{z_0},\end{aligned}$$

which implies that $\gamma_z > \gamma_{z_0}$.

If $\lambda < (1 - \gamma_y)/(2 - \gamma_x - \gamma_y)$, then, noticing that $\gamma_x - 1 < 0$, we have

$$\begin{aligned}(f_0, 1)(z) &= \lambda\gamma_x + (1 - \lambda) = (\gamma_x - 1)\lambda + 1\\ &> (\gamma_x - 1)\frac{1 - \gamma_y}{2 - \gamma_x - \gamma_y} + 1 = \frac{1 - \gamma_x \gamma_y}{2 - \gamma_x - \gamma_y} = \gamma_{z_0},\end{aligned}$$

which also implies that $\gamma_z > \gamma_{z_0}$. Hence (3-3) is confirmed.

Finally, since it is easy to check that $\mu_\infty(C,x) = (1+\gamma(C,x))/(1-\gamma(C,x))$, we can use the fact that the function $(1+t)/(1-t)$ is increasing on [0, 1), to get

$$\begin{aligned}
\mu_\infty(C \amalg D, z_0) &= \min_{z\in(\bar{x},\bar{y})} \mu_\infty(C \amalg D, z) = \frac{1+\gamma_{z_0}}{1-\gamma_{z_0}} \\
&= \Big(1 + \frac{1-\gamma_x\gamma_y}{2-\gamma_x-\gamma_y}\Big)\Big(1 - \frac{1-\gamma_x\gamma_y}{2-\gamma_x-\gamma_y}\Big)^{-1} \\
&= \frac{3-\gamma_x-\gamma_y-\gamma_x\gamma_y}{1-\gamma_x-\gamma_y+\gamma_x\gamma_y} = \frac{1+\gamma_x}{1-\gamma_x} + \frac{1+\gamma_y}{1-\gamma_y} + 1 \\
&= \mu_\infty(C,x) + \mu_\infty(D,y) + 1.
\end{aligned}$$

It follows that

$$\mathrm{as}_\infty(C \amalg D) = \min_{x\in\mathrm{ri}(C),\, y\in\mathrm{ri}(D)} (\mu_\infty(C,x) + \mu_\infty(D,y) + 1) = \mathrm{as}_\infty(C) + \mathrm{as}_\infty(D) + 1$$

and all ∞-critical points z^* of $C \amalg D$ have the form

$$z^* = \frac{1-\gamma_y}{2-\gamma_x-\gamma_y} x^* + \frac{1-\gamma_x}{2-\gamma_x-\gamma_y} y^*,$$

where $x^* = (x, 0, 0)$ with x being an ∞-critical point of C and $y^* = (0, y, 1)$ with y being an ∞-critical point of D. $\square$

Remark 3.3. Let $\mathcal{A} := \{C \in \mathcal{K}^k \mid \mathrm{as}_1(C) = \mathrm{as}_\infty(C),\ k = 0, 1, 2, \dots\}$ be the class of convex bodies whose p-measures coincide for all p, in all dimensions. Then $\mathcal{A}$ is closed under invertible affine transformations and coproducts of convex bodies, as follows from Theorems 2.6 and 3.2. Observe also that a simplex in k dimensions can be considered as the $(k+1)$-fold coproduct of its vertices (trivially symmetric convex bodies in 0 dimensions). Thus, we have naturally the following questions:

Question 1. Is the class of symmetric convex bodies a generating set for $\mathcal{A}$ under invertible affine transformations and coproducts?

Question 2. Does $\mathrm{as}_1(C) = \mathrm{as}_\infty(C)$ hold if $\mathrm{as}_1(C) = \mathrm{as}_2(C)$ (or, generally, if $\mathrm{as}_{p_1}(C) = \mathrm{as}_{p_2}(C)$ for distinct p_1, p_2)?

Acknowledgement

Our sincere thanks go to the referee for his invaluable comments and suggestions, in particular, for his formulating Question 1, which promotes the value of this paper.

References

[Asplund et al. 1962] E. Asplund, E. Grosswald, and B. Grünbaum, "On a measure of asymmetry of convex bodies", *Proc. Cambridge Philos. Soc.* **58** (1962), 217–220. MR 25 #3425 Zbl 0141.39001

[Besicovitch 1948] A. S. Besicovitch, "Measure of asymmetry of convex curves", *J. London Math. Soc.* **23** (1948), 237–240. MR 10,320r Zbl 0035.38402

[Böröczky 2010] K. J. Böröczky, "Stability of the Blaschke–Santaló and the affine isoperimetric inequality", *Adv. Math.* **225**:4 (2010), 1914–1928. MR 2011j:52016 Zbl 1216.52007

[Chakerian and Stein 1964] G. D. Chakerian and S. K. Stein, "On the symmetry of convex bodies", *Bull. Amer. Math. Soc.* **70** (1964), 594–595. MR 29 #517 Zbl 0126.38403

[Dziechcińska-Halamoda and Szwiec 1985] Z. Dziechcińska-Halamoda and W. Szwiec, "On critical sets of convex polyhedra", *Arch. Math.* (*Basel*) **44**:5 (1985), 461–466. MR 86j:52003 Zbl 0554.52004

[Eggleston 1952] H. G. Eggleston, "Measure of asymmetry of convex curves of constant width and restricted radii of curvature", *Quart. J. Math., Oxford Ser.* (2) **3** (1952), 63–72. MR 13,768d Zbl 0046.16003

[Ekström 2000] E. Ekström, "The critical set of a convex body", Project Report 2000:P7, Uppsala University, Department of Mathematics, 2000, available at http://www2.math.uu.se/research/pub/ekstrom.pdf.

[Gluskin and Litvak 2008] E. D. Gluskin and A. E. Litvak, "Asymmetry of convex polytopes and vertex index of symmetric convex bodies", *Discrete Comput. Geom.* **40**:4 (2008), 528–536. MR 2010d:52025 Zbl 1161.52008

[Groemer 2000] H. Groemer, "Stability theorems for two measures of symmetry", *Discrete Comput. Geom.* **24**:2–3 (2000), 301–311. MR 2001e:52002 Zbl 0960.52005

[Groemer and Wallen 2001] H. Groemer and L. J. Wallen, "A measure of asymmetry for domains of constant width", *Beiträge Algebra Geom.* **42**:2 (2001), 517–521. MR 2002i:52003 Zbl 0996.52001

[Grünbaum 1963] B. Grünbaum, "Measures of symmetry for convex sets", pp. 233–270 in *Proceedings of symposia in pure mathematics* (University of Washington, Seattle), vol. VII, edited by V. L. Klee, Amer. Math. Soc., Providence, R.I., 1963. MR 27 #6187 Zbl 0142.20503

[Guo 2005] Q. Guo, "Stability of the Minkowski measure of asymmetry for convex bodies", *Discrete Comput. Geom.* **34**:2 (2005), 351–362. MR 2006d:52003 Zbl 1079.52003

[Guo 2012] Q. Guo, "On p-measures of asymmetry for convex bodies", *Adv. Geom.* **12**:2 (2012), 287–301. MR 2911151 Zbl 1242.52004

[Guo and Kaijser 1999] Q. Guo and S. Kaijser, "On the distance between convex bodies", *Northeast. Math. J.* **15**:3 (1999), 323–331. MR 2001g:52007 Zbl 1011.52002

[Guo and Kaijser 2002] Q. Guo and S. Kaijser, "On asymmetry of some convex bodies", *Discrete Comput. Geom.* **27**:2 (2002), 239–247. MR 2002m:52002 Zbl 1010.52003

[Guo and Kaijser 2003] Q. Guo and S. Kaijser, "Approximation of convex bodies by convex bodies", *Northeast. Math. J.* **19**:4 (2003), 323–332. MR 2004k:52006 Zbl 1071.52010

[Hug and Schneider 2007] D. Hug and R. Schneider, "A stability result for a volume ratio", *Israel J. Math.* **161** (2007), 209–219. MR 2009e:52015 Zbl 1139.52012

[Kaiser 1996] M. J. Kaiser, "The critical point and related symmetry measures of a planar convex set", *Comput. Math. Appl.* **32**:9 (1996), 79–98. MR 98i:52002 Zbl 0871.52002

[Klee 1953] V. L. Klee, Jr., "The critical set of a convex body", *Amer. J. Math.* **75** (1953), 178–188. MR 14,678f Zbl 0050.16604

[Minkowski 1897] H. Minkowski, "Allgemeine Lehrsätze über die convexen Polyeder", *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl.* **1897** (1897), 198–219.

[Mizushima 2000] T. Mizushima, "Estimation of symmetry parameter and tests for symmetry", *Math. Japon.* **52**:3 (2000), 359–376. MR 2002a:62050 Zbl 0965.62032

[Petitjean 2003] M. Petitjean, "Chirality and symmetry measures: a transdisciplinary review", *Entropy* **5**:3 (2003), 271–312. MR 2054025 Zbl 1078.00503

[Rogers and Shephard 1958] C. A. Rogers and G. C. Shephard, "Some extremal problems for convex bodies", *Mathematika* **5** (1958), 93–102. MR 21 #2960 Zbl 0092.15301

[Schneider 1993] R. Schneider, *Convex bodies: the Brunn–Minkowski theory*, Encyclopedia of Mathematics and its Applications **44**, Cambridge University Press, Cambridge, 1993. MR 94d:52007 Zbl 0798.52001

[Schneider 2009] R. Schneider, "Stability for some extremal properties of the simplex", *J. Geom.* **96**:1–2 (2009), 135–148. MR 2011e:52013 Zbl 1201.52008

[Soltan 2005] V. Soltan, "Affine diameters of convex bodies—a survey", *Expo. Math.* **23**:1 (2005), 47–63. MR 2005k:52013 Zbl 1076.52001

[Stein 1956] S. Stein, "The symmetry function in a convex body", *Pacific J. Math.* **6** (1956), 145–148. MR 18,228g Zbl 0071.15701

[Toth 2008] G. Toth, "On the structure of convex sets with applications to the moduli of spherical minimal immersions", *Beiträge Algebra Geom.* **49**:2 (2008), 491–515. MR 2010a:53120 Zbl 1167.52008

[Toth 2009] G. Toth, "Asymmetry of convex sets with isolated extreme points", *Proc. Amer. Math. Soc.* **137**:1 (2009), 287–295. MR 2010f:52004 Zbl 1165.52005

[Tuzikov et al. 1997] A. V. Tuzikov, G. L. Margolin, and A. I. Grenov, "Convex set symmetry measurement via Minkowski addition", *J. Math. Imaging Vision* **7**:1 (1997), 53–68. MR 98a:52004 Zbl 0868.68113

[Tuzikov et al. 2000] A. V. Tuzikov, J. B. Roerdink, and H. J. Heijmans, "Similarity measures for convex polyhedra based on Minkowski addition", *Pattern Recognition* **33**:6 (2000), 979–995.

[Zouaki 2003] H. Zouaki, "Convex set symmetry measurement using Blaschke addition", *Pattern Recognition* **36**:3 (2003), 753–763. Zbl 1029.68128

Received August 8, 2014. Revised March 7, 2015.

QI GUO
DEPARTMENT OF MATHEMATICS
SUZHOU UNIVERSITY OF SCIENCE AND TECHNOLOGY
SUZHOU, JIANGSU 215009
CHINA
guoqi@mail.usts.edu.cn

JINFENG GUO
DEPARTMENT OF MATHEMATICS
SUZHOU UNIVERSITY OF SCIENCE AND TECHNOLOGY
SUZHOU, JIANGSU 215009
CHINA
gjfsx@mail.usts.edu.cn

XUNLI SU
DEPARTMENT OF MATHEMATICS
SUZHOU UNIVERSITY OF SCIENCE AND TECHNOLOGY
SUZHOU, JIANGSU 215009 CHINA
hardysoo@163.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.419

REGULARITY AND ANALYTICITY OF SOLUTIONS IN A DIRECTION FOR ELLIPTIC EQUATIONS

YONGYANG JIN, DONGSHENG LI AND XU-JIA WANG

In this paper, we study the regularity and analyticity of solutions to linear elliptic equations with measurable or continuous coefficients. We prove that if the coefficients and inhomogeneous term are Hölder-continuous in a direction, then the second-order derivative in this direction of the solution is Hölder-continuous, with a different Hölder exponent. We also prove that if the coefficients and the inhomogeneous term are analytic in a direction, then the solution is analytic in that direction.

1. Introduction

We study the regularity and analyticity of solutions in a given direction to the elliptic equation

$$\sum_{i,j=1}^{n} a_{ij}(x)u_{x_i x_j} + \sum_{i=1}^{n} b_i(x)u_{x_i} + c(x)u = f(x) \quad \text{in } \Omega, \tag{1-1}$$

assuming that the coefficients a_{ij}, b_i, c and the inhomogeneous term f are smooth or analytic along the direction, where Ω is a bounded domain in the Euclidean space $\mathbb{R}^n$. We assume that the equation is uniformly elliptic, namely, that there exist positive constants $\Lambda > \lambda > 0$ such that

$$\lambda|\xi|^2 \le \sum a_{ij}(x)\xi_i\xi_j \le \Lambda|\xi|^2 \quad \text{for all } x \in \Omega. \tag{1-2}$$

We also assume that $b_i, c \in L^\infty(\Omega)$, and $f \in L^n(\Omega)$.

The regularity of solutions is a fundamental issue in the study of partial differential equations. Most regularity theories, such as the Schauder estimate and the $W^{2,p}$ estimate, are isotropic; namely, the solution is uniformly regular in all directions. An interesting question is whether the solution to (1-1) is smooth in a direction if the coefficients a_{ij}, b_i, c and the inhomogeneous term f are smooth in this direction only. This question can be asked for more general nonlinear elliptic and parabolic

Jin was supported by ZJNSF LY14A010016 and NSFC 11371323, Li was supported by NSFC 11171266, and Wang was supported by ARC DP1094303 and DP120102718.
MSC2010: primary 35J15; secondary 35B45.
Keywords: Elliptic equation, analyticity, estimates, perturbation method.

equations. One may also consider the regularity when the coefficients a_{ij}, b_i, c and the inhomogeneous term f are smooth in a submanifold of high codimensions.

This is a significant problem in partial differential equations as it is not only stronger than the Schauder estimate but also has applications in areas such as fluid mechanics, partial differential systems, manifolds with nonsmooth metric tensors, and other physical problems such as the propagation of singularities [Taylor 2000; Kukavica and Ziane 2007; Cao and Titi 2008; 2011]. For many PDE systems if one can first prove the regularity of solutions in a direction, one may be able to obtain the full regularity. At a first glance, one may feel that an affirmative answer would be too good to be true, even for an expert in the area. However in this paper we show that this is indeed true at least in dimension two, and also in higher dimensions if the coefficients are continuous. At the moment we are not aware of a counterexample without the continuity. This question is also open for most nonlinear equations and deserves further investigations.

The analyticity of solutions is also an important topic in the regularity theory of partial differential equations. For the linear elliptic equation (1-1), it is well known that if the coefficients a_{ij}, b_i, c and the inhomogeneous term f are analytic, then the solution is also analytic. A similar question is whether the solution is analytic in a direction if a_{ij}, b_i, c and f are analytic only in the given direction.

Let us first state our results on the analyticity of solutions in a given direction:

Theorem 1.1. *Let $u \in W^{2,n}(\Omega)$ be a strong solution to* (1-1). *Assume that the coefficients a_{ij}, b_i, c and the inhomogeneous term f are independent of the variable x_n. Then the solution u is analytic in x_n.*

The proof of Theorem 1.1 is based on the Krylov–Safonov Hölder-continuity of linear elliptic equations. Using the $W^{2,p}$ estimate, we also have:

Theorem 1.2. *Let $u \in W^{2,n}(\Omega)$ be a strong solution to* (1-1). *Assume that the coefficients a_{ij} are continuous, and a_{ij}, b_i, c and f are analytic in the variable x_n. Then the solution u is analytic in x_n.*

In Theorem 1.1, we do not assume the continuity of the coefficients a_{ij}, b_i, c but in Theorem 1.2 we do. An interesting question is whether one can remove the continuity of the a_{ij} in Theorem 1.2. An affirmative answer can be given in dimension two:

Theorem 1.3. *Let $u \in W^{2,2}(\Omega)$ be a strong solution to* (1-1). *Assume that $n = 2$ and a_{ij}, b_i, c and f are analytic in the variable x_2. Then the solution u is analytic in x_2.*

Our results are stronger than the classical results on the analyticity of solutions to linear elliptic equations. In the classical theory the coefficients a_{ij}, b_i, c and the inhomogeneous term f are assumed to be analytic in all directions.

When the coefficients are Hölder-continuous in a given direction, we have the following directional $C^{2,\alpha}$ regularity:

Theorem 1.4. *Let $u \in W^{2,n}(\Omega)$ be a strong solution to* (1-1). *Suppose that a_{ij}, b_i, c are C^α in the ξ-direction for some $0 < \alpha < 1$ and $a_{ij} \in C^0(\Omega)$ and satisfy* (1-2). *Suppose $f \in L^p(\Omega)$ for some $p > n/\alpha$. Then for any $0 < \beta < \alpha - n/p$ and any $y, z \in \Omega_\delta$, we have the estimate*

$$
\begin{aligned}
(1\text{-}3)\quad & |\partial_\xi \partial_x u(y) - \partial_\xi \partial_x u(z)| \\
&\le C d^\beta \left[\sup_\Omega |u| + \|f\|_{L^p(\Omega)} + \int_d^1 \frac{\omega_{f,\xi}(r)}{r^{1+\beta}} \right] + C \int_0^d \frac{\omega_{f,\xi}(r)}{r} \\
&\qquad + C \|a_{ij}\|_{C^\alpha_\xi(\Omega)} (\|f\|_{L^p(\Omega)} + \sup_\Omega |u|) d^{\alpha - n/p},
\end{aligned}
$$

where $\Omega_\delta = \{x \in \Omega \mid \operatorname{dist}(x, \partial\Omega) > \delta\}$ and $d = |y - z|$. The constant C depends on $n, \alpha, \beta, \delta, p, \lambda, \Lambda$ and the modulus of continuity of a_{ij}.

In Theorem 1.4, ξ is a given unit vector, and the notation $\omega_{f,\xi}$ is defined at the beginning of Section 4. The continuity assumption of the a_{ij} is for the use of the $W^{2,p}$ estimate, hence it suffices to assume that the a_{ij} are in the VMO space [Chiarenza et al. 1993], or the a_{ij} are continuous in $n-1$ variables [Kim and Krylov 2007]. In particular, in dimension two, by the $W^{2,p}$ estimate in the latter reference, the continuity of the a_{ij} is not needed. Hence we have:

Corollary 1.5. *Let $u \in W^{2,2}(\Omega)$ be a strong solution to* (1-1). *Assume that $n = 2$ and a_{ij}, b_i, c and f are Hölder-continuous in direction ξ. Then $\partial_\xi \partial_x u$ is Hölder-continuous.*

Note that the Hölder-continuity of $\partial_\xi \partial_x u$ in Theorem 1.4 and Corollary 1.5 is uniform in all directions. But the Hölder exponent of the second derivative is smaller than that of the coefficients and we need to assume $f \in L^p$ for a large p.

Theorem 1.4 improves [Tian and Wang 2010, Theorem 3.2], where the coefficients a_{ij} were assumed to be Lipschitz in ξ, and the directional $C^{2,\alpha}$ regularity was obtained by differentiating (1-1). We point out that Corollary 1.5 was also obtained in [Dong 2012, Section 6]. By the $W^{2,p}$ estimate [Kim and Krylov 2007], related result holds in higher dimension too. That is, if u is a strong solution to (1-1) and if a_{ij}, b_i, c and f are Hölder-continuous in $x' = (x_1, \dots, x_{n-1})$, then $\partial_{x'} \partial_x u$ is Hölder-continuous. The $C^{2,\alpha}$ regularity of solutions in a given direction was also investigated in [Dong and Kim 2011]. See also [Tian and Wang 2010] for discussions.

To prove Theorems 1.1–1.3, we introduce appropriate function spaces and establish related interpolation inequalities. We will prove Theorem 1.1 in Section 2, Theorems 1.2 and 1.3 in Section 3, and Theorem 1.4 in Section 4. In Section 5, we give a brief discussion on equations of divergence form.

2. Proof of Theorem 1.1

For simplicity we assume $b_i = c = 0$; namely, we consider the equation

$$L[u] := \sum_{i,j=1}^{n} a_{i,j}(x')u_{ij} = f(x) \quad \text{in } \Omega, \tag{2-1}$$

where Ω is a bounded domain in $\mathbb{R}^n$, $x' = (x_1, \dots, x_{n-1})$, and $u_{ij} = u_{x_i x_j}$. The proof is similar if $b_i \neq 0$ and $c \neq 0$, provided they satisfy the conditions specified in the introduction. We assume that the coefficients a_{ij} are measurable and satisfy the uniformly elliptic condition (1-2), $f \in L^n(\Omega)$, and the a_{ij} and f are analytic in the x_n variable.

Set $u' = u_{x_n}$, $u'' = u_{x_n x_n}$,

$$u^{(k)} = \frac{\partial^k u}{\partial x_n^k}, \quad k = 1, 2, \dots,$$

$$\langle u\rangle_{\alpha,\Omega} = \sup_{x,y\in\Omega}\left\{\frac{|u(x)-u(y)|}{|x-y|^\alpha} \;\middle|\; (y-x)//e_n\right\},$$

and

$$|u|_{k+\alpha,\Omega} = \sup_\Omega |u| + \langle u^{(k)}\rangle_{\alpha,\Omega}, \quad k = 0, 1, 2, \dots, \tag{2-2}$$

$$\|u\|_{k+\alpha,\Omega} = \sup_\Omega |u| + \sup_{x,y\in\Omega} \frac{|D^k u(x) - D^k u(y)|}{|x-y|^\alpha},$$

where $0 < \alpha \le 1$ and $(y - x)//e_n$ means the vector $y - x$ is parallel to the vector $e_n = (0, \dots, 0, 1)$. We also set

$$\langle u^{(k)}\rangle^{(\beta)}_{\alpha,\Omega} = \sup_{Q_{2r}(x)\subset\Omega} r^{k+\alpha+\beta}\langle u^{(k)}\rangle_{\alpha,Q_r(x)}, \quad \beta \in \mathbb{R},$$

$$|u|^{(\beta)}_{k+\alpha,\Omega} = \sup_{Q_{2r}(x)\subset\Omega}[r^\beta\|u\|_{L^\infty(Q_r(x))} + r^{k+\alpha+\beta}\langle u^{(k)}\rangle_{\alpha,Q_r(x)}], \tag{2-3}$$

and

$$\|u\|^{(\beta)}_{k+\alpha,\Omega} = \sup_{Q_{2r}(x)\subset\Omega}\left[r^\beta\|u\|_{L^\infty(Q_r(x))} + r^{k+\alpha+\beta}\sup_{y,z\in Q_r(x)}\frac{|D^k u(y) - D^k u(z)|}{|y-z|^\alpha}\right],$$

where $Q_r(x)$ denotes the open cube with center x and side-length $2r$. We can extend the above definition to $\alpha = 0$ by letting

$$|u|^{(\beta)}_{k,\Omega} = \sup_{Q_{2r}(x)\subset\Omega}[r^\beta\|u\|_{L^\infty(Q_r(x))} + r^{k+\beta}\langle u^{(k-1)}\rangle_{1,Q_r(x)}] \quad \text{if } k > 0,$$

$$|u|^{(\beta)}_{0,\Omega} = \sup_{Q_{2r}(x)\subset\Omega} r^\beta\|u\|_{L^\infty(Q_r(x))} \quad \text{if } k = 0.$$

We point out the equivalence of the norm $|u|^{(\beta)}_{k+\alpha,\Omega}$ given in (2-3) and the norm

$$[u]^{(\beta)}_{k+\alpha,\Omega} := \sup_{Q_{(1+\sigma)r}(x)\subset\Omega}[r^{\beta}\|u\|_{L^\infty(Q_r(x))} + r^{k+\alpha+\beta}\langle u^{(k)}\rangle_{\alpha,Q_r(x)}],$$

where $\sigma > 0$ is a constant. Namely,

$$C^{-1}|u|^{(\beta)}_{k+\alpha,\Omega} \le [u]^{(\beta)}_{k+\alpha,\Omega} \le C|u|^{(\beta)}_{k+\alpha,\Omega},$$

for some constant C depending only on n, k, α, β and σ. To prove the above inequalities, it suffices to divide the cube $Q_{3r/2}$ into 2^n disjoint smaller cubes if $\sigma \in \left[\frac{1}{2}, 2\right]$, and divide into more, smaller cubes for other σ. Note that if $\beta = -k$, the constant C is independent of k.

We also point out three differences between our definition of the norms $|u|^{(\beta)}_{k+\alpha,\Omega}$ and the usual one [Gilbarg and Trudinger 1998]. That is, (i) the derivative in the former one is taken only on the x_n-direction; (ii) in the Hölder seminorm (2-2) we assume that $(y-x)//e_n$; and (iii) the supremum in (2-3) is taken among all cubes $Q_r(x)$ satisfying the condition $Q_{2r}(x) \subset \Omega$. The reason of choosing the cubes with the property $Q_{2r}(x) \subset \Omega$ is that the norm is homogeneous under rescaling.

First we prove an interpolation inequality for the norm $\|u\|^{(\beta)}_{k+\alpha,\Omega}$:

Lemma 2.1. *Suppose that $j+\beta < k+\alpha$, where $j, k = 0, 1, 2, \ldots$ and $0 \le \alpha, \beta \le 1$. Assume that $u \in C^{k,\alpha}(\Omega)$. Then there exists a positive constant C depending on j, k, α, β, such that*

$$\|u\|^{(\gamma)}_{j+\beta,\Omega} \le C[\|u\|^{(\gamma)}_{k+\alpha,\Omega}]^{(j+\beta)/(k+\alpha)}[\|u\|^{(\gamma)}_{0,\Omega}]^{1-(j+\beta)/(k+\alpha)}. \tag{2-4}$$

Proof. It is well known [Hörmander 1976] that there is a positive constant $C = C(j,k,\alpha,\beta)$ such that

$$\|u\|_{j+\beta,Q_1(0)} \le C(\|u\|_{k+\alpha,Q_1(0)})^{(j+\beta)/(k+\alpha)}(\|u\|_{L^\infty(Q_1(0))})^{1-(j+\beta)/(k+\alpha)}. \tag{2-5}$$

For any $Q_r(x) \subset \Omega$, by rescaling, we obtain

$$\begin{aligned}\|u\|_{L^\infty(Q_r(x))} &+ r^{j+\beta}\langle D^j u\rangle_{\beta,Q_r(x)} \\ &\le C(\|u\|_{L^\infty(Q_r(x))})^{1-(j+\beta)/(k+\alpha)} \\ &\quad\times \left(\|u\|_{L^\infty(Q_r(x))} + r^{k+\alpha}\langle D^k u\rangle_{\alpha,Q_r(x)}\right)^{(j+\beta)/(k+\alpha)}.\end{aligned} \tag{2-6}$$

That is,

$$\begin{aligned}r^{\gamma}\|u\|_{L^\infty(Q_r(x))} &+ r^{j+\beta+\gamma}\langle D^j u\rangle_{\beta,Q_r(x)} \\ &\le C(r^{\gamma}\|u\|_{L^\infty(Q_r(x))})^{1-(j+\beta)/(k+\alpha)} \\ &\quad\times \left(r^{\gamma}\|u\|_{L^\infty(Q_r(x))} + r^{k+\alpha+\gamma}\langle D^k u\rangle_{\alpha,Q_r(x)}\right)^{(j+\beta)/(k+\alpha)}.\end{aligned}$$

Taking the supremum of all cubes $Q_r(x)$ with $Q_{2r}(x) \subset \Omega$, we obtain (2-4). □

Next we extend the inequality (2-4) to the norm $|u|^{(\beta)}_{k+\alpha,\Omega}$:

Lemma 2.2. *Suppose that $j+\beta<k+\alpha$, where $j, k=0, 1, 2, \dots$ and $0\le\alpha, \beta\le 1$. Assume that $u\in L^\infty(\Omega)$ and $u^{(k)}\in C^\alpha(\Omega)$. Then there exists a positive constant C depending on j, k, α, β, such that*

$$|u|^{(\gamma)}_{j+\beta,\Omega}\le C[|u|^{(\gamma)}_{k+\alpha,\Omega}]^{(j+\beta)/(k+\alpha)}[|u|^{(\gamma)}_{0,\Omega}]^{1-(j+\beta)/(k+\alpha)}.$$

Proof. By the rescaling argument in the proof of Lemma 2.1, it suffices to prove

$$|u|_{j+\beta,Q_1(0)}\le C(|u|_{k+\alpha,Q_1(0)})^{(j+\beta)/(k+\alpha)}(\|u\|_{L^\infty(Q_1(0))})^{1-(j+\beta)/(k+\alpha)}. \tag{2-7}$$

By the definition (2-3), it suffices to prove

$$\langle u^{(j)}\rangle_{\beta,Q_1(0)}\le C(|u|_{k+\alpha,Q_1(0)})^{(j+\beta)/(k+\alpha)}(\|u\|_{L^\infty(Q_1(0))})^{1-(j+\beta)/(k+\alpha)}. \tag{2-8}$$

Again, by the definition of (2-3), there exists x_0' such that

$$\begin{aligned}\langle u^{(j)}\rangle_{\beta,Q_1(0)}&\le 2\sup\left\{\frac{|u^{(j)}(x_0',x_n)-u^{(j)}(x_0',y_n)|}{|x_n-y_n|^\beta}\;\Big|\;-1<x_n,y_n<1\right\}\\&=2\langle u^{(j)}(x_0',\cdot)\rangle_{\beta,I},\end{aligned}$$

where $I=(-1,1)\subset\mathbb{R}^1$ is the unit interval. By (2-5) in the one-dimensional case, the right-hand side is bounded by

$$\begin{aligned}\langle u^{(j)}(x_0',\cdot)\rangle_{\beta,I}&\le(\|u(x_0',\cdot)\|_{k+\alpha,I})^{(j+\beta)/(k+\alpha)}(\|u(x_0',\cdot)\|_{L^\infty(I)})^{1-(j+\beta)/(k+\alpha)}\\&\le(|u|_{k+\alpha,Q_1(0)})^{(j+\beta)/(k+\alpha)}(\|u\|_{L^\infty(Q_1(0))})^{1-(j+\beta)/(k+\alpha)}.\end{aligned}$$ □

Theorem 2.3. *Let $u\in W^{2,n}(\Omega)$ be a strong solution of* (2-1), *where the coefficients a_{ij} are measurable and independent of x_n and satisfy the uniformly elliptic condition* (1-2). *Assume that f is analytic in x_n. Then there exists a constant $C=C(n,\lambda,\Lambda)$ such that, for any $Q_R(x_0)\subset\Omega$, the following inequality holds*:

$$|u^{(k)}(x_0)|\le\left(\frac{Ck}{R}\right)^k(\|u\|_{L^\infty(Q_R(x_0))}+1). \tag{2-9}$$

Proof. As the coefficients a_{ij} are independent of x_n and u is a strong solution, one sees that

$$u_\delta':=\frac{1}{\delta}(u(x+\delta e_n)-u(x))$$

is a strong solution to $L[u]=f_\delta'$, where L is the elliptic operator in (2-1). Hence the Krylov–Safonov Hölder estimate holds for u_δ', uniformly in δ. Similarly,

$$u_\delta'':=\frac{1}{\delta^2}(u(x+\delta e_n)+u(x-\delta e_n)-2u(x))$$

is a strong solution to $L[u]=f_\delta''$, and is uniformly Hölder-continuous as $\delta\to 0$. Sending $\delta\to 0$, we see that u'' is Hölder-continuous. By induction, we see that for any $k > 0$, $u^{(k)}$ is Hölder-continuous, and

$$\langle u^{(k)}\rangle_{\alpha, Q_{1/4}(x)} \le C\big(\|u^{(k)}\|_{L^\infty(Q_{1/2}(x))} + \|f^{(k)}\|_{L^\infty(Q_{1/2}(x))}\big) \tag{2-10}$$

for all $k = 1, 2, \dots$, and the constant C is independent of k.

Set $Q_0 = Q_R(x_0)$. Let $Q_{2r}(\hat{x}) \subset Q_R(x_0)$ be any given cube. Then there exist $x_1, x_2 \in Q_r(\hat{x})$ with $(x_2 - x_1)//e_n$ such that

$$r^{1+\alpha}\langle u'\rangle_{\alpha, Q_r(\hat{x})} \le 2r^{1+\alpha}\frac{|u'(x_2) - u'(x_1)|}{|x_2 - x_1|^\alpha}.$$

If $|x_2 - x_1| \ge \frac{1}{4}r$, then, by Lemma 2.2 with $j = 1$, $\beta = 0$, $k = 1$,

$$\begin{aligned}
r^{1+\alpha}&\langle u'\rangle_{\alpha, Q_r(\hat{x})} \\
&\le 2\cdot 4^\alpha r|u'(x_1) - u'(x_2)| \\
&\le 4^{1+\alpha} r\|u'\|_{L^\infty(Q_r(\hat{x}))} \\
&\le C\big(r^{1+\alpha}\langle u'\rangle_{\alpha, Q_r(\hat{x})} + \|u\|_{L^\infty(Q_r(\hat{x}))}\big)^{1/(1+\alpha)}(\|u\|_{L^\infty(Q_r(\hat{x}))})^{\alpha/(1+\alpha)} \\
&\le C\big[(r^{1+\alpha}\langle u'\rangle_{\alpha, Q_r(\hat{x})})^{1/(1+\alpha)}(\|u\|_{L^\infty(Q_r(\hat{x}))})^{\alpha/(1+\alpha)} + \|u\|_{L^\infty(Q_r(\hat{x}))}\big].
\end{aligned} \tag{2-11}$$

If $|x_2 - x_1| < \frac{1}{4}r$, then, by (2-10) and Lemma 2.2,

$$\begin{aligned}
r^{1+\alpha}\langle u'\rangle_{\alpha, Q_r(\hat{x})} &\le 2\cdot r^{1+\alpha}\langle u'\rangle_{\alpha, Q_{r/4}(x_1)} \\
&\le C[r\|u'\|_{L^\infty(Q_{r/2}(x_1))} + r\|f'\|_{L^\infty(Q_{r/2}(x_1))}] \\
&\le C\big\{(r^{1+\alpha}\langle u'\rangle_{\alpha, Q_{r/2}(x_1)})^{1/(1+\alpha)}(\|u\|_{L^\infty(Q_{r/2}(x_1))})^{\alpha/(1+\alpha)} \\
&\qquad + \|u\|_{L^\infty(Q_{r/2}(x_1))} + r\|f'\|_{L^\infty(Q_{r/2}(x_1))}\big\}.
\end{aligned} \tag{2-12}$$

Taking the supremum among all the cubes $Q_r(\hat{x})$ with $Q_{2r}(\hat{x}) \subset Q_R(x_0)$, we obtain from the above estimates (2-11) and (2-12) that

$$\langle u'\rangle^{(0)}_{\alpha, Q_0} \le C\big\{(\langle u'\rangle^{(0)}_{\alpha, Q_0})^{1/(1+\alpha)}(\|u\|_{L^\infty(Q_0)})^{\alpha/(1+\alpha)} + \|u\|_{L^\infty(Q_0)} + R\|f'\|_{L^\infty(Q_0)}\big\},$$

which implies

$$|u|^{(0)}_{1+\alpha, Q_0} \le C(\|u\|_{L^\infty(Q_0)} + R\|f'\|_{L^\infty(Q_0)}).$$

By Lemma 2.2 it follows that

$$\|u'\|_{L^\infty(Q_{R/2}(x_0))} \le \frac{C}{R}(\|u\|_{L^\infty(Q_0)} + R\|f'\|_{L^\infty(Q_0)}).$$

Hence we obtain

$$\begin{aligned}
|u'(x_0)| &\le \frac{C}{R}(\|u\|_{L^\infty(Q_0)} + R\|f'\|_{L^\infty(Q_0)}) \\
&\le \frac{C}{R}(\|u\|_{L^\infty(Q_0)} + 1),
\end{aligned} \tag{2-13}$$

where we used the analyticity of f in x_n.

Next we estimate higher derivatives of u at x_0. Suppose by induction that

$$|u^{(k)}(x_0)| \le \left(\frac{C}{R}\right)^k k^k (\|u\|_{L^\infty(Q_0)} + 1). \tag{2-14}$$

By (2-13), (2-14), and observing that for any $x \in Q_{R/(k+1)}(x_0)$, $Q_{kR/(k+1)}(x) \subset Q_R(x_0)$, we have

$$\begin{aligned}
|u^{(k+1)}(x_0)| &= |(u^{(k)})'(x_0)| \\
&\le \frac{C}{\frac{R}{k+1}} \left(\|u^{(k)}\|_{L^\infty(Q_{R/(k+1)}(x_0))} + \frac{R}{k+1} \|f^{(k+1)}\|_{L^\infty(Q_{R/(k+1)}(x_0))} \right) \\
&\le \frac{C(k+1)}{R} \left\{ \left(\frac{C}{\frac{k}{k+1}R} \right)^k k^k (\|u\|_{L^\infty(Q_0)} + 1) + \frac{R}{k+1} \|f^{(k+1)}\|_{L^\infty(Q_0)} \right\} \\
&\le \left(\frac{C}{R}\right)^{k+1} (k+1)^{k+1} (\|u\|_{L^\infty(Q_0)} + 1).
\end{aligned}$$

In the last inequality we used the analyticity of f in x_n. □

Theorem 2.4. *Let $u \in W^{2,n}(\Omega)$ be a strong solution to* (2-1). *Assume that the coefficients a_{ij} are measurable and independent of x_n and satisfy* (1-2). *Assume that f is analytic in x_n. Then the solution u is analytic in x_n.*

Proof. For any given point $x_0 = (x_0', x_{0,n})$ in Ω, let $r_0 = \frac{1}{4}\operatorname{dist}(x_0, \partial\Omega)$. Consider the Taylor expansion of u in $Q_{r_0}(x_0)$

$$u(x_0', x_n) = \sum_{k=0}^{n} \frac{u^{(k)}(x_0)}{k!}(x_n - x_{0,n})^k + \frac{u^{(n+1)}(x_0', \xi)}{(n+1)!}(x_n - x_{0,n})^{n+1}, \tag{2-15}$$

where $\xi = tx_{0,n} + (1-t)x_n$ for some $t \in (0,1)$. By Theorem 2.3, we know that

$$\begin{aligned}
|u^{(k)}(x_0)| &\le \left(\frac{Ck}{r_0}\right)^k M, \\
|u^{(k+1)}(x_0', \xi)| &\le \left(\frac{C(k+1)}{r_0}\right)^{k+1} M,
\end{aligned}$$

where $M := \|u\|_{L^\infty(Q_{2r_0}(x_0))} + 1$. By Stirling's formula we have

$$(k+1)^{(k+1)} < e^{k+1}(k+1)!.$$

Hence when $|x - x_0| \le r_0/2Ce$ we have

$$\frac{|u^{(k)}(x_0)|}{k!}|x_n - x_{0,n}|^k \le \frac{M}{2^k} \to 0 \quad \text{as } k \to \infty.$$

Hence u is analytic in the x_n direction. □

3. Proof of Theorem 1.2

In this section we prove the analyticity of solutions in x_n to the equation

$$L[u] := \sum_{i,j=1}^{n} a_{ij}(x)u_{ij} = f(x) \quad \text{in } \Omega, \tag{3-1}$$

where the coefficients a_{ij} also depend on x_n. We assume that the a_{ij} are in $C^0(\Omega)$ and satisfy (1-2) and $f \in L^p(\Omega)$ $(p \geq n)$. We also assume that a_{ij} and f are analytic in x_n and satisfy

$$|\partial_{x_n}^k a_{ij}| + |\partial_{x_n}^k f| \leq B^k k! \tag{3-2}$$

for all $k \geq 1$, where $B > 0$ is a constant.

As before, we set $u' = u_{x_n}$, $u'' = u_{x_n x_n}$ and $u^{(k)} = \dfrac{\partial^k u}{\partial x_n^k}$ for all integer $k \geq 1$. In this section we also set

$$\begin{aligned} [u]_{W^{2,p}(\Omega)} &= \sum_{|s|=2} \|D^s u\|_{L^p(\Omega)}, \\ [u^{(\ell)}]^{(\beta)}_{W^{2,p}(\Omega)} &= \sup\nolimits_{Q_r(x)\subset\Omega} d_{Q_r(x)}^{\ell+2-n/p+\beta} [u^{(\ell)}]_{W^{2,p}(Q_r(x))}, \\ \|u^{(\ell)}\|^{(\beta)}_{L^p(\Omega)} &= \sup\nolimits_{Q_r(x)\subset\Omega} d_{Q_r(x)}^{\ell-n/p+\beta} \|u^{(\ell)}\|_{L^p(Q_r(x))}, \end{aligned} \tag{3-3}$$

for $\ell = 0, 1, 2, \ldots$ and $\beta \in \mathbb{R}$, where $d_{Q_r(x)} = \operatorname{dist}(Q_r(x), \partial\Omega)$.

By the $W^{2,p}$ estimate, we have:

Lemma 3.1. *Let $u \in W^{2,n}(\Omega)$ be a strong solution to* (3-1). *Assume that the a_{ij} are in $C^0(\Omega)$ and satisfy* (1-2), *$f \in L^p(\Omega)$ $(p \geq 1)$, and $Q_R(x_0) \subset \Omega$. There exists a constant C such that, if $0 < r < r + \delta < R$, then*

$$\|u\|_{W^{2,p}(Q_r(x_0))} \leq C\left\{\frac{1}{\delta^2}\|u\|_{L^p(Q_{r+\delta}(x_0))} + \|f\|_{L^p(Q_{r+\delta}(x_0))}\right\}, \tag{3-4}$$

where C depends only on n, p, λ, Λ and the moduli of the continuity of the coefficients a_{ij}.

Proof. When $r \leq \delta$, by the $W^{2,p}$ estimate for elliptic equations [Gilbarg and Trudinger 1998] and a rescaling argument, we have

$$\begin{aligned} \|D^2 u\|_{L^p(Q_r(x_0))} &\leq \frac{C}{\delta^2}(\|u\|_{L^p(Q_{r+\delta}(x_0))} + (r+\delta)^2\|f\|_{L^p(Q_{r+\delta}(x_0))}) \\ &\leq C\left(\frac{1}{\delta^2}\|u\|_{L^p(Q_{r+\delta}(x_0))} + \|f\|_{L^p(Q_{r+\delta}(x_0))}\right). \end{aligned} \tag{3-5}$$

When $\delta < r$, we choose $m \geq 2$ such that $r/m \leq \delta < r/(m-1)$, and equally divide the cube $Q_r(x_0)$ into smaller cubes with side-length r/m. Then

$$\|D^2u\|^p_{L^p(Q_r(x_0))} = \sum_i \|D^2u\|^p_{L^p(Q_{r/m}(x_i))}.$$

By (3-5),

$$\|D^2u\|^p_{L^p(Q_{r/m}(x_i))} \leq C\left(\frac{1}{\delta^{2p}}\|u\|^p_{L^p(Q_{2r/m}(x_i))} + \|f\|^p_{L^p(Q_{2r/m}(x_i))}\right). \tag{3-6}$$

Note that for each $Q_{r/m}(x_i)$ there are at most 3^n cubes of the form $Q_{2r/m}(x_j)$ intersecting with it. Hence, summing up, we obtain

$$\|D^2u\|^p_{L^p(Q_r(x_0))} \leq C\left(\frac{1}{\delta^{2p}}\|u\|^p_{L^p(Q_{r+\delta}(x_0))} + \|f\|^p_{L^p(Q_{r+\delta}(x_0))}\right). \tag{3-7}$$

We obtain (3-4). □

We remark that in Lemma 3.1 the assumption $u \in W^{2,n}(\Omega)$ implies that $f \in L^n(\Omega)$. But the inequality (3-4) holds for all $p \geq 1$.

Theorem 3.2. *Let $u \in W^{2,n}(\Omega)$ be a solution to* (3-1). *Assume that the a_{ij} are in $C^0(\Omega)$ and satisfy* (1-2). *Assume also that the a_{ij} and f are analytic in x_n and satisfy* (3-2). *Then u is analytic in x_n.*

Proof. By (3-1), we have

$$\sum a_{ij}(x+\delta e_n)[u'_\delta]_{ij} = -\sum [a_{ij}]'_\delta u_{ij} + f'_\delta, \tag{3-8}$$

where $u'_\delta = (1/\delta)[u(x+\delta e_n) - u(x)]$, $[a_{ij}]'_\delta = (1/\delta)[a_{ij}(x+\delta e_n) - a_{ij}(x)]$, and $e_n = (0, \dots, 0, 1)$ is the unit vector on the x_n-axis. Since the a_{ij} are continuous, by the $W^{2,p}$ estimate, we see that $u'_\delta \in W^{2,p}(\Omega')$ $(p = n)$ for any $\Omega' \subset \Omega$. Sending $\delta \to 0$, we obtain that $u' \in W^{2,p}_{\mathrm{loc}}(\Omega)$ and is a solution to $L[u'] = f' - a'_{ij}u_{ij}$. Similarly $u^{(k)} \in W^{2,p}_{\mathrm{loc}}(\Omega)$ and is a solution to

$$L[u^{(k)}] = f^{(k)} - \sum_{\ell=1}^{k} \binom{\ell}{k} a^{(\ell)}_{ij} u^{(k-\ell)}_{ij} := f^{(k)} - \phi \quad \text{in } \Omega, \tag{3-9}$$

where $\binom{\ell}{k} = k!/(\ell!(k-\ell)!)$.

We will prove Theorem 3.2 by induction. There is no loss of generality in assuming that $\Omega = Q_0$ is the cube of side-length two centered at the origin. By the definition of $[u]^{(n/p)}_{W^{2,p}(Q_0)}$, there exists a cube $Q_{r_0}(x_0) \subset Q_0$ such that

$$[u]^{(n/p)}_{W^{2,p}(Q_0)} \leq 2d_0^2\, [u]_{W^{2,p}(Q_{r_0}(x_0))},$$

where $d_0 = \mathrm{dist}(Q_{r_0}(x_0), \partial Q_0)$. We may assume that the center of Q_{r_0} is the origin; otherwise we may replace $Q_{r_0}(x_0)$ by the larger cube $Q_{1-d_0}(0)$. Therefore the last inequality becomes

$$[u]^{(n/p)}_{W^{2,p}(Q_0)} \le 2(1-r_0)^2 [u]_{W^{2,p}(Q_{r_0})}, \tag{3-10}$$

where Q_{r_0} is centered at the origin. Thanks to Lemma 3.1, there is a constant C independent of r_0 such that

$$\begin{aligned}[u]_{W^{2,p}(Q_{r_0})} &\le C\{4(1-r_0)^{-2}\|u\|_{L^p(Q'_{r_0})} + \|f\|_{L^p(Q'_{r_0})}\} \\ &\le C\{4(1-r_0)^{-2}\|u\|_{L^p(Q_0)} + \|f\|_{L^p(Q_0)}\},\end{aligned}$$

where $Q'_{r_0} = Q_{r_0+(1-r_0)/2} \subset Q_0$. Hence we obtain

$$[u]^{(n/p)}_{W^{2,p}(Q_0)} \le C(\|u\|_{L^p(Q_0)} + \|f\|_{L^p(Q_0)}). \tag{3-11}$$

Next we consider the $W^{2,p}$ estimate for u'. Similarly to (3-10), there exists a cube Q_{r_1}, centered at the origin, such that

$$[u']^{(n/p)}_{W^{2,p}(Q_0)} \le 2(1-r_1)^3 \, [u']_{W^{2,p}(Q_{r_1})}.$$

By (3-9) and Lemma 3.1,

$$[u']_{W^{2,p}(Q_{r_1})} \le C\left\{\frac{9}{(1-r_1)^2}\|u'\|_{L^p(Q'_{r_1})} + \|f'\|_{L^p(Q'_{r_1})} + \sum_{i,j=1}^{n} \|a'_{ij}u_{ij}\|_{L^p(Q'_{r_1})}\right\},$$

where $Q'_{r_1} = Q_{r_1+(1-r_1)/3}$ is a cube centered at the origin. By the interpolation inequality, the right-hand side of the above formula is

$$\begin{aligned}&\le C\big\{(1-r_1)^{-3}\|u\|_{L^p(Q'_{r_1})} + (1-r_1)^{-1}\|D^2u\|_{L^p(Q'_{r_1})} + \|f'\|_{L^p(Q'_{r_1})} + B\|D^2u\|_{L^p(Q'_{r_1})}\big\} \\ &\le CB(1-r_1)^{-3}\{\|u\|_{L^p(Q_0)} + \|f'\|_{L^p(Q_0)} + [u]^{(n/p)}_{W^{2,p}(Q_0)}\}.\end{aligned}$$

Therefore we obtain

$$[u']^{(n/p)}_{W^{2,p}(Q_0)} \le CB(\|u\|_{L^p(Q_0)} + \|f\|_{L^p(Q_0)} + 1),$$

where the number 1 arises in $\|f'\|_{L^p(Q_0)}$.

By induction, let us assume for $\ell = 0, 1, 2, \dots, k$ that

$$[u^{(\ell)}]^{(n/p)}_{W^{2,p}(Q_0)} \le A^\ell \ell!(\|u\|_{L^p(Q_0)} + \|f\|_{L^p(Q_0)} + 1). \tag{3-12}$$

Then, similarly to (3-10), there exists a cube $Q_{r_{k+1}} \subset Q_0$, centered at the origin, such that

$$[u^{(k+1)}]^{(n/p)}_{W^{2,p}(Q_0)} \le 2(1-r_{k+1})^{k+3}[u^{(k+1)}]_{W^{2,p}(Q_{r_{k+1}})}, \tag{3-13}$$

where $Q_{r_{k+1}}$ is a cube with center at the origin. By Lemma 3.1, with $\delta = \dfrac{1-r_{k+1}}{k+3}$,

$$\begin{aligned}(1-r_{k+1})^{k+3}&[u^{(k+1)}]_{W^{2,p}(Q_{r_{k+1}})}\\ &\le C(1-r_{k+1})^{k+3}\bigg\{\frac{(k+3)^2}{(1-r_{k+1})^2}\|u^{(k+1)}\|_{L^p(Q'_{r_{k+1}})}+\|f^{(k+1)}\|_{L^\infty(Q'_{r_{k+1}})}\\ &\qquad\qquad + \sum_{i,j=1}^{n}\sum_{m=0}^{k}\binom{m}{k+1}\|a_{ij}^{(k+1-m)}u_{ij}^{(m)}\|_{L^p(Q'_{r_{k+1}})}\bigg\},\end{aligned}$$

where $Q'_{r_{k+1}} := Q_{r_{k+1}+(1-r_{k+1})/(k+3)}$. Note that $\operatorname{dist}(Q'_{r_{k+1}},\partial Q_0) = \dfrac{k+2}{k+3}(1-r_{k+1})$. We have

$$\begin{aligned}(k+3)^2&(1-r_{k+1})^{k+1}\|u^{(k+1)}\|_{L^p(Q'_{r_{k+1}})}\\ &\le (k+3)^2\Big(\frac{k+3}{k+2}\Big)^{k+1}\Big(\frac{k+2}{k+3}(1-r_{k+1})\Big)^{k+1}[u^{(k-1)}]_{W^{2,p}(Q'_{r_{k+1}})}\\ &\le 4(k+3)^2[u^{(k-1)}]^{(n/p)}_{W^{2,p}(Q_0)}.\end{aligned}$$

Similarly,

$$\begin{aligned}(1-r_{k+1})^{k+3}&\|a_{ij}^{(k+1-m)}u_{ij}^{(m)}\|_{L^p(Q'_{r_{k+1}})}\\ &\le \|a_{ij}^{(k+1-m)}\|_{L^\infty(Q_0)}(1-r_{k+1})^{m+2}[u^{(m)}]_{W^{2,p}(Q'_{r_{k+1}})}\\ &\le 4\|a_{ij}^{(k+1-m)}\|_{L^\infty(Q_0)}[u^{(m)}]^{(n/p)}_{W^{2,p}(Q_0)}.\end{aligned}$$

Hence for fixed i, j, by the induction assumptions,

$$\begin{aligned}(1-r_{k+1})^{k+3}&\sum_{m=0}^{k}\binom{m}{k+1}\|a_{ij}^{(k+1-m)}u_{ij}^{(m)}\|_{L^p(Q'_{r_{k+1}})}\\ &\le 4\sum_{m=0}^{k}\binom{m}{k+1}\|a_{ij}^{(k+1-m)}\|_{L^\infty(Q_0)}[u^{(m)}]^{(n/p)}_{W^{2,p}(Q_0)}\\ &\le 4(k+1)!A^mB^{k+1-m}(\|u\|_{L^p(Q_0)}+\|f\|_{L^p(Q_0)}+1)\\ &\le 4(k+1)!A^kB(\|u\|_{L^p(Q_0)}+\|f\|_{L^p(Q_0)}+1).\end{aligned}$$

Hence by (3-13) we obtain

$$\begin{aligned}[u^{(k+1)}]^{(n/p)}_{W^{2,p}(Q_0)} \le C\big\{&(k+3)^2[u^{(k-1)}]^{(n/p)}_{W^{2,p}(Q_0)}+\|f^{(k+1)}\|_{L^\infty(Q_0)}\\ &+(k+1)!A^kB(\|u\|_{L^p(Q_0)}+\|f\|_{L^p(Q_0)}+1)\big\}.\end{aligned}$$

By (3-2) and the induction assumption (3-12), we then obtain

$$[u^{(k+1)}]^{(n/p)}_{W^{2,p}(Q_0)} \le C(k+1)!(A^{k-1}+A^kB+B^{k+1})(\|u\|_{L^p(Q_0)}+\|f\|_{L^p(Q_0)}+1).$$

Choosing $A \gg B$, we obtain (3-12) for $k+1$.

From (3-12), we obtain that

$$[u^{(k+1)}]_{W^{2,p}(Q_{1/2}(0))} \le 2^{k+1} A^{k+1} (k+1)!(\|u\|_{L^p(Q_0)} + \|f\|_{L^p(Q_0)} + 1).$$

By the Sobolev embedding and since $p > n$, we have

$$|u^{(k+1)}(0)| \le C 2^{k+1} A^{k+1} (k+1)!.$$

Hence u is analytic in x_n at the origin. $\square$

As we remarked in Section 1, the continuity assumption on the a_{ij} can be relaxed. The continuity is used for the $W^{2,p}$ estimate; it suffices to assume that the a_{ij} are continuous in any $n-1$ variables [Kim and Krylov 2007]. In particular, in the dimension-two case, we can remove the continuity of a_{ij} in Theorem 1.2, as the analyticity of a_{ij} automatically implies that they are continuous in one variable. Therefore, for the equation

$$\sum_{i,j=1}^{2} a_{ij}(x) u_{ij} = f(x) \quad \text{in } \Omega, \tag{3-14}$$

where the coefficients a_{ij} satisfy the uniformly elliptic condition (1-2), we have:

Theorem 3.3. *Let $u \in W^{2,2}(\Omega)$ be a strong solution to* (3-14). *Assume that the a_{ij} satisfy* (1-2) *and assume that a_{ij} and f are analytic in x_2. Then under the above conditions*, *u is analytic in x_2.*

4. Proof of Theorem 1.4

Let Ω be a bounded domain in $\mathbb{R}^n$. Let ξ be a unit vector in $\mathbb{R}^n$ and ϕ a function defined in Ω. Set

$$\omega_{\phi,\xi}(r) = \sup\{|\phi(x) - \phi(x + t\xi)| \mid x, x + t\xi \in \Omega,\ |t| \le r\}.$$

We say ϕ is Hölder-continuous in the ξ direction with Hölder exponent α if $\omega_{\phi,\xi} \in C^\alpha$, and write $\phi \in C^\alpha_\xi(\Omega)$, with the norm

$$\|\phi\|_{C^\alpha_\xi(\Omega)} = \sup_{x\in\Omega} |\phi(x)| + \sup_{t>0} \frac{\omega_{\phi,\xi}(t)}{t^\alpha}.$$

To prove Theorem 1.4, we assume for simplicity that $b_i = c = 0$ and consider the equation

$$L[u] := \sum_{i,j=1}^{n} a_{ij}(x) u_{ij} = f(x) \quad \text{in } \Omega, \tag{4-1}$$

where the coefficients a_{ij} satisfies the uniformly elliptic condition (1-2). The proof below is based on a perturbation argument and follows closely that of [Wang 2006].

Proof of Theorem 1.4. Without loss of generality we assume $\xi = e_1 = (1, 0, \dots, 0)$ and $\Omega = B_1(0)$, the unit ball. We set

$$B_k = B_{2^{-k}}(0), \quad \hat{a}_{ij}(x) = a_{ij}(0, x_2, \dots, x_n), \quad \hat{f}(x) = f(0, x_2, \dots, x_n).$$

For $k = 0, 1, 2, \dots$, let u_k be the solution of

$$\begin{aligned} \sum_{i,j=1}^{n} \hat{a}_{ij}(x)(u_k)_{x_i x_j} &= \hat{f}(x) \quad \text{in } B_k, \\ u_k &= u \qquad \text{on } \partial B_k. \end{aligned} \tag{4-2}$$

Then

$$\begin{aligned} \sum_{i,j=1}^{n} \hat{a}_{ij}(x)(u_k - u)_{x_i x_j} &= \sum_{i,j=1}^{n} (a_{ij}(x) - \hat{a}_{ij}(x)) u_{x_i x_j} + \hat{f}(x) - f(x) \quad \text{in } B_k, \\ u_k - u &= 0 \qquad \text{on } \partial B_k. \end{aligned} \tag{4-3}$$

Hence, by the Alexandrov maximum principle, for $k \geq 1$,

$$\begin{aligned} &\sup_{B_k} |u - u_k| \\ &\quad \leq C 2^{-k} \left[\int_{B_k} |(a_{ij}(x) - \hat{a}_{ij}(x)) u_{x_i x_j}|^n \, dx \right]^{1/n} + C 2^{-2k} \omega_{f,\xi}(2^{-k}) \\ &\quad \leq C 2^{-k} \|a_{ij}\|_{C^\alpha_\xi(B_k)} \left[\int_{B_k} |x|^{n\alpha} |u_{x_i x_j}|^n \, dx \right]^{1/n} + C 2^{-2k} \omega_{f,\xi}(2^{-k}) \\ &\quad \leq C 2^{-k} \|a_{ij}\|_{C^\alpha_\xi(B_k)} \left[\left(\int_{B_k} |x|^{n\alpha p/(p-n)} \, dx \right)^{(p-n)/p} \left(\int_{B_k} |u_{x_i x_j}|^p \, dx \right)^{n/p} \right]^{1/n} \\ &\qquad\qquad + C 2^{-2k} \omega_{f,\xi}(2^{-k}) \\ &\quad \leq C 2^{-k} \|a_{ij}\|_{C^\alpha_\xi(B_k)} (2^{-k})^{\alpha + 1 - n/p} \|u\|_{W^{2,p}(B_k)} + C 2^{-2k} \omega_{f,\xi}(2^{-k}) \\ &\quad \leq C (A \cdot (2^{-k})^{2 + \alpha - n/p} + 2^{-2k} \omega_{f,\xi}(2^{-k})), \end{aligned} \tag{4-4}$$

where

$$A = \|u\|_{W^{2,p}(B_1)} \|a_{ij}\|_{C^\alpha_\xi(\Omega)}.$$

Since the a_{ij} are continuous and satisfy the uniformly elliptic condition, by the $W^{2,p}$ estimate,

$$A \leq C (\|u\|_{L^p(\Omega)} + \|f\|_{L^p(\Omega)}) \|a_{ij}\|_{C^\alpha_\xi(\Omega)}.$$

Hence

$$\begin{aligned} \|u_k - u_{k+1}\|_{L^\infty(B_{k+1})} &\leq C \{ A \cdot (2^{-k})^{2 + \alpha - n/p} + 2^{-2k} \omega_{f,\xi}(2^{-k}) \} \\ &= C 2^{-2k} \{ A \cdot (2^{-k})^{\alpha - n/p} + \omega_{f,\xi}(2^{-k}) \}. \end{aligned} \tag{4-5}$$

Since $w_k := u_{k+1} - u_k$ satisfies

$$\hat{a}_{ij}(x) w_{x_i x_j} = 0$$

in B_{k+1}, where the coefficients $\hat{a}_{ij}(x)$ are independent of x_1, by differentiating the equation and by the $W^{2,p}$ estimate, we have

$$\|\partial_\xi w_k\|_{W^{2,p}(B_{k+2})} \le C 2^{3k} \|w_k\|_{L^\infty(B_{k+1})} \quad \text{for all } p > 1.$$

Hence by the Sobolev embedding theorem,

$$\|\partial_\xi w_k\|_{C^{1,\beta}(B_{k+2})} \le C 2^{2k+2\beta} \|w_k\|_{L^\infty(B_{k+1})} \quad \text{for all } \beta \in (0, 1).$$

Therefore by rescaling,

$$\begin{aligned} &\|\partial_\xi \partial_x w_k\|_{L^\infty(B_{k+2})} \le C[A \cdot (2^{-k})^{\alpha - n/p} + \omega_{f,\xi}(2^{-k})], \\ &\|\partial_\xi \partial_x w\|_{C^\beta(B_{k+2})} \le C 2^{k\beta} [A \cdot (2^{-k})^{\alpha - n/p} + \omega_{f,\xi}(2^{-k})]. \end{aligned} \tag{4-6}$$

As the coefficients a_{ij} are continuous, the solution can be approximated by smooth solutions. Hence, to prove Theorem 1.4, we may assume that u is smooth, so that

$$D^2 u_k(0) \to D^2 u(0).$$

For y near 0, let $m \ge 1$ be such that

$$2^{-m-4} \le |y| < 2^{-m-3}.$$

Then

$$\begin{aligned} |\partial_\xi \partial_x u(y) - \partial_\xi \partial_x u(0)| \le |\partial_\xi \partial_x u_m(y) - \partial_\xi \partial_x u_m(0)| + |\partial_\xi \partial_x u_m(0) - \partial_\xi \partial_x u(0)| \\ + |\partial_\xi \partial_x u(y) - \partial_\xi \partial_x u_m(y)|. \end{aligned} \tag{4-7}$$

We have

$$\begin{aligned} |\partial_\xi \partial_x u_m(0) - \partial_\xi \partial_x u(0)| &\le \sum_{k=m}^{\infty} |\partial_\xi \partial_x u_k(0) - \partial_\xi \partial_x u_{k+1}(0)| \\ &\le C \sum_{k=m}^{\infty} [A \cdot (2^{-k})^{\alpha - n/p} + \omega_{f,\xi}(2^{-k})] \\ &\le C \left\{ A \cdot (2^{-m})^{\alpha - n/p} + \int_0^{|y|} \frac{\omega_{f,\xi}(r)}{r} \right\} \\ &\le C \left\{ A \cdot |y|^{\alpha - n/p} + \int_0^{|y|} \frac{\omega_{f,\xi}(r)}{r} \right\}. \end{aligned} \tag{4-8}$$

Similarly,

$$|\partial_\xi \partial_x u(y) - \partial_\xi \partial_x u_m(y)| \le C \left\{ A \cdot |y|^{\alpha - n/p} + \int_0^{|y|} \frac{\omega_{f,\xi}(r)}{r} \right\}.$$

By (4-6) we have

$$
\begin{aligned}
(4\text{-}9) \qquad |\partial_\xi \partial_x w_k(y) - \partial_\xi \partial_x w_k(0)| &\le \|\partial_\xi \partial_x w_k\|_{C^\beta(B_{k+2})} |y|^\beta \\
&\le C|y|^\beta 2^{k\beta} [A \cdot (2^{-k})^{\alpha - n/p} + \omega_{f,\xi}(2^{-k})].
\end{aligned}
$$

Write

$$u_m = u_1 + \sum_{k=1}^{m-1} w_k.$$

We have, for $\beta < \alpha - n/p$,

$$
\begin{aligned}
(4\text{-}10) \quad & |\partial_\xi \partial_x u_m(y) - \partial_\xi \partial_x u_m(0)| \\
&\le |\partial_\xi \partial_x u_1(y) - \partial_\xi \partial_x u_1(0)| + \sum_{k=1}^{m-1} |\partial_\xi \partial_x w_k(y) - \partial_\xi \partial_x w_k(0)| \\
&\le C|y|^\beta \biggl(\|u_1\|_{L^\infty(\Omega)} + \sum_{k=1}^{m-1} 2^{k\beta} \bigl(A \cdot (2^{-k})^{\alpha - n/p} + \omega_{f,\xi}(2^{-k}) \bigr) \biggr) \\
&\le C|y|^\beta \biggl(\|u\|_{L^\infty(\Omega)} + \|f\|_{L^p(\Omega)} + \int_{|y|}^1 \frac{\omega_{f,\xi}(r)}{r^{1+\beta}} \biggr).
\end{aligned}
$$

This completes the proof of Theorem 1.4. □

5. Equation of divergence form

We consider the following linear elliptic equation of divergence form:

$$Lu = \operatorname{div}(A(x)\nabla u(x)) = \operatorname{div} f(x) \quad \text{in } \Omega, \tag{5-1}$$

where the coefficient matrix $A(x) = (a_{ij}(x))_{n\times n}$ satisfies the uniformly elliptic condition (1-2) and $f(x) = (f_1(x), f_2(x), \dots, f_n(x)) \in [L^p(\Omega)]^n$ for $p > 1$. We assume also that a_{ij} and f are analytic in x_n, and that there exists a constant $B > 0$ such that

$$|\partial_{x_n}^k a_{ij}| + |\partial_{x_n}^k f| \le B^k k! \tag{5-2}$$

for all $k \ge 1$.

Definition 5.1. Let $1 < p < \infty$. We say that u is a solution to (5-1) if $u \in W^{1,p}_{\text{loc}}(\Omega)$ and satisfies

$$\int_\Omega a_{ij}(x) u_{x_j} \phi_{x_i}\, dx = \int_\Omega f(x) \phi_{x_i}\, dx$$

for all $\phi \in C_0^\infty(\Omega)$.

As before, we set $u' = u_{x_n}$, $u'' = u_{x_n x_n}$ and $u^{(k)} = \dfrac{\partial^k u}{\partial x_n^k}$ for all integers $k \geq 1$. We also define

$$
\begin{aligned}
[u]_{W^{1,p}(\Omega)} &= \|Du\|_{L^p(\Omega)},\\
\|u^{(k)}\|^{(\beta)}_{W^{1,p}(\Omega)} &= \sup\nolimits_{Q_r(x)\subset\Omega} d_{Q_r(x)}^{k+1-n/p+\beta}[u^{(k)}]_{W^{1,p}(Q_r(x))},\\
\|u^{(k)}\|^{(\beta)}_{L^p(\Omega)} &= \sup\nolimits_{Q_r(x)\subset\Omega} d_{Q_r(x)}^{k-n/p+\beta}\|u^{(k)}\|_{L^p(Q_r(x))},
\end{aligned}
\tag{5-3}
$$

where $d_{Q_r(x)} = \mathrm{dist}(Q_r(x), \partial\Omega)$, k is a nonnegative integer, and $p > 1$ is a constant.

By the $W^{1,p}$ estimate for the divergence form (5-1) in [Di Fazio 1996], we have:

Lemma 5.2. *Let u be a solution to* (5-1). *Assume that the a_{ij} satisfy* (1-2), *$f \in [L^p(\Omega)]^n$ ($p > 1$) and $Q_R(x_0) \subset \Omega$. There exists a constant C such that, if $0 < r < r+\delta < R$, then*

$$
[u]_{W^{1,p}(Q_r(x_0))} \leq C\left\{\frac{1}{\delta}\|u\|_{L^p(Q_{r+\delta}(x_0))} + \sum_{i=1}^{n}\|f_i\|_{L^p(Q_{r+\delta}(x_0))}\right\},
\tag{5-4}
$$

where the constant C depends only on n, p, λ, Λ.

By Lemma 5.2 we then have:

Theorem 5.3. *Let u be a solution to* (5-1). *Assume that the a_{ij} satisfy* (1-2) *and $f \in [L^p(\Omega)]^n$ ($p > n$). Assume that the a_{ij} and f are analytic in the variable x_n. Then u is analytic in x_n.*

The proofs of Lemma 5.2 and Theorem 5.3 are similar to those in Section 3 and are omitted here. Note that the assumption $p > n$ in Theorem 5.3 is for the use of Sobolev embedding; namely, by the estimate $\|u^{(k)}\|_{W^{1,p}(Q_r(0))} \leq C$ one infers that $|u^{(k)}(0)| \leq C_1$.

References

[Cao and Titi 2008] C. Cao and E. S. Titi, "Regularity criteria for the three-dimensional Navier–Stokes equations", *Indiana Univ. Math. J.* **57**:6 (2008), 2643–2661. MR 2010b:35332 Zbl 1159.35053

[Cao and Titi 2011] C. Cao and E. S. Titi, "Global regularity criterion for the 3D Navier–Stokes equations involving one entry of the velocity gradient tensor", *Arch. Ration. Mech. Anal.* **202**:3 (2011), 919–932. MR 2854673 Zbl 1256.35051

[Chiarenza et al. 1993] F. Chiarenza, M. Frasca, and P. Longo, "$W^{2,p}$-solvability of the Dirichlet problem for nondivergence elliptic equations with VMO coefficients", *Trans. Amer. Math. Soc.* **336**:2 (1993), 841–853. MR 93f:35232 Zbl 0818.35023

[Di Fazio 1996] G. Di Fazio, "L^p estimates for divergence form elliptic equations with discontinuous coefficients", *Boll. Un. Mat. Ital. A* (7) **10**:2 (1996), 409–420. MR 97e:35034 Zbl 0865.35048

[Dong 2012] H. J. Dong, "Gradient estimates for parabolic and elliptic systems from linear laminates", *Arch. Ration. Mech. Anal.* **205**:1 (2012), 119–149. MR 2927619 Zbl 1258.35040

[Dong and Kim 2011] H. J. Dong and S. Kim, "Partial Schauder estimates for second-order elliptic and parabolic equations", *Calc. Var. Partial Differential Equations* **40**:3-4 (2011), 481–500. MR 2012d:35025 Zbl 1221.35078

[Gilbarg and Trudinger 1998] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, revised 2nd ed., Grundlehren der Math. Wissenschaften **224**, Springer, Berlin, 1998. MR 2001k:35004 Zbl 1042.35002

[Hörmander 1976] L. Hörmander, "The boundary problems of physical geodesy", *Arch. Ration. Mech. Anal.* **62**:1 (1976), 1–52. MR 58 #29202a Zbl 0331.35020

[Kim and Krylov 2007] D. Kim and N. V. Krylov, "Elliptic differential equations with coefficients measurable with respect to one variable and VMO with respect to the others", *SIAM J. Math. Anal.* **39**:2 (2007), 489–506. MR 2008j:35031 Zbl 1138.35308

[Kukavica and Ziane 2007] I. Kukavica and M. Ziane, "Navier–Stokes equations with regularity in one direction", *J. Math. Phys.* **48**:6 (2007), Article ID #065203. MR 2008g:35155 Zbl 1144.81373

[Taylor 2000] M. E. Taylor, *Tools for PDE: pseudodifferential operators, paradifferential operators, and layer potentials*, Mathematical Surveys and Monographs **81**, American Mathematical Society, Providence, RI, 2000. MR 2001g:35004 Zbl 0963.35211

[Tian and Wang 2010] G. J. Tian and X.-J. Wang, "Partial regularity for elliptic equations", *Discrete Contin. Dyn. Syst.* **28**:3 (2010), 899–913. MR 2011e:35077 Zbl 1193.35025

[Wang 2006] X.-J. Wang, "Schauder estimates for elliptic and parabolic equations", *Chinese Ann. Math. Ser. B* **27**:6 (2006), 637–642. MR 2007f:35106 Zbl 1151.35329

Received July 14, 2014. Revised October 27, 2014.

YONGYANG JIN
DEPARTMENT OF APPLIED MATHEMATICS
ZHEJIANG UNIVERSITY OF TECHNOLOGY
HANGZHOU, 310023
CHINA
yongyang@zjut.edu.cn

DONGSHENG LI
SCHOOL OF MATHEMATICS AND STATISTICS
XI'AN JIAOTONG UNIVERSITY
XI'AN, 710049
CHINA
lidsh@xjtu.edu.cn

XU-JIA WANG
CENTRE FOR MATHEMATICS AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 0200
AUSTRALIA
xu-jia.wang@anu.edu.au

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.437

ON THE DENSITY THEOREM FOR THE SUBDIFFERENTIAL OF CONVEX FUNCTIONS ON HADAMARD SPACES

MINA MOVAHEDI, DARYOUSH BEHMARDI
AND SEYEDEHSOMAYEH HOSSEINI

We introduce a dual space for any geodesically complete Hadamard space. By using this notion we give a new definition of the subdifferential of convex functions on geodesically complete Hadamard spaces. Some properties of this subdifferential, such as a density theorem, are proved.

1. Introduction

Nondifferentiability appears naturally in different areas of mathematics and arises explicitly in the description of various modern technological systems. Nonsmooth analysis studies the local behavior of nondifferentiable functions and sets lacking smooth boundaries. Generalized gradients or subdifferentials refer to several set-valued replacements for the usual derivative which are used in developing differential calculus for nonsmooth functions.

Nondifferentiable functions are often considered on finite-dimensional or infinite-dimensional Banach spaces. Here, the linear structure plays a central role. Attempts have been made to replace Banach spaces with Riemannian manifolds and develop a subdifferential calculus; see [Hosseini and Pouryayevali 2011; 2013a; 2013b; 2013c]. Shafrir [1992] gave a definition of the coaccretive subdifferential of a convex function defined on a Hilbert ball. His approach involves the structure of (B, ρ) as a Hilbert manifold, where ρ is the hyperbolic metric on B; see also [Kopecká and Reich 2010, p. 188].

Unlike Riemannian manifolds, Hadamard spaces are not equipped with a Riemannian metric. Hence, we need new tools to construct a suitable dual space in order to define subdifferentials of functions on Hadamard spaces. B. Ahmadi Kakavandi and M. Amini [2010] defined a dual space for an Hadamard space using the concept of bound vectors. They defined a pseudometric D on $\mathbb{R} \times \mathcal{X} \times \mathcal{X}$, where $\mathcal{X}$ is an Hadamard space, and considered the pseudometric space $(\mathbb{R} \times \mathcal{X} \times \mathcal{X}, D)$ as a subspace of the pseudometric space $(\mathrm{Lip}(\mathcal{X}, \mathbb{R}), L)$ of all real-valued Lipschitz

MSC2010: primary 53C23; secondary 47H05.
Keywords: Hadamard spaces, convex functions, subdifferential.

functions. Then, they defined an equivalence relation on $\mathbb{R} \times \mathcal{X} \times \mathcal{X}$, where the equivalence class of (t, a, b) is

$$[t\,\overrightarrow{ab}] := \left\{ s\,\overrightarrow{cd} : t\langle \overrightarrow{ab}, \overrightarrow{xy} \rangle = s\langle \overrightarrow{cd}, \overrightarrow{xy} \rangle \text{ for } x, y \in \mathcal{X} \right\}.$$

After introducing a dual metric space to $\mathcal{X}$,

$$\mathcal{X}^* := \left\{ [t\,\overrightarrow{ab}] : (t, a, b) \in \mathbb{R} \times \mathcal{X} \times \mathcal{X} \right\},$$

they defined a notion of the subdifferential for a proper function on an Hadamard space.

Here we present a new dual for any Hadamard space and prove a density theorem for the subdifferential of lower semicontinuous convex functions on Hadamard spaces, generalizing the classical one for Hilbert spaces [Clarke et al. 1998]. Our approach differs from the one in [Ahmadi Kakavandi and Amini 2010]: we use the notion of geodesics, defining the dual $\mathcal{X}^*$ as the disjoint union of the sets $\mathcal{X}^*_x$ over $x \in \mathcal{X}$, where $\mathcal{X}^*_x$ contains all unit speed geodesics of $\mathcal{X}$ starting at x. The subdifferential of a function f at a point x is defined as a subset of $\mathcal{X}^*_x$. This property is not visible in Ahmadi Kakavandi and Amini's definition of the subdifferential. This leads us to the claim that the subdifferential of convex functions defined in this paper is an analogue of the concept of the subdifferential of convex functions in Riemannian manifolds and Hilbert balls.

We assume that $\mathcal{X}$ is a geodesically complete Hadamard space with a metric d. Recall that a geodesic in $\mathcal{X}$ is a curve of constant speed which is locally minimizing. We say $\mathcal{X}$ has nonpositive curvature (in the sense of Alexandrov) if every point $p \in \mathcal{X}$ has a neighborhood U with the following properties:

(i) For any two points $x, y \in U$ there is a geodesic $\sigma_x^y : [0, 1] \to U$ from x to y of length $d(x, y)$.

(ii) For any triple of points $x, y, z \in U$, we have

$$d^2(z, m) \leq \tfrac{1}{2}(d^2(z, x) + d^2(z, y)) - \tfrac{1}{4} d^2(x, y),$$

where σ_x^y is as in (i) and $m = \sigma_x^y\left(\tfrac{1}{2}\right)$ is the point halfway between x and y.

We say $\mathcal{X}$ is an Hadamard space if $\mathcal{X}$ is complete and the assertions (i) and (ii) above hold for all points $x, y, z \in \mathcal{X}$. Hadamard spaces are uniquely geodesic, i.e., there exists a unique geodesic between any pair of points.

In this paper, we assume that $\mathcal{X}$ is a geodesically complete Hadamard space, meaning that every geodesic in $\mathcal{X}$ is a subarc of a geodesic which is parametrized on the whole real line. Let $\mathbb{E}^2$ be the Euclidean space equipped with the metric

$$d_{\mathbb{E}^2}((x_1, x_2), (y_1, y_2)) = ((x_1 - y_1)^2 + (x_2 - y_2)^2)^{1/2}.$$

A geodesic triangle $\triangle(x, y, z)$ in $\mathcal{X}$ is the union of three points $x, y, z \in \mathcal{X}$ and the geodesic segments joining them. The comparison triangle for $\triangle(x, y, z)$, is a triangle $\triangle(\bar{x}, \bar{y}, \bar{z})$ in $\mathbb{E}^2$ such that $d(x, y) = d_{\mathbb{E}^2}(\bar{x}, \bar{y})$, $d(x, z) = d_{\mathbb{E}^2}(\bar{x}, \bar{z})$ and $d(z, y) = d_{\mathbb{E}^2}(\bar{z}, \bar{y})$. According to this notation: if a is a point on the geodesic segment joining x, y, then $\bar{a}$ is its comparison point provided that $d(x, a) = d_{\mathbb{E}^2}(\bar{x}, \bar{a})$. Also, the comparison angle $\angle_{\bar{x}}(\bar{y}, \bar{z})$ is the interior angle of the comparison triangle $\triangle(\bar{x}, \bar{y}, \bar{z})$ at $\bar{x}$.

The first step in defining a subdifferential for a function defined on an Hadamard space $\mathcal{X}$ is to introduce a dual space $\mathcal{X}^*$ for $\mathcal{X}$. We denote by $\mathcal{X}^*$ the set of all unit speed geodesics of $\mathcal{X}$, i.e., $\mathcal{X}^* = \coprod_{x \in \mathcal{X}} \mathcal{X}_x^*$ where $\mathcal{X}_x^*$ is the set of all unit speed geodesics of $\mathcal{X}$ starting at x. Consider the map $\langle \cdot, \cdot \rangle : \mathcal{X}_x^* \times \mathcal{X}_x^* \to \mathbb{R}$ defined by

$$\langle \gamma_x^y, \gamma_x^z \rangle = \tfrac{1}{2}[d^2(x, z) + d^2(x, y) - d^2(y, z)].$$

It is clear that $(\langle \gamma_x^y, \gamma_x^y \rangle)^{1/2} = d(x, y)$; see [Berg and Nikolaev 2008] for more details. Let $\gamma_x^y \in X_x^*$, $\sigma_z^w \in X_z^*$ and $D := \mathrm{dom}(\sigma_z^w) = \mathrm{dom}(\gamma_x^y)$. Then we say that γ_x^y is parallel to σ_z^w if there exists $C \in \mathbb{R}$ with $d(\sigma_z^w(t), \gamma_x^y(t)) = C$ for all $t \in D$.

2. The subdifferential of a convex function

In this section, we present a new definition of the subdifferential of a convex function on an Hadamard space. Note that the function $f : \mathcal{X} \to \mathbb{R}$ is called convex if, for any geodesic γ, the composition $f \circ \gamma$ is convex (in the usual sense). Let us start with the definition of the directional derivative for functions on geodesically complete Hadamard spaces.

Definition 2.1. Let $f : \mathcal{X} \to \mathbb{R}$ be a real-valued function. The directional derivative $Df(x; \gamma_x^z)$ of f at $x \in \mathcal{X}$ in the direction $\gamma_x^z \in \mathcal{X}_x^*$ for some $z \in \mathcal{X}$ is defined as

$$Df(x; \gamma_x^z) := \lim_{t \downarrow 0} \frac{f(\gamma_x^z(t)) - f(x)}{t}. \tag{2-1}$$

We will use the following remark in the proof of Theorem 2.4.

Remark 2.2. In the case $\mathcal{X} = \mathbb{R}$, the directional derivative of f at x in the direction of γ_x^{x+b} is defined by

$$Df(x; \gamma_x^{x+b}) = \lim_{t \downarrow 0} \frac{f(x+t) - f(x)}{t}$$

for every $b \in (x, \infty)$. This is the same as the usual directional derivative of f at x in the direction 1, denoted by $Df(x; 1)$.

Theorem 2.3. *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function on $\mathcal{X}$ and consider $\gamma_x^z \in \mathcal{X}_x^*$.*

(i) *The function* $Q : \mathrm{dom}(\gamma_x^z) \cap (0, \infty) \to \mathbb{R}$ *defined by*

$$Q(t) = \frac{f(\gamma_x^z(t)) - f(x)}{t}$$

is increasing.

(ii) $Df(x; \gamma_x^z)$ *exists and is equal to* $\inf_t Q(t)$.

(iii) $Df(x; \gamma_x^x) = 0$.

Proof. (i) Since f is convex, the function $g(t) = f(\gamma_x^z(t))$, defined on $\mathrm{dom}(\gamma_x^z)$, is convex. If $0 < t_1 < t_2$, we have

$$\frac{g(t_1) - g(0)}{t_1} \le \frac{g(t_2) - g(0)}{t_2}.$$

This implies that

$$\frac{f(\gamma_x^z(t_1)) - f(x)}{t_1} \le \frac{f(\gamma_x^z(t_2)) - f(x)}{t_2},$$

which means that Q is increasing.

(ii) Assertion (i) implies that for any decreasing sequence of positive numbers $\{t_n\}$ which converges to zero, the sequence $\{Q(t_n)\}$ is increasing. Hence, $\{Q(t_n)\}$ has a limit, namely $Df(x; \gamma_x^z) = \inf_t Q(t)$.

(iii) For every $x \in \mathcal{X}$ and t, we have $\gamma_x^x(t) = x$. Hence

$$Df(x; \gamma_x^x) = \lim_{t \downarrow 0} \frac{f(\gamma_x^x(t)) - f(x)}{t} = 0. \qquad \square$$

Theorem 2.4 (mean value theorem). *Suppose that* $x, y \in \mathcal{X}$, *and that* $f : \mathcal{X} \longrightarrow \mathbb{R}$ *is convex. Then there exists* $t_0 \in (0, d(x, y))$ *such that*

$$\frac{f(y) - f(x)}{d(x, y)} \le Df(\gamma_x^y(t_0); \sigma^y_{\gamma_x^y(t_0)}).$$

Proof. Let γ_x^y be the unit speed geodesic joining x to y. Then, $f \circ \gamma_x^y$ is a real-valued convex function on $[0, d(x, y)]$. By the mean value theorem for convex functions from $\mathbb{R}$ to $\mathbb{R}$, there exist $t_0 \in (0, d(x, y))$ and $z \in \partial f \circ \gamma_x^y(t_0)$ such that

$$\frac{f \circ \gamma_x^y(d(x, y)) - f \circ \gamma_x^y(0)}{d(x, y)} = z,$$

where $\partial f \circ \gamma_x^y(t_0)$ denotes the subdifferential of the real-valued function $f \circ \gamma_x^y$ at t_0. We set $w = \gamma_x^y(t_0)$. For the unit speed geodesic σ_w^y,

$$Df(w; \sigma_w^y) = \lim_{t \downarrow 0} \frac{f \circ \sigma_w^y(t) - f \circ \sigma_w^y(0)}{t} = Df \circ \sigma_w^y(0; 1).$$

Since the geodesic connecting w and y is unique, we have $\sigma_w^y(t) = \gamma_x^y(t_0 + t)$ for every $t \in [0, d(w, y)]$. Hence, $Df \circ \sigma_w^y(0; 1) = Df \circ \gamma_x^y(t_0; 1)$ and $z \le Df \circ \gamma_x^y(t_0; 1)$. Therefore,

$$\frac{f(y) - f(x)}{d(x, y)} \le Df(\gamma_x^y(t_0); \sigma^y_{\gamma_x^y(t_0)}). \qquad \square$$

Definition 2.5. Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. A geodesic $\gamma_x^z \in \mathcal{X}_x^*$ is called the subgradient of f at x if

$$f(y) \ge f(x) + \langle \gamma_x^z, \sigma_x^y \rangle, \quad \forall y \in \mathcal{X}, \quad \forall \sigma_x^y \in \mathcal{X}_x^*.$$

The set-valued map $\partial f : \mathcal{X} \to \mathcal{X}^*$ is called the subdifferential of f and we call $\partial f(x)$ the subdifferential of f at x: it is the set of all subgradients of f at x.

It is worth pointing out that $\partial f(x) \subset \mathcal{X}_x^*$ for every $x \in \mathcal{X}$. A roughly analogous concept of subdifferential is introduced and investigated on the Hilbert ball in [Reich and Shafrir 1990].

Theorem 2.6. *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. Then $\gamma_x^x \in \partial f(x)$ if and only if x is a minimum point of f.*

Proof. We know that $\langle \gamma_x^x, \sigma_x^y \rangle = 0$ for every $x, y \in \mathcal{X}$ and $\sigma_x^y \in \mathcal{X}_x^*$. Hence, if $\gamma_x^x \in \partial f(x)$, then

$$f(y) \ge f(x) + \langle \gamma_x^x, \sigma_x^y \rangle = f(x), \quad \forall y \in \mathcal{X}, \ \forall \sigma_x^y \in \mathcal{X}_x^*,$$

which means that x is a minimum point of f.

Now assume that x is a minimum point of f, so $f(y) \ge f(x)$ for every $y \in \mathcal{X}$. Then

$$f(y) \ge f(x) + \langle \gamma_x^x, \sigma_x^y \rangle = f(x), \quad \forall y \in \mathcal{X}, \ \forall \sigma_x^y \in \mathcal{X}_x^*,$$

and the proof is complete. $\square$

Theorem 2.7. *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. If $Df(x; \sigma_x^y) \ge \langle \gamma_x^z, \sigma_x^y \rangle$ for all $y \in \mathcal{X}$ and $\sigma_x^y \in \mathcal{X}_x^*$, then $\gamma_x^z \in \partial f(x)$.*

Proof. The relations $Df(x; \sigma_x^y) \ge \langle \gamma_x^z, \sigma_x^y \rangle$ and

$$f(y) - f(x) \ge \frac{f(\sigma_x^y(s)) - f(x)}{s} \ge Df(x; \sigma_x^y)$$

imply $f(y) - f(x) \ge \langle \gamma_x^z, \sigma_x^y \rangle$, and hence $\gamma_x^z \in \partial f(x)$. $\square$

Corollary 2.8. *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. Then x is a minimum point of f if and only if $Df(x; \gamma_x^z) \ge 0$ for each $\gamma_x^z \in \mathcal{X}_x^*$.*

Proof. If x is a minimum point, then $f(\gamma_x^z(t)) \ge f(x)$ for each $z \in \mathcal{X}$ and $t \in \mathrm{dom}\gamma_x^z$. Hence, $Df(x; \gamma_x^z) \ge 0$. The converse is obvious by Theorem 2.7. $\square$

Lemma 2.9. *For each triple of points $x, y, z \in \mathcal{X}$, there exists $w \in \mathcal{X}$ such that $d(x, y) = d(z, w)$ and γ_x^y is parallel to σ_z^w.*

Proof. Since $\mathcal{X}$ is geodesically complete, there is a unit speed geodesic ray γ_x connecting x and y. By Proposition 9.2.28 in [Burago et al. 2001], there exists a unique unit speed geodesic ray σ_z starting at z, parallel to γ_x. Define $w \in \mathcal{X}$ by $w = \sigma_z(d(x, y))$. Then: $d(x, y) = d(w, z)$ and γ_x^y is parallel to σ_z^w. Suppose that σ_z^v is another geodesic segment parallel to γ_x^y. Since it is also parallel to σ_z^w and $d(\sigma_z^w(0), \sigma_z^v(0)) = 0$, we have $d(\sigma_z^w(t), \sigma_z^v(t)) = 0$ for each $t \in [0, d(x, y)]$. $\square$

We use the notation $\gamma_x^y \parallel \gamma_z^w$ when γ_x^y is parallel to γ_z^w for $x, y, z, w \in \mathcal{X}$. We also denote by xy the line segment between $x, y \in \mathbb{E}^2$.

Definition 2.10. (i) The function $P_{xy} : \mathcal{X}_x^* \longrightarrow \mathcal{X}_y^*$ defined by $P_{xy}(\gamma_x^w) = \gamma_y^v$ is called the parallel translation of γ_x^w along γ_x^y. Here, v is selected such that $d(x, w) = d(y, v)$ and γ_x^w is parallel to γ_y^v.

(ii) To define the sum of γ_x^a and γ_x^b, we pick a point c such that by $P_{xa}(\gamma_x^b) = \gamma_x^c$ and put $\gamma_x^a + \gamma_x^b := \gamma_x^c$.

(iii) We define

$$-\gamma_x^y := P_{yx}(\gamma_y^x),$$
$$\gamma_x^a - \gamma_x^b := \gamma_x^a + (-\gamma_x^b).$$

Theorem 2.11. *Suppose that $\gamma_x^y = P_{ax}(\gamma_a^b)$ and $\gamma_x^z = P_{ax}(\gamma_a^c)$. Then:*

(i) $d(b, c) = d(y, z)$,

(ii) $\angle_a(b, c) = \angle_x(y, z)$,

(iii) $\langle \gamma_a^b, \gamma_a^c \rangle = \langle \gamma_x^y, \gamma_x^z \rangle$,

(iv) $\langle -\gamma_x^y, \gamma_x^z \rangle = \langle \gamma_x^y, -\gamma_x^z \rangle$.

Proof. Let $\triangle(\bar{a}, \bar{b}, \bar{c})$ and $\triangle(\bar{x}, \bar{y}, \bar{z})$ be the comparison triangles for $\triangle(a, b, c)$ and $\triangle(x, y, z)$ respectively. By definition, $d(\gamma_a^b(t), \gamma_x^y(t)) = d_{\mathbb{E}^2}(\overline{\gamma_a^b(t)}, \overline{\gamma_x^y(t)}) = C$ where C is constant for each t. We can assume that $\bar{a}\bar{b} \parallel \bar{x}\bar{y}$ and $\bar{a}\bar{c} \parallel \bar{x}\bar{z}$.

This means that $\angle_{\bar{a}}(\bar{b}, \bar{c})$ and $\angle_{\bar{x}}(\bar{y}, \bar{z})$ are two angles with parallel sides. They are therefore congruent or supplementary. But since $d_{\mathbb{E}^2}(\overline{\gamma_a^b(t)}, \overline{\gamma_x^y(t)})$ is constant for each t, the two angles are congruent.

By a similar argument, we get $\angle_{\bar{a}}(\overline{\gamma_a^b(t)}, \overline{\gamma_a^c(t)}) = \angle_{\bar{x}}(\overline{\gamma_x^y(t)}, \overline{\gamma_x^z(t)})$ for each t. Thus, by definition, $\angle_a(b, c) = \angle_x(y, z)$. Moreover, $\triangle(\bar{a}, \bar{b}, \bar{c})$ is congruent to $\triangle(\bar{x}, \bar{y}, \bar{z})$. Then: $d_{\mathbb{E}^2}(\bar{b}, \bar{c}) = d_{\mathbb{E}^2}(\bar{y}, \bar{z})$ and hence $d(b, c) = d(y, z)$. Now by (i) and the definition of $\langle \cdot, \cdot \rangle$, (iii) is obvious.

To prove (iv), suppose that $-\gamma_x^z = \gamma_x^{z'}$ and $-\gamma_x^y = \gamma_x^{y'}$. Let $\triangle_1 = \triangle(\overline{x_1}, \overline{y'}, \bar{z})$ and $\triangle_2 = \triangle(\overline{x_2}, \bar{y}, \overline{z'})$ be the comparison triangles for $\triangle(x, y', z)$ and $\triangle(x, y, z')$ respectively. Since $\gamma_x^{y'} \parallel \gamma_y^x$ and $\gamma_x^{z'} \parallel \gamma_z^x$, we can consider $\triangle_1$ and $\triangle_2$ such that

$\overline{x_1}\,\overline{y'}$ is parallel to $\overline{y}\overline{x_2}$ and $\overline{x_2}\overline{z'}$ is parallel to $\overline{z}\overline{x_1}$. Then: $\angle_{\overline{x_1}}(\overline{z}, \overline{y'}) = \angle_{\overline{x_2}}(\overline{y}, \overline{z'})$. Therefore, $\triangle_1$ and $\triangle_2$ are congruent. Hence, $d_{\mathbb{E}^2}(\overline{z'}, \overline{y}) = d_{\mathbb{E}^2}(\overline{y'}, \overline{z})$. It means that $d(z', y) = d(y', z)$. Now we have

$$\begin{aligned}\langle -\gamma_x^y, \gamma_x^z\rangle &= \langle \gamma_x^{y'}, \gamma_x^z\rangle = \tfrac{1}{2}[d^2(x,z) + d^2(y',x) - d^2(y',z)]\\ &= \tfrac{1}{2}[d^2(x,z') + d^2(x,y) - d^2(z',y)] = \langle \gamma_x^y, \gamma_x^{z'}\rangle = \langle \gamma_x^y, -\gamma_x^z\rangle. \qquad \square\end{aligned}$$

Lemma 2.12. *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. Then $\partial f : \mathcal{X} \to \mathcal{X}^*$ is monotone; that is*

$$\langle \gamma_x^y, \sigma_x^z - P_{yx}(\eta_y^n)\rangle \le 0, \quad \forall x, y \in \mathcal{X}, \quad \forall \eta_y^n \in \partial f(y), \quad \forall \sigma_x^z \in \partial f(x).$$

Proof. Suppose that $\eta_y^n \in \partial f(y)$ and $\sigma_x^z \in \partial f(x)$. Thus $f(y) - f(x) \ge \langle \gamma_x^y, \sigma_x^z\rangle$ and $f(x) - f(y) \ge \langle \gamma_y^x, \eta_y^n\rangle$. Note that

$$\langle \gamma_y^x, \eta_y^n\rangle = \langle P_{yx}(\eta_y^n), P_{yx}(\gamma_y^x)\rangle = \langle -P_{yx}(\eta_y^n), \gamma_x^y\rangle.$$

Therefore

$$\langle \gamma_x^y, \sigma_x^z - P_{yx}(\eta_y^n)\rangle \le 0, \quad \forall x, y \in \mathcal{X}, \quad \forall \eta_y^n \in \partial f(y), \quad \forall \sigma_x^z \in \partial f(x). \qquad \square$$

Let S be a nonempty closed convex subset of $\mathcal{X}$ and $\pi_S : \mathcal{X} \longrightarrow S$ be the nearest point map onto S.

Now we need some lemmas to prove the density theorem for the subdifferential of a convex lower semicontinuous function on $\mathcal{X}$.

Lemma 2.13. *Let $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ be a proper, convex and lower semicontinuous function. Suppose that $(e, r_e) \in (\operatorname{epi}(f))^c$ and $X_0 = (x_0, f(x_0)) = \pi_{\operatorname{epi}(f)}(e, r_e)$ with $f(x_0) - r_e = 1$. Then: $\partial f(x_0) \ne \varnothing$.*

Proof. Set $E = (e, r_e)$. By Proposition 2.4 in [Bridson and Haefliger 1999], for each $A = (a, r_a) \in \operatorname{epi}(f)$ not equal to X_0, we have $\angle_{X_0}(E, A) \ge \frac{\pi}{2}$. Consequently: $\rho^2(A, X_0) + \rho^2(X_0, E) \le \rho^2(A, E)$, where ρ is the metric of the space $\mathcal{X} \times \mathbb{R}$ defined by

$$\rho^2((x_1, r_1), (x_2, r_2)) = d^2(x_1, x_2) + (r_2 - r_1)^2.$$

Thus

$$d^2(a, x_0) + d^2(x_0, e) + (f(x_0) - r_a)^2 + (f(x_0) - r_e)^2 \le d^2(a, e) + (r_e - r_a)^2.$$

Therefore, we can easily find

$$\tfrac{1}{2}[d^2(a, x_0) + d^2(x_0, e) - d^2(a, e)] \le (r_a - f(x_0))(f(x_0) - r_e). \tag{2-2}$$

Since $f(x_0) - r_e = 1$, we get

$$\langle \gamma_{x_0}^e, \gamma_{x_0}^a\rangle \le r_a - f(x_0)$$

for all $a \in \operatorname{dom} f$. Put $r_a = f(a)$. Clearly, the above inequality holds for each $a \notin \operatorname{dom} f$. Hence, $\gamma^e_{x_0} \in \partial f(x_0)$. □

It is worth pointing out that since r_a in (2-2) (in the proof of Lemma 2.13) can be selected large enough, we get $f(x_0) \geq r_e$.

Remark 2.14. The notation $(1-t)a \oplus tb$ is used for some results on Hilbert balls in [Shafrir 1992], on hyperbolic spaces in [Goebel and Reich 1984; Reich and Shafrir 1990] and on Hadamard spaces in [Dhompongsa and Panyanak 2008] to denote the unique point a_t such that $d(a, a_t) = td(a, b)$ and $d(a_t, b) = (1-t)d(a, b)$. Now, if (x_0, y_0) and (x_1, y_1) are two points in $\mathcal{X} \times \mathcal{Y}$ and (x, y) is a point on the unique geodesic joining them, (x, y) is the unique point satisfying the equations

$$\rho((x_0, y_0), (x, y)) = t\rho((x_0, y_0), (x_1, y_1)),$$
$$\rho((x_1, y_1), (x, y)) = (1-t)\rho((x_0, y_0), (x_1, y_1))$$

for some $t \in [0, 1]$. The point

$$(\gamma^{x_1}_{x_0}(td(x_0, x_1)), \gamma^{y_1}_{y_0}(td(y_0, y_1))) = ((1-t)x_0 \oplus tx_1, (1-t)y_0 \oplus ty_1)$$

has the same property. Hence

$$(1-t)(x_0, y_0) \oplus t(x_1, y_1) = ((1-t)x_0 \oplus tx_1, (1-t)y_0 \oplus ty_1)$$

for all $t \in [0, 1]$.

If $x, y \in \mathcal{X}$, we denote by $[\![x, y]\!]$ the set $\{\gamma^y_x(t) : t \in \operatorname{dom}\gamma^y_x\}$.

Lemma 2.15. *Let $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ be a proper, convex and lower semicontinuous function. Suppose that $(y_0, r_0) \in (\operatorname{epi}(f))^c$ and $X_0 = (x_0, f(x_0)) = \pi_{\operatorname{epi}(f)}((y_0, r_0))$ and $x_0 \in \operatorname{int}(\operatorname{dom} f)$, where $\operatorname{dom} f = \{x \in \mathcal{X} \mid f(x) < \infty\}$. Then: $r_0 \neq f(x_0)$.*

Proof. Assume by contradiction that $r_0 = f(x_0)$. Put $Y_0 = (y_0, r_0)$. Let r be a positive number so that $B(x_0, r) \subseteq \operatorname{dom} f$. There is a $\lambda_0 \in [0, 1]$ such that $\gamma^{y_0}_{x_0}(\lambda d(x_0, y_0)) \in B(x_0, r)$ for the unit speed geodesic $\gamma^{y_0}_{x_0}$ and for each $\lambda \in [0, \lambda_0]$. First suppose that there exists $x_1 \in B(x_0, r) \cap [\![x_0, y_0]\!]$ such that $f(x_0) < f(x_1)$. Hence, $x_1 = \gamma^{y_0}_{x_0}(\lambda_1 d(x_0, y_0))$ for some $\lambda_1 \in (0, \lambda_0)$. Put $X_1 = (x_1, f(x_1)) \in \operatorname{epi} f$. Then,

$$\rho^2(X_1, X_0) + \rho^2(X_0, Y_0) \leq \rho^2(X_1, Y_0).$$

Putting $\alpha = (f(x_1) - f(x_0))^2$, we have

$$\rho^2(X_0, X_1) = d^2(x_0, x_1) + \alpha = \lambda_1^2 d^2(x_0, y_0) + \alpha, \tag{2-3}$$

$$\rho^2(Y_0, X_1) = d^2(y_0, x_1) + \alpha = (1-\lambda_1)^2 d^2(x_0, y_0) + \alpha, \tag{2-4}$$

$$\rho^2(X_0, Y_0) = d^2(x_0, y_0). \tag{2-5}$$

Hence, by (2-3), (2-4) and (2-5), we have

$$\lambda_1^2 d^2(x_0, y_0) + \alpha + d^2(x_0, y_0) \le (1-\lambda_1)^2 d^2(x_0, y_0) + \alpha.$$

Thus $\lambda_1^2 + 1 \le (1-\lambda_1)^2$, and we get the contradiction $\lambda_1 \le 0$. Next, consider the case that $f(x) \le f(x_0)$ for each $x \in B(x_0, r) \cap [\![x_0, y_0]\!]$. Let

$$Y_n = (1 - \tfrac{1}{n}) X_0 \oplus \tfrac{1}{n} Y_0 = (y_n, r_n).$$

By Proposition 2.4 in [Bridson and Haefliger 1999], X_0 is the nearest point of $\operatorname{epi}(f)$ to each Y_n, and $\{Y_n\}$ is a sequence converging to X_0. If $y_0 \in B(x_0, r)$, then $f(y_0) \le f(x_0)$. Thus $(y_0, r_0) = (y_0, f(x_0)) \in \operatorname{epi}(f)$, which is a contradiction. Therefore, $y_0 \in (B(x_0, r))^c$. By Remark 2.14 we have $r_n = f(x_0)$ for every n, so a similar argument for each Y_n shows that $y_n \in (B(x_0, r))^c$. This means that $\{y_n\}$ is a sequence in $(B(x_0, r))^c$ converging to x_0. Thus we get the contradiction $x_0 \notin B(x_0, r)$. □

Lemma 2.16. *Let $E' \in (\operatorname{epi}(f))^c$ and $X_0 = (x_0, f(x_0)) = \pi_{\operatorname{epi}(f)}(E')$. Then there exists $E = (e, r_e) \in (\operatorname{epi}(f))^c$ such that $f(x_0) - r_e = 1$ and $X_0 = \pi_{\operatorname{epi}(f)}(E)$.*

Proof. Let γ be the geodesic joining X_0 to E'. Put $E' = (e', r_{e'})$. First suppose that $f(x_0) - r_{e'} \ge 1$. Since γ is continuous by the intermediate value theorem, the assertion is obvious.

Next, suppose that $f(x_0) - r_{e'} < 1$. Put

$$l = \rho(X_0, E') \quad \text{and} \quad s = \frac{l}{f(x_0) - r_{e'}}.$$

Let $\overline{\gamma}$ be the extension of γ to $[0, \infty)$ that is the unit speed geodesic ray emanating from X_0. Put $E = \overline{\gamma}(s)$. We claim that E is the desired point. If $E = (e, r_e)$, then one has $E' = (1 - \frac{l}{s}) X_0 \oplus \frac{l}{s} E$. By Remark 2.14, $e' = (1 - \frac{l}{s}) x_0 \oplus \frac{l}{s} e$ and $r_{e'} = (1 - \frac{l}{s}) f(x_0) + \frac{l}{s} r_e$. Hence, $f(x_0) - r_{e'} = \frac{l}{s}(f(x_0) - r_e)$. Therefore,

$$f(x_0) - r_e = \frac{s}{l}(f(x_0) - r_{e'}) = s \times \frac{f(x_0) - r_{e'}}{l} = 1.$$

Now we prove that $\pi_{\operatorname{epi}(f)}(E) = X_0$. Suppose for a contradiction that $\pi_{\operatorname{epi}(f)}(E) = X'$ and $X_0 \ne X'$. Then $\angle_{X_0}(X', E') \ge \frac{\pi}{2}$ and $\angle_{X'}(X_0, E) \ge \frac{\pi}{2}$. Then the sum of the angles of $\triangle(X', X_0, E)$ is more than π, which is a contradiction. □

The next theorem is a generalization of the density theorem on geodesically complete Hadamard spaces. For a density theorem on Hilbert spaces see [Clarke et al. 1998].

Theorem 2.17. *Suppose that f is a proper, convex and lower semicontinuous function. Then* $\operatorname{dom}(\partial f(x))$ *is dense in* $\operatorname{int}(\operatorname{dom} f)$.

Proof. Given $x_0 \in \operatorname{int}(\operatorname{dom} f)$, the point $X_0 = (x_0, f(x_0))$ is a boundary point of $\operatorname{epi}(f)$. So, there exists a sequence $Y_n = (y_n, r_n)$ in the complement of $\operatorname{epi}(f)$ that converges to X_0. Since $\operatorname{epi}(f)$ is convex and closed in $\mathcal{X} \times \mathbb{R}$ there exists a unique point $X_n = (x_n, f(x_n)) \in \operatorname{epi}(f)$ such that $\pi_{\operatorname{epi}(f)}(Y_n) = X_n$ for each Y_n. Moreover,

$$\rho(X_n, X_0) \le \rho(X_n, Y_n) + \rho(Y_n, X_0) \le 2\rho(Y_n, X_0),$$

which implies that X_n converges to X_0. Therefore, the sequence $\{x_n\}$ converges to x_0 and for every neighborhood U of x_0, there exists $x_n \in U$. By Lemma 2.16, one can assume that $f(x_n) - r_n = 1$, so by Lemma 2.13, $\partial f(x_n) \neq \varnothing$. □

References

[Ahmadi Kakavandi and Amini 2010] B. Ahmadi Kakavandi and M. Amini, "Duality and subdifferential for convex functions on complete CAT(0) metric spaces", *Nonlinear Anal.* **73**:10 (2010), 3450–3455. MR 2012b:47131 Zbl 1200.53045

[Berg and Nikolaev 2008] I. D. Berg and I. G. Nikolaev, "Quasilinearization and curvature of Aleksandrov spaces", *Geom. Dedicata* **133** (2008), 195–218. MR 2008m:53167 Zbl 1144.53045

[Bridson and Haefliger 1999] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften **319**, Springer, Berlin, 1999. MR 2000k:53038 Zbl 0988.53001

[Burago et al. 2001] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics **33**, Amer. Math. Soc., Providence, RI, 2001. MR 2002e:53053 Zbl 0981.51016

[Clarke et al. 1998] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth analysis and control theory*, Graduate Texts in Mathematics **178**, Springer, New York, 1998. MR 99a:49001 Zbl 1047.49500

[Dhompongsa and Panyanak 2008] S. Dhompongsa and B. Panyanak, "On Δ-convergence theorems in CAT(0) spaces", *Comput. Math. Appl.* **56**:10 (2008), 2572–2579. MR 2009k:47161 Zbl 1165.65351

[Goebel and Reich 1984] K. Goebel and S. Reich, *Uniform convexity, hyperbolic geometry, and nonexpansive mappings*, Monographs and Textbooks in Pure and Applied Mathematics **83**, Marcel Dekker, New York, 1984. MR 86d:58012 Zbl 0537.46001

[Hosseini and Pouryayevali 2011] S. Hosseini and M. R. Pouryayevali, "Generalized gradients and characterization of epi–Lipschitz sets in Riemannian manifolds", *Nonlinear Anal.* **74**:12 (2011), 3884–3895. MR 2012g:49034 Zbl 1225.49046

[Hosseini and Pouryayevali 2013a] S. Hosseini and M. R. Pouryayevali, "Euler characteristic of epi–Lipschitz subsets of Riemannian manifolds", *J. Convex Anal.* **20**:1 (2013), 67–91. MR 3086441 Zbl 1261.49002

[Hosseini and Pouryayevali 2013b] S. Hosseini and M. R. Pouryayevali, "Nonsmooth optimization techniques on Riemannian manifolds", *J. Optim. Theory Appl.* **158**:2 (2013), 328–342. MR 3084379 Zbl 1273.90110

[Hosseini and Pouryayevali 2013c] S. Hosseini and M. R. Pouryayevali, "On the metric projection onto prox-regular subsets of Riemannian manifolds", *Proc. Amer. Math. Soc.* **141**:1 (2013), 233–244. MR 2988725 Zbl 1277.58005

[Kopecká and Reich 2010] E. Kopecká and S. Reich, "A mean ergodic theorem for nonlinear semigroups on the Hilbert ball", *J. Nonlinear Convex Anal.* **11**:2 (2010), 185–197. MR 2011j:47178 Zbl 1209.47028

[Reich and Shafrir 1990] S. Reich and I. Shafrir, "Nonexpansive iterations in hyperbolic spaces", *Nonlinear Anal.* **15**:6 (1990), 537–558. MR 91k:47135 Zbl 0728.47043

[Shafrir 1992] I. Shafrir, "Coaccretive operators and firmly nonexpansive mappings in the Hilbert ball", *Nonlinear Anal.* **18**:7 (1992), 637–648. MR 93d:47100 Zbl 0752.47018

Received April 30, 2014. Revised January 19, 2015.

MINA MOVAHEDI
DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCES
ALZAHRA UNIVERSITY
TEHRAN
IRAN
m.movahedi18@yahoo.com
m.movahedi@alzahra.ac.ir

DARYOUSH BEHMARDI
DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCES
ALZAHRA UNIVERSITY
TEHRAN
IRAN
d2beh@yahoo.com
behmardi@alzahra.ac.ir

SEYEDEHSOMAYEH HOSSEINI
SEMINAR OF APPLIED MATHEMATICS
EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE
CH-8092 ZÜRICH
SWITZERLAND
seyedehsomayeh.hosseini@math.ethz.ch

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.449

L^p REGULARITY OF WEIGHTED SZEGŐ PROJECTIONS ON THE UNIT DISC

SAMANGI MUNASINGHE AND YUNUS E. ZEYTUNCU

We present a family of weights on the unit disc for which the corresponding weighted Szegő projection operators are irregular on L^p spaces. We further investigate the dual spaces of weighted Hardy spaces corresponding to this family.

1. Introduction

1.1. ***Classical setting.*** Let $\mathbb{D}$ denote the unit disc in $\mathbb{C}$ and $\mathbb{T}$ the unit circle. Let $\mathcal{O}(\mathbb{D})$ denote the set of holomorphic functions on $\mathbb{D}$. For $1 \leq p < \infty$, the ordinary Hardy space is defined as

$$\mathcal{H}^p(\mathbb{T}) = \{f \in \mathcal{O}(\mathbb{D}) \text{ and } \|f\|_{\mathcal{H}^p} < \infty\},$$

where

$$\|f\|_{\mathcal{H}^p}^p = \sup_{0 \leq r < 1} \int_0^{2\pi} |f(re^{i\theta})|^p \, d\theta.$$

It is known (see [Duren 1970]) that functions in $\mathcal{H}^p(\mathbb{T})$ have boundary limits almost everywhere, i.e., for almost every $\theta \in [0, 2\pi]$

$$f(e^{i\theta}) = \lim_{r \to 1^-} f(re^{i\theta})$$

exists. Moreover,

$$\|f\|_{L^p(\mathbb{T})} = \|f\|_{\mathcal{H}^p(\mathbb{T})},$$

where $L^p(\mathbb{T})$ is defined using the standard Lebesgue measure (denoted by $d\theta$) on the unit circle. It is also known that $\mathcal{H}^p(\mathbb{T})$ is a closed subspace of $L^p(\mathbb{T})$. In particular, for $p = 2$, the orthogonal projection operator, called the Szegő projection operator exists;

$$\boldsymbol{S} : L^2(\mathbb{T}) \longrightarrow \mathcal{H}^2(\mathbb{T}).$$

MSC2010: primary 32A25; secondary 32A35, 42B30.
Keywords: Szegő projection, A_p weights, Weighted Hardy spaces, Dual spaces.

The operator $\boldsymbol{S}$ is an integral operator with the kernel $S(z, w)$ (called the Szegő kernel), and for $f \in L^2(\mathbb{T})$,

$$\boldsymbol{S} f(z) = \int_{\mathbb{T}} S(z, w) f(w)\, d\theta.$$

It follows from the general theory of reproducing kernels that for any orthonormal basis $\{e_n(z)\}_{n=0}^{\infty}$ for $\mathcal{H}^2(\mathbb{T})$, the Szegő kernel is given by

$$S(z, w) = \sum_{n=0}^{\infty} e_n(z) \overline{e_n(w)}.$$

1.2. ***Weighted setting.*** Let $g(z)$ be a holomorphic function on $\mathbb{D}$ that is continuous on $\overline{\mathbb{D}}$ and has no zeros inside $\mathbb{D}$. We set $\mu(z) = |g(z)|^2$ and define weighted Hardy spaces and weighted Szegő projections using the function $\mu(z)$ as a weight on $\mathbb{T}$.

For $1 \le p < \infty$, we define the weighted Lebesgue and Hardy spaces with respect to μ as

$$L^p(\mathbb{T}, \mu) = \{f \text{ measurable function on } \mathbb{D} \text{ and } \|f\|_{p,\mu} < \infty\},$$

where

$$\|f\|_{p,\mu}^p = \int_{\mathbb{T}} |f(w)|^p \mu(w) d\theta = \int_{\mathbb{T}} |f(w)(g(w))^{2/p}|^p\, d\theta,$$

and

$$\mathcal{H}^p(\mathbb{T}, \mu) = \{f \in \mathcal{O}(\mathbb{D}) \text{ such that } \|f\|_{\mathcal{H}^p,\mu} < \infty\},$$

where

$$\|f\|_{\mathcal{H}^p,\mu}^p = \sup_{0 \le r < 1} \int_0^{2\pi} |f(re^{i\theta})(g(re^{i\theta}))^{2/p}|^p\, d\theta.$$

Note that, $f \in \mathcal{H}^p(\mathbb{T}, \mu)$ implies $f(z)(g(z))^{2/p} \in \mathcal{H}^p(\mathbb{T})$, which in turn gives that $f(z)(g(z))^{2/p}$ has almost everywhere boundary limits. Hence so does $f(z)$. Additionally, $\|f\|_{\mathcal{H}^p,\mu} = \|f\|_{p,\mu}$. Furthermore, $L^p(\mathbb{T}, \mu)$ is a Banach space and $\mathcal{H}^p(\mathbb{T}, \mu)$ is a closed subspace of $L^p(\mathbb{T}, \mu)$.

In particular, again when $p = 2$, we obtain the weighted Szegő projection

$$\boldsymbol{S}_\mu : L^2(\mathbb{T}, \mu) \longrightarrow \mathcal{H}^2(\mathbb{T}, \mu).$$

Following the similar theory, we note that $\boldsymbol{S}_\mu$ is an integral operator

$$\boldsymbol{S}_\mu f(z) = \int_{\mathbb{T}} S_\mu(z, w) f(w) \mu(w)\, d\theta.$$

If $\{f_n(z)\}_{n=0}^{\infty}$ is an orthonormal basis for $\mathcal{H}^2(\mathbb{T},\mu)$ then

$$S_\mu(z,w)=\sum_{n=0}^{\infty} f_n(z)\overline{f_n(w)}.$$

We are interested in the action of $\boldsymbol{S}_\mu$ on $L^p(\mathbb{T},\mu)$. By definition, $\boldsymbol{S}_\mu$ is a bounded operator from $L^2(\mathbb{T},\mu)$ to $L^2(\mathbb{T},\mu)$. The problem we investigate is the boundedness of $\boldsymbol{S}_\mu$ from $L^p(\mathbb{T},\mu)$ to $L^p(\mathbb{T},\mu)$ for other values of $p\in(1,\infty)$. Note that for any given weight μ as above, we can associate an interval $I_\mu\subset(1,\infty)$ such that $\boldsymbol{S}_\mu$ is bounded from $L^p(\mathbb{T},\mu)$ to $L^p(\mathbb{T},\mu)$ if and only if $p\in I_\mu$. By definition, $2\in I_\mu$, and by duality and interpolation, I_μ is a conjugate symmetric interval around 2. Namely, if some $p_0>2$ is in I_μ, so is q_0 where $1/q_0+1/p_0=1$.

In the classical setting, i.e., $\mu\equiv 1$, the Szegő projection operator is bounded from $L^p(\mathbb{T})$ to $L^p(\mathbb{T})$ for any $1<p<\infty$, see [Zhu 2007, page 257].

The purpose of this note is to construct weights μ on $\mathbb{T}$ for which the corresponding interval I_μ can be any open interval larger than $\{2\}$ but smaller than $(1,\infty)$.

Theorem 1. *For any given $p_0>2$, there exists a weight μ on $\mathbb{T}$ such that $I_\mu=(q_0,p_0)$ where $1/q_0+1/p_0=1$, i.e., the weighted Szegő projection $\boldsymbol{S}_\mu$ is bounded on $L^p(\mathbb{T},\mu)$ if and only if $q_0<p<p_0$.*

Our proof of this theorem is similar to the proof of the analogous statement for weighted Bergman projections in [Zeytuncu 2013] with modifications from Bergman kernels to Szegő kernels. The main ingredient is the theory of A_p weights on $\mathbb{T}$.

When the weighted Szegő projection $\boldsymbol{S}_\mu$ is bounded on $L^p(\mathbb{T},\mu)$ for some p, one can identify the dual space of the weighted Hardy space $\mathcal{H}^p(\mathbb{T},\mu)$. However, when $\boldsymbol{S}_\mu$ fails to be bounded, a different approach is needed to identify the dual spaces. In the third section, we address this issue and describe the dual spaces of weighted Hardy spaces.

The following notation is used in the rest of the note. We write $f(z)\simeq g(z)$ when $c\cdot g(z)\le f(z)\le C\cdot g(z)$ for some positive constants c and C which are independent of z. Similarly we write $f(z)\lesssim g(z)$ when $f(z)\le C\cdot g(z)$ for some positive constant C. We use $d\theta$ for the Lebesgue measure on the unit circle $\mathbb{T}$. When we integrate functions (that are also defined on the unit disc) on $\mathbb{T}$, instead of writing $e^{i\theta}$, we keep z and w as the variables.

2. Proof of Theorem 1

2.1. ***Relation between weighted kernels.*** The particular choice of $\mu(z)$ indicates the following relation between the weighted Szegő kernels $S_\mu(z,w)$ and the ordinary Szegő kernel $S(z,w)$.

Proposition 2. *For $\mu(z) = |g(z)|^2$ as above, the following relation holds*

$$S(z, w) = g(z) S_\mu(z, w) \overline{g(w)}. \tag{1}$$

Proof. Let $\{e_n(z)\}_{n=0}^{\infty}$ be an orthonormal basis for $\mathcal{H}^2(\mathbb{T})$. Since $g(z)$ does not vanish inside $\mathbb{D}$, each $e_n(z)/g(z)$ is a holomorphic function on $\mathbb{D}$ and is in $\mathcal{H}^2(\mathbb{T}, |g|^2)$ by construction. Following the orthonormal properties of the $e_n(z)$ we have

$$\left\langle \frac{e_n(z)}{g(z)}, \frac{e_m(z)}{g(z)} \right\rangle_\mu = \langle e_n(z), e_m(z) \rangle = \delta_{n,m},$$

where $\delta_{n,m}$ is the Kronecker delta.

Also for any f in $\mathcal{H}^2(\mathbb{T}, |g|^2)$, $(f \cdot g)$ is in $\mathcal{H}^2(\mathbb{T})$ and hence can be written as a linear combination of the $e_n(z)$. Consequently so can f, using the quotients $e_n(z)/g(z)$. Hence, $\{e_n(z)/g(z)\}_{n=0}^{\infty}$ is an orthonormal basis for $\mathcal{H}^2(\mathbb{T}, |g|^2)$.

Therefore, using the basis representation of the Szegő kernels we obtain

$$\begin{aligned} S(z, w) &= \sum_{n=0}^{\infty} e_n(z) \overline{e_n(w)} = g(z) \left(\sum_{n=0}^{\infty} \frac{e_n(z)}{g(z)} \overline{\frac{e_n(w)}{g(w)}} \right) \overline{g(w)} \\ &= g(z) S_\mu(z, w) \overline{g(w)}. \end{aligned}$$

□

2.2. A_p *weights on* $\mathbb{T}$. For $p \in (1, \infty)$, a weight μ on $\mathbb{T}$ is said to be in $A_p(\mathbb{T})$ if

$$\sup_{\substack{I \\ I \subset \mathbb{T}}} \left(\frac{1}{|I|} \int_I \mu(\theta) \, d\theta \right) \left(\frac{1}{|I|} \int_I \mu(\theta)^{\frac{-1}{p-1}} \, d\theta \right)^{p-1} < \infty,$$

where I denotes intervals in $\mathbb{T}$.

These weights are used to characterize the L^p regularity of the ordinary Szegő projection on weighted spaces. The following result appears in [Garnett 1981] and is used in [Lanzani and Stein 2004, Equation (2.3)] in connection with a conformal map based approach to the investigation of the unweighted Szegő projection for a general domain.

Theorem 3. *The ordinary Szegő projection* $\mathbf{S}$ *is bounded from* $L^p(\mathbb{T}, \mu)$ *to* $L^p(\mathbb{T}, \mu)$ *if and only if* $\mu \in A_p(\mathbb{T})$.

Proof. This result is an immediate consequence of the fact that the Szegő kernel of the unit disc agrees with the Cauchy kernel (see [Kerzman and Stein 1978]) together with the classical weighted theory for the latter, see also [Garnett 1981]. □

The following theorem follows from Equation (1) and Theorem 3.

Proposition 4. *For $1 < p < \infty$ and $\mu(z) = |g(z)|^2$ as above, the following are equivalent.*

(1) $\mathbf{S}_\mu$ *is bounded from* $L^p(\mathbb{T}, |g|^2)$ *to* $L^p(\mathbb{T}, |g|^2)$.

(2) *$\boldsymbol{S}$ is bounded from $L^p(\mathbb{T}, |g|^{2-p})$ to $L^p(\mathbb{T}, |g|^{2-p})$.*

(3) $|g|^{2-p} \in A_p(\mathbb{T})$.

Proof. Theorem 3 gives the equivalence of (2) and (3). We show the equivalence of (1) and (2). Using the relation between the kernels from the previous proposition, we obtain the following relation between the corresponding operators:

$$g(z)\big(\boldsymbol{S}_\mu f\big)(z) = \big(\boldsymbol{S}(f \cdot g)\big)(z) \quad \text{for } f \in L^2(\mathbb{T}, |g|^2).$$

Indeed, suppose (2) is true. Then

$$\begin{aligned}\|\boldsymbol{S}_\mu f\|_{p,|g|^2}^p &= \int_{\mathbb{T}} |(\boldsymbol{S}_\mu f)(w)|^p |g(w)|^2 \, d\theta \\ &= \int_{\mathbb{T}} |\big(\boldsymbol{S}(f \cdot g)\big)(w)|^p |g(w)|^{2-p} \, d\theta = \|\boldsymbol{S}(f \cdot g)\|_{p,|g|^{2-p}}^p \\ &\lesssim \|f \cdot g\|_{p,|g|^{2-p}}^p = \|f\|_{p,|g|^2}^p \,,\end{aligned}$$

which proves (1).

Now when (1) is true,

$$\begin{aligned}\|\boldsymbol{S} f\|_{p,|g|^{2-p}}^p &= \int_{\mathbb{T}} |(\boldsymbol{S} f)(w)|^p |g(w)|^{2-p} d\theta \\ &= \int_{\mathbb{T}} |(\boldsymbol{S}_\mu(f/g))(w)|^p |g(w)|^2 d\theta = \|\boldsymbol{S}_\mu(f/g)\|_{p,|g|^2}^p \\ &\lesssim \|f/g\|_{p,|g|^2}^p = \|f\|_{p,|g|^{2-p}}^p\end{aligned}$$

and hence (2) is true. ☐

We can now present a family of weights that behave as claimed in Theorem 1.

Theorem 5. *For $\alpha \geq 0$, let $g_\alpha(z) = (z-1)^\alpha$ and $\mu_\alpha(z) = |g_\alpha(z)|^2$. Then the weighted Szegő projection operator $\boldsymbol{S}_{\mu_\alpha}$ is bounded on $L^p(\mathbb{T}, \mu_\alpha)$ if and only if $p \in \big(\frac{2\alpha+1}{\alpha+1}, \frac{2\alpha+1}{\alpha}\big)$.*

Remark 6. Theorem 5 is a quantitative version of Theorem 1 and therefore we obtain a proof of Theorem 1 when we prove Theorem 5.

Remark 7. Note that as $\alpha \to 0^+$ the interval $\big(\frac{2\alpha+1}{\alpha+1}, \frac{2\alpha+1}{\alpha}\big)$ approaches $(1, \infty)$ and as $\alpha \to \infty$ the interval $\big(\frac{2\alpha+1}{\alpha+1}, \frac{2\alpha+1}{\alpha}\big)$ approaches $\{2\}$. Hence, any conjugate symmetric interval around 2 can be achieved as the boundedness range of a weighted Szegő projection.

Proof of Theorem 5. First note that on intervals I with $\theta = 0 \notin I$, the weight $|g_\alpha(z)|^{2-p} = |z-1|^{\alpha(2-p)} \simeq C$. Therefore, both integrals in the $A_p(\mathbb{T})$ condition are finite and hence so is the supremum over all such intervals when p is in the given range. On intervals that contain $z = 0$ we have the following.

Step 1. We show that for the weights $\omega(z) = |g_\alpha(z)|^{2-p} = |z-1|^{\alpha(2-p)}$, the second integral in the $A_p(\mathbb{T})$ condition diverges for arcs $I = (-\epsilon, \epsilon)$ if and only if p is outside the given region.

For intervals $I = (-\epsilon, \epsilon)$ with small ϵ and $p \le \frac{2\alpha+1}{\alpha+1}$,

$$\begin{aligned}\int_I \omega(z)^{\frac{1}{1-p}} d\theta &= \int_{-\epsilon}^{\epsilon} |e^{i\theta} - 1|^{\frac{\alpha(2-p)}{(1-p)}} d\theta \\ &= \int_{-\epsilon}^{\epsilon} \left(\sqrt{2(1-\cos(\theta))}\right)^{\frac{\alpha(2-p)}{(1-p)}} d\theta \\ &\simeq \int_{-\epsilon}^{\epsilon} \theta^{\frac{\alpha(2-p)}{(1-p)}} d\theta = \infty,\end{aligned}$$

because $\alpha(2-p)/(1-p) \le -1$. Hence $\omega \notin A_p(\mathbb{T})$ for such p.

Also, when $p \ge \frac{2\alpha+1}{\alpha}$,

$$\int_I \omega(z)\, d\theta \simeq \int_{-\epsilon}^{\epsilon} \theta^{\alpha(2-p)}\, d\theta = \infty,$$

because $\alpha(2-p) \le -1$. Hence $\omega \notin A_p(\mathbb{T})$ for $p \ge \frac{2\alpha+1}{\alpha}$ either.

The same calculations show convergence of all integrals for p in the desired range.

Step 2. We show that for $p \in \left(\frac{2\alpha+1}{\alpha+1}, \frac{2\alpha+1}{\alpha}\right)$ the integrals in the A_p condition are finite over any (general) interval $I = (\theta_0 - R, \theta_0 + R)$ with $\theta_0 \neq 0$. We consider two cases.

Case 1. $I \cap \operatorname{Arc}(0, 2R) = \varnothing$.

On such intervals I, $3R < \theta_0$ and so $2\theta_0/3 \le \theta_0 - R \le \theta \le \theta_0 + R \le 4\theta_0/3$ giving $\theta \simeq \theta_0$. So, $\omega = |z-1|^{\alpha(2-p)} \simeq \theta_0^{\alpha(2-p)}$. Therefore,

$$\frac{1}{|I|}\int_I \omega(z)\, d\theta \lesssim \frac{1}{2R}\int_I \theta_0^{\alpha(2-p)}\, d\theta = \theta_0^{\alpha(2-p)}.$$

and

$$\begin{aligned}\left(\frac{1}{|I|}\int_I \omega(z)^{\frac{1}{1-p}}\, d\theta\right)^{p-1} &\lesssim \left(\frac{1}{2R}\int_I \theta_0^{\frac{\alpha(2-p)}{1-p}}\, d\theta\right)^{p-1} \\ &= \theta_0^{-\alpha(2-p)}.\end{aligned}$$

Hence the supremum over all such intervals is finite.

Case 2. $I \cap \operatorname{Arc}(0, 2R) \neq \varnothing$.

In this case, since $I \subset \operatorname{Arc}(0, 4R)$ and $\alpha(2-p) + 1 > 0$ when $2\alpha + 1/\alpha > p$, we have

$$\frac{1}{|I|}\int_I \omega(z)\, d\theta \simeq \frac{2}{8R}\int_0^{4R} \theta^{\alpha(2-p)}\, d\theta = \frac{1}{4R}\frac{\theta^{\alpha(2-p)+1}}{\alpha(2-p)+1}\bigg|_0^{4R} = \frac{4R^{\alpha(2-p)}}{\alpha(2-p)+1}.$$

Also since $\alpha(2-p)/(1-p)+1>0$ when $2\alpha+1/(\alpha+1)<p$,

$$\left(\frac{1}{|I|}\int_I \omega(z)^{\frac{1}{1-p}}\,d\theta\right)^{p-1} \simeq \left(\frac{2}{8R}\int_0^{4R}\theta^{\frac{\alpha(2-p)}{1-p}}\,d\theta\right)^{p-1} = \left(\frac{1}{4R}\frac{\theta^{\frac{\alpha(2-p)}{1-p}+1}}{\frac{\alpha(2-p)}{1-p}+1}\right)^{p-1}$$

$$\simeq \left(\frac{R^{\frac{\alpha(2-p)}{1-p}}}{\frac{\alpha(2-p)}{1-p}+1}\right)^{p-1} = \frac{2R^{-\alpha(2-p)}}{\left(\frac{\alpha(2-p)}{1-p}+1\right)^{p-1}}.$$

Therefore, the supremum over all intervals of the type in case 2 are also finite and $\omega=|g|^{2-p}\in A^p(\mathbb{T})$ if and only if $p\in\left(\frac{2\alpha+1}{\alpha+1},\frac{2\alpha+1}{\alpha}\right)$. □

Remark 8. The analog of Theorem 1 for domains in $\mathbb{C}^n$ ($n\geq 2$) is an open problem. See [Békollé and Bonami 1995] for a partial result. Also see [Lanzani and Stein 2013] for the regularity on strongly pseudoconvex domains.

3. Duality

In this section, we investigate the duals of Hardy spaces corresponding to weights from the previous section. For $\alpha\geq 0$ and $\mu_\alpha(z)=|z-1|^{2\alpha}$, a consequence of Theorem 5 is the following.

Theorem 9. *Let $\alpha\geq 0$ and $\mu_\alpha(z)=|z-1|^{2\alpha}$. For any $p\in\left(\frac{2\alpha+1}{\alpha+1},\frac{2\alpha+1}{\alpha}\right)$, the dual space of the weighted Hardy space $\mathcal{H}^p(\mathbb{T},|z-1|^{2\alpha})$ can be identified with $\mathcal{H}^q(\mathbb{T},|z-1|^{2\alpha})$, where $1/p+1/q=1$, under the pairing*

$$\langle f,h\rangle=\int_{\mathbb{T}} f(z)\overline{h(z)}|z-1|^{2\alpha}\,d\theta.$$

Proof. This is a standard argument; however, we present a proof here for completeness. For a given function $h\in\mathcal{H}^q(\mathbb{T},|z-1|^{2\alpha})$, we define a linear functional on $\mathcal{H}^p(\mathbb{T},|z-1|^{2\alpha})$ by

$$\boldsymbol{G}(f)=\int_{\mathbb{T}} f(z)\overline{h(z)}|z-1|^{2\alpha}\,d\theta.$$

It is clear that, by Hölder's inequality, $\boldsymbol{G}$ is a bounded functional with operator norm less than $\|h\|_{\mathcal{H}^q(\mathbb{T},|z-1|^{2\alpha})}$.

Conversely, let $\boldsymbol{G}$ be a bounded linear functional on $\mathcal{H}^p(\mathbb{T},|z-1|^{2\alpha})$. By the Hahn–Banach theorem, $\boldsymbol{G}$ extends to a bounded linear functional on $L^p(\mathbb{T},|z-1|^{2\alpha})$. Now using the duality of L^p spaces, we find a function $h\in L^q(\mathbb{T},|z-1|^{2\alpha})$ such that

$$\boldsymbol{G}(f)=\int_{\mathbb{T}} f(z)\overline{h(z)}|z-1|^{2\alpha}dz \quad \text{for } f\in L^p(\mathbb{T},|z-1|^{2\alpha}).$$

When we restrict $\boldsymbol{G}$ to $L^p(\mathbb{T}, |z-1|^{2\alpha}) \cap \mathcal{H}^2(\mathbb{T}, |z-1|^{2\alpha})$ and use self-adjointness of $\boldsymbol{S}_{\mu_\alpha}$ we get the following.

$$\begin{aligned}\boldsymbol{G}(f) &= \int_{\mathbb{T}} f(z)\overline{h(z)}|z-1|^{2\alpha}\,d\theta \\ &= \int_{\mathbb{T}} (\boldsymbol{S}_{\mu_\alpha} f)(z)\overline{h(z)}|z-1|^{2\alpha}\,d\theta \\ &= \int_{\mathbb{T}} f(z)\overline{(\boldsymbol{S}_{\mu_\alpha} h)(z)}|z-1|^{2\alpha}\,d\theta\end{aligned}$$

for $f \in L^p(\mathbb{T}, |z-1|^{2\alpha}) \cap \mathcal{H}^2(\mathbb{T}, |z-1|^{2\alpha})$.

Since the intersection of these two spaces is dense in $\mathcal{H}^p(\mathbb{T}, |z-1|^{2\alpha})$, we note that $\boldsymbol{G}$ is represented by the function $(\boldsymbol{S}_{\mu_\alpha} h)(z)$ and $\boldsymbol{S}_{\mu_\alpha} h \in \mathcal{H}^q(\mathbb{T}, |z-1|^{2\alpha})$ by Theorem 5. □

A natural question arises after this statement. How can we identify the dual space of the weighted Hardy space, $\mathcal{H}^p(\mathbb{T}, |z-1|^{2\alpha})$, when $p \notin \left(\frac{2\alpha+1}{\alpha+1}, \frac{2\alpha+1}{\alpha}\right)$?

The answer to this question follows from the following result on the boundedness of the weighted Szegő projection, $\boldsymbol{S}_{\mu_\alpha}$. Similar results for weighted Bergman projections have been presented recently in [Arroussi and Pau 2014] and [Constantin and Peláez 2015].

Proposition 10. *Let $\alpha \geq 0$ and $\mu_\alpha = |z-1|^{2\alpha}$. For any $1 < p < \infty$, the weighted Szegő projection $\boldsymbol{S}_{\mu_\alpha}$ is bounded on $L^p(\mathbb{T}, |z-1|^{\alpha p})$.*

Remark 11. Note that as p varies, changes occur not only in the integrability scale but also in the measure.

Proof. The proof follows from the relation between the kernels in Proposition 2 and the fact that the unweighted Szegő projection $\boldsymbol{S}$ is bounded on $L^p(\mathbb{T})$ for $1 < p < \infty$.

Let us take $f(z) \in L^p(\mathbb{T}, |z-1|^{\alpha p})$ and set

$$\tilde{f}(z) = f(z)\frac{|z-1|^{2\alpha}}{(z-1)^\alpha},$$

then we have $\tilde{f} \in L^p(\mathbb{T})$. Using this notation, we notice

$$\begin{aligned}\boldsymbol{S}_{\mu_\alpha} f(z) &= \int_{\mathbb{T}} S_{\mu_\alpha}(z, w) f(w)|w-1|^{2\alpha}\,d\theta \\ &= \frac{(z-1)^\alpha}{(z-1)^\alpha}\int_{\mathbb{T}} S_{\mu_\alpha}(z, w)(w-1)^\alpha f(w)\frac{|w-1|^{2\alpha}}{(w-1)^\alpha}\,d\theta \\ &= \frac{1}{(z-1)^\alpha}\int_{\mathbb{T}} S(z, w)\tilde{f}(w)\,d\theta \\ &= \frac{1}{(z-1)^\alpha}\boldsymbol{S}\big(\tilde{f}(w)\big)(z),\end{aligned}$$

where we invoke Proposition 2 when we pass from the second to the third line. Next by using the fact that the unweighted Szegő projection operator $\boldsymbol{S}$ is bounded on $L^p(\mathbb{T})$, we obtain the following.

$$\begin{aligned}
\|\boldsymbol{S}_{\mu_\alpha} f\|_{L^p(\mathbb{T},|z-1|^{\alpha p})} &= \int_{\mathbb{T}} |z-1|^{\alpha p} \frac{1}{|z-1|^{\alpha p}} \big|\boldsymbol{S}\big(\tilde{f}(w)\big)(z)\big|^p \, d\theta \\
&= \big\|\boldsymbol{S}\big(\tilde{f}(w)\big)\big\|^p_{L^p(\mathbb{T})} \lesssim \|\tilde{f}(w)\|^p_{L^p(\mathbb{T})} \\
&= \left\| f(w) \frac{|w-1|^{2\alpha}}{(w-1)^\alpha} \right\|^p_{L^p(\mathbb{T})} \\
&= \|f\|^p_{L^p(\mathbb{T},|z-1|^{\alpha p})}.
\end{aligned}$$

This finishes the proof of the proposition. □

Now we can answer the duality question by using Proposition 10. Following the same argument as in the proof of Theorem 9, we obtain the following statement.

Theorem 12. *Let $\alpha \geq 0$ and $\mu_\alpha = |z-1|^{2\alpha}$. Then for any $p \in (1, \infty)$, the dual space of the weighted Hardy space $\mathcal{H}^p(\mathbb{T}, |z-1|^{\alpha p})$ can be identified with $\mathcal{H}^q(\mathbb{T}, |z-1|^{\alpha q})$, where $1/p + 1/q = 1$, under the pairing*

$$\langle f, h \rangle = \int_{\mathbb{T}} f(z)\overline{h(z)}|z-1|^{2\alpha} \, d\theta.$$

At first, the two duality results in Theorem 9 and Theorem 12 may seem confusing for $p \in \left(\frac{2\alpha+1}{\alpha+1}, \frac{2\alpha+1}{\alpha}\right)$. However, the main point is to note the difference in the exponents of the weights and the way the pairing is defined. We illustrate these two results in the following example.

Example 13. Let us take $\alpha = 1/2$. Then $\boldsymbol{S}_{|z-1|}$ is bounded on $L^p(\mathbb{T}, |z-1|)$ for $p \in (4/3, 4)$. In particular, for any $p \in (4/3, 4)$, the dual space of $\mathcal{H}^p(\mathbb{T}, |z-1|)$ can be identified with $\mathcal{H}^q(\mathbb{T}, |z-1|)$, where $1/p + 1/q = 1$, under the pairing

$$\langle f, h \rangle_{|z-1|} = \int_{\mathbb{T}} f(z)\overline{h(z)}|z-1| \, d\theta.$$

On the other hand, using the second duality result for any $p > 1$, the dual space of $\mathcal{H}^p(\mathbb{T}, |z-1|)$ can be identified with $\mathcal{H}^q(\mathbb{T}, |z-1|^{q/p})$ when $1/p + 1/q = 1$, under the pairing

$$\langle f, h \rangle_{|z-1|^{2/p}} = \int_{\mathbb{T}} f(z)\overline{h(z)}|z-1|^{2/p} \, d\theta.$$

Acknowledgments

We thank the anonymous referee for constructive comments on the proofs of Proposition 10 and Theorem 12 and also for the useful editorial remarks on the exposition of the article. This study started when the first author visited the second author at Texas A&M University; we thank the Department of Mathematics for the hospitality. The first author also wishes to acknowledge the support from the WKU Office of Research for this visit.

References

[Arroussi and Pau 2014] H. Arroussi and J. Pau, "Reproducing kernel estimates, bounded projections and duality on large weighted Bergman spaces", *J. Geom. Anal.* (online publication July 2014).

[Békollé and Bonami 1995] D. Békollé and A. Bonami, "Estimates for the Bergman and Szegő projections in two symmetric domains of $\mathbf{C}^n$", *Colloq. Math.* **68**:1 (1995), 81–100. MR 95k:32022 Zbl 0863.47018

[Constantin and Peláez 2015] O. Constantin and J. Á. Peláez, "Boundedness of the Bergman projection on L^p-spaces with exponential weights", *Bull. Sci. Math.* **139**:3 (2015), 245–268. MR 3336683 Zbl 06352015

[Duren 1970] P. L. Duren, *Theory of H^p spaces*, Pure and Applied Mathematics **38**, Academic Press, New York, 1970. MR 42 #3552 Zbl 0215.20203

[Garnett 1981] J. B. Garnett, *Bounded analytic functions*, Pure and Applied Mathematics **96**, Academic Press, New York, 1981. MR 83g:30037 Zbl 0469.30024

[Kerzman and Stein 1978] N. Kerzman and E. M. Stein, "The Cauchy kernel, the Szegö kernel, and the Riemann mapping function", *Math. Ann.* **236**:1 (1978), 85–93. MR 58 #6199 Zbl 0419.30012

[Lanzani and Stein 2004] L. Lanzani and E. M. Stein, "Szegö and Bergman projections on non-smooth planar domains", *J. Geom. Anal.* **14**:1 (2004), 63–86. MR 2004m:30063 Zbl 1046.30023

[Lanzani and Stein 2013] L. Lanzani and E. M. Stein, "Cauchy-type integrals in several complex variables", *Bull. Math. Sci.* **3**:2 (2013), 241–285. MR 3084008 Zbl 1277.32002

[Zeytuncu 2013] Y. E. Zeytuncu, "L^p regularity of weighted Bergman projections", *Trans. Amer. Math. Soc.* **365**:6 (2013), 2959–2976. MR 3034455 Zbl 1278.32007

[Zhu 2007] K. Zhu, *Operator theory in function spaces*, 2nd ed., Mathematical Surveys and Monographs **138**, American Mathematical Society, Providence, RI, 2007. MR 2008i:47064 Zbl 1123.47001

Received April 14, 2014. Revised October 29, 2014.

SAMANGI MUNASINGHE
DEPARTMENT OF MATHEMATICS
WESTERN KENTUCKY UNIVERSITY
BOWLING GREEN, KY 42101
UNITED STATES
samangi.munasinghe@wku.edu

YUNUS E. ZEYTUNCU
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MICHIGAN-DEARBORN
DEARBORN, MI 48128
UNITED STATES
zeytuncu@umich.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.459

TOPOLOGY OF COMPLETE FINSLER MANIFOLDS ADMITTING CONVEX FUNCTIONS

SORIN V. SABAU AND KATSUHIRO SHIOHAMA

We investigate the topology of a complete Finsler manifold (M, F) admitting a locally nonconstant convex function.

1. Introduction

Let (M, F) be an n-dimensional Finsler manifold. The well-known Hopf–Rinow theorem (see for example [Bao et al. 2000]) states that M is complete if and only if the exponential map $\exp_p$ at some point $p \in M$ (and hence for every point on M) is defined on the whole tangent space T_pM to M at that point. This is equivalent to saying that (M, F) is geodesically complete with respect to forward geodesics at every point on M. Throughout this article we assume that (M, F) *is geodesically complete with respect to forward geodesics*.

A function $\varphi : (M, F) \to \mathbb{R}$ is said to be *convex* if and only if along every (forward and backward) geodesic $\gamma : [a, b] \to (M, F)$, the restriction $\varphi \circ \gamma : [a, b] \to \mathbb{R}$ is a convex function, that is,

$$\varphi \circ \gamma((1-\lambda)a + \lambda\, b) \leq (1-\lambda)\varphi \circ \gamma(a) + \lambda\varphi \circ \gamma(b), \quad 0 \leq \lambda \leq 1. \tag{1-1}$$

If the inequality in the above relation is strict for all γ and for all $\lambda \in (0, 1)$, then φ is called *strictly convex*. If the second order difference quotient, namely the quantity $\{\varphi \circ \gamma(h) - \varphi \circ \gamma(-h) - 2\varphi \circ \gamma(0)\}/h^2$ is bounded away from zero on every compact set on M along all γ, then φ is called *strongly convex*. In the case where φ is at least C^2, its convexity can be written in terms of the Finslerian Hessian of φ, but we do not need to do this in the present paper.

If $\varphi \circ \gamma$ is a convex function of one variable, then the function $\varphi \circ \bar{\gamma}$ is also convex, where $\bar{\gamma}$ is the reverse curve of γ. We recall that in general, if γ is a Finslerian geodesic, it does not mean that the inverse curve $\bar{\gamma}$ is also a geodesic.

Every noncompact manifold admits a complete (Riemannian or Finslerian) metric and a nontrivial smooth function which is convex with respect to this metric (see [Greene and Shiohama 1981b]).

MSC2010: primary 53C60; secondary 53C22.
Keywords: Finsler manifolds, geodesics, convex functions.

If a nontrivial convex function $\varphi : (M, F) \to \mathbb{R}$ is constant on an open set, then φ assumes its minimum on this open set and the number of components of *a level set* $M_a^a(\varphi) := \varphi^{-1}(\{a\})$, $a \geq \inf_M \varphi$ is equal to that of the boundary components of the minimum set of φ. Here we denote $\inf_M \varphi := \inf\{\varphi(x) \mid x \in M\}$.

A convex function φ is said to be *locally nonconstant* if it is not constant on any open set of M. From now on *we always assume that a convex function is locally nonconstant.*

The purpose of this article is to investigate the topology of complete Finsler manifolds admitting (locally nonconstant) convex functions $\varphi : (M, F) \to \mathbb{R}$. Convex functions on complete Riemannian manifolds have been fully discussed in [Greene and Shiohama 1981b] and elsewhere. Although the distance function on (M, F) is not symmetric and the backward geodesics do not necessarily coincide with the forward geodesics, we prove that most of the Riemannian results in [Greene and Shiohama 1981b] have Finsler extensions, as stated below.

We first discuss the topology of a Finsler manifold (M, F) admitting a convex function φ.

Theorem 1.1 (compare [Greene and Shiohama 1981b, Theorem F]). *Assume we have a convex function $\varphi : (M, F) \to \mathbb{R}$ all of whose level sets are compact.*

(1) *If* $\inf_M \varphi$ *is not attained, there exists a homeomorphism*

$$H : M_a^a(\varphi) \times (\inf_M \varphi, \infty) \to M,$$

for an arbitrary fixed number $a \in (\inf_M \varphi, \infty)$, *such that*

$$\varphi(H(y, t)) = t, \quad \forall y \in M_a^a(\varphi), \quad \forall t \in (\inf_M \varphi, \infty).$$

(2) *If* $\lambda := \inf_M \varphi$ *is attained, then* M *is homeomorphic to the normal bundle over* $M_\lambda^\lambda(\varphi)$ *in* M.

Next, we discuss the case where φ has a disconnected level.

Theorem 1.2 (compare [Greene and Shiohama 1981b, Theorem A]). *Assume the convex function $\varphi : (M, F) \to \mathbb{R}$ has a disconnected level set $M_c^c(\varphi)$ for some $c \in \varphi(M)$.*

(1) *The infimum* $\inf_M \varphi$ *is attained.*

(2) *If* $\lambda := \inf_M \varphi$, *then* $M_\lambda^\lambda(\varphi)$ *is a totally geodesic smooth hypersurface which is totally convex without boundary.*

(3) *The normal bundle of* $M_\lambda^\lambda(\varphi)$ *in* M *is trivial.*

(4) *If* $b > \lambda$, *then the boundary of the* b*-sublevel set* $M^b(\varphi) := \{x \in M \mid \varphi(x) \leq b\}$ *has exactly two components.*

The diameter function $\delta : \varphi(M) \to \mathbb{R}_+$ plays an important role in this article and is defined by

$$\delta(t) := \sup\{d(x, y) \mid x, y \in M_t^t(\varphi)\}. \tag{1-2}$$

Sharafutdinov [1978] had proved earlier the existence of a distance nonincreasing map $M_b^b(\varphi) \to M_a^a(\varphi)$, $b \geq a$, between two compact levels of a convex function φ on a complete Riemannian manifold (M, g).

It is known from [Sharafutdinov 1978] and [Greene and Shiohama 1981b] that the diameter function δ of a complete Riemannian manifold admitting a convex function is monotone nondecreasing. However it is not certain if it is monotone on a Finsler manifold. In Theorem 1.1, we do not use the monotone property but only the local Lipschitz property of δ which is proved in Proposition 3.3.

We finally discuss the number of ends of a Finsler manifold (M, F) admitting a convex function φ. As stated above, the diameter function δ, defined on the image of the convex function φ, may not be monotone. It might occur that a convex function defined on a Finsler manifold (M, F) may simultaneously admit both compact and noncompact levels. This fact makes it difficult to study the number of ends of the manifold (M, F). However, we shall discuss all the possible cases and prove:

Theorem 1.3 (compare [Greene and Shiohama 1981b, Theorems C, D and G]). *Let $\varphi : (M, F) \to \mathbb{R}$ be a convex function.*

(A) *Assume that φ admits a disconnected level.*
  - (A1) *If all the levels of φ are compact, then M has two ends.*
  - (A2) *If all the levels of φ are noncompact, then M has one end.*
  - (A3) *If both compact and noncompact levels of φ exist simultaneously, then M has at least three ends.*

(B) *Assume that all the levels of φ are connected and compact.*
  - (B1) *If $\inf_M \varphi$ is attained, then M has one end.*
  - (B2) *If $\inf_M \varphi$ is not attained, then M has two ends.*

(C) *If all the levels are connected and noncompact, then M has one end.*

(D) *Assume that all the levels of φ are connected and that φ admits both compact and noncompact levels simultaneously. Then we have:*
  - (D1) *If $\inf_M \varphi$ is not attained, then M has two ends.*
  - (D2) *If $\inf_M \varphi$ is attained, then M has at least two ends.*

(E) *Finally, if M has two ends, then all the levels of φ are compact.*

Remark 1.4. The supplementary condition that all of the levels of φ are simultaneously compact or noncompact in the hypothesis of Theorem 1.1 is necessary because we have not proved that the diameter function δ is monotone nondecreasing for a

Finsler manifold. If this property of monotonicity were true, then this assumption could be removed.

We summarize some historical background of convex and related functions on manifolds, G-spaces and Alexandrov spaces. Locally nonconstant convex functions, affine functions and peakless functions have been investigated on complete Riemannian manifolds and complete noncompact Busemann G-spaces and Alexandrov spaces in various ways. The topology of Riemannian manifolds admitting convex functions was investigated in [Bangert 1978; Greene and Shiohama 1981b; 1981a; 1987], and that of Busemann G-surfaces in [Innami 1982a; Mashiko 1999b]. It should be noted that convex functions on complete Alexandrov surfaces are *not continuous*.

A weaker notion than convex functions similar to quasiconvex functions, namely *peakless functions*, has been introduced by Busemann [1955], and studied later on in [Busemann and Phadke 1983] and [Innami 1983]. The topology of complete manifolds admitting locally geodesically (strictly) quasiconvex and uniformly locally convex filtrations have been investigated by Yamaguchi [1986a; 1986b; 1988]. The isometry groups of complete Riemannian manifolds (N, g) admitting strictly convex functions have been discussed in [Yamaguchi 1982] and other places. A well known classical theorem due to Cartan states that every compact isometry group on an Hadamard manifold H has a fixed point. This follows from the simple fact that the distance function to every point on H is strictly convex. Peakless functions and totally geodesic filtrations on complete manifolds have been discussed in [Innami 1983; Busemann and Phadke 1983; Yamaguchi 1986a; 1986b; 1988] and others.

A convex function is said to be *affine* if and only if the equality in (1-1) holds for all γ and for all $\lambda \in (0, 1)$. A splitting theorem for Riemannian manifolds admitting affine functions has been investigated in [Innami 1982b], while Alexandrov spaces admitting affine functions have been studied in [Innami 1982b; Mashiko 1999a; Mashiko 2002]. An overview on the convexity of Riemannian manifolds can be found in [Burago and Zalgaller 1977].

The properties of isometry groups on Finsler manifolds admitting convex functions will be discussed elsewhere. For basic facts on Finsler and Riemannian geometry, we refer to [Bao et al. 2000; Chern et al. 1999; Cheeger and Ebin 2008; Sakai 1992].

2. Fundamental facts

We summarize some fundamental facts on convex sets and convex functions on a Finsler manifold (M, F). Most of these are trivial in the Riemannian case, but we consider it useful to formulate and prove them in the more general Finslerian setting.

Let (M, F) be a complete Finsler manifold. At each point $p \in M$, the indicatrix $\Sigma_p \subset T_pM$ at p is defined by $\Sigma_p := \{u \in T_pM \mid F(p, u) = 1\}$. The *reversibility function* $\lambda : (M, F) \to \mathbb{R}^+$ of (M, F) is given as

$$\lambda(p) := \sup\{F(p, -X) \mid X \in \Sigma_p\}.$$

Clearly, λ is continuous on M and

$$\lambda(p) = \max\left\{\frac{F(p, -X)}{F(p, X)} \mid X \in T_pM \setminus \{0\}\right\}.$$

Let $C \subset M$ be a compact set. There exists a constant $\lambda(C) > 0$ depending on C such that if $p \in C$ and $X \in \Sigma_p$, then

$$\frac{1}{\lambda(C)} F(p, X) \leq F(p, -X) \leq \lambda(C) \cdot F(p, X).$$

In particular, if $\sigma : [0, 1] \to C$ is a smooth curve, then the integral length

$$L(\sigma) := \int_0^1 F(\sigma(t), \dot{\sigma}(t))\, dt$$

of σ satisfies

$$\frac{1}{\lambda(C)} L(\sigma) \leq L(\sigma^{-1}) \leq \lambda(C) \cdot L(\sigma).$$

Here we set $\sigma^{-1}(t) := \sigma(1 - t)$, where $t \in [0, 1]$ is the reverse curve of σ.

It is well known that the topology of (M, F) as an inner metric space is equivalent to that of M as a manifold. For a compact set $C \subset M$, the inner metric d_F of (M, F) induced from the Finslerian fundamental function has the property

$$\frac{1}{\lambda(C)} d_F(p, q) \leq d_F(q, p) \leq \lambda(C) \cdot d_F(p, q), \quad \forall p, q \in C.$$

Let $\mathrm{inj} : (M, F) \to \mathbb{R}_+$ be the *injectivity radius function* of the exponential map. Namely, for a point $p \in M$, $\mathrm{inj}(p)$ is the maximal radius of a ball, centered at the origin of the tangent space T_pM at p, on which $\exp_p$ is injective.

A classical result due to J. H. C. Whitehead [1935] states that there exists a *convexity radius function* $r : (M, F) \to \mathbb{R}$ such that if

$$B(p, r) := \{x \in M \mid d(p, x) < r\}$$

is an r-ball centered at p, then for every $q \in B(p, r(p))$ and for every $r' \in (0, r(p))$, $B(q, r') \subset B(p, r)$ is *strongly convex*. Namely, the distance function to q is strongly convex along every geodesic in $B(q, r')$ with $r' \in (0, r(p))$ if its extension does not pass through q.

A closed set $U \subset M$ is called *locally convex* if and only if $U \cap B(p, r)$ is convex for every $x \in U$ and for some $r \in (0, r(p))$. Notice that this definition is stated only for closed sets, since every open set is obviously locally convex.

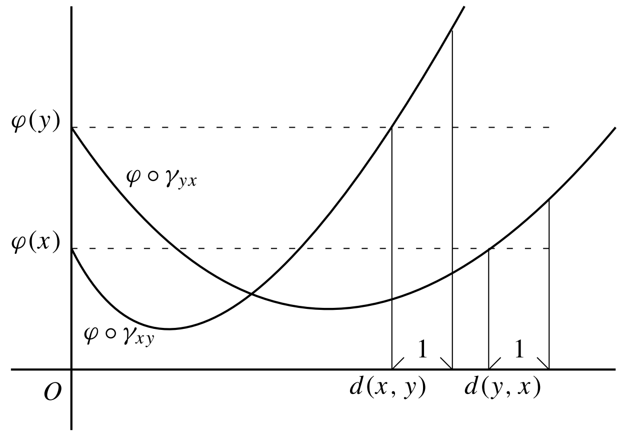

Figure 1. A convex function is locally Lipschitz.

A set $V \subset M$ is called *totally convex* if and only if every geodesic joining two points in V is contained entirely in V. A closed hemisphere in the standard sphere $\mathbb{S}^n$ is locally convex and an open hemisphere is strongly convex, while $\mathbb{S}^n$ itself is the only totally convex set in it. If it exists, the minimum set of a convex function on (M, F) is totally convex.

Proposition 2.1. *A convex function* $\varphi : (M, F) \to \mathbb{R}$ *defined as in* (1-1) *is locally Lipschitz.*

Proof. Let $C \subset M$ be an arbitrary fixed compact set and $C_1 := \{x \in M \mid d(C, x) \leq 1\}$. Here we set $d(C, x) := \min\{d(y, x) \mid y \in C\}$. For points $x, y \in C_1$ we denote by

$$\gamma_{xy} : [0, d(x, y)] \to M, \quad \gamma_{yx} : [0, d(y, x)] \to M$$

minimizing geodesics with

$$\gamma_{xy}(0) = x, \quad \gamma_{xy}(d(x, y)) = y,$$
$$\gamma_{yx}(0) = y, \quad \gamma_{yx}(d(y, x)) = x.$$

The slope inequalities along the two convex functions $\varphi \circ \gamma_{xy}|_{[0,d(x,y)+1]}$ and $\varphi \circ \gamma_{yx}|_{[0,d(y,x)+1]}$ imply that, if $\Lambda := \sup_{C_1} \varphi$ and $\lambda := \inf_{C_1} \varphi$ (see Figure 1), then

$$\frac{\varphi(y) - \varphi(x)}{d(x, y)} \leq \Lambda - \lambda, \quad \frac{\varphi(x) - \varphi(y)}{d(y, x)} \leq \Lambda - \lambda.$$

It follows that there exists a constant $L = L(C) > 0$ such that

$$\sup\left\{\frac{d(x, y)}{d(y, x)} \,\Big|\, x, y \in C\right\} \leq L,$$

and therefore we have

$$\left|\frac{\varphi(x)-\varphi(y)}{d(x,y)}\right|,\ \left|\frac{\varphi(y)-\varphi(x)}{d(y,x)}\right| \le L(\Lambda-\lambda). \qquad \square$$

Proposition 2.2. *If $C \subset (M, F)$ is a closed locally convex set, then there exists a k-dimensional totally geodesic submanifold W of M contained in C, and its closure coincides with C.*

Proof. Let $r : (M, F) \to \mathbb{R}$ be the convexity radius function. For every point $p \in C$ there exists a $k(p)$-dimensional smooth submanifold of M which is contained entirely in C and such that $k(p)$ is the maximal dimension of all such submanifolds in C, where $0 \le k(p) \le n$. At least $\{p\}$ is such a submanifold, with dimension 0.

Let $K \subset M$ be a large compact set containing p and $r(K)$ the convexity radius of K, namely $r(K) := \min\{r(x) \mid x \in K\}$. We also put $k := \max\{k(p) \mid p \in C\}$.

Let $W(p) \subset C$ be a k-dimensional smooth submanifold of M. Suppose that $W(p) \cap B(p; r) \subsetneqq C \cap B(p; r)$ for a sufficiently small $r \in (0, r(K))$. Then there exists a point $q \in B(p; r) \cap (C \setminus W(p))$. Clearly $\dot\gamma_{pq}(0)$ is transversal to $T_pW(p)$, and hence a family of minimizing geodesics

$$\{\gamma_{xq} : [0, d(x, q)] \to B(p; r) \mid x \in W(p) \cap B(p; r)\}$$

with $\gamma_{xq}(0) = x$, $\gamma_{xq}(d(x, q)) = q$ has the property that every $\dot\gamma_{xq}(0)$ is transversal to $T_xW(p)$. Therefore, this family of geodesics forms a $(k+1)$-dimensional submanifold contained in C, a contradiction to the choice of k. This proves $W(p) \cap B(p; r) = C \cap B(p; r)$ for a sufficiently small $r \in (0, r(K))$. We then observe that $\bigcup_{p\in C} W(p) =: W \subset C$ forms a k-dimensional smooth submanifold which is totally geodesic. Indeed, for any tangent vector v to W, there exists $p \in C$ such that $v \in T_pW(p)$, and due to the convexity of C, the geodesic $\gamma_v : [0, \varepsilon] \to M$ cannot leave the submanifold W.

We finally prove that the closure $\overline{W}$ of W coincides with C. Indeed, suppose that there exists a point $x \in C \setminus \overline{W}$. We then find a point $y \in \overline{W} \setminus W$ such that $d(x, y) = d(x, \overline{W}) < r(K)$. If

$$\dot\gamma_{xy}(d(x, y)) \in T_y\overline{W} := \lim_{y_j \to y} T_{y_j}M,$$

then $\gamma_{xy}(d(x, y) + \varepsilon) \in W$ for a sufficiently small $\varepsilon > 0$. Let $U \subset W \cap B(x, r(K))$ be an open set around $\gamma_{xy}(d(x, y) + \varepsilon)$.

Then a family of geodesics

$$\{\gamma_{xz} : [0, d(x, z)] \to B(x; r(K)) \mid z \in U\} \tag{2-1}$$

forms a k-dimensional submanifold contained in W and hence $y \in W$, a contradiction to $y \in \overline{W} \setminus W$. Therefore, $\dot\gamma_{xy}(d(x, y))$ does not belong to $T_y\overline{W}$, and (2-1) again

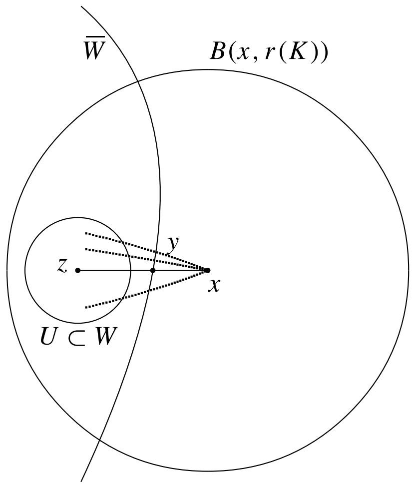

Figure 2. The closure $\overline{W}$ of W coincides with C.

forms a $(k+1)$-dimensional submanifold in C, a contradiction to the choice of k (see Figure 2). □

Let $C \subset M$ be a closed locally convex set and $p \in C$. There exists a totally geodesic submanifold $W \subset C$ as stated in Proposition 2.2. We call W the *interior of C* and denote it by $\mathrm{Int}(C)$. The *boundary of C* is defined by $\partial C := C \setminus \mathrm{Int}(C)$, and *the dimension of C* is defined by $\dim C := \dim \mathrm{Int}(C)$. The *tangent cone* $\mathscr{C}_p(C) \subset T_pM$ of C at a point $p \in C$ is defined by

(2-2) $$\mathscr{C}_p(C) := \{\xi \in T_pM \mid \exp_p t\xi \in \mathrm{Int}(C) \text{ for some } t > 0\}.$$

Clearly, $\mathscr{C}_p(C) = T_p \mathrm{Int}(C) \setminus \{0\}$ for $p \in \mathrm{Int}(C)$.

We also define the tangent space T_pC of C at a point $p \in \partial C$ as $\lim_{q\to p} T_q \mathrm{Int}(C)$. We claim that there exists for every point $p \in \partial C$ an open half space $T_pC_+ \subset T_pC$ containing $\mathscr{C}_p(C)$:

(2-3) $$\mathscr{C}_p(C) \subset T_pC_+ \subset T_pC := \lim_{q\to p} T_q \mathrm{Int}(C),\ q \in \mathrm{Int}(C).$$

Indeed, for any points $p \in \partial C$ and $q \in B(p; r(K)) \cap \mathrm{Int}(C)$, consider a minimizing geodesic $\gamma_{qp} : [0, d(q, p)] \to B(p; r(K))$. Suppose that there is a point $q \in \mathrm{Int}(C)$ such that $z := \gamma_{qp}(d(q, p) + \varepsilon) \in C$ for a sufficiently small $\varepsilon > 0$. We then have $\dot\gamma_{qp}(d(q, p)) \in T_pC$, and hence the tangent cone $\mathscr{C}_p(C)$ as obtained in (2-2) is contained entirely in T_pC, a contradiction to the choice of $p \in \partial C$. From the above argument we observe that if $p \in \partial C$, then there exists a hyperplane $H_p \subset T_pC$ such that $\mathscr{C}_p(C)$ is contained in a half space $T_p(C)_+ \subset T_pC$ bounded by H_p.

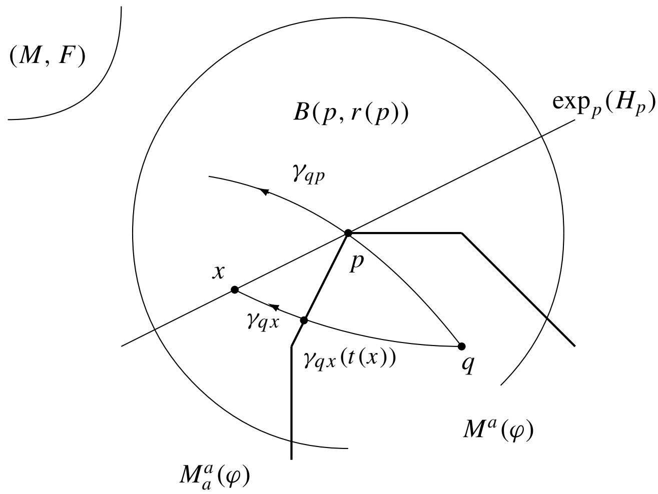

Figure 3. An atlas of local charts at an arbitrary point $p \in M_a^a(\varphi)$.

Proposition 2.3. *Let $\varphi : (M, F) \to \mathbb{R}$ be a convex function. Then, $M_a^a(\varphi)$ is an embedded topological submanifold of dimension $n-1$ for every $a > \inf_M \varphi$.*

Proof. Let $p \in M_a^a(\varphi)$ and $q \in B(p; r(p)) \cap \mathrm{Int}(M^a(\varphi))$. There exists a hyperplane $H_p \subset T_p M$ such that

$$H_p = \partial T_p(M^a(\varphi))_+ \quad \text{and} \quad \mathscr{C}_p(M^a(\varphi)) \subset T_p(M^a(\varphi))_+.$$

Every point $x \in \exp_p(H_p) \cap B(p; r(p))$ is joined to q by a unique minimizing geodesic $\gamma_{qx} : [0, d(q,x)] \to M$ such that $\gamma_{qx}(0) = q$, $\gamma_{qx}(d(q,x)) = x$. Then there exists a unique parameter $t(x) \in (0, d(q,x)]$ such that $M_a^a(\varphi) \cap B(p; r(p))$ contains $\gamma_{qx}(t(x))$. Let $B_H(O; r(p))$ be the open $r(p)$-ball in H_p centered at the origin O of M_p. We then have a map $\alpha_p : B_H(O; r(p)) \to M_a^a(\varphi)$ such that

$$\alpha_p(u) := \gamma_{qx}(t(x)), \quad u \in B_H(O; r(p)), \quad \exp_p u = x.$$

Clearly, α_p gives a homeomorphism between $B_H(O; r(p))$ and its image in $M_a^a(\varphi)$. Thus the family of maps $\{(B_H(O; r(p)), \alpha_p) \mid p \in M_a^a(\varphi)\}$ forms an atlas of $M_a^a(\varphi)$ (see Figure 3). □

3. Level sets configuration

We shall give the proofs of Theorems 1.2 and 1.3. The following lemma is elementary and useful for our discussion.

Lemma 3.1. *Let $\varphi : (M, F) \to \mathbb{R}$ be a convex function. If $M_a^a(\varphi)$ is compact, then so is $M_b^b(\varphi)$ for all $b \geq a$. If $M_a^a(\varphi)$ is noncompact, then so is $M_b^b(\varphi)$ for all $b \leq a$.*

Proof. First of all we prove that if $M_a^a(\varphi)$ is compact, then so is $M_b^b(\varphi)$ for all $b \geq a$.

Suppose that $M_b^b(\varphi)$ is noncompact for some $b > a$. Take a point $p \in M_a^a(\varphi)$ and a divergent sequence $\{q_j\}_{j\geq 1}$ on $M_b^b(\varphi)$. Since $M_a^a(\varphi)$ is compact, there is a positive number L such that $d(p, x) < L$ for all $x \in M_a^a(\varphi)$. Let $\gamma_j : [0, d(p, q_j)] \to M$ be a minimizing geodesic with $\gamma_j(0) = p$, $\gamma_j(d(p, q_j)) = q$ for $j \geq 1$. Compactness of $M_a^a(\varphi)$ implies that each $\varphi \circ \gamma_j|_{[L, d(p,q_j)-L]}$ is monotone and nondecreasing for all large numbers j.

Choosing a subsequence $\{\gamma_i\}$ of $\{\gamma_j\}$ if necessary, we find a ray $\gamma_\infty : [0, \infty) \to M$ emanating from p such that $\varphi \circ \gamma_\infty$ is monotone, nondecreasing and bounded above, and hence is identically equal to a. This contradicts the assumption that $M_a^a(\varphi)$ is compact. □

The following Proposition 3.2 is the basic piece in the proof of Theorem 1.1. Under the assumptions in Theorem 1.1, we divide M into countable compact sets such that

$$M = \bigcup_{j=-\infty}^{\infty} \varphi^{-1}[t_{j-1}, t_j],$$

where $\{t_j\}$ is monotone increasing and $\lim_{j\to-\infty} t_j = \inf_M \varphi$ (if $\inf_M \varphi$ is not attained) and $\lim_{j\to\infty} t_j = \infty$. In applying Proposition 3.2 to each $\varphi^{-1}[t_{j-1}, t_j]$, the undefined numbers b_{k+1} and b_0 appearing in the proof of the proposition play the role of margins to be pasted with $\varphi^{-1}[t_j, t_{j+1}]$ (using b_{k+1}) and with $\varphi^{-1}[t_{j-2}, t_{j-1}]$ (using b_{-1}), respectively.

Proposition 3.2. *Let $M_a^a(\varphi) \subset M$ be a connected and compact level set and $b > a$ a fixed value. Then there exists a homeomorphism $\Phi_a^b : M_b^b(\varphi) \times [a, b] \to M_a^b(\varphi)$ such that*

$$\varphi \circ \Phi_a^b(x, t) = t, \quad (x, t) \in M_b^b(\varphi) \times [a, b]. \tag{3-1}$$

Proof. Let $K \subset M$ be a compact set with $M_a^b(\varphi) \subset \mathrm{Int}(K)$ and $r := r(K)$ the convexity radius over K. We define two divisions as follows. Let $a = a_0 < a_1 < \cdots < a_k = b$ and $b_{-1} < b_0 < \cdots < b_k$ be given such that $\varphi^{-1}[b_{-1}, b_k] \subset \mathrm{Int}(K)$ and

(1) $b_{-1} < a_0 < b_0 < a_1 < \cdots < a_{k-1} < b_{k-1} < a_k = b < b_k$,

(2) $b_j := \frac{1}{2}(a_j + a_{j+1})$, $j = 0, 1, \ldots, k-1$,

(3) $\varphi^{-1}(\{a_{j-1}\}) \subset \bigcup\{B(x, r) \mid x \in \varphi^{-1}(\{a_{j+1}\})\}$, $j = 1, \ldots, k-1$,

(4) $\varphi^{-1}(\{b_{-1}\}) \subset \bigcup\{B(y, r) \mid y \in \varphi^{-1}(\{a_1\})\}$,

(5) $\varphi^{-1}(\{a_{k-1}\}) \subset \bigcup\{B(z, r) \mid z \in \varphi^{-1}(\{b_k\})\}$.

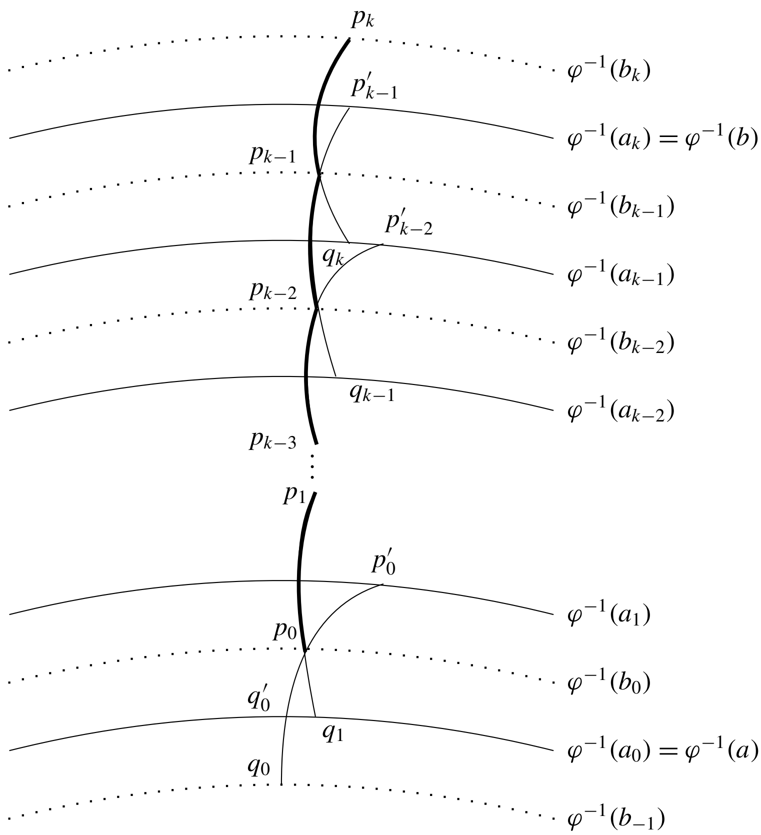

Figure 4. The broken geodesic $T(p_k)$.

Obviously we have $[a, b] \subset (b_{-1}, b_k)$.

For an arbitrary fixed point $p'_j \in \varphi^{-1}(\{a_{j+1}\})$, we have a minimizing geodesic $T(p'_j, q_j)$ realizing the distance $d(p'_j, \varphi^{-1}(-\infty, a_{j-1}])$ and q_j the foot of p'_j on $\varphi^{-1}(-\infty, a_{j-1}]$. Then the family of all such minimizing geodesics emanating from all the points on $\varphi^{-1}(\{a_{j+1}\})$ to the points on $\varphi^{-1}(\{a_{j-1}\})$ simply covers the set $\varphi^{-1}[b_{j-1}, b_j]$, $j = 1, 2, \dots, k$. We define $p_j := T(p'_j, q_j) \cap \varphi^{-1}(\{b_j\})$ and $p_{j-1} := T(p'_j, q_j) \cap \varphi^{-1}(\{b_{j-1}\})$. With this point p_{j-1}, we then choose $p'_{j-1} \in \varphi^{-1}(\{a_j\})$ and $q_{j-1} \in \varphi^{-1}(\{a_{j-2}\})$ in such a way that $T(p'_{j-1}, q_{j-1})$ realizes the distance $d(p'_{j-1}, q_{j-1}) = d(p_{j-1}, \varphi^{-1}(-\infty, a_{j-2}])$ and contains p_{j-1} in its interior. We thus obtain the inductive construction of a sequence $\{T(p'_j, q_j) \mid j = 1, \dots, k\}$ of minimizing geodesics.

We finally choose a point $p'_0 \in \varphi^{-1}(\{a_1\})$ and $q_0 \in \varphi^{-1}(\{b_{-1}\})$ such that $T(p'_0, q_0)$ is a unique minimizing geodesic, with q_0 being the foot of p'_0 on $\varphi^{-1}(-\infty, b_{-1}]$. If we set $q'_0 := \varphi^{-1}(\{a_0\}) \cap T(p'_0, q_0)$, then $d(p_0, q_1) \leq d(p_0, q'_0)$ follows from the fact that q_1 is the foot of p_0 on $\varphi^{-1}(-\infty, a_0]$. Therefore the slope inequality along

$T(p_0', q_0)$ implies

$$\frac{a_0 - b_0}{d(p_0, q_1)} \le \frac{a_0 - b_0}{d(p_0, q_0')} \le \frac{b_{-1} - a_0}{d(q_0', q_0)},$$

and hence there exists a positive number

$$\Delta_a^b(K) := \min\Big\{ \frac{a - b_{-1}}{d(q_0', \varphi^{-1}(-\infty, b_{-1}])} \mid q_0' \in \varphi^{-1}(\{a\}) \Big\}$$

with the property that all the slopes of $\varphi \circ T(p_j', q_j)$ for every $j = 0, 1, \dots, k$ are negative and bounded above by $-\Delta_a^b(K)$.

Next, we define a broken geodesic $T(p_k) := T(p_k, p_{k-1}) \cup \cdots \cup T(p_1, p_0)$ for $p_k \in \varphi^{-1}(\{b_k\})$ with its break points at $p_j \in \varphi^{-1}(\{b_j\})$, $j = 0, 1, \dots, k-1$ in such a way that each $T(p_j, p_{j-1})$ is a proper subarc of a unique minimizing geodesic $T(p_j', q_j)$, where q_j is the foot of p_j' on $\varphi^{-1}(-\infty, a_{j-1}]$ (see Figure 4). Then $T(p_{j-1}, p_{j-2})$ is a proper subarc of $T(p_{j-1}', q_{j-1})$. Clearly, the convex function along $T(p_k)$ is monotone strictly decreasing, since the slopes along $\varphi \circ T(p_k)$ are all bounded above by $-\Delta_a^b(K)$. We then observe from the construction that the family of all the broken geodesics emanating from all points on $\varphi^{-1}(\{b_k\})$ and ending at points on $\varphi^{-1}(\{b_{-1}\})$ simply covers $\varphi^{-1}[a, b]$. The desired homeomorphism Φ_a^b is now obtained by defining $\Phi_a^b(x, t)$ as the intersection of a $T(x)$ emanating from x:

$$\Phi_a^b(x, t) = T(x) \cap \varphi^{-1}(\{t\}). \qquad \square$$

Proposition 3.3. *Assume that all the levels of φ are compact. Then the diameter function $\delta : \varphi(M) \to \mathbb{R}$ defined by*

$$\delta(a) := \sup\{d(x, y) \mid x, y \in \varphi^{-1}(\{a\}),\ a \in \varphi(M)\}$$

is locally Lipschitz.

Proof. Let $\inf_M \varphi < a < b < \infty$, and let $r = r(M_a^b(\varphi))$ be the convexity radius over $M_a^b(\varphi)$. Let $x, y \in \varphi^{-1}(\{s\})$ for $s \in [a, b)$ be such that $d(x, y) = \delta(s)$. Proposition 3.2 then implies that there are points $x', y' \in M_b^b(\varphi)$ such that $\Phi_a^b(x', s) = x$ and $\Phi_a^b(y', s) = y$. Moreover, we have $\Phi_a^b(x', t) = T(x') \cap M_t^t(\varphi)$ and $\Phi_a^b(y', t) = T(y') \cap M_t^t(\varphi)$, and the length $L(T(p)|_{[s,t]})$ of $T(p)|_{[s,t]}$, for $p \in M_b^b(\varphi)$ and for every $a \le s < t \le b$, is bounded above by

$$L(T(p)|_{[s,t]}) \le |t - s| / \Delta_a^b(M_a^b(\varphi)). \tag{3-2}$$

Therefore, by setting $\lambda = \lambda(M_a^b(\varphi))$ the reversibility constant on $M_a^b(\varphi)$, we have

$$\begin{aligned}\delta(s) = d(x, y) &\le d(x, \Phi_a^b(x', t)) + d(\Phi_a^b(x', t), \Phi_a^b(y', t)) + d(\Phi_a^b(y', t), y)\\ &\le \lambda|t-s|/\Delta_a^b(M_a^b(\varphi)) + \delta(t) + |t-s|/\Delta_a^b(M_a^b(\varphi))\\ &= (1+\lambda)|t-s|/\Delta_a^b(M_a^b(\varphi)) + \delta(t).\end{aligned}$$

Similarly, by choosing $x, y \in \varphi^{-1}(\{t\})$, $d(x, y) = \delta(t)$, we obtain

$$\begin{aligned}\delta(t) = d(x, y) &\le d(x, \Phi_a^b(x', s)) + d(\Phi_a^b(x', s), \Phi_a^b(y', s)) + d(\Phi_a^b(y', s), y)\\ &\le (1+\lambda)|t-s|/\Delta_a^b(M_a^b(\varphi)) + \delta(s),\end{aligned}$$

and hence,

$$|\delta(t) - \delta(s)| \le (1+\lambda)|t-s|/\Delta_a^b(M_a^b(\varphi)). \qquad \square$$

Proof of Theorem 1.1. We first assume that $\inf_M \varphi$ is not attained. Let $\{a_j\}_{j\in\mathbb{Z}}$ be a monotone increasing sequence of real numbers with $\lim_{j\to-\infty} a_j = \inf_M \varphi$ and $\lim_{j\to\infty} a_j = \infty$. We then apply Proposition 3.2 to each integer j and obtain a homeomorphism $\Phi_j^{j+1} : \varphi^{-1}(\{a_{j+1}\}) \times (a_j, a_{j+1}] \to M_{a_j}^{a_{j+1}}$ such that

$$\varphi \circ \Phi_j^{j+1}(x, t) = t, \quad x \in \varphi^{-1}(\{a_{j+1}\}), \quad t \in (a_j, a_{j+1}].$$

The composition of these homeomorphisms gives the desired homeomorphism $\varphi : \varphi^{-1}(\{a\}) \times (\inf_M \varphi, \infty) \to M$.

If $\lambda := \inf_M \varphi$ is attained, then $M_\lambda^\lambda(\varphi)$ is a k-dimensional totally geodesic submanifold which is totally convex with $0 \le k \le \dim M - 1$. A tubular neighborhood $B\big(M_\lambda^\lambda(\varphi), r(M_\lambda^\lambda(\varphi))\big)$ around the minimum set is a normal bundle over $M_\lambda^\lambda(\varphi)$ in M and its boundary $\partial B\big(M_\lambda^\lambda(\varphi), r(M_\lambda^\lambda(\varphi))\big)$ is homeomorphic to a level of φ. Therefore M is homeomorphic to the normal bundle over the minimum set in M. This proves Theorem 1.1. $\square$

Remark 3.4. Under the assumption in Theorem 1.1, it is not certain whether or not $\lim_{t\to\inf_M \varphi} \delta(t) = \infty$. It might happen that every level set above the infimum is compact but the minimum set is noncompact. However, we do not know such an example on a Finsler manifold.

Remark 3.5. The basic difference of treatments of convex functions between Riemannian and Finsler geometry can be interpreted as follows.

In the case where $\varphi : (M, g) \to \mathbb{R}$ is a convex function with noncompact levels, on a Riemannian manifold, the homeomorphism $\Phi_a^b : M_b^b(\varphi) \times [a, b] \to M_a^b(\varphi)$ is obtained as follows. Fix a point $p \in M_a^a(\varphi)$ and a sequence of R_j-balls centered at p, $\{B(p, R_j)\}_{j\ge1}$ with $\lim_{j\to\infty} R_j = \infty$. Setting K_j to be the closure of $B(p, R_j)$, for $j \ge 1$, we find a sequence of constants $\Delta_j := \Delta_a^b(K_j)$. If

$x \in K_j \cap M_b^b(\varphi)$ is a fixed point, we then have a broken geodesic $T(x) := T(x_k, x_{k-1}) \cup \cdots \cup T(x_1, x_0)$ as obtained in the proof of Proposition 3.2, where $x_0 \in M_a^a(\varphi)$. The properties of the Riemannian distance function now apply to $T(x_j, x_{j-1}) : [0, d(x_j, x_{j-1})] \to (M, g)$. Consequently, the distance function from $p \in M_a^a(\varphi)$, namely $t \mapsto d(p, T(x_j, x_{j-1}))(t)$, is strictly monotone decreasing. Here $T(x_j, x_{j-1})$ is parameterized by arc-length such that $T(x_j, x_{j-1})(0) = x_j$ and $T(x_j, x_{j-1})(d(x_j, x_{j-1})) = x_{j-1}$. Therefore, $T(x)$ is contained entirely in K_j, and moreover, the length $L(T(x))$ of $T(x)$ satisfies

$$L(T(x)) \leq (b-a)/\Delta_j, \quad \forall x \in K_j \cap M_a^b(\varphi).$$

If $y_0 \in M_a^a(\varphi) \cap K_j$ is an arbitrary fixed point, then Proposition 3.2 again implies that there exists a point $y = y_m \in M_b^b(\varphi)$ such that $T(y) = T(y_k, y_{k-1}) \cup \cdots \cup T(y_1, y_0)$ has length at most $(b-a)/\Delta_j$. Hence we have

$$d(p, y) < R_j + (b-a)/\Delta_j + 1.$$

We therefore observe that the correspondence $x \mapsto x_0$ between $M_b^b(\varphi)$ and $M_a^a(\varphi)$ through $T(x)$ is bijective, and the desired homeomorphism is constructed.

However, in the Finslerian case where all the levels of a convex function $\varphi : (M, F) \to \mathbb{R}$ are noncompact, the correspondence $x \mapsto x_0$ between $M_b^b(\varphi)$ and $M_a^a(\varphi)$ through $T(x) = T(x_k, x_{k-1}) \cup \cdots \cup T(x_1, x_0)$ may not be obtained. In fact, the monotone decreasing property of $t \mapsto d(p, T(x_j, x_{j-1}))$ might not hold for a Finsler metric. Therefore, for a point $x \in K_j \cap M_b^b(\varphi)$, $T(x)$ may not necessarily be contained in K_j. Hence, we may fail in controlling the length of $T(x)$ in terms of Δ_j. By the same reason, we cannot prove the monotone nondecreasing property of the diameter function for compact levels of a convex function $\varphi : (M, F) \to \mathbb{R}$.

4. Proof of Theorem 1.2

We take a minimizing geodesic $\sigma : [0, \ell] \to M$ such that $\sigma(0)$ and $\sigma(\ell)$ belong to distinct components of $M_c^c(\varphi)$.

For the proof of (1), we assert that $\inf_M \varphi = \inf_{0 \leq t \leq \ell} \varphi \circ \sigma(t)$. Suppose that $b := \inf_{0 \leq t \leq \ell} \varphi \circ \sigma(t) > \inf_M \varphi$. Since φ is locally nonconstant, we may assume without loss of generality that $b := \inf_{0 \leq t \leq \ell} \varphi \circ \sigma(t)$ is attained at a unique point, say, $q = \sigma(\ell_0)$.

Setting $r = r(\sigma(\ell_0))$, we find a number $a \in (\inf_M \varphi, b)$ such that there is a unique foot $p \in M_a^a(\varphi)$ of q on $M_a^a(\varphi)$, namely $d(\sigma(\ell_0), M_a^a(\varphi)) = d(\sigma(\ell_0), p)$.

Let $\alpha : [0, d(q, p)] \to M$ be the unique minimizing geodesic with $\alpha(0) = q$, $\alpha(d(q, p)) = p$. The points on $\alpha(t)$, for $0 \leq t \leq d(q, p)$, can be joined to $q_\pm := \sigma(\ell_0 \pm r)$ by a unique minimizing geodesic $\gamma_{\alpha(t)q_\pm} : [0, d(\alpha(t), q_\pm)] \to B(q; r)$ with $\gamma_{\alpha(t)q_\pm}(0) = \alpha(t)$, $\gamma_{\alpha(t)q_\pm}(d(\alpha(t)), q_\pm) = q_\pm$.

Since $\varphi(q_\pm) > b$, the right-hand derivative of $\varphi \circ \gamma_{\alpha(t)q_\pm}$ at $d(\alpha(t), q_\pm)$ is bounded below by

$$(\varphi \circ \gamma_{\alpha(t)q_\pm})'_+(\varphi \circ \gamma_{\alpha(t)q_\pm}(d(\alpha(t)), q_\pm)) > \frac{\varphi(q_\pm) - b}{2r} > 0.$$

Thus, for every $t \in [0, d(q, p)]$, $\gamma_{\alpha(t)q_\pm}$ meets $M_c^c(\varphi)$ at $\gamma_{\alpha(t)q_\pm}(u^\pm(t))$ with

$$u^\pm(t) \leq \frac{2r(c-a)}{\varphi(q_\pm) - b} + 2r,$$

and hence there are curves $C_0^\pm : [0, d(q, p)] \to M_c^c(\varphi)$ with

$$C_0^+(0) = \sigma(\ell), \quad C_0^-(0) = \sigma(0),$$
$$C_0^+(d(q, p)) = \gamma_{pq_+}(u^+(d(q, p))), \quad C_0^-(d(q, p)) = \gamma_{pq_-}(u^-(d(q, p))).$$

Let $\tau_t : [0, d(p, \sigma(t))] \to M$ for $t \in [\ell_0 - r, \ell_0 + r]$ be a minimizing geodesic with $\tau_t(0) = p$, $\tau_t(d(p, \sigma(t))) = \sigma(t)$. Every τ_t meets $M_c^c(\varphi)$ at a parameter value $\leq 2rc/(b-a)$, and hence we have a curve $C_1 : [\ell_0 - r, \ell_0 + r] \to M_c^c(\varphi)$ such that

$$C_1(t) = \tau_t[0, 2rc/(b-a)] \cap M_c^c(\varphi).$$

Thus, considering the union $C_0^- \cup C_1 \cup (C_0^+)^{-1}$, it follows that $\sigma(0)$ can be joined to $\sigma(\ell)$ in $M_c^c(\varphi)$, a contradiction. This proves (1) (see Figure 5).

We next prove (2). Let $\lambda := \inf_M \varphi$. Clearly $M_\lambda^\lambda(\varphi)$ is totally convex, and hence Proposition 2.2 implies that $M_\lambda^\lambda(\varphi)$ carries the structure of a smooth totally geodesic submanifold.

Suppose that $\dim M_\lambda^\lambda(\varphi) < n-1$. Then the normal bundle is connected, and at each point $p \in M_\lambda^\lambda(\varphi)$ the indicatrix $\Sigma_p \subset T_pM$ has the property that $\Sigma_p \setminus \Sigma_p(M_\lambda^\lambda \varphi))$ is arcwise connected. Here, $\Sigma_p(M_\lambda^\lambda(\varphi)) \subset \Sigma_p$ is the indicatrix at p of $M_\lambda^\lambda(\varphi)$. Choose points q_0 and q_1 on distinct components of $M_c^c(\varphi)$, and an interior point $p \in M_\lambda^\lambda(\varphi)$. If $\gamma_i : [0, d(p, q_i)] \to M$ for $i = 0, 1$ is a minimizing geodesic with $\gamma_i(0) = p$, $\gamma_i(d(p, q_i)) = q_i$, then $\dot\gamma_0(0)$ and $\dot\gamma_1(0)$ are joined by a curve $\Gamma : [0, 1] \to \Sigma_p \setminus \Sigma_p(M_\lambda^\lambda(\varphi))$ such that $\Gamma(0) = \dot\gamma_0(0)$, $\Gamma(1) = \dot\gamma_1(0)$. The same method as developed in the proof of (1) yields a continuous 1-parameter family of geodesics $\gamma_t : [0, \ell_t] \to M$ with $\gamma_t(0) = p$, $\dot\gamma_t(0) = \Gamma(t)$ and $\gamma_t(\ell_t) \in M_c^c(\varphi)$ for all $t \in [0, 1]$. Thus we have a curve $t \mapsto \gamma_t(\ell_t)$ in $M_c^c(\varphi)$ joining q_0 to q_1, a contradiction. This proves $\dim M_\lambda^\lambda(\varphi) = n-1$.

We use the same idea to prove that $M_\lambda^\lambda(\varphi)$ has no boundary. In fact, supposing that the boundary is nonempty, the tangent cone of $M_\lambda^\lambda(\varphi)$ at a boundary point x is contained entirely in a closed half space of $T_xM_\lambda^\lambda(\varphi)$, and hence $\Sigma_x \setminus \Sigma_x(M_\lambda^\lambda(\varphi))$ is arcwise connected. A contradiction is derived by constructing a curve in $M_c^c(\varphi)$ joining q_0 to q_1. This proves (2).

The triviality of the normal bundle over $M_\lambda^\lambda(\varphi)$ in M is now clear, giving (3).

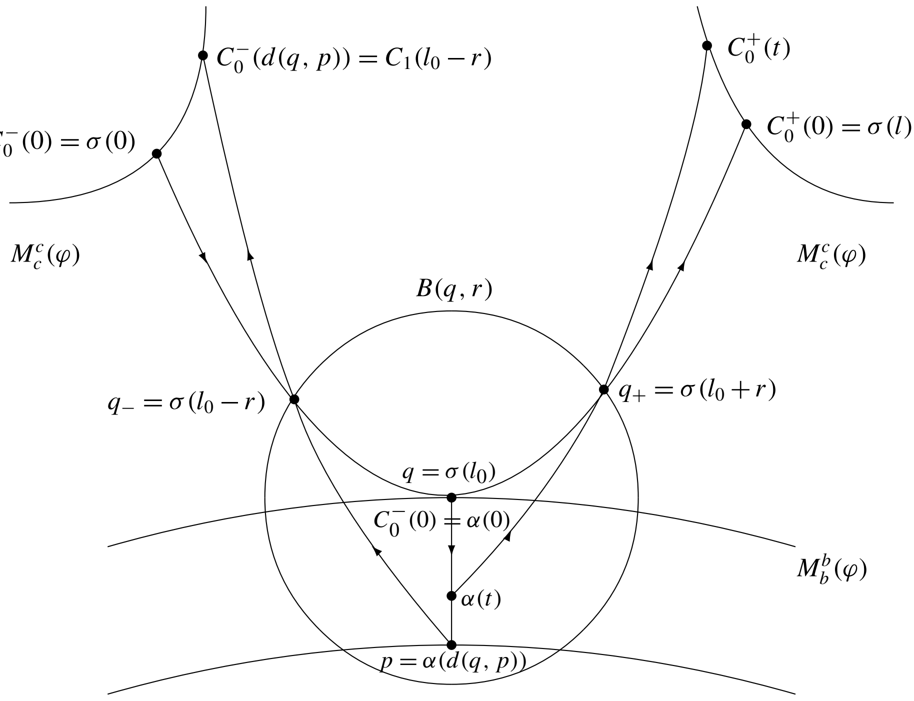

Figure 5. The proof of Theorem 1.2.

To prove (4), suppose that $M_a^a(\varphi)$ for some $a \in \varphi(M)$ has more than two components. Let $q_1, q_2, q_3 \in M_a^a(\varphi)$ be in distinct components, and take $p \in M_\lambda^\lambda(\varphi)$. Let $\gamma_i : [0, d(p, q_i)] \to M$ for $i = 1, 2, 3$ be minimizing geodesics with $\gamma_i(0) = p$, $\gamma_i(d(p, q_i)) = q_i$. Since the normal bundle over $M_\lambda^\lambda(\varphi)$ in M is trivial by (3), it follows that $\Sigma_p \setminus \Sigma_p(M_\lambda^\lambda(\varphi))$ has exactly two components. Two of the three initial vectors, say $\dot\gamma_1(0)$ and $\dot\gamma_2(0)$, belong to the same component of $\Sigma_p \setminus \Sigma_p(M_\lambda^\lambda(\varphi))$. Then the same technique as developed in the proof that $\dim M_\lambda^\lambda(\varphi) = n - 1$ applies, and q_1 is joined to q_2 by a curve in $M_a^a(\varphi)$. This contradiction proves (4). $\square$

5. Ends of (M, F)

An *end* ε of a noncompact manifold X is an assignment to each compact set $K \subset X$ a component $\varepsilon(K)$ of $X \setminus K$ such that $\varepsilon(K_1) \supset \varepsilon(K_2)$ if $K_1 \subset K_2$. Every noncompact manifold has at least one end. For instance, $\mathbb{R}^n$ has one end if $n > 1$ and two ends if $n = 1$.

In the present section we discuss the number of ends of (M, F) admitting a convex function, namely we will prove Theorem 1.3. As is seen in the previous section, it may happen that a convex function $\varphi : (M, F) \to \mathbb{R}$ has both compact

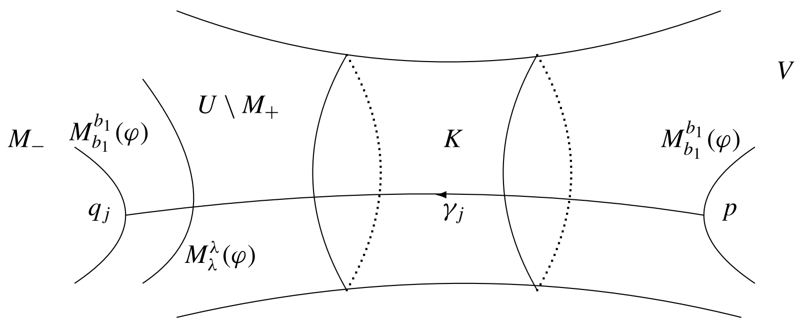

Figure 6. The proof of Theorem 1.3 (A2).

and noncompact levels simultaneously. In this section let $\{K_j\}_{j\geq 1}$ be an increasing sequence of compact sets such that $\lim_{j\to\infty} K_j = M$.

Proof of Theorem 1.3. We first prove (A1).

Theorem 1.2 (1) implies that φ attains its infimum $\lambda := \inf_M \varphi$. Given an arbitrary compact set $A \subset M$, there exists a number $a \in \varphi(M)$ such that $M_a^a(\varphi)$ has two components and $A \subset \varphi^{-1}[\lambda, a]$. Then $M \setminus A$ contains two unbounded open sets $\varphi^{-1}(a, \infty)$, proving (A1).

We next prove (A2). Suppose that M has more than one end. There is a compact set $K \subset M$ such that $M \setminus K$ has at least two unbounded components, say U and V. Setting $a := \min_K \varphi$ and $b := \max_K \varphi$, we have

$$\lambda \leq a < b < \infty.$$

We assert that

$$M_\lambda^\lambda(\varphi) \cap U \neq \varnothing, \quad M_\lambda^\lambda(\varphi) \cap V \neq \varnothing, \quad M_\lambda^\lambda(\varphi) \cap K \neq \varnothing.$$

In order to prove that $M_\lambda^\lambda(\varphi) \cap K \neq \varnothing$, we suppose that $\lambda < a$. Once $M_\lambda^\lambda(\varphi) \cap K \neq \varnothing$ has been established, it will turn out that $M_\lambda^\lambda(\varphi)$ intersects all the unbounded components of $M \setminus K$.

Suppose the contrary, namely $M_\lambda^\lambda(\varphi) \cap K = \varnothing$. Without loss of generality we may assume $M_\lambda^\lambda(\varphi) \subset U$. From Theorem 1.2 (3) it follows that $M \setminus M_\lambda^\lambda(\varphi) = M_- \cup M_+$ (a disjoint union with $\partial M_+ = \partial M_- = M_\lambda^\lambda(\varphi)$).

Setting $M_- \subset U$, we observe that $K \cup V \subset M_+$.

If $b_1 > b$, then M_- contains a component of $M_{b_1}^{b_1}(\varphi)$ and another component of $M_{b_1}^{b_1}(\varphi)$ is contained entirely in V. We then observe that if $\sup_{U\setminus M_-} \varphi = \infty$, then $U \setminus M_-$ contains a component of $M_{b_1}^{b_1}(\varphi)$, for φ takes values $\leq b$ on $\partial(U \setminus M_-)$ and $M_{b_1}^{b_1}(\varphi)$ does not meet the boundary of $U \setminus M_-$. This contradicts Theorem 1.2 (4), for $\partial M_{b_1}^{b_1}(\varphi)$ has at least three components. Therefore we have $\sup_{U\setminus M_-} \varphi < \infty$.

Let $\{q_j\} \subset M_{b_1}^{b_1}(\varphi)$ be a divergent sequence of points, and fix $p \in M_{b_1}^{b_1}(\varphi) \subset V$. Let $\gamma_j : [0, d(p, q_j)] \to M \setminus M_-$ be a minimizing geodesic with $\gamma_j(0) = p$, $\gamma_j(d(p, q_j)) = q_j$ for $j = 1, 2, \ldots$. Clearly γ_j passes through a point on K and $\varphi \circ \gamma_j$ is bounded above by b_1. If $\gamma : [0, \infty) \to M \setminus M_-$ is a ray with $\dot\gamma(0) = \lim_{j\to\infty} \dot\gamma_j(0)$, then $\varphi \circ \gamma$ is constant on $[0, \infty)$ and $\varphi \circ \gamma(t) = b_1$ for all $t > b_1$. This is a contradiction to the choice of $b = \max_K \varphi$, for γ passes through a point on K at which φ takes the value b_1. This proves the assertion (see Figure 6).

We next assert that if $b_1 > b$ is fixed, then $M_{b_1}^{b_1}(\varphi)$ has at least four components. In fact, we observe from $M_{b_1}^{b_1}(\varphi) \cap K = \varnothing$ that each unbounded component of $(M \setminus M_\lambda^\lambda(\varphi)) \cap (M \setminus K)$ contains a component of $M_{b_1}^{b_1}(\varphi)$. This contradicts Theorem 1.2 (4), and (A2) is proved.

The proof of (A3) is a consequence of (D2), and given after the proof of (D2).

We now prove (B1). From the assumption that $\inf_M \varphi$ is attained, it follows from Lemma 3.1 that $\varphi^{-1}[\inf_M \varphi, b_j]$ is compact for all j, where $\{b_j\}$ is a monotone divergent sequence. Then $K_j := \varphi^{-1}[\lambda, b_j]$ is monotone increasing and $\lim_{j\to\infty} K_j = M$. Clearly $M \setminus K_j$ contains a unique unbounded domain $\varphi^{-1}(b_j, \infty)$ for every j. This proves Theorem 1.3 (B1).

For (B2), if $\inf_M \varphi$ is not attained, we have monotone sequences $\{a_j\}$ and $\{b_j\}$ such that

$$\lim_{j\to\infty} a_j = \inf_M \varphi, \quad \lim_{j\to\infty} b_j = \infty,$$
$$[a_j, b_j] = \varphi(K_j), \quad j = 1, 2, \ldots.$$

Then for all large numbers j, $M \setminus K_j$ contains two unbounded domains

$$M \setminus K_j \supset \varphi^{-1}(b_j, \infty) \cup \varphi^{-1}(\inf_M \varphi, a_j).$$

This proves that M has exactly two ends.

We first prove (C) under an additional assumption that $\lambda := \inf_M \varphi$ is attained. Suppose that M has more than one end. Using the same notation as in the proof of (A2),

$$\lambda := \inf_M \varphi, \quad a := \min_K \varphi, \quad b := \max_K \varphi,$$

where $K \subset M$ is a compact set such that $M \setminus K$ has at least two unbounded components U and V.

We first assert that $K \cap M_\lambda^\lambda(\varphi) \neq \varnothing$. In fact, supposing that $K \cap M_\lambda^\lambda(\varphi) = \varnothing$ we find a component V of $M \setminus K$ such that if $b' > b$ then $M_{b'}^{b'}(\varphi) \subset V$ and $M_\lambda^\lambda \subset U$. Here the assumption that all the levels of φ are connected is essential. As is seen in the proof of (A2), there exist at least two components of $M_{b'}^{b'}$ for $b' > b$ such that one component lies in U and another in V. This contradicts the assumption in (C), and the first assertion is proved.

The same proof technique as developed in (A2) implies that $M_\lambda^\lambda(\varphi)$ passes through points on K, U and V. Fix a point $p \in V \cap M_\lambda^\lambda(\varphi)$, a divergent sequence $\{q_j\}$ of points in $U \setminus M_\lambda^\lambda(\varphi)$ and $\gamma_j : [0, d(p, q_j)] \to M$ a minimizing geodesic with $\gamma_j(0) = p$, $\gamma_j(d(p, q_j)) = q_j$.

From the construction of γ_j, we observe that $\varphi \circ \gamma$ is strictly increasing, and hence we find a number $t_j > 0$ such that $\gamma_j(t_j) \in M_{b'}^{b'}(\varphi) \cap U$. More precisely, $\gamma_j[0, d(p, q_j)]$ meets $M_\lambda^\lambda(\varphi)$ only at the origin, for $M_\lambda^\lambda(\varphi)$ is totally convex and hence if $\gamma(t_0) \in M_\lambda^\lambda(\varphi)$, for some $t_0 \in [0, d(p, q_j))$, then $\gamma_j[0, d(p, q_j)]$ is contained entirely in $M_\lambda^\lambda(\varphi)$.

Therefore $M_{b'}^{b'}(\varphi)$ has more than one component (one in U and another in V), a contradiction to the assumption in (C). This concludes the proof of (C) in this case.

We next prove (C) in the case where $\inf_M \varphi$ is not attained. Assume again that M has more than one end. We then have

$$\inf_M \varphi < a < b < \infty, \quad a := \min_K \varphi, \quad b := \max_K \varphi.$$

Since all the levels are connected, we find $\inf_M \varphi < a' < a$ and $b < b'$ such that $M_{a'}^{a'}(\varphi) \subset U$ and $M_{b'}^{b'}(\varphi) \subset V$. Let $\{y_j\} \subset M_{b'}^{b'}(\varphi)$ be a divergent sequence of points and fix a point $x \in M_{a'}^{a'}(\varphi)$. Let $\gamma_j : [0, d(x, y_j)] \to M$ for $j = 1, 2, \ldots$ be a minimizing geodesic with $\gamma_j(0) = x$ and $\gamma_j(d(x, y_j)) = y_j$. There exists a ray $\gamma : [0, \infty) \to M$ emanating from x such that $\dot\gamma(0) = \lim_{j\to\infty} \dot\gamma_j(0)$. Clearly, every γ_j passes through a point on K and hence, so does γ. From construction, $\varphi \circ \gamma : [0, \infty) \to \mathbb{R}$ is bounded from above by b', and therefore is constant. However it is impossible, for $\varphi(x) = a'$ and $\varphi \circ \gamma(t_0) \geq a > a'$ at a point $\gamma(t_0) \in K$. This completes the proof of (C).

For the proof of (D1), suppose that M has more than two ends.

Let $K \subset M$ be a connected compact subset such that $M \setminus K$ contains at least three unbounded components, say U, V and W. We may consider that U contains $\varphi^{-1}[b', \infty)$, for all $b' > b$. Since all the levels of φ are connected, we have

$$\sup_{M \setminus U} \varphi \leq b.$$

In fact, suppose that there exists a point $x \in M \setminus U$ such that $\varphi(x) = b'$ for some $b' > b$. Then $M_{b'}^{b'}(\varphi) \cap K = \varnothing$ and hence $M_{b'}^{b'}(\varphi)$ is disconnected, a contradiction to the assumption of (D).

Let $\{x_j\} \subset V$ and $\{y_j\} \subset W$ be two divergent sequences of points, and let $\gamma_j : [0, d(x_j, y_j)] \to M \setminus U$ be a minimizing geodesic joining x_j to y_j. Since γ_j passes through a point on K, there exists a straight line $\gamma : \mathbb{R} \to M \setminus U$ such that $\dot\gamma(0)$ is obtained as the limit of a converging sequence of vectors $\dot\gamma_j(t_j) \in K$ for $j = 1, 2, \ldots$. Clearly, $\varphi \circ \gamma : \mathbb{R} \to \mathbb{R}$ is bounded above, and hence constant taking

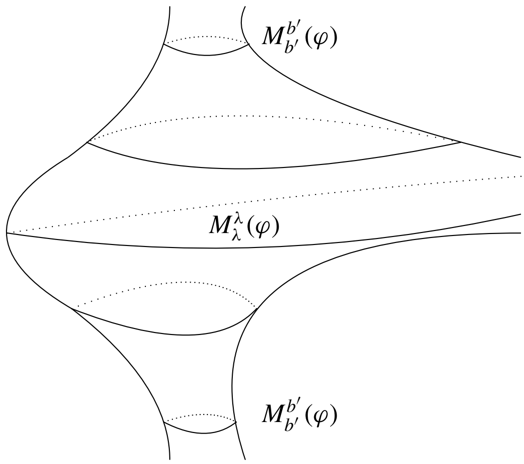

Figure 7. The proof of Theorem 1.3 (D1).

a value $\mu = \varphi \circ \gamma(0) \in [a, b]$. We therefore observe that

$$M_\mu^\mu(\varphi) \cap K \neq \varnothing, \quad M_\mu^\mu(\varphi) \cap W \neq \varnothing \quad \text{and} \quad M_\mu^\mu(\varphi) \cap V \neq \varnothing.$$

We next choose a value $a' \in (\inf_M \varphi, a)$. We may assume without loss of generality that $M_{a'}^{a'}(\varphi) \subset V$. Let $\{z_j\} \subset M_\mu^\mu(\varphi) \cap W$ be a divergent sequence of points and $x \in M_{a'}^{a'}(\varphi)$ an arbitrary fixed point. Let $\sigma_j : [0, d(x, z_j)] \to M \setminus U$ be a minimizing geodesic with $\sigma_j(0) = x$, $\sigma(d(x, z_j)) = z_j$ for all $j = 1, 2, \dots$. Clearly, $\varphi \circ \sigma_j$ is monotone increasing in W. Let $\sigma : [0, \infty) \to M$ be a ray such that $\dot\sigma(0) = \lim_{j\to\infty} \dot\sigma_j(0)$. We then observe that $\varphi \circ \sigma$ is monotone increasing on an unbounded interval $[\bar b, \infty)$ for some $\bar b > 0$, and bounded above by μ. Thus, it is identically equal to a'. Recall that $\varphi \circ \sigma(0) = \varphi(x) = a'$. However this is impossible since $a' < \min_K \varphi = a$ and $\sigma[0, \infty)$ passes through a point on K. We therefore observe that $M \setminus (K \cup U)$ has exactly one end. This proves (D1).

The proof of (D2) is now clear and omitted.

The proof of (A3) is now a straightforward consequence of (D2). See Figure 8. If $M_b^b(\varphi)$ is compact for some $b \in \varphi(M)$, then $\varphi^{-1}[b, \infty)$ has two ends. From the assumption and Theorem 1.2 (1), we observe that $M_\lambda^\lambda(\varphi)$ is noncompact. Therefore $M^b(\varphi) = \varphi^{-1}[\lambda, b]$ is noncompact and so has at least one end. This proves (A3).

Finally, we prove (E). Suppose that φ admits both compact and noncompact levels simultaneously. The same notation as in the proof of (D) will be used. If φ admits a disconnected level, then $\varphi^{-1}[b', \infty)$ consists of two unbounded components for all $b' > b$.

Then Theorem 1.2 (1) and Lemma 3.1 imply that $\lambda := \inf_M \varphi$ is attained and $M_\lambda^\lambda(\varphi)$ is connected and noncompact. Therefore, every compact set K containing

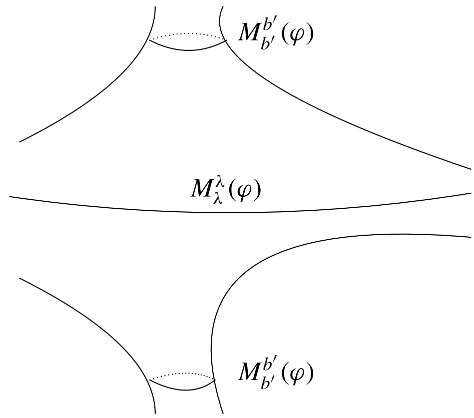

Figure 8. The proof of Theorem 1.3 (A3).

$M_{b'}^{b'}(\varphi)$ has the property that $M \setminus K$ has more than two unbounded components. In fact, two components of $M \setminus K$ contain $\varphi^{-1}[b', \infty)$ and the other component intersects with $M_{\lambda}^{\lambda}(\varphi)$ outside K. This proves that M has at least three ends, a contradiction to the assumption of (E).

If all the levels of φ are connected and noncompact, then M has one end by (C), a contradiction to the assumption of (E). This completes the proof of (E) and hence of Theorem 1.3. □

References

[Bangert 1978] V. Bangert, "Riemannsche Mannigfaltigkeiten mit nicht-konstanter konvexer Funktion", *Arch. Math.* (*Basel*) **31** (1978), 163–170. MR 80e:53040 Zbl 0372.53021

[Bao et al. 2000] D. Bao, S.-S. Chern, and Z. Shen, *An introduction to Riemann–Finsler geometry*, Graduate Texts in Mathematics **200**, Springer, New York, 2000. MR 2001g:53130 Zbl 0954.53001

[Burago and Zalgaller 1977] Y. D. Burago and V. A. Zalgaller, "Выпуклые множества в румановых пространствах неотрицательной кривизны", *Uspekhi Mat. Nauk* **32**:3(195) (1977), 3–55. Translated as "Convex sets in Riemannian spaces of non-negative curvature" in *Russ. Math. Surv.* **32**:3 (1977), 1–57. MR 57 #4054 Zbl 0397.53031

[Busemann 1955] H. Busemann, *The geometry of geodesics*, Academic Press, New York, 1955. MR 17,779a Zbl 0112.37002

[Busemann and Phadke 1983] H. Busemann and B. B. Phadke, "Peakless and monotone functions on G-spaces", *Tsukuba J. Math.* **7**:1 (1983), 105–135. MR 84m:53073 Zbl 0526.53064

[Cheeger and Ebin 2008] J. Cheeger and D. G. Ebin, *Comparison theorems in Riemannian geometry*, AMS Chelsea, Providence, RI, 2008. MR 2009c:53043 Zbl 1142.53003

[Chern et al. 1999] S.-S. Chern, W. H. Chen, and K. S. Lam, *Lectures on differential geometry*, Series on University Mathematics **1**, World Scientific, River Edge, NJ, 1999. MR 2000m:53001 Zbl 0940.53001

[Greene and Shiohama 1981a] R. E. Greene and K. Shiohama, "Convex functions on complete noncompact manifolds: differentiable structure", *Ann. Sci. École Norm. Sup.* (4) **14**:4 (1981), 357–367. MR 83m:53057 Zbl 0488.57012

[Greene and Shiohama 1981b] R. E. Greene and K. Shiohama, "Convex functions on complete noncompact manifolds: topological structure", *Invent. Math.* **63**:1 (1981), 129–157. MR 82e:53065 Zbl 0468.53033

[Greene and Shiohama 1987] R. E. Greene and K. Shiohama, "The isometry groups of manifolds admitting nonconstant convex functions", *J. Math. Soc. Japan* **39**:1 (1987), 1–16. MR 88a:53032 Zbl 0611.53039

[Innami 1982a] N. Innami, "A classification of Busemann G-surfaces which possess convex functions", *Acta Math.* **148** (1982), 15–29. MR 84f:53063 Zbl 0559.53051

[Innami 1982b] N. Innami, "Splitting theorems of Riemannian manifolds", *Compositio Math.* **47**:3 (1982), 237–247. MR 84e:53053 Zbl 0514.53040

[Innami 1983] N. Innami, "Totally flat foliations and peakless functions", *Arch. Math.* (*Basel*) **41**:5 (1983), 464–471. MR 85i:53076 Zbl 0554.53045

[Mashiko 1999a] Y. Mashiko, "Affine functions on Alexandrov surfaces", *Osaka J. Math.* **36**:4 (1999), 853–859. MR 2001a:53057 Zbl 1037.53503

[Mashiko 1999b] Y. Mashiko, "Convex functions on Alexandrov surfaces", *Trans. Amer. Math. Soc.* **351**:9 (1999), 3549–3567. MR 2000g:53082 Zbl 0962.53041

[Mashiko 2002] Y. Mashiko, "A splitting theorem for Alexandrov spaces", *Pacific J. Math.* **204**:2 (2002), 445–458. MR 2003b:53044 Zbl 1058.53036

[Sakai 1992] T. Sakai, *Riemannian geometry*, Mohanabo, Tokyo, 1992. In Japanese; translated in Translations of Mathematical Monographs **149**, American Mathematical Society, Providence, RI, 1996. MR 97f:53001 Zbl 0886.53002

[Sharafutdinov 1978] V. A. Sharafutdinov, "The Pogorelov–Klingenberg theorem for manifolds homeomorphic to $\mathbb{R}^n$", *Sibir. Mat. Zh.* **18** (1978), 649–657. In Russian; translated in *Sib. Math. J.* **18** (1978), 915–925. MR 0487896 Zbl 0411.53031

[Whitehead 1935] J. H. C. Whitehead, "On the covering of a complete space by the geodesics through a point", *Ann. of Math.* (2) **36**:3 (1935), 679–704. MR 1503245 Zbl 0012.27802

[Yamaguchi 1982] T. Yamaguchi, "The isometry groups of Riemannian manifolds admitting strictly convex functions", *Ann. Sci. École Norm. Sup.* (4) **15**:1 (1982), 205–212. MR 84d:53045 Zbl 0501.53028

[Yamaguchi 1986a] T. Yamaguchi, "Locally geodesically quasiconvex functions on complete Riemannian manifolds", *Trans. Amer. Math. Soc.* **298**:1 (1986), 307–330. MR 87j:53069 Zbl 0611.53040

[Yamaguchi 1986b] T. Yamaguchi, "Uniformly locally convex filtrations on complete Riemannian manifolds", pp. 308–318 in *Curvature and topology of Riemannian manifolds* (Katata, 1985), edited by K. Shiohama et al., Lecture Notes in Mathematics **1201**, Springer, Berlin, 1986. MR 87m:53051 Zbl 0592.53033

[Yamaguchi 1988] T. Yamaguchi, "On the structure of locally convex filtrations on complete manifolds", *J. Math. Soc. Japan* **40**:2 (1988), 221–234. MR 89e:53067 Zbl 0667.53036

Received January 7, 2014.

SORIN V. SABAU
TOKAI UNIVERSITY
SCHOOL OF SCIENCE
DEPARTMENT OF MATHEMATICS
SAPPORO 005-8601
JAPAN
sorin@tokai.ac.jp

KATSUHIRO SHIOHAMA
FUKUOKA INSTITUTE OF TECHNOLOGY
INSTITUTE OF INFORMATION SCIENCE
FUKUOKA 841-0108
JAPAN
shiohama_k@tw.drive-net.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 276, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.276.483

VARIATIONS OF THE TELESCOPE CONJECTURE AND BOUSFIELD LATTICES FOR LOCALIZED CATEGORIES OF SPECTRA

F. LUKE WOLCOTT

We investigate several versions of the telescope conjecture on localized categories of spectra and implications between them. Generalizing the "finite localization" construction, we show that, on such categories, localizing away from a set of strongly dualizable objects is smashing. We classify all smashing localizations on the harmonic category, $H\mathbb{F}_p$-local category and I-local category, where I is the Brown–Comenetz dual of the sphere spectrum; all are localizations away from strongly dualizable objects, although these categories have no nonzero compact objects. The Bousfield lattices of the harmonic, $E(n)$-local, $K(n)$-local, $H\mathbb{F}_p$-local and I-local categories are described, along with some lattice maps between them. One consequence is that in none of these categories is there a nonzero object that squares to zero. Another is that the $H\mathbb{F}_p$-local category has localizing subcategories that are not Bousfield classes.

1. Introduction

The telescope conjecture, first stated by Ravenel [1984, Conjecture 10.5], is a claim about two classes of localization functors in the p-local stable homotopy category of spectra. First, one can localize away from a finite type $n+1$ spectrum $F(n+1)$; the acyclics are the smallest localizing subcategory containing $F(n+1)$, and we denote this functor by L_n^f. Second, one can localize at the wedge of the first $n+1$ Morava K-theories $K(0) \vee \cdots \vee K(n)$; the acyclics are all spectra that smash with $K(0) \vee \cdots \vee K(n)$ to zero and this is denoted L_n. Both these localizations are smashing, i.e., they commute with coproducts. The telescope conjecture (TC_n) claims that L_n^f and L_n are isomorphic. In fact, here we consider three slightly different versions, $\mathrm{TC1}_n$, $\mathrm{TC2}_n$, and $\mathrm{TC3}_n$, of the telescope conjecture. In Section 3 we articulate them carefully and show implications between them.

The conjecture is known to hold for $n=0$ [Ravenel 1992, p. 79] and for $n=1$ when $p=2$ [Mahowald 1982] and $p>2$ [Miller 1981]. A valiant but unsuccessful

MSC2010: 55P42, 18E30, 55U35.
Keywords: telescope conjecture, Bousfield localization, Bousfield lattice, harmonic category, smashing localization.

effort at a counterexample, for $n \geq 2$, was undertaken by Mahowald, Ravenel, and Shick, as outlined in [Mahowald et al. 2001]. Since then little progress has been made, and the original conjecture remains open.

A generalization of the telescope conjecture can be stated for spectra, as well as other triangulated categories. Localization away from a finite spectrum, i.e, a compact object of the category, always yields a smashing localization functor (see, e.g., [Bousfield 1979a, Proposition 2.9] or [Miller 1992] or [Hovey et al. 1997, Theorem 3.3.3]). The Generalized Smashing Conjecture (GSC) is that every smashing localization arises in this way. If true, then every smashing localization is determined by its compact acyclics; if the GSC holds in spectra, then so must the TC_n for all n.

The GSC, essentially stated for spectra decades ago by Bousfield [1979b, Conjecture 3.4], has been formulated in many other triangulated categories, in many cases labeled as the telescope conjecture, and in many cases proven to hold. Neeman [1992] made the conjecture for the derived category $D(R)$ of a commutative ring R and showed that it holds when the ring is Noetherian. See also [Hovey et al. 1997, Theorem 6.3.7] or [Krause and Šťovíček 2010] for a generalization. On the other hand, Keller [1994] gave an example of a non-Noetherian ring for which the GSC fails. Benson, Iyengar, and Krause have shown that the GSC holds in a stratified category [Benson et al. 2011a], such as the stable module category of a finite group [Benson et al. 2011b]. Balmer and Favi [2011] showed that in a tensor triangulated category with a good notion of support, the GSC is a "local" question.

It is worth noting that there are further variations of the GSC that we will not consider here. Krause [2000] formulated a variation of the GSC, in terms of subcategories generated by sets of maps, that makes sense (and holds) for any compactly generated triangulated category. Krause and Solberg [2003] gave a variation for stable module categories, stated in terms of cotorsion pairs. See also [Krause 2005; Angeleri Hügel et al. 2008; Brüning 2007; Šťovíček 2010].

To date, Keller's ring yields the only category where the GSC is known to fail. In this paper we give several more examples. Incidentally, each is a well generated triangulated category that is not compactly generated.

One of our main results is the following. We weaken the assumptions for "finite localization" and show that, in many categories, localization away from any set of strongly dualizable objects yields a smashing localization. (Recall that an object X is strongly dualizable if $F(X, Y) \cong F(X, \mathbb{1}) \wedge Y$ for all Y, where $\mathbb{1}$ is the tensor unit and $F(-,-)$ the function object bifunctor.) Let $\mathrm{loc}(X)$ denote the smallest localizing subcategory containing X. We prove the following as Theorem 3.5.

Theorem (A). *Let* T *be a well generated tensor triangulated category such that* $\mathrm{loc}(\mathbb{1}) = \mathsf{T}$. *Let* $A = \{B_\alpha\}$ *be a (possibly infinite) set of strongly dualizable objects. Then there exists a smashing localization functor* $L : \mathsf{T} \to \mathsf{T}$ *with* $\mathrm{Ker}\, L = \mathrm{loc}(A)$.

Thus we are led to conjecture the following.

Conjecture (Strongly Dualizable Generalized Smashing Conjecture (SDGSC)). *Every smashing localization is localization away from a set of strongly dualizable objects.*

We give several examples of categories where the GSC fails but the SDGSC holds. In fact, we consider a topological setting, where one can also formulate a version (or versions, rather) of the original telescope conjecture.

Specifically, we consider localized categories of spectra. Let $\mathcal{S}$ be the p-local stable homotopy category, and let $\wedge$ denote the smash (i.e., tensor) product. Take Z to be an object of $\mathcal{S}$, and let $L = L_Z : \mathcal{S} \to \mathcal{S}$ be the localization functor that annihilates Z_*-acyclic objects. The full subcategory of L-local objects, that is, objects X for which $X \to LX$ is an equivalence, has a tensor triangulated structure induced by that of $\mathcal{S}$. Let $\mathcal{L}$ denote this category; the triangles are the same as in $\mathcal{S}$, the coproduct is $X \coprod_{\mathcal{L}} Y = L(X \coprod Y)$ and the tensor is $X \wedge_{\mathcal{L}} Y = L(X \wedge Y)$.

In Definition 3.6, we define localization functors l_n^f and l_n on $\mathcal{L}$ that are localized versions of L_n^f and L_n. The localized telescope conjecture (LTC) claims that l_n^f and l_n are isomorphic. In fact, we give three versions of the LTC and in Theorems 3.12 and 3.13 establish implications between them. Then, examining specific examples of localized categories of spectra, we conclude the following in Theorems 4.3, 5.11, 6.1, 6.5 and 6.9 and Corollary 5.6.

Theorem (B). *All versions of the localized telescope conjecture*, LTC1$_i$, LTC2$_i$, *and* LTC3$_i$ *hold for all $i \geq 0$, in the $\bigvee_{n\geq 0} K(n)$-local* (*i.e.*, *harmonic*), *$K(n)$-local*, *$H\mathbb{F}_p$-local*, *BP-local*, *and I-local categories*, *where I is the Brown–Comenetz dual of the sphere spectrum.*

In order to consider the GSC and SDGSC in $\mathcal{L}$, we must classify the smashing localizations on $\mathcal{L}$. We are able to do this in several examples.

Theorem (C). *In the harmonic category*, *the* GSC *fails but the* SDGSC *holds*, *and likewise in the $H\mathbb{F}_p$-local and I-local categories. In the BP-local category the* GSC *fails but the* SDGSC *is open. In the $E(n)$-local and $K(n)$-local categories the* GSC *and* SDGSC *both hold.*

Proof. This theorem is concluded from Theorems 4.4 and 5.11, Propositions 6.3 and 6.10, and Corollaries 5.6 and 6.7. □

One novelty in our approach is our use of Bousfield lattice arguments. Given an object X in a tensor triangulated category T, the Bousfield class of X is defined by $\langle X\rangle = \{W \mid W \wedge X = 0\}$. It is now known [Iyengar and Krause 2013] that every well generated tensor triangulated category has a set of Bousfield classes. This set has the structure of a lattice and is called the Bousfield lattice of T. One can now attempt to calculate the Bousfield lattices of categories of localized spectra. Furthermore, every smashing localization yields a pair of so-called complemented Bousfield classes. Information about the Bousfield lattice of a category gives information about its complemented classes, which gives information about the smashing localization functors on the category.

Moreover, the first version of the telescope conjecture $\mathrm{TC1}_n$ is that two spectra $T(n)$ and $K(n)$ have the same Bousfield class. In the localized version this becomes ($\mathrm{LTC1}_n$) the claim that $\langle LT(n)\rangle = \langle LK(n)\rangle$ in the Bousfield lattice of $\mathcal{L}$. One is thus led to investigating Bousfield lattices of localized spectra.

Corollary 2.7 gives an upper bound, $2^{2^{\aleph_0}}$, on the cardinality of such lattices. Jon Beardsley has calculated the Bousfield lattice of the harmonic category to be isomorphic to the power set of $\mathbb{N}$; we give this calculation in Proposition 4.2. In Corollary 5.4 and Proposition 5.7 we show that one can realize this lattice as an inverse limit of the Bousfield lattices of $E(n)$-local categories, as n ranges over $\mathbb{N}$. Then in Corollary 5.10 and Propositions 6.2 and 6.6, we show that the $K(n)$-local, $H\mathbb{F}_p$-local, and I-local categories all have two-element Bousfield lattices. In Proposition 6.11 we give a lower bound, $2^{\aleph_0}$, on the cardinality of the Bousfield lattice of the BP-local category.

One immediate object-level application of these Bousfield lattice calculations is the following. Call an object X *square-zero* if it is nonzero but $X \wedge X = 0$. Then Proposition 2.9 shows that there are no square-zero objects in the harmonic, $E(n)$-, $K(n)$-, $H\mathbb{F}_p$-, or I-local categories.

We are also able to answer the analogue of a conjecture by Hovey and Palmieri, originally stated for the stable homotopy category. Conjecture 9.1 in [Hovey and Palmieri 1999] is that every localizing subcategory is a Bousfield lattice. Proposition 6.4 demonstrates that this fails in the $H\mathbb{F}_p$-local category by giving two localizing subcategories that are not Bousfield classes.

Section 2 establishes the categorical setting, and provides background on localization, Bousfield lattices, and stable homotopy theory. Section 3 defines the various versions of the telescope conjecture, for spectra and for localized spectra, and establishes implications among them. The remainder of the paper is devoted to examining specific examples: the harmonic category (Section 4), the $E(n)$-local and $K(n)$-local categories (Section 5), and the $H\mathbb{F}_p$-local, I-local, BP-local, and $F(n)$-local categories (Section 6). All results are new unless cited. Most of the results on the $E(n)$-local and $K(n)$-local categories in Section 5 follow in a straightforward way from Hovey and Strickland's work [1999] and are included for completeness.

2. Preliminaries

2A. ***Categorical setting.*** We start with the notion of a tensor triangulated category C, i.e., a triangulated category with set-indexed coproducts and a closed symmetric monoidal structure compatible with the triangulation [Hovey et al. 1997, Appendix A]. Let $\Sigma : \mathsf{C} \to \mathsf{C}$ denote the shift and $[X, Y]$ the morphisms from X to Y, and let $[X, Y]_n = [\Sigma^n X, Y]$ for any $n \in \mathbb{Z}$.

Let $- \wedge -$ denote the smash (tensor) product, $\mathbb{1}$ the unit, and $F(-,-)$ the function object bifunctor; $F(X, -)$ is the right adjoint to $X \wedge -$. Recall that an object X in C is said to be *strongly dualizable* if the natural map $DX \wedge Y \to F(X, Y)$ is an isomorphism for all Y, where $DX = F(X, \mathbb{1})$ is the Spanier–Whitehead dual. Since $F(\mathbb{1}, X) \cong X$ for all X, the map $F(\mathbb{1}, \mathbb{1}) \wedge Y \to F(\mathbb{1}, Y)$ is an equivalence and $\mathbb{1}$ is always strongly dualizable.

For a regular cardinal α, we say an object X is *α-small* if every morphism $X \to \coprod_{i\in I} Y_i$ factors through $\coprod_{i\in J} Y_i$ for some $J \subseteq I$ with $|J| < \alpha$. If X is $\aleph_0$-small we say X is *compact* ([Hovey et al. 1997] calls this *small*); this is equivalent to the condition that the natural map $\bigoplus_{i\in K}[X, Z_i] \to [X, \coprod_{i\in K} Z_i]$ is an isomorphism for any set-indexed coproduct $\coprod_{i\in K} Z_i$. We say C is *α-well generated* if it has a set of perfect generators [Krause 2010, Section 5.1] which are α-small, and C is *well generated* if it is α-well generated for some α. See [Krause 2010] for more details.

A *localizing* subcategory is a triangulated subcategory of C that is closed under retracts and coproducts; a *thick* subcategory is a triangulated subcategory that is closed under retracts. Given an object or set of objects X, let $\mathrm{loc}(X)$ (resp. $\mathrm{th}(X)$) denote the smallest localizing (resp. thick) subcategory containing X. We say that $\mathrm{loc}(X)$ is *generated by X*.

Notation 2.1. Throughout this paper let T be a well generated tensor triangulated category such that $\mathrm{loc}(\mathbb{1}) = \mathsf{T}$.

In the language of [Hovey et al. 1997], such a T is almost a "monogenic stable homotopy category", except that we do not insist that the unit $\mathbb{1}$ is compact.

In practice, in this paper T will always be either the p-local stable homotopy category of spectra $\mathcal{S}$ or the category $\mathcal{L}_Z$ of L_Z-local objects derived from a localization functor $L_Z : \mathcal{S} \to \mathcal{S}$. The former satisfies Notation 2.1 by [Hovey et al. 1997, Example 1.2.3(a)], and the latter by Theorem 2.3 and Lemma 2.4 below.

2B. *Background on localization.* Recall that a *localization functor* (or simply *localization*) on a tensor triangulated category C is an exact functor $L : \mathsf{C} \to \mathsf{C}$, along with a natural transformation $\eta : 1 \to L$ such that $L\eta$ is an equivalence and $L\eta = \eta L$. We call $\operatorname{Ker} L$ the *L-acyclics*. It follows that there is an exact functor $C : \mathsf{C} \to \mathsf{C}$, called *colocalization*, such that every X in C fits into an exact triangle $CX \to X \to LX$, with CX L-acyclic. An object Y is *L-local* if it is in the essential image of L, and this is equivalent to satisfying $[Z, Y] = 0$ for all L-acyclic Z. See [Hovey et al. 1997, Chapter 3] or [Krause 2010] for further background.

We also recall two special types of localizations. A localization $L : \mathsf{C} \to \mathsf{C}$ is said to be *smashing* if L preserves coproducts, equivalently if $LX \cong L\mathbb{1} \wedge X$ for all X.

Given a set A of objects of C, we say that a localization functor $L : \mathsf{C} \to \mathsf{C}$ is *localization away from A* if the L-acyclics are precisely $\operatorname{loc}(A)$. If such a localization exists, we also say it is *generated by A*. When $\mathsf{C} = \mathcal{S}$, it is well known (e.g., [Miller 1992; Mahowald and Sadofsky 1995]) that localization away from a set of compact objects exists and yields a smashing localization functor. As mentioned in the introduction, this result has been generalized to other categories as well (e.g., [Hovey et al. 1997, Theorem 3.3.3], [Balmer and Favi 2011, Theorem 4.1]). We present a further generalization in Theorem 3.5.

In this paper we will restrict our attention to homological localizations, which we now describe. Given an object Z in a tensor triangulated category C, the *Bousfield class* of Z is defined to be

$$\langle Z \rangle = \{W \in \mathsf{C} \mid W \wedge Z = 0\}.$$

Extending a classical result of Bousfield's for $\mathcal{S}$, Iyengar and Krause recently showed [2013, Proposition 2.1] that for every object Z in a well generated tensor triangulated category C, there is a localization functor $L_Z : \mathsf{C} \to \mathsf{C}$ with L_Z-acyclics precisely $\langle Z \rangle$. We call such an L_Z a *homological localization at Z*.

Notation 2.2. Let T be as in Notation 2.1, with tensor unit $\mathbb{1}$. For an object Z in T, let $L_Z : \mathsf{T} \to \mathsf{T}$ be homological localization at Z. Let $\mathcal{L}_Z$ denote the category of L_Z-local objects, the essential image of L_Z.

Theorem 2.3 [Hovey et al. 1997, 3.5.1, 3.5.2]. *Let $L = L_Z : \mathsf{T} \to \mathsf{T}$ be a localization, and $\mathcal{L}_Z$ the category of L_Z-local objects. Then $\mathcal{L}_Z$ has a natural structure as a tensor triangulated category, generated by $L_Z\mathbb{1}$, which is the unit. Considered as a functor from T to $\mathcal{L}_Z$, L preserves triangles, the tensor product and its unit, coproducts, and strong dualizability. Furthermore, L preserves compactness if and only if L is a smashing localization.*

Explicitly, for L-local objects X, X_i and Y, in $\mathcal{L}$ we have $\coprod_{\mathcal{L}} X_i = L\big(\coprod_{\mathsf{T}} X_i\big)$ and $X \wedge_{\mathcal{L}} Y = L(X \wedge_{\mathsf{T}} Y)$ and $F_{\mathcal{L}}(X, Y) = F(X, Y)$. Note that $L_Z\mathbb{1}$ is strongly dualizable but may not be compact in $\mathcal{L}_Z$.

Lemma 2.4. *The category $\mathcal{L}_Z$ is well generated.*

Proof. By Proposition 2.1 of [Iyengar and Krause 2013], the L_Z-acyclics $\langle Z \rangle$ form a well generated localizing subcategory of T. Then by [Krause 2010, Theorem 7.2.1], the Verdier quotient $\mathsf{T}/\langle Z \rangle$, which is equivalent to the local category $\mathcal{L}_Z$, is well generated. □

We conclude this subsection with a lemma containing four useful well-known facts. Recall that a *ring object* in a tensor triangulated category is an object R with an associative multiplication map $\mu : R \wedge R \to R$ and a unit $\iota : \mathbb{1} \to R$, making the evident diagrams commute. If R is a ring object, then an *R-module object* is an object M with a map $m : R \wedge M \to M$ along with evident commutative diagrams. Note that $R \wedge X$ is an R-module object for every X. A *skew field object* is a ring object such that every R-module object is *free*, i.e., isomorphic to a coproduct of suspensions of R [Hovey et al. 1997, Definition 3.7.1].

Lemma 2.5. *Let* C *be a tensor triangulated category with* $\operatorname{loc}(\mathbb{1}) = \mathsf{C}$, *and* $L : \mathsf{C} \to \mathsf{C}$ *a localization.*

(1) *Every localizing subcategory* S *of* C *is tensor-closed*; *that is*, *if* $X \in \mathsf{S}$ *and* $Y \in \mathsf{C}$, *then* $X \wedge Y \in \mathsf{S}$.

(2) *For all* X *and* Y *in* C, $L(X \wedge Y) = L(LX \wedge LY)$.

(3) *Considered as a functor from* C *to* $\mathcal{L}$, L *also preserves ring objects and module objects.*

(4) *If* R *is a ring object and* M *is an* R*-module object* (*in particular*, *if* $M = R$), *then* M *is* R*-local.*

Proof. For (1), note that $Y \in \operatorname{loc}(\mathbb{1}) = \mathsf{C}$, so $X \wedge Y \in \operatorname{loc}(X \wedge \mathbb{1}) = \operatorname{loc}(X) \subseteq \mathsf{S}$.

For (2), consider the exact triangle $X \wedge CY \to X \wedge Y \to X \wedge LY$. Since CY is L-acyclic and these form a localizing subcategory, $L(X \wedge CY) = 0$, so $L(X \wedge Y) = L(X \wedge LY)$. Using the same reasoning with the triangle $CX \wedge LY \to X \wedge LY \to LX \wedge LY$, the result follows.

If $R \in \mathsf{C}$ is a ring object, then $L(\mu) : L(R \wedge R) = L(LR \wedge LR) = LR \wedge_{\mathcal{L}} LR \to LR$, and all the localized diagrams commute. A similar argument applies to module objects, completing part (3).

Part (4) is [Ravenel 1984, Proposition 1.17(a)]. □

2C. ***Background on Bousfield lattices.*** Every well generated tensor triangulated category, and hence every localized category of spectra, has a set (rather than a proper class) of Bousfield classes [Iyengar and Krause 2013, Theorem 3.1]. This was also recently shown for every tensor triangulated category with a combinatorial model [Casacuberta et al. 2014]. This set is called the *Bousfield lattice* $\mathrm{BL}(\mathsf{T})$ and has a lattice structure which we now recall. Refer to [Hovey and Palmieri 1999; Wolcott 2014] for more details.

The partial ordering is given by reverse inclusion: we say $\langle X \rangle \le \langle Y \rangle$ when $W \wedge Y = 0$ implies $W \wedge X = 0$. It is also helpful to remember that, unwinding definitions, $\langle X \rangle \le \langle Y \rangle$ precisely when every L_X-local object is also L_Y-local. Clearly $\langle 0 \rangle$ is the minimum and $\langle \mathbb{1} \rangle$ is the maximum class. The join of any set of classes is $\bigvee_{i \in I} \langle X_i \rangle = \left\langle \coprod_{i \in I} X_i \right\rangle$, and the meet is defined to be the join of all lower bounds.

The smash product induces an operation on Bousfield classes, where $\langle X \rangle \wedge \langle Y \rangle$ is equal to $\langle X \wedge Y \rangle$. This is a lower bound, but in general not the meet. However, if we restrict to the subposet $\mathrm{DL} = \{\langle W \rangle \mid \langle W \wedge W \rangle = \langle W \rangle\}$, then the meet and smash agree. Since coproducts distribute across the smash product, DL is a distributive lattice.

We say a class $\langle X\rangle$ is *complemented* if there exists a class $\langle X^c\rangle$ such that $\langle X\rangle \wedge \langle X^c\rangle = \langle 0\rangle$ and $\langle X\rangle \vee \langle X^c\rangle = \langle \mathbb{1}\rangle$. The collection of complemented classes is denoted BA. For example, every smashing localization $L : \mathsf{T} \to \mathsf{T}$ gives a pair of complemented classes, namely $\langle C\mathbb{1}\rangle$ and $\langle L\mathbb{1}\rangle$. Because every complemented class is also in DL, BA is a Boolean algebra.

Proposition 2.6. *Let* T *be as in Notation 2.1, and* $L_Z : \mathsf{T} \to \mathsf{T}$ *a localization functor as in Notation 2.2. Then* L_Z *induces a well-defined order-preserving map of lattices* $\mathrm{BL}(\mathsf{T}) \to \mathrm{BL}(\mathcal{L}_Z)$, *where* $\langle X\rangle \mapsto \langle L_Z X\rangle$. *This map is surjective and sends* $\mathrm{DL}(\mathsf{T})$ *onto* $\mathrm{DL}(\mathcal{L}_Z)$ *and* $\mathrm{BA}(\mathsf{T})$ *onto* $\mathrm{BA}(\mathcal{L}_Z)$.

Proof. Most of this is proved in Lemma 3.1 of [Wolcott 2014]. For $\langle X\rangle \in \mathrm{DL}(\mathsf{T})$, using Lemma 3.10 we get

$$\langle LX\rangle = \langle L(X \wedge X)\rangle = \langle L(LX \wedge LX)\rangle = \langle LX \wedge_{\mathcal{L}} LX\rangle.$$

Likewise, one can check that for $\langle X\rangle \in \mathrm{BA}(\mathsf{T})$, the class $\langle LX\rangle \in \mathrm{BL}(\mathcal{L})$ is complemented by $\langle L(X^c)\rangle$, keeping in mind that $\langle L\mathbb{1}\rangle$ is the top class in $\mathrm{BL}(\mathcal{L})$. $\square$

Corollary 2.7. *For any* $Z \in \mathcal{S}$, *we have* $|\mathrm{BL}(\mathcal{L}_Z)| \le 2^{2^{\aleph_0}}$.

Proof. Observe that $|\mathrm{BL}(\mathcal{L}_Z)| \le |\mathrm{BL}(\mathcal{S})| \le 2^{2^{\aleph_0}}$, where the second inequality is proved in [Ohkawa 1989]. $\square$

Lemma 2.8. *Let* T *be as in Notation 2.1, and X and Y objects of* T. *Then* $\langle X\rangle \le \langle Y\rangle$ *if and only if* $L_X = L_X L_Y = L_Y L_X$ *and in this case the following diagram commutes (also with* BL *replaced by* DL *or* BA).

$$\begin{tikzcd}
 & \mathrm{BL}(\mathsf{T}) \arrow[dl, two heads, "L_Y"'] \arrow[dr, two heads, "L_X"] & \\
\mathrm{BL}(\mathcal{L}_Y) \arrow[rr, two heads, "L_X"] & & \mathrm{BL}(\mathcal{L}_X)
\end{tikzcd}$$

Proof. The first equivalence is straightforward; it follows from [Ravenel 1984, Proposition 1.22] and the observation that $\langle X\rangle \le \langle Y\rangle$ precisely when all L_X-locals are L_Y-locals. The last remark follows from Proposition 2.6. $\square$

Here we mention one object-level application of the Bousfield lattice calculations of Sections 5 and 6.

Proposition 2.9. *There are no square-zero objects in the harmonic*, $E(n)$-, $K(n)$-, $H\mathbb{F}_p$-, *or* I-*local categories.*

Proof. In Corollary 2.8 of [Wolcott 2014], we show that in a well generated tensor triangulated category, there are no square-zero objects if and only if BA = DL = BL. The claim follows from Corollaries 5.4 and 5.10 and Propositions 6.2 and 6.6. $\square$

2D. ***Background on spectra.*** We quickly review some relevant background on the stable homotopy category. See [Ravenel 1992; 1993; Hovey 1995a; Mahowald and Sadofsky 1995] for more details. Fix a prime p and let $\mathcal{S}$ denote the p-local stable homotopy category of spectra. Let S^0 denote the sphere spectrum. The *finite* spectra $\mathcal{F}$ are the compact objects of $\mathcal{S}$, and $\mathcal{F} = \mathrm{th}(S^0)$. The structure of $\mathcal{F}$ is determined by the Morava K-theories $K(i)$. For each $i \geq 0$, $K(i)$ is a skew field object in $\mathcal{S}$, such that $K(i) \wedge K(j) = 0$ when $i \neq j$. If X is a finite spectrum and $K(j) \wedge X = 0$, then $K(j-1) \wedge X = 0$. We say a finite spectrum X is *type n* if n is the smallest integer such that $K(n) \wedge X \neq 0$. Define $\mathcal{C}_n = \langle K(n-1) \rangle \cap \mathcal{F}$. Then every thick subcategory of $\mathcal{F}$ is $\mathcal{C}_n$ for some n. It follows that any two spectra of type n generate the same thick subcategory, and hence Bousfield class; let $F(n)$ denote a generic type n spectrum.

Given a type n spectrum X, there is a v_n self-map $f : \Sigma^d X \to X$ for which $[S^0, K(n) \wedge f]_i$ is an isomorphism for all i and for which $[S^0, K(m) \wedge f]_j = 0$ for all j and $m \neq n$. We define $f^{-1}X$ to be the telescope, i.e., sequential or homotopy colimit, of the diagram $X \to \Sigma^{-d} X \to \cdots$. By the periodicity theorem, any choice of v_n self-map f yields an isomorphic telescope. The telescopes of different type n spectra are Bousfield equivalent; denote this class by $\langle T(n) \rangle$.

As mentioned above, localization away from a finite spectrum $F(n+1)$ exists and is smashing. This localization functor is denoted L_n^f and is the same as homological localization at $T(0) \vee \cdots \vee T(n)$.

Let $E(n)$ denote the Johnson–Wilson spectrum; this is a ring spectrum with $\langle E(n) \rangle = \langle K(0) \vee \cdots \vee K(n) \rangle$. Define $L_n : \mathcal{S} \to \mathcal{S}$ to be homological localization at $E(n)$. A deep theorem of Ravenel [1992, Theorem 7.5.6] shows that L_n is smashing for all n. The functors L_n^f and L_n are the only known smashing localization functors on $\mathcal{S}$.

The L_n^f-acyclics are given by $\mathrm{loc}(F(n+1)) = \mathrm{loc}(\mathcal{C}_{n+1}) = \mathrm{loc}(\langle K(n) \rangle \cap \mathcal{F}) = \mathrm{loc}(\langle E(n) \rangle \cap \mathcal{F})$. The L_n-acyclics are $\langle E(n) \rangle = \mathrm{loc}(\langle E(n) \rangle)$. Thus every L_n^f-acyclic is L_n-acyclic, and we have $\langle K(0) \vee \cdots \vee K(n) \rangle \leq \langle T(0) \vee \cdots \vee T(n) \rangle$ for all n. It follows that there is a natural map $L_n^f \to L_n$.

For convenience later, we collect some calculations in $\mathcal{S}$.

Lemma 2.10. *In* $\mathrm{BL}(\mathcal{S})$ *we have the following.*

(1) $\langle F(m) \rangle \leq \langle F(n) \rangle$ *if and only if* $m \geq n$. *For all* n *and* m, $\langle F(m) \wedge F(n) \rangle \neq \langle 0 \rangle$. *Furthermore*, $\langle F(n) \wedge F(n) \rangle = \langle F(n) \rangle$ *for all* n.

(2) $\langle F(m) \wedge T(n) \rangle = \langle 0 \rangle$ *when* $m > n$, *and* $\langle F(m) \wedge T(n) \rangle = \langle T(n) \rangle$ *when* $m \leq n$.

(3) $\langle T(m) \wedge T(n) \rangle = \langle 0 \rangle$ *when* $m \neq n$, *and* $\langle T(n) \wedge T(n) \rangle = \langle T(n) \rangle$.

(4) $\langle F(m) \wedge K(n) \rangle = \langle 0 \rangle$ *when* $m > n$, *and* $\langle K(n) \rangle = \langle F(m) \wedge K(n) \rangle \leq \langle F(m) \rangle$ *when* $m \leq n$.

(5) $\langle T(m) \wedge K(n)\rangle = \langle 0\rangle$ *when* $m \neq n$, *and* $\langle K(n)\rangle = \langle T(n) \wedge K(n)\rangle \leq \langle T(n)\rangle$.

(6) $\langle K(m) \wedge K(n)\rangle = \langle 0\rangle$ *when* $m \neq n$, *and* $\langle K(n) \wedge K(n)\rangle = \langle K(n)\rangle$.

Proof. Part (1) is Theorem 14 of [Hopkins and Smith 1998], along with the observation that $\langle F(n)\rangle$ is complemented by $\langle L_{n-1}^f S^0\rangle$, and hence is in DL.

Part (2) is in [Ravenel 1993, 2.8(i)], [Mahowald and Sadofsky 1995, 6.2], and [Hovey and Palmieri 1999, Section 5]. Part (3) is also in [Hovey and Palmieri 1999, Section 5].

Part (4) follows from the definition of type m spectra. Since each $K(i)$ is a skew field object, $F(m) \wedge K(i) \neq \langle 0\rangle$ implies this $K(i)$-module object $F(m) \wedge K(i)$ is a wedge of suspensions of $K(i)$.

From the periodicity theorem, $T(m)$ has nonzero $K(m)$ homology, and therefore $T(m) \wedge K(m) \neq 0$. The rest of Part (5) is in [Ravenel 1992, Proposition A.2.13]. Finally, Part (6) is well known. □

Let I denote the Brown–Comenetz dual of the sphere spectrum. Recall that a harmonic spectrum is one that is local with respect to $\bigvee_{i\geq 0} K(i)$. The following theorem plays a large role in the results in this paper.

Theorem 2.11. *There are no nonzero compact objects in the following categories.*

(1) *The* BP*-local category*

(2) *The harmonic category*

(3) *The* $H\mathbb{Z}$*-local category*

(4) *The* $H\mathbb{F}_p$*-local category*

(5) *The* I*-local category*

Proof. This is Corollary B.13 in [Hovey and Strickland 1999]. □

3. Local versions of the telescope conjecture

In this section, let $L = L_Z : \mathcal{S} \to \mathcal{S}$ be a localization functor for some $Z \in \mathcal{S}$, and let $\mathcal{L} = \mathcal{L}_Z$ denote the category of L-locals. First we state the various versions of the original telescope conjecture on $\mathcal{S}$.

Definition 3.1. Fix an integer $n \geq 0$. On $\mathcal{S}$, we have the following versions of the telescope conjecture.

$\mathrm{TC1}_n$	$\langle T(n)\rangle = \langle K(n)\rangle$.
$\mathrm{TC2}_n$	$L_n^f X \xrightarrow{\sim} L_n X$ for all X.
$\mathrm{TC3}_n$	If X is type n and f is a v_n self-map, then $L_n X \cong f^{-1} X$.
GSC	Every smashing localization is generated by a set of compact objects.
SDGSC	Every smashing localization is generated by a set of strongly dualizable objects.

Theorem 3.2. *On the category $\mathcal{S}$, we have the following statements.*

(1) $\mathrm{TC1}_n$ *holds if and only if* $\mathrm{TC3}_n$ *holds.*

(2) $\mathrm{TC2}_n$ *holds if and only if* $\mathrm{TC1}_i$ *holds for all* $i \le n$.

(3) *If* $\mathrm{TC2}_{n-1}$ *and* $\mathrm{TC1}_n$ *hold, then* $\mathrm{TC2}_n$ *holds.*

(4) GSC *holds if and only if* SDGSC *holds, and this implies* $\mathrm{TC2}_n$ *for all* n.

Remark 3.3. Note that if we quantify over all n, the first three versions of the telescope conjecture are equivalent. That is,

$$\mathrm{TC1}_n \text{ for all } n \iff \mathrm{TC2}_n \text{ for all } n \iff \mathrm{TC3}_n \text{ for all } n.$$

Remark 3.4. $\mathrm{TC2}_n$ holds if and only if $L_n^f S^0 \xrightarrow{\sim} L_n S^0$. Indeed, since both L_n^f and L_n are smashing, the subcategory of objects W such that $L_n^f W \xrightarrow{\sim} L_n W$ is localizing. Thus if it contains S^0, it contains $\mathrm{loc}(S^0) = \mathcal{S}$.

Proof. First we show the equivalence of $\mathrm{TC1}_n$ and $\mathrm{TC3}_n$. This is also sketched in [Mahowald et al. 2001, 1.13]. For any type n spectrum Y, $\mathrm{th}(Y) = \mathrm{th}(F(n))$ and so we have $\mathrm{th}(L_n Y) = \mathrm{th}(L_n F(n))$, and $\langle L_n Y\rangle = \langle L_n F(n)\rangle$. A construction in [Ravenel 1992, 8.3] gives a type n spectrum Y with $L_n Y \in \mathrm{th}(K(n))$. Thus $\langle L_n Y\rangle \le \langle K(n)\rangle$, and $0 \ne \langle L_n F(n)\rangle = \langle K(n)\rangle$. Suppose $\mathrm{TC3}_n$ holds. Then $\langle L_n Y\rangle = \langle f^{-1} Y\rangle = \langle T(n)\rangle$, and so $\langle T(n)\rangle = \langle K(n)\rangle$.

If X is type n and f is a v_n self-map, then [Mahowald and Sadofsky 1995, Proposition 3.2] implies that $L_n^f X \cong L_{T(n)} X \cong f^{-1} X$. Thus assuming $\mathrm{TC1}_n$, we have $L_{K(n)} X \cong f^{-1} X$, and to prove $\mathrm{TC3}_n$ it suffices to show that $L_n X \cong L_{K(n)} X$. This is known (see, e.g., [Hovey 1995a]), but we will give a proof that extends well to the localized setting. Since $\langle K(n)\rangle \le \langle E(n)\rangle$, localization at $K(n)$ gives a map $L_n X \to L_{K(n)} X$. It suffices to show that this is an L_n-equivalence. The fiber is $K(n)$ acyclic, so $L_n X \wedge K(n) \to L_{K(n)} X \wedge K(n)$ is an isomorphism. Consider $i < n$. The triangle $C_n X \wedge K(i) \to X \wedge K(i) \to L_n X \wedge K(i)$ shows that $L_n X \wedge K(i)$ is zero, because X is type n and $C_n X$ is $K(i)$ acyclic. Lemma 3.3.1 in [Hovey et al. 1997] states that $LW = LS^0 \wedge W$ for any localization L and strongly dualizable W. Since every finite spectrum is strongly dualizable, $L_{K(n)} X \wedge K(i) = L_{K(n)} S^0 \wedge X \wedge K(i) = 0$. Thus $L_n X \wedge K(i) \to L_{K(n)} X \wedge K(i)$ is an isomorphism for all $i \le n$, and hence $L_n X \to L_{K(n)} X$ is an L_n-equivalence.

For the second statement, we note that $\mathrm{TC2}_n$ is equivalent to the statement $\langle T(0) \vee \cdots \vee T(n)\rangle = \langle K(0) \vee \cdots \vee K(n)\rangle$. Smashing this with $\langle T(i)\rangle$, for $0 \le i \le n$, and using Lemma 2.10, yields $\mathrm{TC1}_i$ for each i. The third statement is also clear from this observation.

Finally, GSC holds if and only if SDGSC holds because objects in $\mathcal{S}$ are compact if and only if they are strongly dualizable. Given GSC, consider L_n. The GSC would imply that the L_n-acyclics are $\mathrm{loc}(\langle E(n)\rangle \cap \mathcal{F})$. As observed earlier, this is the same as $\mathrm{loc}(F(n+1))$, the L_n^f-acyclics. Therefore we would have $L_n^f \cong L_n$. □

The SDGSC is new, and we will discuss it first. As mentioned in the above proof, in $\mathcal{S}$ compactness is equivalent to strong dualizability. It is well known that localization away from a set of compact objects is smashing. The GSC is precisely the statement that the converse holds. However, as we will show next, one only needs strong dualizability to generate a smashing localization functor. We will state our result in slightly more general terms.

Theorem 3.5. *Let* T *be a well generated tensor triangulated category such that* $\mathrm{loc}(\mathbb{1}) = \mathsf{T}$, *as in Notation 2.1. Let* $A = \{B_\alpha\}$ *be a (possibly infinite) set of strongly dualizable objects. Then there exists a smashing localization functor* $L : \mathsf{T} \to \mathsf{T}$ *with* $\mathrm{Ker}\, L = \mathrm{loc}(A)$.

Proof. Let $E = \vee_\alpha B_\alpha$ and note that $\mathrm{loc}(E) = \mathrm{loc}(A)$. The category T is well generated by hypothesis. The localizing subcategory $\mathsf{S} = \mathrm{loc}(E)$ is also well generated, by [Iyengar and Krause 2013, Remark 2.2], and is tensor-closed by Lemma 2.5.

By [Iyengar and Krause 2013, Proposition 2.1] there exists a localization functor $L : \mathsf{T} \to \mathsf{T}$ with $\mathrm{Ker}\, L = \mathsf{S}$. We will show that L is a smashing localization.

First we claim that the L-locals are tensor-closed. For any $Y \in \mathsf{T}$, we have

$$\begin{aligned} Y \text{ is } L\text{-local} &\iff [W, Y]_n = 0 \text{ for all } W \in \mathsf{S} \text{ and all } n \in \mathbb{Z} \\ &\iff [E, Y]_n = \prod [B_\alpha, Y]_n = 0 \text{ for all } n \in \mathbb{Z} \\ &\iff [B_\alpha, Y]_n = 0 \text{ for all } \alpha \text{ and } n \in \mathbb{Z} \\ &\iff DB_\alpha \wedge Y = 0 \text{ for all } \alpha. \end{aligned}$$

The second equivalence follows from the fact that $\{X \mid [X, Y]_n = 0 \text{ for all } n \in \mathbb{Z}\}$ is a localizing subcategory containing E, and hence all of S. The final equivalence uses the fact that the B_α are strongly dualizable.

Now suppose Y is L-local and X is arbitrary. Then $DB_\alpha \wedge Y = 0$ for all α, so $DB_\alpha \wedge Y \wedge X = 0$ for all α, and thus $Y \wedge X$ is L-local. This shows that the L-locals are tensor-closed.

Consider the localization triangle $C\mathbb{1} \to \mathbb{1} \to L\mathbb{1}$, where $L\mathbb{1}$ is L-local and $C\mathbb{1} \in \mathsf{S}$. For arbitrary $X \in \mathsf{T}$, tensoring gives an exact triangle,

$$C\mathbb{1} \wedge X \to X \to L\mathbb{1} \wedge X.$$

The object $L\mathbb{1} \wedge X$ is L-local, since the locals are tensor-closed. Likewise, $C\mathbb{1} \wedge X \in \mathsf{S}$, since S is tensor-closed and so $L(C\mathbb{1} \wedge X) = 0$. Therefore $X \to L\mathbb{1} \wedge X$ is an L-equivalence from X to an L-local object, and it follows that $LX \cong L\mathbb{1} \wedge X$. This shows that L is a smashing localization. □

In the stable homotopy category, and more generally whenever $\mathbb{1} \in \mathsf{T}$ is compact, this gives nothing new; by [Hovey et al. 1997, Theorem 2.1.3(d)] compact and strongly dualizable are equivalent. Consider, however, the harmonic category, which has no nonzero compact objects. In Section 4 we classify all smashing localizations on the harmonic category; they are indexed by $\mathbb{N}$. Thus the GSC fails in the harmonic category but, as we show in Theorem 4.4, the SDGSC holds. In fact, in the following sections we will give several examples of categories where the GSC fails but the SDGSC holds.

On the other hand, we do not expect the SDGSC to hold in complete generality, since Keller's counterexample [1994] to the GSC is also a counterexample to the SDGSC; in the derived category of a ring R, the unit R is compact and strongly dualizable, so the GSC and SDGSC are equivalent.

The GSC and SDGSC make sense in any localized category, but $\mathrm{TC1}_n$, $\mathrm{TC2}_n$, and $\mathrm{TC3}_n$ may not, since $T(n)$ and $K(n)$ may not be objects in $\mathcal{L}$. Instead we present the following definitions.

Definition 3.6. Let $L : \mathcal{S} \to \mathcal{S}$ be a localization, and $\mathcal{L}$ the category of L-locals.

(1) Let $l_n^f : \mathcal{L} \to \mathcal{L}$ denote localization at $\langle LT(0) \vee LT(1) \vee \cdots \vee LT(n)\rangle$.

(2) Let $l_n : \mathcal{L} \to \mathcal{L}$ denote localization at $\langle LK(0) \vee LK(1) \vee \cdots \vee LK(n)\rangle$.

Before stating and proving a local version of Theorem 3.2, we establish some results about l_n^f and l_n. First we make an observation about calculations in $\mathrm{BL}(\mathcal{L})$.

Lemma 3.7. *All the calculations in Lemma 2.10 are valid in* $\mathrm{BL}(\mathcal{L})$ *if we replace* $F(n)$, $T(n)$, *and* $K(n)$ *with* $LF(n)$, $LT(n)$, *and* $LK(n)$.

Proof. This follows from Theorem 2.3 and the statements in Lemma 2.10. □

Proposition 3.8. *The functor* l_n^f *is localization away from* $LF(n+1)$, *and hence is smashing.*

Proof. By Theorem 3.5 we know that there is some smashing localization functor $l : \mathcal{L} \to \mathcal{L}$ that is localization away from $LF(n+1)$; we wish to show $l = l_n^f$. Let $\mathbb{1} = LS^0$ for simplicity of notation, and let c denote the colocalization corresponding to l. We claim that the l-acyclics are precisely $\mathrm{loc}(c\mathbb{1})$. Clearly $c\mathbb{1}$ is l-acyclic, and these are a localizing subcategory, so $\mathrm{loc}(c\mathbb{1}) \subseteq \{l\text{-acyclics}\}$. On the other hand, suppose W is l-acyclic. Because l is smashing, $LW = W \wedge l\mathbb{1} = 0$, so $W = W \wedge c\mathbb{1}$. Then since $W \in \mathrm{loc}(\mathbb{1}) = \mathcal{L}$, we have $W = W \wedge c\mathbb{1} \in \mathrm{loc}(\mathbb{1} \wedge c\mathbb{1}) = \mathrm{loc}(c\mathbb{1})$, proving the claim.

By definition, the l-acyclics are also given by $\mathrm{loc}(LF(n+1))$. Therefore, we have $\langle LF(n+1)\rangle = \langle c\mathbb{1}\rangle$.

The class $\langle F(n+1)\rangle$ is complemented by $\langle T(0) \vee \cdots \vee T(n)\rangle$ in $\mathrm{BL}(\mathcal{S})$ [Hovey and Palmieri 1999, Section 5], and therefore $\langle LF(n+1)\rangle$ is complemented by

$\langle LT(0) \vee \cdots \vee LT(n)\rangle$ in $\mathrm{BL}(\mathcal{L})$. At the same time, $\langle c\mathbb{1}\rangle$ is complemented by $\langle l\mathbb{1}\rangle$, and complements are unique. We conclude that $\{l\text{-acyclics}\} = \langle l\mathbb{1}\rangle = \langle LT(0) \vee \cdots \vee LT(n)\rangle$. Since l and l_n^f are two localizations on $\mathcal{L}$ with the same acyclics, they are equal. □

Lemma 3.9. *If L is smashing, then $l_n^f = LL_n^f = L_n^f L$ and $l_n = LL_n = L_n L$, and both are smashing.*

Proof. Smashing localization functors always commute, and they compose to give a smashing localization. The functor $LL_n : \mathcal{S} \to \mathcal{S}$, sending $X \mapsto L(L_n S^0 \wedge X) = LS^0 \wedge L_n S^0 \wedge X$ is a smashing localization. Since LL_n-locals are L-local, it also gives a smashing localization on $\mathcal{L}$. The acyclics of this functor are $\langle LL_n S^0\rangle$ in $\mathrm{BL}(\mathcal{L})$, which is $\langle LK(0) \vee \cdots \vee LK(n)\rangle$. Thus LL_n and l_n are localizations on $\mathcal{L}$ with the same acyclics, and hence isomorphic. The same proof works for $l_n^f = LL_n^f$. □

In the category $\mathcal{S}$, for a type n finite spectrum X with a v_n map $f : \Sigma^d X \to X$ and telescope $f^{-1}X$, it is known [Mahowald and Sadofsky 1995, Proposition 3.2] that $L_n^f X \cong L_{T(n)} X \cong f^{-1}X$. The following proposition shows that the local version of this result holds as well.

Lemma 3.10. *Let $L : \mathcal{S} \to \mathcal{S}$ be a localization, and l_n^f, X and $f^{-1}X$ as above. Then*

$$l_n^f(LX) \cong L_{LT(n)}(LX) \cong L(f^{-1}X).$$

Proof. The proof parallels the [Mahowald and Sadofsky 1995] result; one must only check that everything works when localized. If $L_{LT(n)}(LX) \cong L(f^{-1}X)$ holds for a single type n spectrum, then it holds for all type n spectra. So without loss of generality, we can choose X to be a type n spectrum that is a ring object in $\mathcal{S}$. Then for any v_n self-map f, the telescope $f^{-1}X$ is also a ring object [Mahowald and Sadofsky 1995, Lemma 2.2]. By Lemma 2.5, $L(f^{-1}X)$ is a ring object in $\mathcal{L}$, and hence is local with respect to itself.

Lemma 2.2 in [Mahowald and Sadofsky 1995] shows that we have $X \wedge f^{-1}X \cong f^{-1}X \wedge f^{-1}X$ in $\mathcal{S}$, so $LX \wedge_{\mathcal{L}} L(f^{-1}X) \cong L(f^{-1}X) \wedge_{\mathcal{L}} L(f^{-1}X)$ in $\mathcal{L}$ and the canonical map $LX \to L(f^{-1}X)$ is an $L(f^{-1}X)$-equivalence. It follows that $L_{LT(n)}(LX) \cong L(f^{-1}X)$.

Since $\langle LT(n)\rangle \le \langle LT(0) \vee \cdots \vee LT(n)\rangle$, we have that $L(f^{-1}X)$ is l_n^f-local. One then uses Lemma 3.7 to see that $LX \to L(f^{-1}X)$ is a l_n^f-equivalence, and so $l_n^f(LX) = l_n^f(L(f^{-1}X)) = L(f^{-1}X)$. □

Definition 3.11. Let $L : \mathcal{S} \to \mathcal{S}$ be a localization, and consider the category $\mathcal{L}$ of locals. Fix an $n \ge 0$. We have the following versions of the telescope conjecture on $\mathcal{L}$.

LTC1$_n$	$\langle LT(n)\rangle = \langle LK(n)\rangle$.
LTC2$_n$	$l_n^f X \xrightarrow{\sim} l_n X$ for all X.
LTC3$_n$	If $X \in \mathcal{S}$ is type n and f is a v_n self-map, then $l_n(LX) \cong L(f^{-1}X)$.
GSC	Every smashing localization is generated by a set of compact objects.
SDGSC	Every smashing localization is generated by a set of strongly dualizable objects.

Theorem 3.12. *On the category $\mathcal{L}$, we have the following statements.*

(1) LTC1$_n$ *implies* LTC3$_n$.

(2) LTC2$_n$ *holds if and only if* LTC1$_i$ *holds for all $i \le n$.*

(3) *If* LTC2$_{n-1}$ *and* LTC1$_n$ *hold*, *then* LTC2$_n$ *holds.*

Proof. Note that LTC2$_n$ is equivalent to the statement $\langle LT(0) \vee \cdots \vee LT(n)\rangle = \langle LK(0) \vee \cdots \vee LK(n)\rangle$, so the last two statements are clear. We will show that LTC1$_n$ implies LTC3$_n$ by mimicking the proof in Theorem 3.2.

If X is type n and f is a v_n self-map, Lemma 3.10 shows that $l_n^f(LX) \cong L_{LT(n)}(LX) \cong L(f^{-1}X)$. Then LTC1$_n$ implies $L_{LK(n)}LX \cong L(f^{-1}X)$. So it suffices to show that $l_n(LX) = L_{LK(n)}LX$. Now we must show that the map $L_{LK(n)} : l_n(LX) \to L_{LK(n)}LX$ is an l_n-equivalence. The same reasoning as in Theorem 3.2, along with the computations of Lemma 3.7 and some definition unwinding, gives us that $l_n(LX) \wedge LK(i) \to L_{LK(n)}LX \wedge LK(i)$ is an equivalence for all $i \le n$; we only need to notice that Lemma 3.3.1 in [Hovey et al. 1997] applies to strongly dualizable objects, and that LX is strongly dualizable. □

Theorem 3.13. *If*, *furthermore*, *$L : \mathcal{S} \to \mathcal{S}$ is a smashing localization*, *then on the category $\mathcal{L}$ of locals we have that* LTC3$_n$ *implies* LTC1$_n$, *and*

$$\text{GSC} \iff \text{SDGSC} \Rightarrow \text{LTC2}_n \textit{ for all } n.$$

Remark 3.14. In this case, LTC2$_n$ is equivalent to $l_n^f(LS^0) \xrightarrow{\sim} l_n(LS^0)$, since by Lemma 3.9 both l_n^f and l_n are smashing, so the argument in Remark 3.4 applies.

Proof. By [Hovey et al. 1997, Theorem 2.1.3(d)], the compact objects and strongly dualizable objects in $\mathcal{L}$ coincide. Thus GSC holds if and only if SDGSC holds, and this implies LTC2$_n$ just as in Theorem 3.2.

Suppose X has type n and LTC3$_n$ holds. As in the proof of Theorem 3.2, $\langle L_nF(n)\rangle = \langle K(n)\rangle$ in BL($\mathcal{S}$), and so $\langle LL_nLF(n)\rangle = \langle LK(n)\rangle$ in BL($\mathcal{L}$). By Lemma 3.9, we have $\langle l_nLF(n)\rangle = \langle LK(n)\rangle$. Then LTC3$_n$ implies that $\langle LT(n)\rangle = \langle L(f^{-1}X\rangle = \langle l_n(LX)\rangle = \langle l_nLF(n)\rangle$, so LTC1$_n$ holds. □

Question 3.15. Is $l_n : \mathcal{L} \to \mathcal{L}$ always a smashing localization?

This is the case in all the local categories investigated in this paper, whether or not $L : \mathcal{S} \to \mathcal{S}$ is a smashing localization. If one could show l_n is always smashing, then most likely on $\mathcal{L}$ one would have that SDGSC implies LTC2$_n$ for all n.

We would of course like to know if and when information on localized telescope conjectures can help with those in the original category $\mathcal{S}$, where all versions remain open.

Proposition 3.16. *Let $L : \mathcal{S} \to \mathcal{S}$ be a localization, with localized category $\mathcal{L}$.*

(1) *If* $\mathrm{TC1}_n$ *holds on $\mathcal{S}$, then* $\mathrm{LTC1}_n$ *holds on $\mathcal{L}$.*

(2) *If* $\mathrm{TC2}_n$ *holds on $\mathcal{S}$, then* $\mathrm{LTC2}_n$ *holds on $\mathcal{L}$.*

(3) *If* $\mathrm{TC3}_n$ *holds on $\mathcal{S}$, then* $\mathrm{LTC3}_n$ *holds on $\mathcal{L}$.*

Furthermore, if L is a smashing localization, then we have the following.

(4) *If* GSC *holds on $\mathcal{S}$, then* GSC *holds on $\mathcal{L}$.*

(5) *If* SDGSC *holds on $\mathcal{S}$, then* SDGSC *holds on $\mathcal{L}$.*

Proof. Part (1) follows immediately from Proposition 2.6. So does Part (2), since $\mathrm{TC2}_n$ is equivalent to the statement $\langle T(0) \vee \cdots \vee T(n) \rangle = \langle K(0) \vee \cdots \vee K(n) \rangle$, and similarly for $\mathrm{LTC2}_n$. From this and Theorems 3.2 and 3.12 we have

$$\mathrm{TC3}_n \iff \mathrm{TC1}_n \Rightarrow \mathrm{LTC1}_n \Rightarrow \mathrm{LTC3}_n.$$

Now suppose L is smashing and the GSC holds on $\mathcal{S}$. Let $l : \mathcal{L} \to \mathcal{L}$ be a smashing localization. Thus l is defined by $l(LY) = l(LS^0) \wedge_{\mathcal{L}} LY = lS^0 \wedge LS^0 \wedge Y$. We can therefore extend l to be a smashing localization on all of $\mathcal{S}$, with the map $X \mapsto lS^0 \wedge LS^0 \wedge X = lLS^0 \wedge X$. Since the GSC holds on $\mathcal{S}$ by assumption, the acyclics of this functor are $\langle lLS^0 \rangle = \mathrm{loc}(A)$, for some set of compact objects A in $\mathcal{S}$. Here $\langle lLS^0 \rangle$ refers to the Bousfield class in $\mathrm{BL}(\mathcal{S})$.

We must show that $\langle lLS^0 \rangle$ in $\mathrm{BL}(\mathcal{L})$ is generated by a set of objects that are compact in $\mathcal{L}$. Note that $\langle lLS^0 \rangle$ in $\mathrm{BL}(\mathcal{L})$ is $\{LW \mid LW \wedge_{\mathcal{L}} lLS^0 = 0\} = \{LW \mid LW \wedge_{\mathcal{S}} lLS^0 = 0\} = \langle lLS^0 \rangle \cap \mathcal{L}$, where the latter $\langle lLS^0 \rangle$ is in $\mathrm{BL}(\mathcal{S})$. Therefore $\langle lLS^0 \rangle$ in $\mathrm{BL}(\mathcal{L})$ is $\mathrm{loc}(A) \cap \mathcal{L}$. We claim that this is $\mathrm{loc}(L(A))$. Since L sends compacts to compacts, this will show that l is generated by a set of compacts.

If $X \in \mathrm{loc}(A)$, then $LX \in \mathrm{loc}(L(A))$. If $X \in \mathcal{L}$ in addition, then we have $X \cong LX \in \mathrm{loc}(L(A))$. For the other inclusion, note that the intersection of two localizing subcategories is a localizing subcategory, and $\mathcal{L}$ is a localizing subcategory of $\mathcal{S}$ because L is smashing. For $Y \in A$, LY is in $\mathcal{L}$, and $LY = LS^0 \wedge Y \in \mathrm{loc}(Y) \subseteq \mathrm{loc}(A)$. Therefore $L(A) \subseteq \mathrm{loc}(A) \cap \mathcal{L}$, and $\mathrm{loc}(L(A)) = \mathrm{loc}(A) \cap \mathcal{L}$.

Part (5) follows immediately, since if L is smashing then GSC holds if and only if SDGSC in both $\mathcal{S}$ and $\mathcal{L}$. □

Balmer and Favi [2011, Proposition 4.4] have also recently proved Part (4) in the slightly more general setting of a smashing localization on a unital algebraic stable homotopy category; the above proof would apply there as well. One would like to prove Part (5) without the assumption that L is smashing, but it's not clear if this is possible.

Letting $L = L_Z$, for $Z = \bigvee_{i\geq 0} K(i)$, $E(n)$, $K(n)$, BP, $H\mathbb{F}_p$, or I provides interesting examples of categories $\mathcal{L}$ on which to investigate these telescope conjectures. Furthermore, $\mathrm{LTC1}_n$ suggests the relevance of Bousfield lattices to understanding these questions. In the remaining sections, we focus on specific localized categories.

4. The Harmonic category

Let $Q = \bigvee_{i\geq 0} K(i)$ and $L = L_Q : \mathcal{S} \to \mathcal{S}$, and consider the harmonic category $\mathcal{H}$ of L-locals. Harmonic localization is not smashing. An object is called *harmonic* if it is L-local and *dissonant* if it is L-acyclic. For example, finite spectra, suspension spectra, finite torsion spectra, and BP are known to be harmonic [Hovey 1995a; Ravenel 1984]. On the other hand, I and $H\mathbb{F}_p$ are dissonant.

In order to answer the telescope conjectures in $\mathcal{H}$, we will first calculate the Bousfield lattice of $\mathcal{H}$. In this section all smash products are in $\mathcal{H}$ unless otherwise noted. Given any set P, let 2^P denote the power set of P.

Definition 4.1. Given $X \in \mathcal{H}$, define the *support* of X to be

$$\mathrm{supp}(X) = \{i \mid X \wedge K(i) \neq 0\} \subseteq \mathbb{N}.$$

The following result and proof was pointed out to us by Jon Beardsley.

Proposition 4.2. *The Bousfield lattice of* $\mathcal{H}$ *is* $2^{\mathbb{N}}$.

Proof. Each $K(n)$ is a ring object, and hence $K(n)$-local by Lemma 2.5. Because $\langle K(n)\rangle \leq \langle Q\rangle$, $K(n)$-locals are harmonic, thus each $K(n)$ is harmonic. The argument hinges on the fact that $K(n)$ is a skew field object in $\mathcal{H}$: for $X = LX$ in $\mathcal{H}$, if $X \wedge K(n) \neq 0$ then $X \wedge K(n) = L(X \wedge_{\mathcal{S}} K(n))$ so $X \wedge_{\mathcal{S}} K(n) \neq 0$, and $X \wedge_{\mathcal{S}} K(n)$ must be a nonempty wedge of suspensions of $K(n)$s. Thus

$$\begin{aligned} X \wedge K(n) &= L(X \wedge_{\mathcal{S}} K(n)) = L(\vee \Sigma^i K(n)) = L(\vee \Sigma^i LK(n)) \\ &= \coprod_{\mathcal{L}} \Sigma^i LK(n) = \coprod_{\mathcal{L}} \Sigma^i K(n). \end{aligned}$$

It follows that $LX \wedge K(n) = 0$ if and only if $LX \wedge_{\mathcal{S}} K(n) = 0$. Furthermore, if $LX \wedge K(n) \neq 0$, then $\langle LX \wedge K(n)\rangle = \langle K(n)\rangle$, where these are Bousfield classes in $\mathrm{BL}(\mathcal{H})$.

By the definition of L, for any $W \in \mathcal{S}$, if $W \wedge_{\mathcal{S}} K(n) = 0$ for all n, then $LW = 0$. Combining this with the above observation, we get that a local object $W = LW$ has $W \wedge K(n) = 0$ in $\mathcal{H}$ for all n if and only if $W = 0$.

Therefore, for any $X, Y \in \mathcal{H}$, we have

$$Y \wedge X = 0 \iff Y \wedge X \wedge K(n) = 0 \ \text{ for all } n \iff Y \wedge K(n) = 0 \ \text{ for all } n \in \mathrm{supp}(X).$$

We conclude that there is a lattice isomorphism $F : \mathrm{BL}(\mathcal{H}) \to 2^{\mathbb{N}}$, given by

$$\langle X\rangle = \bigvee_{\mathrm{supp}(X)} \langle K(i)\rangle \mapsto \mathrm{supp}(X), \quad N \subseteq \mathbb{N} \mapsto \bigvee_{i\in N} \langle K(i)\rangle. \qquad \square$$

Theorem 4.3. *On $\mathcal{H}$, for all $n \geq 0$, we have that* $\mathrm{LTC1}_n$, $\mathrm{LTC2}_n$, *and* $\mathrm{LTC3}_n$ *hold.*

Proof. By Lemmas 2.10 and 3.7, $LT(n)$ and $LK(n) = K(n)$ have the same support. The above theorem then implies that $\langle LT(n)\rangle = \langle LK(n)\rangle$. Thus $\mathrm{LTC1}_n$ holds for all n, and the claim follows from Theorem 3.12. $\square$

Next, we classify all smashing localizations on $\mathcal{H}$, and show that the GSC fails but the SDGSC holds. The proof is based on that of [Hovey and Strickland 1999, Theorem 6.14], which classifies smashing localizations in the $E(n)$-local category.

Theorem 4.4. *If $L' : \mathcal{H} \to \mathcal{H}$ is a smashing localization functor, then $L' = l_n^f$ for some $n \geq 0$, or $L' = 0$ or $L' = \mathrm{id}$. Therefore the* GSC *fails but the* SDGSC *holds on $\mathcal{H}$.*

Proof. Let $L' : \mathcal{H} \to \mathcal{H}$ be a smashing localization functor, and let $\mathbb{1} = LS^0$ be the unit in $\mathcal{H}$. The acyclics of L' are given by $\langle L'\mathbb{1}\rangle$. From Proposition 4.2, $\langle L'\mathbb{1}\rangle$ is equal to the wedge of $\langle K(i)\rangle$ for all $i \in \mathrm{supp}(L'\mathbb{1})$. If $\mathrm{supp}(L'\mathbb{1}) = \varnothing$ then $\langle L'\mathbb{1}\rangle = \langle 0\rangle$ and $L' = 0$.

Assume now that $\mathrm{supp}(L'\mathbb{1})$ is not empty, and take $j \in \mathrm{supp}(L'\mathbb{1})$. We will show that $\langle L'\mathbb{1}\rangle \geq \langle K(0) \vee \cdots \vee K(j)\rangle$. It follows that either $\langle L'\mathbb{1}\rangle = \bigvee_{i\geq 0}\langle K(i)\rangle = \langle \mathbb{1}\rangle$ and $L' = \mathrm{id}$, or $L' = l_n = l_n^f$ for $n = \max(\mathrm{supp}(L'\mathbb{1}))$.

Since $\langle K(j)\rangle \leq \langle L'\mathbb{1}\rangle$, from Lemma 2.8 we have $L_{K(j)}L' = L'L_{K(j)} = L_{K(j)}$. Therefore $\langle L_{K(j)}\mathbb{1}\rangle = \langle L'\mathbb{1} \wedge L_{K(j)}\mathbb{1}\rangle \leq \langle L'\mathbb{1}\rangle$. Proposition 5.3 of [Hovey and Strickland 1999] shows that, in $\mathcal{S}$, $L_{K(j)}S^0 \wedge_{\mathcal{S}} K(i)$ is nonzero for $0 \leq i \leq j$ and zero for $i > j$. Note that $L_{K(j)}S^0 = L_{K(j)}LS^0 = L_{K(j)}\mathbb{1}$ and, as remarked in the proof of Proposition 4.2, $LX \wedge K(i) = 0$ if and only if $LX \wedge_{\mathcal{S}} K(i) = 0$. Therefore, in $\mathrm{BL}(\mathcal{H})$ we have $\langle L_{K(j)}\mathbb{1}\rangle = \langle K(0) \vee \cdots \vee K(j)\rangle$, and so $\langle L'\mathbb{1}\rangle \geq \langle K(0) \vee \cdots \vee K(j)\rangle$ as desired.

Each l_n^f is localization away from $LF(n+1)$ by Proposition 3.8, which is strongly dualizable by Theorem 2.3. The identity is localization away from zero, and the zero functor is localization away from LS^0; these are both strongly dualizable. Therefore the SDGSC holds. On the other hand, Corollary B.13 in [Hovey and Strickland 1999] shows that there are no nonzero compact objects in $\mathcal{H}$, so the GSC fails. $\square$

Question 4.5. Classify localizing subcategories of $\mathcal{H}$.

It seems likely that every localizing subcategory of $\mathcal{H}$ is a Bousfield class, and so these are in bijection with $2^{\mathbb{N}}$, but we have been unable to prove this.

5. The $E(n)$- and $K(n)$- local categories

5A. ***The $E(n)$-local category.*** Recall that $\langle E(n)\rangle = \langle K(0) \vee K(1) \vee \cdots \vee K(n)\rangle$. In this section, fix $L = L_n = L_{E(n)} : \mathcal{S} \to \mathcal{S}$ and let $\mathcal{L}_n$ denote the local category. The functor L_n is smashing, and so by Notation 2.1 each $LF(i)$ is compact in $\mathcal{L}_n$. Hovey and Strickland [1999] have studied $\mathcal{L}_n$ in detail and determined the localizing subcategories, smashing localizations, and Bousfield lattice of $\mathcal{L}_n$. We begin by recalling these results.

Lemma 5.1. *For $0 \le i \le n$ we have $LK(i) = K(i)$, and for $i > n$ we have $LK(i) = 0$.*

Proof. This follows from $\langle E(n)\rangle = \langle K(0) \vee K(1) \vee \cdots \vee K(n)\rangle$. □

Theorem 5.2 [Hovey and Strickland 1999, Theorem 6.14]. *The lattice of localizing subcategories of $\mathcal{L}_n$, ordered by inclusion, is in bijection with the lattice of subsets of the set $\{0, 1, \ldots, n\}$, where a localizing subcategory S corresponds to*

$$\{i \mid K(i) \in \mathsf{S}\}.$$

Corollary 5.3. *Every localizing subcategory of $\mathcal{L}_n$ is a Bousfield class, in particular a localizing subcategory S is the Bousfield class*

$$\bigvee \langle K(j) \mid K(j) \notin \mathsf{S}, 0 \le j \le n\rangle.$$

Corollary 5.4. *For every $n \ge 0$, there is a lattice isomorphism*

$$f_n : \mathrm{BL}(\mathcal{L}_n) \xrightarrow{\sim} 2^{\{0,1,\ldots,n\}}.$$

Proof. The isomorphism is given by

$$\langle X\rangle = \textstyle\bigvee_{X \wedge K(i) \neq 0} \langle K(i)\rangle \mapsto \{i \mid X \wedge K(i) \neq 0\},$$

$$N \subseteq \{0, 1, \ldots, n\} \mapsto \textstyle\bigvee_{i \in N} \langle K(i)\rangle.$$ □

Theorem 5.5 [Hovey and Strickland 1999, Corollary 6.10]. *If $L' : \mathcal{L}_n \to \mathcal{L}_n$ is a smashing localization, then $L' = L_i = L_i^f$ for some $0 \le i \le n$ or $L' = 0$. Thus the* GSC *holds on $\mathcal{L}_n$.*

Corollary 5.6. *On $\mathcal{L}_n$, all of* $\mathrm{LTC1}_i$, $\mathrm{LTC2}_i$, $\mathrm{LTC3}_i$ *hold for all i, and* GSC *and* SDGSC *also hold.*

Proof. This follows from Theorems 5.5 and 3.13. Note that for $i > n$, we have $LT(i) = 0 = LK(i)$ by Lemma 2.10, and so $l_i = l_n = l_n^f = l_i^f$. □

Recall that there is a natural map $L_n X \to L_{n-1} X$ for all X in $\mathcal{S}$ and n, and by Proposition 2.6 this induces a surjective lattice map $\mathrm{BL}(\mathcal{L}_n) \to \mathrm{BL}(\mathcal{L}_{n-1})$ and an inverse system of lattice maps.

$$\cdots \to \mathrm{BL}(\mathcal{L}_n) \to \mathrm{BL}(\mathcal{L}_{n-1}) \to \cdots \to \mathrm{BL}(\mathcal{L}_1) \to \mathrm{BL}(\mathcal{L}_0)$$

Proposition 5.7. *The lattice isomorphisms F and f_n from Proposition 4.2 and Corollary 5.4 realize* $\mathrm{BL}(\mathcal{H})$ *as the inverse limit of the maps* $\mathrm{BL}(\mathcal{L}_n) \to \mathrm{BL}(\mathcal{L}_{n-1})$.

Proof. By Lemma 2.8 and the facts that $L_Q K(i) = K(i)$ for all i, and $L_n K(i) = K(i)$ for $i \le n$ and $L_n K(i) = 0$ for $i > n$, we get the following diagram for all n. The map $2^{\{0,1,\dots,n\}} \to 2^{\{0,1,\dots,n-1\}}$ is induced by sending $m \mapsto m$ for $m < n$ but $n \mapsto 0$, and the maps $2^{\mathbb{N}} \to 2^{\{0,1,\dots,i\}}$ are defined similarly.

$$\begin{tikzcd}
 & \mathrm{BL}(\mathcal{H}) \arrow[rr, leftrightarrow, "F"'] \arrow[dl, two heads] \arrow[dd, two heads] & & 2^{\mathbb{N}} \arrow[dl, two heads] \arrow[dd, two heads] \\
\mathrm{BL}(\mathcal{L}_n) \arrow[rr, leftrightarrow, "f_n"'] \arrow[dr, two heads] & & 2^{\{0,1,\dots,n\}} \arrow[dr, two heads] & \\
 & \mathrm{BL}(\mathcal{L}_{n-1}) \arrow[rr, leftrightarrow, "f_{n-1}"'] & & 2^{\{0,1,\dots,n-1\}}
\end{tikzcd}$$ □

5B. ***The $K(n)$-local category.*** Although an incredibly complicated category in its own right, the $K(n)$-local category is quite basic from the perspective of localizing subcategories, Bousfield lattices, and telescope conjectures. In this subsection, let $L = L_{K(n)} : \mathcal{S} \to \mathcal{S}$ be localization at $K(n)$, and let $\mathcal{K}_n$ denote the category of locals. Hovey and Strickland classify the localizing subcategories of $\mathcal{K}_n$, and there are not many of them.

Proposition 5.8 [Hovey and Strickland 1999, Theorem 7.5]. *There are no nonzero proper localizing subcategories of $\mathcal{K}_n$.*

This Proposition implies that the Bousfield lattice of $\mathcal{K}_n$ is the two-element lattice $\{\langle 0\rangle, \langle K(n)\rangle\}$. We will prove a slightly more general result that will be used again in Section 6A.

Proposition 5.9. *Consider a category T as in Notation 2.1, an object Z in T, and localization $L_Z : \mathsf{T} \to \mathsf{T}$ with localized category $\mathcal{L}_Z$.*

(1) *If Z is a ring object, then $\langle L_Z Z\rangle = \langle Z\rangle$ is the maximum class in* $\mathrm{BL}(\mathcal{L}_Z)$.

(2) *If Z is a skew field object, then* $\mathrm{BL}(\mathcal{L}_Z)$ *is the two-element lattice* $\{\langle 0\rangle, \langle Z\rangle\}$.

Proof. For (1), note that Lemma 2.5 implies $L_Z Z = Z$. Consider $\langle Z\rangle$ in $\mathrm{BL}(\mathcal{L}_Z)$. By definition, this is the collection of all $W \in \mathcal{L}_Z$ with $L(Z \wedge_{\mathsf{T}} W) = 0$. But $Z \wedge_{\mathsf{T}} W$ is a Z-module object in T, and hence L_Z-local. The only object that is both local and acyclic with respect to any localization is zero, so $Z \wedge_{\mathsf{T}} W = 0$. But this says that W is L_Z-acyclic, and hence zero in $\mathcal{L}_Z$. Therefore, in $\mathrm{BL}(\mathcal{L}_Z)$ we have $\langle Z\rangle = \{0\}$.

Now suppose Z is a skew field object in T. In particular, it is a ring object, so $\langle Z\rangle$ is the maximum class in $\mathrm{BL}(\mathcal{L}_Z)$. Consider $\langle LX\rangle$ in $\mathrm{BL}(\mathcal{L}_Z)$, for arbitrary $X \in \mathsf{T}$. If $X \wedge_{\mathsf{T}} Z = 0$, then $LX = 0$. Otherwise, $X \wedge_{\mathsf{T}} Z$ is a wedge of suspensions of Z, so $\langle Z\rangle = \langle X \wedge_{\mathsf{T}} Z\rangle \leq \langle X\rangle$ in $\mathrm{BL}(\mathsf{T})$. Then $\langle Z\rangle = \langle L_Z Z\rangle \leq \langle L_Z X\rangle$ in $\mathrm{BL}(\mathcal{L}_Z)$, so $\langle L_Z X\rangle = \langle Z\rangle$. □

Corollary 5.10. *The Bousfield lattice of $\mathcal{K}_n$ is* $\{\langle 0\rangle, \langle K(n)\rangle\}$.

Theorem 5.11. *In $\mathcal{K}_n$, all of* $\mathrm{LTC1}_i$, $\mathrm{LTC2}_i$, $\mathrm{LTC3}_i$ *hold for all i, and* GSC *and* SDGSC *also hold.*

Proof. In light of Theorem 3.12, we will show that $\mathrm{LTC1}_i$ holds for all i. This follows from Lemma 2.10: for $i \neq n$ we have $LT(i) = 0 = LK(i)$, but $LT(n) \neq 0$ so by the last corollary $\langle LT(n)\rangle = \langle K(n)\rangle = \langle LK(n)\rangle$.

There are exactly two smashing localizations on $\mathcal{K}_n$. The identity functor is smashing and is localization away from 0, which is compact and strongly dualizable. The zero functor is smashing and is localization away from LS^0, which is strongly dualizable. It is not compact, but by Theorem 7.3 in [Hovey and Strickland 1999] $LF(n)$ is compact in $\mathcal{K}_n$ and $\mathrm{loc}(LF(n)) = \mathrm{loc}(LS^0) = \mathcal{K}_n$. Therefore the zero functor is also generated by a compact object. This shows that both the GSC and SDGSC hold. □

6. Other localized categories

In this section we will consider several other localized categories. In each case, let $L_Z : \mathcal{S} \to \mathcal{S}$ denote the localization functor that annihilates $\langle Z\rangle$, and let $\mathcal{L}_Z$ denote the category of L_Z-locals.

6A. ***The $H\mathbb{F}_p$-local category.*** The Eilenberg–MacLane spectrum $H\mathbb{F}_p$ is a skew field object in $\mathcal{S}$; in fact, every skew field object in $\mathcal{S}$ is isomorphic to either $H\mathbb{F}_p$ or a $K(n)$. Unlike the $\langle K(n)\rangle$, it is not complemented; for example, $\langle I\rangle \leq \langle H\mathbb{F}_p\rangle$ but $I \wedge H\mathbb{F}_p = 0$. So $\langle H\mathbb{F}_p\rangle \in \mathrm{DL}\backslash\mathrm{BA}$. Hovey and Palmieri [1999] have conjectured several results about the collection of classes less than $\langle H\mathbb{F}_p\rangle$ in $\mathrm{BL}(\mathcal{S})$. The telescope conjectures and Bousfield lattice of $\mathcal{L}_{H\mathbb{F}_p}$ are quite simple.

Theorem 6.1. *In $\mathcal{L}_{H\mathbb{F}_p}$, all of* $\mathrm{LTC1}_n$, $\mathrm{LTC2}_n$, $\mathrm{LTC3}_n$ *hold for all n.*

Proof. For all n, $K(n) \wedge H\mathbb{F}_p = 0$ and $T(n) \wedge H\mathbb{F}_p = 0$, by [Hovey and Palmieri 1999, p. 16]. Therefore $LK(n) = 0 = LT(n)$ and $\mathrm{LTC1}_n$ holds for all n. Note that $l_n = l_n^f$ is the zero functor for all n. □

In order to discuss the GSC and SDGSC in this category, we must classify the smashing localizations. We will do this by using what we know about the Bousfield lattice.

Proposition 6.2. *The Bousfield lattice of $\mathcal{L}_{H\mathbb{F}_p}$ is the two-element lattice* $\{\langle 0\rangle, \langle H\mathbb{F}_p\rangle\}$.

Proof. This follows immediately from Proposition 5.9 because $H\mathbb{F}_p$ is a skew field object in $\mathcal{S}$. □

Recall that every smashing localization gives a pair of complemented classes in BA $\subseteq$ BL. Thus in $\mathcal{L}_{H\mathbb{F}_p}$ there are exactly two smashing localizations, the trivial ones given by smashing with zero and with the unit.

Proposition 6.3. *In $\mathcal{L}_{H\mathbb{F}_p}$, the* GSC *fails but the* SDGSC *holds.*

Proof. The identity functor is smashing and is localization away from 0, which is compact and strongly dualizable. By [Hovey and Strickland 1999, Corollary B. 13], there are no nonzero compact objects in $\mathcal{L}_{H\mathbb{F}_p}$. So the zero functor, which is localization away from LS^0, is generated by a strongly dualizable object but not a compact one. □

One application of this Bousfield lattice calculation is to the question of classifying localizing subcategories. Every Bousfield class is a localizing subcategory. Hovey and Palmieri [1999, Conjecture 9.1] suggest that the converse holds in the p-local stable homotopy category. The original conjecture is still open, but the question can be asked in any well-generated tensor triangulated category. For example, in a stratified category every localizing subcategory is a Bousfield class. The question is interesting, since in general localizing subcategories are hard to classify. In many cases, including $\mathcal{S}$, it is not even known if there is a set of localizing subcategories. Recently Stevenson [2014] found the first counterexample, in an algebraic setting: in the derived category of an absolutely flat ring that is not semiartinian, there are localizing subcategories that are not Bousfield classes. Now we show that $\mathcal{L}_{H\mathbb{F}_p}$ provides another counterexample.

Proposition 6.4. *In $\mathcal{L}_{H\mathbb{F}_p}$ there are localizing subcategories that are not Bousfield classes.*

Proof. The following counterexample was suggested to us by Mark Hovey. The Bousfield lattice of $\mathcal{L}_{H\mathbb{F}_p}$ has only two elements: $\langle 0\rangle = \mathcal{L}_{H\mathbb{F}_p}$ and $\mathbb{1} = \{0\}$. It suffices to find a proper nonzero localizing subcategory in $\mathcal{L}_{H\mathbb{F}_p}$.

Consider the Moore spectrum $M(p)$, defined by the triangle $S^0 \xrightarrow{p} S^0 \to M(p)$; this spectrum is $H\mathbb{F}_p$-local. Consider the following full subcategory in $\mathcal{L}_{H\mathbb{F}_p}$.

$$\mathcal{A} = \{X \in \mathcal{L}_{H\mathbb{F}_p} \mid [X, M(p)]_n = 0 \text{ for all } n \in \mathbb{Z}\}.$$

This is a localizing subcategory, called the cohomological Bousfield class of $M(p)$ and denoted $\langle M(p)^*\rangle$ in [Hovey 1995b]. The spectrum $H\mathbb{F}_p$ is a ring object, and hence local with respect to itself. As mentioned in Section 4, it is known that $H\mathbb{F}_p$ is dissonant and $M(p)$ is harmonic, so $[H\mathbb{F}_p, M(p)]_n = 0$ for all n, and $H\mathbb{F}_p \in \mathcal{A}$. On the other hand, the identity on $M(p)$ is nonzero, so $M(p) \notin \mathcal{A}$. This shows that $\mathcal{A}$ is a localizing subcategory that is not a Bousfield class.

Another example comes from $Z = L_{H\mathbb{F}_p}(BP)$. Clearly $Z \notin \langle Z^*\rangle$. But BP is also harmonic, so $[H\mathbb{F}_p, BP]_n = 0$ and $[H\mathbb{F}_p, Z]_n = 0$ for all n, and $H\mathbb{F}_p \in \langle Z^*\rangle$. Since $Z \in \langle M(p)^*\rangle$, we know that $\langle M(p)^*\rangle \neq \langle Z^*\rangle$. □

Both these counterexamples are cohomological Bousfield classes. It would be interesting to find a localizing subcategory in $\mathcal{L}_{H\mathbb{F}_p}$ that is not a cohomological Bousfield class, or show there are none. Also, it is not clear what, if anything, the previous proposition might tell us about the original conjecture in $\mathcal{S}$. For example, as localizing subcategories in $\mathcal{S}$, we have that $\langle M(p)^* \rangle = \langle I \rangle$ [Hovey 1995b, 3.3].

6B. ***The I-local category.*** Recall that by I we mean the Brown–Comenetz dual of the sphere spectrum. It is a rare example of a nonzero spectrum that squares to zero. Hovey and Palmieri [1999, Lemma 7.8] conjecture that $\langle I \rangle$ is minimal in $\mathrm{BL}(\mathcal{S})$.

Theorem 6.5. *On $\mathcal{L}_I$, for all n we have that* $\mathrm{LTC1}_n$, $\mathrm{LTC2}_n$, *and* $\mathrm{LTC3}_n$ *all hold.*

Proof. By Lemma 7.1(c) of [Hovey and Palmieri 1999], $T(n) \wedge I = 0$ for all n, so $LT(n) = 0$. Since $K(i)$ is a BP-module, and $BP \wedge I = 0$ by [Hovey and Strickland 1999, Corollary B.11], we also have $K(n) \wedge I = 0$ for all n. Therefore $\langle LT(n) \rangle = \langle 0 \rangle = \langle LK(n) \rangle$ for all n, and the rest follows from Theorem 3.12. □

Proposition 6.6. *The Bousfield lattice of $\mathcal{L}_I$ is the two-element lattice* $\{\langle 0 \rangle, \langle L_I S^0 \rangle\}$.

Proof. By [Hovey and Palmieri 1999, 7.1(c)], $\langle I \rangle < \langle H\mathbb{F}_p \rangle$. Then Proposition 2.6 implies that there is a surjective lattice map from $\mathrm{BL}(\mathcal{L}_{H\mathbb{F}_p}) = \{\langle 0 \rangle, \langle H\mathbb{F}_p \rangle\}$ onto $\mathrm{BL}(\mathcal{L}_I)$. Note that, by Lemma 2.8, we have $\langle L_I H\mathbb{F}_p \rangle = \langle L_I L_{H\mathbb{F}_p} S^0 \rangle = \langle L_I S^0 \rangle$.

It remains to show that $\langle L_I S^0 \rangle \neq \langle 0 \rangle$. But any X in $\mathcal{S}$ with $X \wedge I \neq 0$ in $\mathcal{S}$ will have $L_I X \neq 0$ and $L_I X \notin \langle L_I S^0 \rangle$ in $\mathrm{BL}(\mathcal{L}_I)$; this is due to the fact that $L_I X \wedge_{\mathcal{L}_I} L_I S^0 = L_I(L_I X \wedge_{\mathcal{S}} L_I S^0) = L_I(X \wedge_{\mathcal{S}} S^0) = L_I(X)$. For example, $F(n) \wedge I \neq 0$ for all n [Hovey and Palmieri 1999, 7.1(e)]. □

Corollary 6.7. *In $\mathcal{L}_I$, the* GSC *fails but the* SDGSC *holds.*

Proof. Corollary B.13 of [Hovey and Strickland 1999] also shows that $\mathcal{L}_I$ has no nonzero compacts, so the proof is the same as for $\mathcal{L}_{H\mathbb{F}_p}$. □

Hovey states the Dichotomy Conjecture in [1995a, Conjecture 3.10]: In $\mathcal{S}$, every spectrum has either a finite local or a finite acyclic. In [Hovey and Palmieri 1999] the authors discuss several equivalent formulations and some implications. We briefly point out a relationship between this conjecture and Proposition 6.6.

Proposition 6.8. *If the Dichotomy Conjecture holds, then the cardinality of* $\mathrm{BL}(\mathcal{L}_I)$ *is at most two.*

Proof. Lemma 7.8 of [Hovey and Palmieri 1999] shows that if the Dichotomy Conjecture holds, then $\langle I \rangle$ is minimal among nonzero classes in $\mathrm{BL}(\mathcal{S})$. This is the case if and only if $a\langle I \rangle$ is maximal among non-top classes in $\mathrm{BL}(\mathcal{S})$, where $a(-)$ denotes the complementation operation first studied by Bousfield [1979b]. Let $a\langle I \rangle\!\uparrow$ denote the sublattice $\{\langle X \rangle \mid \langle X \rangle \geq a\langle I \rangle\} \subseteq \mathrm{BL}(\mathcal{S})$. In [Wolcott 2014, Proposition 3.2] we show that there is a surjective lattice map from $a\langle I \rangle\!\uparrow$ onto $\mathrm{BL}(\mathcal{L}_I)$. Thus, if the

Dichotomy Conjecture holds, $a\langle I\rangle\uparrow$ has cardinality two and $\mathrm{BL}(\mathcal{L}_I)$ has cardinality at most two. $\square$

As for classifying localizing subcategories of $\mathcal{L}_I$, or at least perhaps finding a proper nonzero localizing subcategory, we must get around the fact that so many spectra are I-acyclic. We know that $LF(n)\neq 0$ for all n, however $\mathrm{loc}(LF(n))$ is the acyclics of $l_{n-1}^f:\mathcal{L}_I\to\mathcal{L}_I$ and Theorem 6.5 shows that $l_n^f=0$ for all n. Thus $\mathrm{loc}(LF(n))=\mathrm{loc}(LS^0)$ in $\mathcal{L}_I$ for each n.

6C. *The BP-local category.*

Theorem 6.9. *On $\mathcal{L}_{BP}$, for all n we have that* $\mathrm{LTC1}_n$, $\mathrm{LTC2}_n$, *and* $\mathrm{LTC3}_n$ *all hold.*

Proof. We will show that $\mathrm{LTC2}_n$ holds for all n, and the rest follows from Theorem 3.12. Since each $K(i)$ is a BP-module spectrum, $\langle K(i)\rangle\leq\langle BP\rangle$, and since $K(i)$ is local with respect to itself this implies that $K(i)$ is BP-local. Furthermore, this implies $\langle E(n)\rangle\leq\langle BP\rangle$, so from Lemma 2.8 $L_n=L_nL=LL_n$ as functors on $\mathcal{S}$.

We claim that $L_n:\mathcal{L}_{BP}\to\mathcal{L}_{BP}$, taking $LY\mapsto L_nLY=L_nY$, is a smashing localization functor on $\mathcal{L}_{BP}$. We have $L_n(LY)=L(L_nY)=L(L_nS^0\wedge_{\mathcal{S}}Y)=L(LL_nS^0\wedge_{\mathcal{S}}LY)=L(L_nLS^0\wedge_{\mathcal{S}}LY)=(L_nLS^0)\wedge_{\mathcal{L}_{BP}}(LY)$. This shows that on $\mathcal{L}_{BP}$ the localization functor L_n is also given by smashing with the localization of the unit, L_nLS^0, and thus is smashing.

We know that L_n and l_n are isomorphic since both are localization functors on $\mathcal{L}_{BP}$ that annihilate $\langle K(0)\vee\cdots\vee K(n)\rangle=\langle LK(0)\vee\cdots\vee LK(n)\rangle$.

On $\mathcal{S}$, the natural map $L_n^fX\to L_nX$ is a BP-equivalence [Ravenel 1993, Theorem 2.7(iii)]. This means that $LL_n^fX=LL_nX$ for all objects X in $\mathcal{S}$, in particular for all BP-local objects. Therefore $LL_n^f=L_n=l_n$ is a smashing localization functor on $\mathcal{L}_{BP}$. The acyclics are $\langle LL_n^f(LS^0)\rangle=\langle LL_n^fS^0\rangle=\langle LT(0)\vee\cdots\vee LT(n)\rangle$. These are the same acyclics as for l_n^f, and so we conclude that l_n^f and l_n are isomorphic, and the natural map $l_n^fX\to l_nX$ is an isomorphism. $\square$

Proposition 6.10. *The* GSC *fails in $\mathcal{L}_{BP}$.*

Proof. The proof of the last theorem showed that $L_n:\mathcal{L}_{BP}\to\mathcal{L}_{BP}$ is a (different) smashing localization for each n. However, by [Hovey and Strickland 1999, Corollary B.13] the category $\mathcal{L}_{BP}$ has no nonzero compact objects. $\square$

Note that the SDGSC could still hold, since all the smashing localizations we have identified on $\mathcal{L}_{BP}$ are of the form $L_n=l_n=l_n^f$, and so are generated by strongly dualizable objects. The question of finding any other smashing localizations on $\mathcal{L}_{BP}$ is probably at least as hard as doing so on $\mathcal{S}$, in light of Proposition 3.16.

All of $\langle E(n)\rangle$, $\langle K(n)\rangle$, $\langle H\mathbb{F}_p\rangle$, and $\langle I\rangle$ are "small" in $\mathrm{BL}(\mathcal{S})$, so by Lemma 2.8 it is not surprising that the Bousfield lattices of their localized categories are not very large; this is not true of $\langle BP\rangle$ in $\mathrm{BL}(\mathcal{S})$. We have the following bounds on the Bousfield lattice of the local category.

Proposition 6.11. *The Bousfield lattice of* $\mathcal{L}_{BP}$ *has* $2^{\aleph_0} \le |\mathrm{BL}(\mathcal{L}_{BP})| \le 2^{2^{\aleph_0}}$.

Proof. The second inequality is Corollary 2.7. Since $\langle K(i)\rangle \le \langle BP\rangle$ for all i, we have $\langle Q\rangle = \langle \bigvee_{i\ge 0} K(i)\rangle \le \langle BP\rangle$, and so by Propositions 2.6 and 4.2 we have $|\mathrm{BL}(\mathcal{L}_{BP})| \ge |\mathrm{BL}(\mathcal{H})| = 2^{\aleph_0}$. □

6D. ***The $F(n)$-local category.*** We conclude with a discussion of the $F(n)$-local category.

Any smashing localization $L : \mathcal{S} \to \mathcal{S}$ gives a splitting of the Bousfield lattice

$$\mathrm{BL}(\mathcal{S}) \xrightarrow{\sim} \mathrm{BL}(\mathcal{L}_{LS^0}) \times \mathrm{BL}(\mathcal{L}_{CS^0}),$$

where $\langle X\rangle \mapsto (\langle X \wedge LS^0\rangle, \langle X \wedge CS^0\rangle)$. See [Iyengar and Krause 2013, Proposition 6.12] or [Wolcott 2014, Theorem 5.14] for more details. Taking $L = L_n^f : \mathcal{S} \to \mathcal{S}$, we have $\langle LS^0\rangle = \langle T(0) \vee \cdots \vee T(n)\rangle$ and $\langle CS^0\rangle = \langle F(n+1)\rangle$. Of course, the relationship between $\mathcal{L}_{T(0)\vee\cdots\vee T(n)}$ and $\mathcal{L}_{E(n)}$ of Section 5A is immediately related to the original $\mathrm{TC1}_n$ in $\mathcal{S}$. However, this suggests that $\mathcal{L}_{F(n)}$ is worth investigating further.

By Lemma 2.10, in $\mathrm{BL}(\mathcal{S})$ there is a chain

$$\langle S^0\rangle = \langle F(0)\rangle \ge \langle F(1)\rangle \ge \langle F(2)\rangle \ge \cdots,$$

and by Lemma 2.8 this gives a chain of lattice surjections

$$\mathrm{BL}(\mathcal{S}) = \mathrm{BL}(\mathcal{L}_{F(0)}) \twoheadrightarrow \mathrm{BL}(\mathcal{L}_{F(1)}) \twoheadrightarrow \mathrm{BL}(\mathcal{L}_{F(2)}) \twoheadrightarrow \cdots.$$

From the above observations, we expect $\mathrm{BL}(\mathcal{L}_{F(n)})$ to be about as complicated as $\mathrm{BL}(\mathcal{S})$. For example, $F(n) \wedge I \ne 0$ for all n, and so $L_{F(n)}I$ is a square-zero object in $\mathcal{L}_{F(n)}$. This means that, unlike in most of the localized categories discussed throughout this paper, we know that $\mathrm{BA}(\mathcal{L}_{F(n)}) \ne \mathrm{BL}(\mathcal{L}_{F(n)})$.

Acknowledgements

We would like to thank Dan Christensen for extensive discussions and suggestions and Jon Beardsley for Proposition 4.2.

References

[Angeleri Hügel et al. 2008] L. Angeleri Hügel, J. Šaroch, and J. Trlifaj, "On the telescope conjecture for module categories", *J. Pure Appl. Algebra* **212**:2 (2008), 297–310. MR 2008m:16015 Zbl 1141.16010

[Balmer and Favi 2011] P. Balmer and G. Favi, "Generalized tensor idempotents and the telescope conjecture", *Proc. Lond. Math. Soc.* (3) **102**:6 (2011), 1161–1185. MR 2012d:18010 Zbl 1220.18009

[Benson et al. 2011a] D. Benson, S. B. Iyengar, and H. Krause, "Stratifying triangulated categories", *J. Topol.* **4**:3 (2011), 641–666. MR 2012m:18016 Zbl 1239.18013

[Benson et al. 2011b] D. J. Benson, S. B. Iyengar, and H. Krause, "Stratifying modular representations of finite groups", *Ann. of Math.* (2) **174**:3 (2011), 1643–1684. MR 2846489 Zbl 1261.20057

[Bousfield 1979a] A. K. Bousfield, "The Boolean algebra of spectra", *Comment. Math. Helv.* **54**:3 (1979), 368–377. MR 81a:55015 Zbl 0421.55002

[Bousfield 1979b] A. K. Bousfield, "The localization of spectra with respect to homology", *Topology* **18**:4 (1979), 257–281. MR 80m:55006 Zbl 0417.55007

[Brüning 2007] K. Brüning, "Thick subcategories of the derived category of a hereditary algebra", *Homology, Homotopy Appl.* **9**:2 (2007), 165–176. MR 2009d:18018 Zbl 1142.18008

[Casacuberta et al. 2014] C. Casacuberta, J. J. Gutiérrez, and J. Rosický, "A generalization of Ohkawa's theorem", *Compos. Math.* **150**:5 (2014), 893–902. MR 3209799 Zbl 1300.55007

[Hopkins and Smith 1998] M. J. Hopkins and J. H. Smith, "Nilpotence and stable homotopy theory, II", *Ann. of Math.* (2) **148**:1 (1998), 1–49. MR 99h:55009 Zbl 0927.55015

[Hovey 1995a] M. Hovey, "Bousfield localization functors and Hopkins' chromatic splitting conjecture", pp. 225–250 in *The Čech centennial* (Boston, MA, 1993), edited by M. Cenkl and H. Miller, Contemp. Math. **181**, Amer. Math. Soc., Providence, RI, 1995. MR 96m:55010 Zbl 0830.55004

[Hovey 1995b] M. Hovey, "Cohomological Bousfield classes", *J. Pure Appl. Algebra* **103**:1 (1995), 45–59. MR 96g:55008 Zbl 0841.55008

[Hovey and Palmieri 1999] M. Hovey and J. H. Palmieri, "The structure of the Bousfield lattice", pp. 175–196 in *Homotopy invariant algebraic structures* (Baltimore, MD, 1998), edited by J.-P. Meyer et al., Contemp. Math. **239**, Amer. Math. Soc., Providence, RI, 1999. MR 2000j:55033 Zbl 0947.55014

[Hovey and Strickland 1999] M. Hovey and N. P. Strickland, *Morava K-theories and localisation*, vol. 139, Mem. Amer. Math. Soc. **666**, Amer. Math. Soc., Providence, RI, 1999. MR 99b:55017 Zbl 0929.55010

[Hovey et al. 1997] M. Hovey, J. H. Palmieri, and N. P. Strickland, *Axiomatic stable homotopy theory*, vol. 128, Mem. Amer. Math. Soc. **610**, Amer. Math. Soc., Providence, RI, 1997. MR 98a:55017 Zbl 0881.55001

[Iyengar and Krause 2013] S. B. Iyengar and H. Krause, "The Bousfield lattice of a triangulated category and stratification", *Math. Z.* **273**:3-4 (2013), 1215–1241. MR 3030697 Zbl 06149061

[Keller 1994] B. Keller, "A remark on the generalized smashing conjecture", *Manuscripta Math.* **84**:2 (1994), 193–198. MR 95h:18014 Zbl 0826.18004

[Krause 2000] H. Krause, "Smashing subcategories and the telescope conjecture—an algebraic approach", *Invent. Math.* **139**:1 (2000), 99–133. MR 2000k:55016 Zbl 0937.18013

[Krause 2005] H. Krause, "Cohomological quotients and smashing localizations", *Amer. J. Math.* **127**:6 (2005), 1191–1246. MR 2007d:18018 Zbl 1090.18007

[Krause 2010] H. Krause, "Localization theory for triangulated categories", pp. 161–235 in *Triangulated categories*, edited by T. Holm et al., London Math. Soc. Lecture Note Ser. **375**, Cambridge Univ. Press, 2010. MR 2012e:18026 Zbl 1232.18012

[Krause and Solberg 2003] H. Krause and Ø. Solberg, "Applications of cotorsion pairs", *J. London Math. Soc.* (2) **68**:3 (2003), 631–650. MR 2004k:16028 Zbl 1061.16006

[Krause and Šťovíček 2010] H. Krause and J. Šťovíček, "The telescope conjecture for hereditary rings via Ext-orthogonal pairs", *Adv. Math.* **225**:5 (2010), 2341–2364. MR 2011j:16013 Zbl 1242.16007

[Mahowald 1982] M. Mahowald, "The image of J in the EHP sequence", *Ann. of Math.* (2) **116**:1 (1982), 65–112. MR 83i:55019 Zbl 0504.55010

[Mahowald and Sadofsky 1995] M. Mahowald and H. Sadofsky, "v_n telescopes and the Adams spectral sequence", *Duke Math. J.* **78**:1 (1995), 101–129. MR 96h:55006 Zbl 0984.55008

[Mahowald et al. 2001] M. Mahowald, D. Ravenel, and P. Shick, "The triple loop space approach to the telescope conjecture", pp. 217–284 in *Homotopy methods in algebraic topology* (Boulder, CO, 1999), edited by J. P. C. Greenlees et al., Contemp. Math. **271**, Amer. Math. Soc., Providence, RI, 2001. MR 2002g:55014 Zbl 0984.55009

[Miller 1981] H. R. Miller, "On relations between Adams spectral sequences, with an application to the stable homotopy of a Moore space", *J. Pure Appl. Algebra* **20**:3 (1981), 287–312. MR 82f:55029 Zbl 0459.55012

[Miller 1992] H. Miller, "Finite localizations", *Bol. Soc. Mat. Mexicana* (2) **37**:1-2 (1992), 383–389. MR 96h:55009 Zbl 0852.55015

[Neeman 1992] A. Neeman, "The chromatic tower for $D(R)$", *Topology* **31**:3 (1992), 519–532. MR 93h:18018 Zbl 0793.18008

[Ohkawa 1989] T. Ohkawa, "The injective hull of homotopy types with respect to generalized homology functors", *Hiroshima Math. J.* **19**:3 (1989), 631–639. MR 90j:55013 Zbl 0697.55008

[Ravenel 1984] D. C. Ravenel, "Localization with respect to certain periodic homology theories", *Amer. J. Math.* **106**:2 (1984), 351–414. MR 85k:55009 Zbl 0586.55003

[Ravenel 1992] D. C. Ravenel, *Nilpotence and periodicity in stable homotopy theory*, Annals of Mathematics Studies **128**, Princeton University Press, 1992. MR 94b:55015 Zbl 0774.55001

[Ravenel 1993] D. C. Ravenel, "Life after the telescope conjecture", pp. 205–222 in *Algebraic K-theory and algebraic topology* (Lake Louise, AB, 1991), edited by P. G. Goerss and J. F. Jardine, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. **407**, Kluwer Acad. Publ., Dordrecht, 1993. MR 96i:55016 Zbl 0899.55009

[Stevenson 2014] G. Stevenson, "Derived categories of absolutely flat rings", *Homology Homotopy Appl.* **16**:2 (2014), 45–64. MR 3234500 Zbl 06416954

[Šťovíček 2010] J. Šťovíček, "Telescope conjecture, idempotent ideals, and the transfinite radical", *Trans. Amer. Math. Soc.* **362**:3 (2010), 1475–1489. MR 2010j:18024 Zbl 1197.18005

[Wolcott 2014] F. L. Wolcott, "Bousfield lattices of non–Noetherian rings: Some quotients and products", *Homology Homotopy Appl.* **16**:2 (2014), 205–229. MR 3257574 Zbl 06416962

Received July 14, 2014. Revised January 13, 2015.

F. LUKE WOLCOTT
DEPARTMENT OF MATHEMATICS
LAWRENCE UNIVERSITY
711 E BOLDT WAY
SPC 24
APPLETON, WI 54911
UNITED STATES

luke.wolcott@gmail.com

CONTENTS

Volume 276, 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 276 No. 2 August 2015

0030-8730(201508)276:2;1-W