A diffusion approach to Stein's method on Riemannian manifolds

HUILING LE[a], ALEXANDER LEWIS[b], KARTHIK BHARATH[c] and CHRISTOPHER FALLAIZE[d]

School of Mathematical Sciences, University Park, University of Nottingham, Nottingham, UK,
[a]*Huiling.Le@nottingham.ac.uk,* [b]*pmxal12@exmail.nottingham.ac.uk,* [c]*Karthik.Bharath@nottingham.ac.uk,*
[d]*Chris.Fallaize@nottingham.ac.uk*

We detail an approach to developing Stein's method for bounding integral metrics on probability measures defined on a Riemannian manifold M. Our approach exploits the relationship between the generator of a diffusion on M having a target invariant measure and its characterising Stein operator. We consider a pair of such diffusions with different starting points, and through analysis of the distance process between the pair, derive Stein factors, which bound the solution to the Stein equation and its derivatives. The Stein factors contain curvature-dependent terms and reduce to those currently available for R^m, and moreover imply that the bounds for R^m remain valid when M is a flat manifold.

Keywords: Coupling; integral metrics; Stein equation; stochastic flow; Wasserstein distance

1. Introduction

The eponymous method to estimate integral metrics and semi-metrics on spaces of probability measures proposed by Charles Stein (Stein, 1972) has led to tremendous improvements in distributional approximation techniques. See, for example, the surveys by Barbour and Chen (2014) and Ross (2011). The method has mainly been developed for probability measures on R^m or N^m for $m \geq 1$. The focus of this paper is on developing a version of the method that can be employed to approximate probability measures on an *m*-dimensional Riemannian manifold.

Abstracting Stein's method to a general space X in a heuristic manner is useful for elucidating its key ingredients and the ensuing challenges in developing a corresponding version on manifolds. The goal is to bound an integral (semi-)metric

$$d_{\mathsf{H}}(\mu, \nu) := \sup_{h \in \mathsf{H}} \left| \int h \, d\mu - \int h \, d\nu \right|$$

between a probability measure ν and a target probability measure μ on X with respect to a class H of real-valued test functions on X. Stein's method is centred around the construction and study of an operator L that maps functions $f : \mathsf{X} \to \mathsf{R}$ in a certain class F into mean-zero functions under μ: if $X \sim \mu$, then $\mathrm{E}[Lf(X)] = 0$ for every $f \in \mathsf{F}$. The operator L thus encodes information about μ and, when F is sufficiently large, one may determine a function $f_h \in \mathsf{F}$ associated with every $h \in \mathsf{H}$ that solves the *Stein equation* (or the Poisson equation in PDE literature)

$$h(x) - \mathrm{E}[h(X)] = Lf_h(x).$$

As a consequence, bounding $d_{\mathsf{H}}(\mu, \nu)$ reduces to bounding the term $\sup_{f_h \in \mathsf{F}} \mathrm{E}[Lf_h(Z)]$, where $Z \sim \nu$, achieved in

application-specific ways. An important implication, profitably used in some applications, is that the need to compute an expectation with respect to µ in dH is circumvented; an example is when

ν is the empirical measure based on points $x_1,...,x_n$ on X and μ represents a conjectured limit probability measure. Upper bounds on d_H then depend explicitly on the smoothness of the functions in F. Hence, integral to the success of Stein's method in upper bounding d_H are the following requirements:
(a) construction of the operator L and identifying its domain F; and, (b) determination of the solution f_h and its regularity properties.

An introductory account on choices of L satisfying requirement (a) for various probability measures μ on R (or some subset thereof) is available in Ross (2011). When $X = R^m$, $m > 1$, focus has mainly been restricted to the case when μ is a Gaussian measure (see Barbour (1988), Chatterjee and Meckes (2008), Meckes (2009)) although more recently results on extensions to non-Gaussian measures have appeared in Chen et al. (2019), Fang, Shao and Xu (2019), Mackey and Gorham (2016), Mijoule, Reinert and Swan (2019).

An important observation by Barbour (1988) relates the operator L to the infinitesimal generator of a diffusion process on R^m that solves an SDE with invariant measure μ. This observation enables identification, and examination, of the solution to the Stein equation with the transition semigroup associated with L. The diffusion approach hence opens up the possibility of defining L for μ on a manifold M by considering an SDE on M whose solution is a diffusion with invariant measure μ.

Broadly, this is the approach we adopt in this paper. On a complete Riemannian manifold (M,g) without boundary, we consider approximating probability measures of the form μ_φ with density, up to a normalisation constant, $e^{-\varphi}$ with respect to the volume measure dvol for a smooth φ. Under some conditions on φ and the geometry of M, the diffusion with infinitesimal generator

$$1$$

$$L_\varphi := \{\Delta - \langle \nabla\varphi, \nabla \rangle\}$$

$$2$$

has μ_φ as its invariant measure, where ∇ and Δ are the (Riemannian) gradient and Laplace-Beltrami operators, respectively. The operator L_φ generates mean-zero functions under μ_φ.

We address requirement (b) for generalising Stein's method to M by adapting the approach in Mackey and Gorham (2016) for log-concave measures μ on R^m to the manifold setting. In their paper, bounds on lower-order derivatives of the solution f_h, known as *Stein factors* (see Röllin (2012)), were derived by studying the distance between a pair of coupled diffusions X_t and Y_t with same invariant measure μ starting at distinct points. Analogously, we construct a pair of diffusions $X_{t,x}$ and $Y_{t,y}$ on M starting at x and y with identical generator L_φ, and study the distance process $\rho(X_{t,x}, Y_{t,y})$ around neighbourhoods of non-empty cut loci. In particular, when there is no first conjugate point contained in the cut locus to any given point in M we establish exponential *pathwise* contraction for trajectories of the two diffusions towards their initial points; on the other hand when first conjugate points are present, we establish a similar contraction property that holds on *average*.

The study of the distance process enables the determination of Stein factors which bound the Lipschitz constants of the solution f_h, and its first and second derivatives, where the geometry of M manifests itself through curvature-dependent terms in the factors. The derived bounds on f_h, as well as on its first and second derivatives, reduce to the ones of Mackey and Gorham (2016) for R^m, which we show remain valid for complete, connected flat manifolds. The Stein factors are then used to construct upper bounds on integral (semi-)metrics between μ_φ and another probability measure on M for specific choices of the class of test functions H. In particular, using the first order bound on f_h, we derive an upper bound on the Wasserstein distance between μ_φ and μ_ψ. A related generalisation of Stein's method to manifolds, based on the approach of Fang, Shao and Xu (2019), can be found in Thompson (2020).

The paper is organised as follows. In Section 2.1 we define relevant quantities and introduce notation, and in Section 2.2 we describe assumptions on probability measures and diffusions under consideration, and the key condition (3) and assumption (A1) for the derivation of our results; the conditions are explicated with some examples. In Section 3 we describe the coupling of a pair of diffusions on M and analyse their distance process, when conjugate points are absent (Section 3.1) and present (Section 3.2). In Section 4 we consider the Stein equation and its solution and derive Stein factor bounds that depend on the curvature of M. In Section 5, using the Stein factors, we derive bounds for integral (semi-)metrics: in Section 5.1 we derive an upper bound on the Wasserstein distance between μ_φ and a probability measure of similar type, and in Section 5.2 we do the same for an integral semi-metric between μ_φ and an arbitrary probability measure on M.

2. Preliminaries

2.1. Notation and definitions

We assume throughout that (M,g) is a complete and connected Riemannian manifold without boundary of dimension m and with covariant derivative D; by D^i, $i > 1$ we then denote higher orders of D. We shall denote by $\rho(x,y)$ the Riemannian distance between any two points x and y in M, and by dvol the Riemannian volume measure of (M,g). We denote by $T_x(M)$ the tangent space to M at $x \in M$ and by TM the tangent bundle of M. For $k \geq 1$, $\mathsf{C}^k(M)$ denotes the class of k-times continuously differentiable real-valued functions on M, $\mathsf{C}(M)$ denotes the set of continuous functions, and $\mathsf{C}_0(M)$ denotes continuous functions vanishing at infinity. The Lipschitz constant $C_0(h)$ of a Lipschitz function $h \in \mathsf{C}(M)$ is defined as

$$C_0(h) := \sup_{x \ast y \in M} \frac{|h(x) - h(y)|}{\rho(x,y)}.$$

Higher-order Lipschitz constants of a function depend on bounding tensor fields. Accordingly, for each $x \in M$ define the operator norm at x for a tensor field T on M, based on n-fold tangent vectors at x, as

$$|T|_{op} := \sup_{\substack{v_1, \cdots, v_n \in T_x(M), |v_i| \ast 0 \\ \prod_{i=1}^{n} |v_i|}} \frac{|T(v_1, \cdots, v_n)|}{}.$$

Then, if $h \in \mathsf{C}^k(M)$, for $k \geq 1$, we may define

$$C_i(h) := \sup_{\gamma_{x,y}, x \ast y \in M} \frac{|D^i h(x) - \Pi_{\gamma_{x,y}}(D^i h(y))|_{op}}{\rho(x,y)}, \quad i = 1, \ldots, k, \quad (1)$$

and call them the Lipschitz constants of $D^i h$, where $\gamma_{x,y}$ denotes any possible minimal geodesic from y to x and $\Pi_{\gamma_{x,y}}$ denotes the parallel transport from $T_y(M)$ to $T_x(M)$ along $\gamma_{x,y}$. Note that $Dh = dh$ and that $\mathrm{Hess}^h = D^2 h$, where Hess^h is the Hessian of h. Note also that $\sup_{x \in M} |D^{i+1} h(x)|_{op} = C_i(h)$ for $i = 0, \cdots, k-1$. Finally, if X and Z are two random variables on M with $X \sim \mu$ and $Z \sim \nu$, abusing notation, we interchangeably use $d_{\mathsf{H}}(\nu, \mu)$ and $d_{\mathsf{H}}(Z, X)$ to denote the integral (semi-)metric between the two probability measures, where

$$d_{\mathsf{H}}(Z,X) := \sup_{h \in \mathsf{H}} |\mathrm{E}[h(Z)] - \mathrm{E}[h(X)]|$$

with respect to a set of real-valued test functions H.

2.2. Key assumptions

On M, we consider probability measures of the form

$$d\mu_\varphi = \frac{1}{c(\varphi)} e^{-\varphi} \, d\text{vol},$$

with $c(\varphi) = \int_M e^{-\varphi} \, d\text{vol} < \infty$ and with support on the entire space M. We assume that $\varphi \in C^2(M)$ is such that $\nabla\varphi$ is Lipschitz; specifically, we assume that $D\varphi$ has finite Lipschitz constant $C_1(\varphi)$. Throughout, X denotes a random variable with $X \sim \mu_\varphi$.

The uniformly elliptic operator $L_\varphi = 1/2 \{\Delta - \langle \nabla\varphi, \nabla \cdot \rangle\}$ then is the infinitesimal generator of a Feller diffusion process that solves the Itô stochastic differential equation

$$dX_t = dB_t^M - \frac{1}{2}\nabla\varphi(X_t)dt, \tag{2}$$

where B_t^M is a Brownian motion on M. If there is a constant $\kappa > 0$ such that

$$\text{Ric}(x) + \text{Hess}^\varphi(x) \geq -\kappa g(x), \ \forall x \in M, \tag{3}$$

where Ric is the Ricci curvature tensor, then the corresponding semigroup $P_t = e^{tL_\varphi}$ is conservative (see Bakry (1986)), i.e., $P_t 1 \equiv 1$ for all $t > 0$ or, equivalently, X_t will, with probability one, not leave M in finite time. For a successful development of the Stein's method on M, we need the *Bakry-Emery curvature criterion*: there is a constant $\kappa > 0$ such that,

$$\text{(A1): } \text{Ric}(x) + \text{Hess}^\varphi(x) \geq 2\kappa g(x) \forall x \in M \ .$$

Evidently, assumption (A1) implies the condition in (3).

Remark 1. When M = $\mathbb{R}^m$, (A1) simplifies to $v^\top \text{Hess}^\varphi v \geq 2\kappa$ for any unit (column) vector v in $\mathbb{R}^m$ where, as usual, Hess^φ is treated as an $m \times m$ matrix. Hence, (A1) reduces to the requirement in Mackey and Gorham (2016) that $-\varphi$ is 2κ-strongly concave, noting that in their notation, φ here is $-\log p$, up to a constant. This is also true if the Ricci curvature of M is always non-positive. In general, (A1) is weaker than the requirement that $-\varphi$ is c-strongly concave for some $c > 0$.

Example 1. In order to elucidate condition (3) and assumption (A1) we look at a few example manifolds and probability measures μ_φ.

(i) M *is the standard sphere* S^m *of dimension m*. The function $\varphi(x)$ corresponding to the von Mises-Fisher distribution $M_m(x_0, c)$ takes the form $\varphi(x) = -c \cos(r(x))$, with $r(x) = \rho(x_0, x)$ for $c > 0$ and a fixed point $x_0 \in M$. Since $D^2 f(r) = f''(r)dr \times dr + f'(r)D^2 r$ on general manifolds and since $D^2 r = \cot(r)\{g - dr \times dr\}$ on S^m (see Greene and Wu (1979)), it follows that

$$\text{Hess}^\varphi(x) = -cD^2 \cos(r(x)) = c \cos(r(x))g(x);$$

this ensures that

$$\text{Ric}(x) + \text{Hess}^\varphi(x) \geq \{(m-1) - c\}g(x),$$

and condition (3) holds for the von Mises-Fisher distribution with $\kappa > \max\{-(m-1) + c, 0\}$. However, if

there is a $\kappa > 0$ such that assumption (A1) holds, then we must have $0 < c < m - 1$. This requires in particular $m > 1$, and thus any von Mises-Fisher distribution on the circle fails to satisfy (A1).

(ii)   M *is hyperbolic space* H^m *with sectional curvature* -1. For $\varphi(x) = c\rho(o,x)^2$ where $c > 0$ and

o is a fixed point in M, we have $\int_M e^{-\varphi} d\mathrm{vol} < \infty$ as, in terms of normal coordinates at o,

$d\mathrm{vol} = \sinh(\rho)^{m-1} d\rho d\theta$. On the other hand, $\mathrm{Hess}^{\varphi}(x) \geq 2cg(x)$ by the Hessian Comparison Theorem and $\mathrm{Ric}(x) = -(m-1)g(x)$. Hence condition (3) holds with $\kappa > \max\{(m-1)-2c, 0\}$ for such a φ. Moreover, if $c > (m-1)/2$, then there is a $\kappa > 0$ such that assumption (A1) holds.

(iii) M *is the complex projective space* CP^m *equipped with the Fubini-Study metric. This is also the Kendall shape space of configurations in* R^2 *with* $m+1$ *labelled landmarks.* Let A be an $(m+1)\times(m+1)$ Hermitian matrix, i.e. $A = A^*$ and $\varphi(z) = -z^*Az$, for $z = x + iy \in \mathsf{C}^{m+1}$ (column vectors) and $|z| = 1$, where A^* denotes the complex conjugate transpose of A. Without loss of generality, we may assume that the smallest eigenvalue of A is zero. The corresponding μ_φ is the complex Bingham distribution on $\mathsf{C}S^m = S^{2m+1}$. Since $\varphi(z) = \varphi(e^{i\theta}z)$, μ_φ can be regarded as a distribution on M (see Kent (1994)). It can be shown that $\mathrm{Hess}^{\varphi}(w,w) = 2\{\varphi(z)-\varphi(w)\} \geq -2\lambda_{max}$ for a horizontal (with respect to the projection from S^{2m+1} to CP^m) unit vector $w \in T_z(S^{2m+1})$, where $\lambda_{max} > 0$ is the largest eigenvalue of A.

The complex projective space CP^m equipped with the Fubini-Study metric is an Einstein manifold with its Ricci curvature tensor equal to $2(m+1)$ times the metric tensor. Thus,

$$\mathrm{Ric} + \mathrm{Hess}^{\varphi} \geq 2\{m+1-\lambda_{max}\}\, g,$$

and so, for the complex Bingham distribution on CP^m, condition (3) holds with $\kappa > 2\max\{\lambda_{max}-(m+1),0\}$ and assumption (A1) holds if $\lambda_{max} < m+1$.

(iv) M *is the rotation group* $SO(m)$ *with the bi-invariant metric determined by* $g(E_1,E_2) := -\frac{1}{2}tr(E_1E_2)$ *for skew-symmetric* E_1, E_2, *where* $m > 2$. Assume that, for $S \in M$, $\varphi(S) = -c\, tr(S_0S)$ with $S_0 \in SO(m)$ and a constant $c > 0$. Then, the corresponding μ_φ is a von Mises-Fisher distribution on $SO(m)$. It can be shown that $\mathrm{Hess}^{\varphi} \geq -c\, g$.

Recall that the Killing form of M is $B(E_1,E_2) = (m-2)tr(E_1E_2)$ and the Ricci curvature $\mathrm{Ric}(E_1,E_2) = -\frac{1}{4}B(E_1,E_2) = \frac{m-2}{2}g(E_1,E_2)$. Thus, in this case,

$$\mathrm{Ric} + \mathrm{Hess}^{\varphi} \geq \frac{m-2}{2} - cg,$$

and so, for the von Mises-Fisher distribution on $SO(m)$, condition (3) holds with $\kappa > \max\{c-(m-2)/2, 0\}$ and assumption (A1) holds if $c < (m-2)/2$.

3. The distance between coupled diffusions

Our approach to define the Stein equation on M and analyse properties of its solution rests on the construction of a pair of diffusions (X_t, Y_t), and handling of the distance process $\rho(X_t, Y_t)$ between the pair. In particular, we prove exponential contraction of $\rho(X_t, Y_t)$ towards the initial points, and thus extend the approach used by Mackey and Gorham (2016) on R^m to the manifold setting. In contrast to the Euclidean setting, since the distance function $(x,y) \to \rho(x,y)$ is not in $C^2(M \times M)$ if the cut locus of a point in M is not empty, analysis of the

distance process $\rho(X_t, Y_t)$ requires additional care.

3.1. When no conjugate points are present in cut loci

We first consider the relatively simple situation where there is no conjugate point in the cut locus of any given point in M. In this setting, by modifying the arguments in Kendall (1986a) and Kendall (1986b), we are able to establish exponential pathwise contraction of distance between the diffusions, aided by a key result given in Lemma 3 in Appendix A of Supplementary Material (Le et al., 2024), which expresses the distance function in terms of finitely many smooth functions in neighbourhoods of cut point, despite it not belonging to $C^2(M \times M)$.

Note first that, in terms of a Brownian motion B_t on $\mathbb{R}^m$ starting from the origin, the Itô differential equation (2) with initial condition $X_0 = x_0$ is equivalent to

$$d_s X_t = \Xi_t \, d_s B_t - \frac{1}{2} \nabla\varphi(X_t) \, dt, \; X_0 = x_0; \; d_s\Xi_t = H\Xi \, d_s X_t, \; \Xi(X_0) = \xi_0, \qquad (4)$$

where d_s denotes the Stratonovich differential, H the horizontal lift from TM to the tangent bundle of the orthonormal frame bundle $O(M)$, where ξ_0 sits above x_0. For an introduction to horizontal lifts and orthonormal frame bundles, see for example, Kobayashi and Nomizu (1963).

Theorem 1. *Assume that* M *has the property that there is no conjugate point to any given point in* M*, and that the Bakry-Emery curvature criterion (A1) holds for a constant* $\kappa > 0$*. Then, for any* $x_0, y_0 \in$ M*, there is a pair of coupled diffusions* (X_t, Y_t) *starting from* (x_0, y_0) *such that both* X_t *and* Y_t *satisfy* (2) *and, for any* $e \geq 1$,

$$\rho(X_t, Y_t) \leq \rho(x_0, y_0) e^{-\kappa t}, \; t \geq 0. \quad (5)$$

Proof. Consider the map

$$\text{Exp} : T\,M \to M \times M; \; (x, v) \to (x, \exp_x(v)).$$

For any $(x, v) \in T\,M$, this map provides an intervening geodesic $s \to \exp_x(sv), 0 \leq s \leq 1$, connecting x and $\exp_x(v)$. The length of this geodesic is at least the distance between x and $\exp_x(v)$. If the interior of this geodesic does not intersect the cut locus of x, then it is also a minimal geodesic between its two end points. Denote by $\tilde{\Pi}_{(x,v)}$ the parallel transport along this *intervening* geodesic from x to $\exp_x(v)$ where, for our purpose, $\tilde{\Pi}_{(x,v)}$ is taken to be the identity map on $T_x(M)$ if $x = \exp_x(v)$ even though this may imply a discontinuity. For any given $(x_0, y_0) \in$ M $\times$ M, we take $v_0 \in T_{x_0}(M)$ such that

$$\exp_{x_0}(v_0) = y_0 \text{ and } |v_0| = \rho(x_0, y_0). \quad (6)$$

Under the given assumptions, y_0 is not conjugate to x_0. Then, if y_0 is a cut point of x_0, a consequence of the proof of Lemma 3 in Appendix A of Supplementary Material is that there is a neighbourhood N of (x_0, y_0) such that $\text{Exp}^{-1}(N)$ is a disjoint union of a finite number of open sets on TM and, restricted to each such set, Exp is a diffeomorphism from that set onto N. If y_0 is not a cut point of x_0, then v_0 is uniquely determined by $v_0 = \exp_{x_0}^{-1}(y_0)$ and a similar result holds with just one component in $\text{Exp}^{-1}(N)$.

Hence, in particular, TM is locally a covering space of M × M. Within such a neighbourhood N of a given (x_0, y_0), we can determine a continuous process $(X_t, V_t) \in$ TM starting from (x_0, v_0) associated with (2), by solving the following coupled diffusions X_t and $Y_t = \exp_{X_t}(V_t)$:

$$d_s X_t = \Xi_t \, d_s B_t - \tfrac{1}{2} \nabla\varphi(X_t)dt; \quad X_0 = x_0;$$

$$d_s Y_t = Y_t \, d_s B' - \tfrac{1}{2}\nabla\varphi(Y_t)dt, \quad Y_0 = y_0;$$

$$d_s\Xi_t = H\Xi \, d_s X_t, \quad \Xi(X_0) = \xi_0; \quad (7)$$

$$d_s Y_t = HY \, d_s Y_t, \quad Y(Y_0) = \eta_0;$$

$$dB' = (Y^{-1}\tilde{\Xi}_t)dB_t,$$

$$\Pi_{X_t, V_t}$$

where, similarly to Ξ and ξ_0 for X, Y and η_0 are respectively a lift of Y to the orthonormal frame bundle $O(M)$ and η_0 sits above y_0. Since B' is also a Brownian motion on $\mathbf{R}^m$, both X_t and Y_t are diffusions satisfying (2) before they leave N.

When (X_t, Y_t) hits the boundary of N, we can find a neighbourhood N' of (X_t, Y_t) satisfying the above properties of N. Then, allowing V_t to move discontinuously without altering (X_t, Y_t) such that, after the jump, it satisfies (6), we can continue to run (X_t, Y_t) within N' so defined. Note that, if $X_{t_0} = Y_{t_0}$ for some $t_0 \geq 0$, then $X_t = Y_t$ for $t \geq t_0$.

For (X_t, Y_t) constructed as above, denote by $\tilde{\rho}(X_t, Y_t)$ the length of the intervening geodesic $\exp_{X_t}(sV_t)$ between X_t and $Y_t = \exp_{X_t}(V_t)$; and write γ_t for the unit speed intervening geodesic from X_t to Y_t, that is, $\gamma_t(s) = \exp_{X_t}(sV_t/|V_t|)$. Note that $\tilde{\rho}(X_t, Y_t)$ depends implicitly on the choice of v_0, which is not unique when y_0 is a cut point of x_0. On the other hand, for any given v_0 which satisfies (6), $\tilde{\rho}$ is a smooth function of (x, y) within the neighbourhood N chosen as above. However, the change of neighbourhood from N to N' usually results in a discontinuity for the process $\tilde{\rho}(X_t, Y_t)$. Nevertheless, $\rho(X_t, Y_t)$ is always continuous and

$$\rho(X_t, Y_t) \leq \tilde{\rho}(X_t, Y_t), \quad t \geq 0,$$

where the latter becomes an equality immediately after the jump. Hence, to find an upper bound for $\rho(X_t, Y_t)$, it is sufficient to find an upper bound for $\tilde{\rho}(X_t, Y_t)$.

To bound $\tilde{\rho}(X_t, Y_t)$ we may assume, without loss of generality, that (X_t, Y_t) lies in N for all $t \geq 0$. Write $u_0, u_1, \cdots, u_{m-1}$ for an orthonormal base in $\mathbf{R}^m$ such that $\Xi_t u_0 = \gamma_t'(0)$, and, for $i = 0, 1, \cdots, m-1$, let $v_i = (Y^{-1}\Pi_{(X_t, V_t)}\Xi_t)u_i$. Then, the Itô formula for $\tilde{\rho}(X_t, Y_t)$ is given by

$$d\tilde{\rho}(X_t, Y_t) = (\Xi_t u_0)\tilde{\rho}(X_t, Y_t)d\langle u_0, B_t\rangle + (Y_t v_0)\tilde{\rho}(X_t, Y_t)d\langle v_0, B'_t\rangle$$

$$+ \frac{1}{2}\sum_{i=0}^{m-1}(\Xi_t u_i + Y_t v_i)^2 \tilde{\rho}(X_t, Y_t)dt \quad (8)$$

$$+ \frac{1}{2}\{\langle \nabla\varphi(X_t), \gamma_t'(0)\rangle - \langle \nabla\varphi(Y_t), \gamma_t'(\tilde{\rho}(X_t, Y_t))\rangle\}dt.$$

Since $\langle u_0, B_t\rangle = \langle v_0, B'_t\rangle$, since $(\Xi_t u_0)\tilde{\rho}(X_t, Y_t) = -(Y_t v_0)\tilde{\rho}(X_t, Y_t)$ and since

$$(\Xi_t\, u_0)^2\, \tilde{\rho}(X_t, Y_t) = (Y_t\, v_0)^2\, \tilde{\rho}(X_t, Y_t) = (\Xi_t\, u_0)(Y_t\, v_0)\tilde{\rho}(X_t, Y_t) = 0,$$

(8) simplifies to

$$2\mathrm{d}\tilde{\rho}(X_t, Y_t) = \sum_{i=1}^{m-1} (\Xi_t\, u_i + Y_t\, v_i)^2\, \tilde{\rho}(X_t, Y_t)\mathrm{d}t$$

$$+ \{\Diamond\nabla\varphi(X_t), \gamma/t(0)\Diamond - \nabla\varphi(Y_t), \gamma/t\,(\tilde{\rho}(X_t, Y_t))\Diamond\}\mathrm{d} \tag{9}$$

Denote by $\dot{J}_t$ the Jacobi vector field along γ_t with $\dot{J}_{tt}(0) = \Xi_t u_i$ and $\dot{J}(1) = Y_t v_i$. Then, since $\tilde{\rho}$ is

smooth under the assumption that (X_t, Y_t) lies in a given neighbourhood of (x_0, y_0), using the second-variation formula (see Cheeger and Ebin (1975)), a modification of the argument by Kendall (1986a) shows that the right hand side of (9) is given by

$$\int_0^{\tilde{\rho}(X_t,Y_t)} \sum_{i=1}^{m-1} \left\{ \left| D_{\gamma I_t(s)}(\dot{J}(s)) \right|^2 - \langle R(\dot{J}(s), \gamma I_t(s)) I \gamma_t(s), \dot{J}(s) \rangle \right\} ds \, dt$$

$$(10)$$

$$+ \left\{ \langle \nabla\varphi(X_t), \gamma I_t(0) \rangle - \nabla\varphi(Y_t), \gamma I_t(\tilde{\rho}(X_t,Y_t)) \rangle \, dt \right\},$$

where the integral is along γ_t and R denotes the curvature tensor of M. To analyse the first term of (10), we use a modified form of the argument in Cheeger and Ebin (1975), the proof of Lemma 1.21. It shows that, for each $i = 1,...,m-1$,

$$\int_0^{\tilde{\rho}(X_t,Y_t)} \left| D_{\gamma I_t(s)}(\dot{J}(s)) \right|^2 - \langle R(\dot{J}(s), \gamma I_t(s)) I \gamma_t(s), \dot{J}(s) \rangle \, ds$$

$$\leq \int_0^{\tilde{\rho}(X_t,Y_t)} \left| D_{\gamma I_t(s)}(V^i(s)) \right|^2 - \langle R(V^i(s), \gamma I_t(s)) I \gamma_t(s), V^i(s) \rangle \, ds,$$

where $V^i(s) := (\tilde{\Pi}(X_t, sV_t/|V_t|) \Xi_t) u_i$. Now, since V^i is parallel along γ_t, it follows that $D_{\gamma I_t(s)}(V^i(s)) = 0$. As a consequence, since $\{I\gamma_t(s), V^1(s),...,V^{m-1}(s)\}$ forms an orthonormal base of $T_{\gamma_t(s)}(M)$ and $\langle R(I\gamma_t(s), \gamma I_t(s)) I \gamma_t(s), \gamma I_t(s) \rangle \equiv 0$, we have

$$\int_0^{\tilde{\rho}(X_t,Y_t)} \sum_{i=1}^{m-1} \left| D_{\gamma I_t(s)}(\dot{J}(s)) \right|^2 - \langle R(\dot{J}(s), \gamma I_t(s)) I \gamma_t(s), \dot{J}(s) \rangle \, ds$$

$$\leq \int_0^{\tilde{\rho}(X_t,Y_t)} \sum_{i=1}^{m-1} - \langle R(V^i(s), \gamma I_t(s)) I \gamma_t(s), V^i(s) \rangle \, ds \quad (11)$$

$$\int_0^0 \tilde{\rho}(X_t, Y_t) = -\operatorname{Ric}(\gamma_t(s))(\,/\gamma_t(s), \gamma/_t(s))\, ds. \qquad ^{i=1}$$

For the remaining two terms of (10), we note that

$$\frac{d}{ds}\langle\nabla\varphi(\gamma_t(s)),\gamma/_t(s)\rangle = \langle D_{\gamma/_t(s)}(\nabla\varphi(\gamma_t(s))),\gamma/_t(s)\rangle + \langle\nabla\varphi(\gamma_t(s)),D_{\gamma/_t(s)}\gamma/_t(s)\rangle$$

$$= \langle D_{\gamma/_t(s)}(\nabla\varphi(\gamma_t(s))),\gamma/_t(s)\rangle$$

$$= \operatorname{Hess}^\varphi(\,/\gamma_t(s),\gamma/_t(s)),$$

as γ_t is a geodesic. From this, we deduce that

$$\tilde{\rho}(X_t,Y_t)\langle\nabla\varphi(Y_t),\gamma/_t(\tilde{\rho}(X_t,Y_t))\rangle - \langle\nabla\varphi(X_t),\gamma/_t(0)\rangle = \int_0 \operatorname{Hess}^\varphi(\,/\gamma_t(s),\gamma/_t(s))\, ds. \quad (12)$$

Thus, under the Bakry-Emery curvature criterion (A1) condition, (10), (11) and (12) together give that

$$2d\tilde{\rho}(X_t,Y_t) \leq -\int_0^{\tilde{\rho}(X_t,Y_t)} \mathrm{Ric}(/\gamma_t(s),\gamma/_t(s)) + \mathrm{Hess}^\varphi(/\gamma_t(s),\gamma/_t(s))\, ds\, dt \quad (13)$$

$$\leq -2\kappa\tilde{\rho}(X_t,V_t)\,dt.$$

Now, for any $e \geq 1$, it follows from (13) that

$$d\tilde{\rho}(X_t,Y_t) \leq -e\kappa\tilde{\rho}(X_t,Y_t)\, dt,$$

so that

$$e\rho(X_t,Y_t) = \tilde{\rho}(X_0,Y_0) + \int_0^t e e\kappa\, \tilde{\rho}(X_s,Y_s)\, ds + d\tilde{\rho}(X_s,Y_s)$$

$$\leq \tilde{\rho}(X_0,Y_0)\,.$$

Finally, by recalling that $\tilde{\rho}(X_0,Y_0) = \rho(X_0,Y_0)$, we have

$$\rho(X_t,Y_t) \leq \tilde{\rho}(X_t,Y_t) \leq \rho(X_0,Y_0)\, e^{-\kappa t}$$

as required.

3.2. When conjugate points are present in cut loci

When conjugate points are present in cut loci in M, the construction of a pair of diffusions in the proof of Theorem 1 fails at such points. More precisely, if y_0 is a (first) conjugate point of x_0 along the geodesic $\exp(sv)$, which also lies in the cut locus of x_0, then $D\exp_{x_0}(v)$ is singular. This means that it would be impossible to find a neighbourhood N of (x_0,y_0) that has the properties described above following (6). In particular, it would be impossible to find a subset of TM, as specified there, such that Exp is a diffeomorphism from that subset onto N. It is evident from the proof of Theorem 1 that the existence of such a diffeomorphism offers a way to couple $(X_{x,t},Y_{y,t})$ at, and beyond, cut points.

   Nevertheless, we now show that it is still possible to construct a pair of diffusions on M with properties that (i) they both satisfy (2) and (ii) the *expected* distance between them contracts at least exponentially. This relies on a generalisation of the technique used in Theorem 5 of Kendall (1986b) to deal with the presence of conjugate points. In the non-conjugate part of the cut locus of M analysis proceeds as with Theorem 1. To warn us of when the diffusions get close to the first conjugate locus, we use the operator L_φ, and monitor the value of its action on the distance function ρ; this value decays towards $-\infty$ when the points approach the first conjugate locus. Effectively, we determine a neighbourhood $N_{2\delta} \subset$ M $\times$M of the first conjugate locus in M $\times$M for a constant δ that depends on κ and the injectivity radius of M. Once the coupled diffusions enter $N_{2\delta}$, the closure of $N_{2\delta}$, we decouple them, run independent diffusions until they hit M $\setminus \overline{W_\delta}$, where $N_\delta \supset N_{2\delta}$, and then return to coupling again. We first need two preliminary results before stating and proving the main result in this section. Observe that the set

$$\tilde{E} := \{(x,v) \in TM \mid \text{the geodesic } \exp_x(sv),\ 0 \leq s \leq 1,$$

contains no conjugate point of $x\}$ is an open set in TM. The map

Exp: $(x, v) \to (x, \exp(v))$ maps $\tilde{\mathsf{E}}$ surjectively to its image

$$E := \{(x, y) \in M \times M \mid \text{there is a geodesic from } x \text{ to } y \text{ containing no conjugate point}\}. \tag{14}$$

Then, the construction (7) of (X_t, Y_t) can be applied to the case when the starting point (x_0, y_0) is in E and it remains valid until the first exit of (X_t, V_t) from $\tilde{E}$. We now modify the construction by Kendall (1986b): combine the coupled diffusions (X_t, Y_t) defined by (7), while the corresponding (X_t, V_t) is not too close to the boundary of $\tilde{E}$, with X_t, Y_t evolving independently.

For this, we first need a result on the distance function of two independent diffusions on M specified by (2). Lemma 3 in Appendix A of the Supplementary Material ensures the following property of $\rho(x, y)$ on neighbourhoods of the cut locus

$$C := \{(x, y) \in M \times M \mid y \text{ lies in the cut locus of } x\}$$

of $M \times M$: there is a set $C_0 \subset C$ such that

(i) C_0 contains the (first)-conjugate part of C;
(ii) for any $(x, y) \in C \backslash C_0$, there is a neighbourhood N of (x, y) in $M \times M$ and two smooth functions e_1 and e_2 on N such that

$$\rho(x', y') = \min\{e_1(x', y'), e_2(x', y')\}, \ \forall (x', y') \in N.$$

Since the (first)-conjugate part of C has co-dimension 2 in $M \times M$ (see Barden and Le (1997)), the result of that Lemma also implies that C_0 can be chosen to have co-dimension 2. Also, similarly to the argument at the beginning of the proof of Theorem 1, N in (ii) above can be chosen such that $\mathrm{Exp}^{-1}(N)$ is a disjoint union of two open sets V_1, V_2 in TM and, restricted to each V_i, Exp is a diffeo

morphism from that set to N. Then, the smooth function $e_i(x', y')$ constructed in the proof of Lemma

3 in Appendix A of the Supplementary Material is in fact the length of the geodesic from x' and y', the initial tangent vector v_i to which lies in V_i. That is, using our notation for the length of intervening

geodesics, we have $e_i(x', y') = \tilde{\rho}(x', \mathrm{exp}'(v_i))$. This leads to the following generalisation of Theorem

5 of Kendall (1986b) and of Theorem 3 of Barden and Le (1997). The proof of this generalisation is a slight modification of the proof for Theorem 3 of Barden and Le (1997) (see also Le and Barden (1995) for more detailed derivations), and we hence omit it here.

Lemma 1. *Suppose that X_t and Y_t are independent diffusions on M, both satisfying* (2). *Then, the distance $\rho(X_t, Y_t)$ is a semimartingale and, before the first time that $X_t = Y_t$,*

$$d\rho(X_t, Y_t) = \sqrt{2} dB_t + \frac{1}{2} \left[L_{\varphi,1} \rho(X_t, Y_t) + L_{\varphi,2} \rho(X_t, Y_t) \right] dt - dL_t,$$

where B_t is a Brownian motion on $\mathbb{R}$; L is a non-decreasing process that is locally constant outside C; and, for fixed x_0 and $x * x_0$,

$$L_{\varphi,1} \rho(x, x_0) := \begin{cases} & \\ & \end{cases}$$

0 *if* $(x, x_0) \in C_0$; 1

$$\{L_\varphi \tilde{\rho}(\exp(v_1), x_0) + L_\varphi \tilde{\rho}(\exp(v_2), x_0)\} \text{ if } (x, x_0) \in C \backslash C_0;$$

$$L_\varphi \rho(x, x_0)^{x_0 x_0} \text{ otherwise,}$$

and $\mathsf{L}_{\varphi,2} \rho$ *is similarly defined with respect to the second argument of* ρ*, and where the operator* L_φ *is defined by* $L_\varphi = \frac{1}{2}\{\Delta - \langle \nabla\varphi, \nabla \square \rangle\}$.

To detect that the coupled (X_t, Y_t), constructed by (7), is close to the boundary of E and to control the independent diffusions X_t and Y_t, we need the following generalisation of a geometric description (see Kendall (1986b)), wherein we replace the Laplacian operator considered there with L_φ, and replace the lower bound constant c determining the set O_c (which was denoted by U_c by Kendall (1986b)) by $c\rho(x,y)$. Since φ is in $C^2(M)$, the proof for our result is analogous to that for the lemma in Kendall (1986b), and we omit it here.

Lemma 2. *For any c >0,*

$$O_c \subset \bar{O}_c \subset \tilde{E},$$

where

$$\tilde{O}_c := \{(x,v) \in E | L_{\varphi,1}\, \tilde{\rho}(x,\exp_x(v)) + L_{\varphi,2}\, \tilde{\rho}(x,\exp_x(v)) > -2c\rho(x,\exp_x(v))\}$$

and, as before, $\tilde{\rho}(x,\exp_x(v))$ denotes the length of the intervening geodesic $\gamma(t) = \exp(tv)$, $0 \le t \le 1$.

We are now ready to prove the following result for Riemannian manifolds M with non-empty conjugate locus (e.g., spheres), which is weaker than Theorem 1 in that the exponential contraction between the diffusions towards their initial points is in expectation and not pathwise.

Theorem 2. *Assume that the Bakry-Emery curvature criterion (A1) holds for a constant $\kappa > 0$. Then, for any $e \ge 1$ and for any $x_0, y_0 \in M$, there is a pair of diffusions (X_t, Y_t) starting from (x_0, y_0) such that both X_t and Y_t satisfy (2) and*

$$E\rho(X_t, Y_t) \le \rho(x_0, y_0)\, e^{-\kappa t}, \quad t \ge 0. \quad (15)$$

Note that, unlike the result of Theorem 1, the (X_t, Y_t) constructed here will depend on e.

Proof. Let $\kappa > 0$ be the constant in Bakry-Emery curvature criterion (A1). For given $e \in [1, n]$, fix $\delta_n > 0$ sufficiently large such that

(i) $\delta_n > \kappa + 4(n-1)/r_0^2$, where $r_0 > 0$ is the minimum of the injectivity radius and a fixed positive constant r_0 say;

(ii) $O_{\delta_n} \supset \{(x,y) \in M \times M \mid \rho(x,y) < r_0/2\}$, where $O_\delta = \text{Exp}(\bar{O}_\delta)$ and where O_c is the subset of TM as defined in Lemma 2 above.

We now construct diffusions X_t and Y_t, both satisfying (2), as follows. For given $(x_0, y_0) \in M \times M$, if there is a minimal geodesic between them which contains no conjugate point, we construct diffusions X_t and Y_t by solving (7) beginning at (x_0, y_0). By allowing the corresponding (X_t, V_t) to jump if necessary, as commented following the construction (7), we continue such a construction for (X_t, Y_t) until the first time that (X_t, V_t) leaves $O_{2\delta_n}$. Suppose that (X_t, V_t) leaves $O_{2\delta_n}$ at time τ. We then consider all minimal geodesics between X_τ and Y_τ containing no conjugate point and, if possible, choose one

for which the corresponding (X_τ, V_τ) lies in $\bar{O}_{\delta_n}$. We then repeat the construction as before with the

chosen new starting point. This iterated construction continues until the choice of such (X_τ, V_τ) in $\bar{O}_{\delta_n}$ is no longer possible. If it is not possible initially to choose a minimal geodesic containing no conjugate point, or if at some

stage a choice of the above (X_τ, V_τ) in $O\delta_n$ is impossible, then we continue the construction of X_t and Y_t by evolving them independently until (X_t, Y_t) hits $O\delta_n$.

To show that the required result holds for (X_t, Y_t) constructed in such a way, it is sufficient by Theorem 1 to restrict to the case when X_t and Y_t evolve independently. Then, (X_t, Y_t) is not in $O\delta_n$. Recalling

that a co-dimension 2 set in $M \times M$ is a polar set of a non-degenerate diffusion on $M \times M$ it follows from Lemmas 1 and 2 and from the choice of δ_n that

$$d\, e^{\kappa t}\rho(X_t,Y_t)$$

$$\leq dM_t + \epsilon\kappa e^{\kappa t}\rho(X_t,Y_t)\,dt$$

$$e e^{\kappa t}\rho(X_t,Y_t)^{-1}\left[L_{\varphi,1}\rho(X_t,Y_t) + L_{\varphi,2}\rho(X_t,Y_t)\right]dt$$

$$e(e-1)e^{\kappa t}\rho(X_t,Y_t)^{-2}\,dt$$

$$\leq dM_t + e e^{\kappa t}\rho(X_t,Y_t)\left[\kappa - \delta_n + (e-1)\rho(X_t,Y_t)^{-2}\right]dt$$

$$\leq dM_t + e e^{\kappa t}\rho(X_t,Y_t)\left[\kappa - \delta_n + 4(n-1)/r^2\right]dt$$

$$\leq dM_t,$$

where M_t is a martingale. Hence, we have $E\,\rho(X_t,Y_t)\leq\rho(x_0,y_0)\,e^{-\kappa t}$ as required.

Remark 2. In the literature, there are several ways to construct couplings for proving the existence of contractivity. For example, in the curvature setting, the framework of weighted Riemannian manifolds is now part of a broader one for CD-spaces (see e.g., Sturm (2006a,b)). In this context, the existence of contractive couplings was treated by Kuwada (2010), von Renesse and Sturm (2005). In particular, the Kuwada duality theorem (see Kuwada (2010), Theorem 2.2), in conjunction with the implication of contractivity of the heat flow under Curvature-Dimension condition, implies the existence of a contractive coupling such as in the proof of Corollary 1 in von Renesse and Sturm (2005). The coupling we construct here, in addition to proving the required contractivity, will also be employed in the Supplementary Material to study certain stochastic vector fields along the paths $X_{x,t}$ and $Y_{y,t}$, which play important roles in obtaining the Stein factors.

4. Solution to the Stein equation and Stein factors

We are now ready to turn our attention to the *Stein equation*

$$h(x) - E\,[h(X)] = L_{\varphi}f_h(x), \quad (16)$$

where h belongs to a suitable class of real-valued test functions on M. Using the distance process $\rho(X_{x,t},Y_{y,t})$ for a pair of diffusions $(X_{x,t},Y_{y,t})$ constructed above, in this Section we determine the solution f_h to the Stein equation (16) and examine its properties.

4.1. The solution f_h

Let

$$H_0 := \{h \in C_0(M) \mid h \text{ is Lipschitz with } C_0(h) < \infty\}. \quad (17)$$

Proposition 1. *Let* M *be a complete and connected Riemannian manifold. Assume that the Bakry-Emery curvature criterion (A1) holds for a constant* $\kappa > 0$ *and that* X *is a random variable on* M *with distribution* μ_φ *such that* $E[\rho(X,x)] < \infty$ *for some* $x \in$ M. *For every* $h \in H_0$ *the function*

$$f_h(x) := \int_0^\infty E[h(X)] - E\, h(X_{x,t})\, dt \quad (18)$$

is (i) well-defined; (ii) Lipschitz with constant $C_0(f_h) \le C_0(h)/\kappa$.

"

Remark 3. If M $= R^m$, $\mathrm{Ric}(u,u) + \mathrm{Hess}^\varphi(u,u) = \mathrm{Hess}^\varphi(u,u)$. Thus, Proposition 1(ii) recovers the corresponding result in Mackey and Gorham (2016), as the constant 2κ here corresponds to constant k there. Moreover, the result of Proposition 1(ii) is equivalent to that of Proposition 6.1 in Thompson (2020).

Proof. Let $(X_{x,t}, Y_{y,t})$ be the pair of diffusions in Theorem 2 with $e = 1$, starting from (x,y). Then, both $X_{x,t}$ and $Y_{y,t}$ satisfy (2). Since μ_φ is the invariant measure for Y_t, using the Lipschitz property of h and Theorem 2,

$$\int_0^\infty E[h(X)] - E\, h(X_{x,t})\, dt$$

$$= \int_0^\infty \int_M E\, h(Y_{y,t}) - E\, h(X_{x,t})\, d\mu_\varphi(y)dt$$

$$\le \int_0^\infty \int_M C_0(h)\, E\, \rho(X_{x,t}, Y_{y,t})\, d\mu_\varphi(y)dt$$

$$\le \int_0^\infty C_0(h)E[\rho(X,x)]\, e^{-\kappa t}\, dt$$

$$< \infty.$$

This proves that f_h is well-defined. Now, for any $x,y \in$ M,

$$\int_0^\infty |f_h(y) - f_h(x)| \le E\, h(Y_{y,t}) - E\, h(X_{x,t})\, dt$$

$$\int_0^\infty \le C_0(h)\, E\, \rho(X_{x,t}, Y_{y,t})\, dt$$

$$\le \int_0^\infty C_0(h)\rho(x,y)\, e^{-\kappa t}\, dt = C_0(h)\rho(x,y). \frac{1}{\kappa}$$

The next result shows that the function f_h defined by (18) solves the Stein equation for the probability measure μ_φ.

Theorem 3. *Assume that* M *is a complete and connected Riemannian manifold and that Bakry-Emery curvature criterion (A1) holds for a constant* $\kappa > 0$. *Let* X *be a random variable on* M *with distribution* μ_φ *such that* $E[\rho(X,x)] < \infty$ *for some* $x \in$ M. *For* $h \in H_0$, *the function* f_h *in (18) solves the Stein equation (16).*

Remark 4. When $M = \mathbb{R}^m$ this result recovers the result by Mackey and Gorham (2016); in particular, $\mathrm{E}\, L_\varphi f_h(X) = 0$. On the other hand, the Bakry-Emery curvature criterion (A1) implies certain restrictions on the probability measures to which we can apply Theorem 3. For example, as noted in Example 1(i), one cannot apply it to von Mises-Fisher distributions on the circle. In this case, using direct integration by parts, for probability measures μ_φ with $X \sim \mu_\varphi$ on S^1, the function

$$\int_{-\pi}^{x \varphi(x)} (h(y) - \mathrm{E}\,[h(X)]) \mathrm{d}\mu_\varphi(y)\,,$$

for a constant a, solves the Stein equation $h(x) - \mathrm{E}\,[h(X)] = g'(x) - \varphi'(x)g_h(x)$ associated with first-order Stein operator $A_\varphi g - \varphi g = g' - \varphi' g$ (see Lewis (2021)).

Proof. Let $X_{x,t}$ be a diffusion starting from x and satisfying (2). Since the corresponding semigroup $\{P_t \,|\, t \geq 0\}$ is strongly continuous on $C_0(M)$ and L_φ is the infinitesimal generator of $X_{x,t}$, we have

$$(P_t h)(x) - h(x) = L_\varphi \mathrm{E} \int_0^t h(X_{x,s})\, \mathrm{d}s$$

for $h \in C_0(M)$ (Ethier and Kurtz, 1986, Prop. 1.5). However, for $\tilde{h}(x) = h(x) + a$ where $a \in \mathbb{R}$, $h(x) - \mathrm{E}\,\tilde{h}(X) = h(x) - \mathrm{E}\,[h(X)]$. Then, by taking $a = \mathrm{E}\,[h(X)]$ and noting $L_\varphi(a) = 0$, we can also write the above as

$$(P_t h)(x) - h(x) = -L_\varphi \mathrm{E} \int_0^t [h(X)] - \mathrm{E}\,\tilde{h}(X_{x,s})\, \mathrm{d}s. \quad (19)$$

Now, take $(X_{x,t}, Y_{y,t})$ to be the pair of diffusions, starting from (x, y), as Theorem 2 with $e = 1$. Since Y_t satisfies (2), the fact that μ_φ is the invariant measure of Y_t gives that

$$\mathrm{E}\,[h(X)] - (P_t h)(x) = \int_M \mathrm{E}\,h(Y_{y,t}) - \mathrm{E}\,h(X_{x,t})\, \mathrm{d}\mu_\varphi(y)$$

$$\leq C_0(h) \int_M \mathrm{E}\,\rho(Y_{y,t}, X_{x,t})\, \mathrm{d}\mu_\varphi(y) \leq C_0(h)\mathrm{E}\,[\rho(X, x)]e^{-\kappa t},$$

where the last inequality follows from Theorem 2 and where $C_0(h)$ is the Lipschitz constant for h. Thus,

$$\lim_{t \to \infty} (P_t h)(x) = \mathrm{E}\,[h(X)].$$

On the other hand, the result of Theorem 2 implies that we may apply the Dominated Convergence Theorem to obtain that, as $t \to \infty$, the right hand side of (19) tends to $-L_\varphi f_h(x)$, so that

$$h(x) - \mathrm{E}\,[h(X)] = L_\varphi f_h(x)$$

as required.

4.2. Stein factors

In the literature, Stein factors refer to bounds on solutions f_h of the Stein equation (16). A direct consequence

of Proposition 1 and Theorem 3 is that f_h defined by (18) is differentiable and Df_h is bounded.

Proposition 2. *Under the conditions of Theorem* 3, Df_h *exists and*

$$\sup_{x\in M} |Df_h(x)|_{op} \le C_0(h)/\kappa$$
where f_h is defined by (18).

We will see later in Section 5.1 that the bound on Df_h given above suffices to bound the Wasserstein distance between the probability measure μ_φ and another $\mu_\psi \propto e^{-\psi}$. However, for bounding more general integral (semi-)metrics, bounds on first-and second-order derivatives of f_h, known as Stein factors, are needed. Accordingly, denote by $\overset{!}{\mathrm{Ric}}$ the tensor equivalent to $\mathrm{Ric} + \mathrm{Hess}^\varphi$ in the sense that, for any $x \in M$,

and for any $u, u' \in T_x(M)$

$$\langle \overset{!}{\mathrm{Ric}}_\varphi(u), u' \rangle = \mathrm{Ric}(u, u') + \mathrm{Hess}_\varphi(u, u'). \qquad (20)$$

Recall that (see O'Neill (1983))

$$\mathrm{Hess}_\varphi(u, u') = \langle D_u(\nabla\varphi), u' \rangle, \qquad (21)$$

and that, in terms of a (local) frame field $e_1, \cdots, e_m$,

$$\mathrm{Ric}(u, u') = \sum_{i=1}^m \langle R(u, e_i)e_i, u' \rangle,$$

Stein's method on manifolds 1095

$$+ \int_0^\infty \mathrm{E}\left[Dh(Y_{y,t})(\tilde{v}_{x\,t} - v_{y\,t}) \right] dt$$

φ

where R denotes the Riemannian curvature tensor. Thus, it is possible to express $\overset{!}{\mathrm{Ric}}$ explicitly in

terms of the frame field as

$$\overset{!}{\mathrm{Ric}}_\varphi(u) = \sum_{i=1}^m R(u, e_i)e_i + D_u(\nabla\varphi). \qquad (22)$$

We can define the Lipschitz constant for $\overset{!}{\mathrm{Ric}}_\varphi$ in a similar way to the definition of the Lipschitz constant

given in (1). Let

$$\mathcal{H}_1 := \{ h \in C_0(M) \cap C(M) \,|\, C_0(h) < \infty, C_1(h) < \infty \}. \qquad (23)$$

Proposition 3. *Assume that the conditions of Theorem 3 hold. Assume further that $\overset{!}{\mathrm{Ric}}_\varphi$ is Lipschitz with finite Lipschitz constant $L(\overset{!}{\mathrm{Ric}}_\varphi)$. For every $h \in \mathcal{H}_1$ with f_h defined in (18), Df_h is Lipschitz with constant*

$$C_1(f_h) \le \frac{C_1(h)}{2\kappa} + \frac{L(\overset{!}{\mathrm{Ric}}_\varphi) C_0(h)}{2\kappa^2} .$$

Remark 5. As noted in Remark 3, if $M = \mathbb{R}^m$, $\mathrm{Ric}^{!}(u)$, $u' = \mathrm{Hess}^{\varphi}(u, u')$. Then, since $\mathrm{Hess}^{\varphi} = D^2\varphi$, $L(\mathrm{Ric}^{!}) = C_2(\varphi)$. Thus, Proposition 3 recovers the corresponding result in Mackey and Gorham (2016). On the other hand, the result of Proposition 3 differs from the corresponding Proposition 6.2 in Thompson (2020): in theirs, the relationship between the constant c_1 obtained and those given in the assumptions is not specified; using our notation, the upper bound for $C_1(f_h)$ there would depend only on $C_0(h)$ while ours depends on both $C_0(h)$ and $C_1(h)$.

Proof. The proof uses Lemmas 4 and 5 given in Appendix B of Supplementary Material. For any $x \in M$ and $v \in T_x(M)$, consider the vector field v_t^x along the path $X_{x,t}$ which solves the differential equation

$$\frac{D v_t^x}{dt} = -\frac{1}{2} \operatorname{Ric}_{\varphi}^!(v_t^x) \quad (24)$$

with $v_0^x = v$, where $X_{x,t}$ is the solution to (2). It is known that, for any fixed $t > 0$ and under the given condition for h, $N_s = D\,E\,h(X_{X_{x,s}},t-s)(v_s^x)$ is a local martingale for $0 \le s \le t$ (see Thalmaier (1997)). Since

$$|N_s| \le |D\,E\,h(X_{X_{x,s}},t-s)|_{op}|v_s^x| \le C_0(h)|v_s^x|,$$

using Lemma 4 (Appendix B of Supplementary Material) with $q = 1$, we see that $E\,[|N_s|] < \infty$. Hence, N_s is in fact a martingale on $[0,t]$, and so $E\,[N_0] = E\,[N_t]$, which in turn gives

$$D\,E\,h(X_{x,t})(v) = E\,Dh(X_{x,t})(v_t^x).$$

(See also Thompson (2020, Theorem 11.2), where the Z there corresponds to $-2\nabla\varphi$ here.) Thus, from the definition of f_h, the Dominated Convergence Theorem and Theorem 2, it follows that, for any $v \in T_x(M)$,

$$Df_h(x)(v) = D\int_0^\infty E\,h(X_{x,t})(v)\,dt = \int_0^\infty E\,Dh(X_{x,t})(v_t^x)\,dt. \quad (25)$$

Now, consider the pair of diffusions $(X_{x,t}, Y_{y,t})$, starting from (x, y), in Theorem 2 with $e = 2$. First, by applying the Hölder inequality, Theorem 2 and Lemma 4 (Appendix B of Supplementary Material), we have that

$$E\left|Dh(X_{x,t})(v_{x,t}^x) - \Pi_\gamma Dh(Y_{y,t})(v_{y,t}^Y)\right| \quad (26)$$

$$\le C_1(h)\,E\,\rho(X_{x,t}, Y_{y,t})|v_t^x| \le C_1(h)\rho(x, y)|v| \, e^{-2\kappa t}.$$

Moreover, writing v_t^y for the solution of (24) with the underlying path $X_{x,t}$ replaced by $Y_{y,t}$ and with the initial condition $v_0^y = \Pi_{\gamma_{x,y}}(v)$, and denoting $\Pi_\gamma(v_t^{xx})$ by $\tilde{v}_{tt}$, we also have

and for any $u, u' \in T_x(M)$

$$\text{Ric}!_\varphi(u), u' = \text{Ric}(u, u') + \text{Hess}_\varphi(u, u'). \qquad (20)$$

Recall that (see O'Neill (1983))

$$\text{Hess}_\varphi(u, u') = D_u(\nabla \varphi), u', \qquad (21)$$

$$X_{x,t}, Y_{y,t}$$

$$L(\text{Ric}^!)$$
$$\varphi$$

$$xy \quad xy \quad -\kappa t$$
$$\mathbb{E} Dh(Y_{y,t})(\tilde{v} - v) \le C_0(h) \, \mathbb{E} \, \tilde{v} - v \le C_0(h) \, \rho(x, y)|v| \, e, \qquad (27)$$
$$tt \quad tt$$

$$2\kappa$$

where the second inequality follows from Lemma 5 (Appendix B of Supplementary Material) with $q = 1$.

Finally, noting that $\Pi_{\gamma_{x,y}}(Df_h(y))(v) = Df_h(y)(\Pi_{\gamma_{x,y}}(v))$, together with (25), (26) and (27), implies that

$$|(Df_h(x) - \Pi_{\gamma_{x,y}} Df_h(y))(v)| = |Df_h(x)(v) - Df_h(y)(\Pi_{\gamma_{x,y}}(v))| \int^\infty$$
$$xy$$

$$\le \mathbb{E} \, Dh(X_{x,t})(v) - Dh(Y_{y,t})(v) \, dt$$
$$tt$$
$$0$$

$$\int_\infty$$

$$x$$

$$\le \mathbb{E} Dh(X_{x,t}) - \Pi_\gamma Dh(Y_{y,t})(v) \, dt$$
$$t$$
$$X_{x,t}, Y_{y,t}$$
$$0$$

$$L(\text{Ric}^!)$$

$$1 \; \varphi$$
$$C_1(h) + C_0(h)$$
$$2\kappa \; 2\kappa^2$$

$$\rho(x, y)|v|,$$

i.e. Df_h is Lipschitz with the required constant.

The argument in Remark 5 regarding the case when $M = \mathbb{R}^m$ can be extended to the case when M has constant Ricci curvature, which implies that the bounds in Mackey and Gorham (2016) continue to hold for such M. This gives the following corollary.

Corollary 1. *Assume that the conditions of Theorem* 3 *hold. Assume further that* M *is* Ric *flat and* φ *has finite Lipschitz constant* $C_2(\varphi)$. *Then, for every* $h \in H_1$ *and* f_h *as defined in* (18), Df_h *is Lipschitz with constant*

$$1 \; \underline{C_2(\varphi)}$$
$$C_1(f_h) \le C_1(h) + C_0(h) \, .$$

$$2\kappa \; 2\kappa^2 \text{ The curvature of the manifold plays a more explicit role in the}$$

Lipschitz constant for $D^2 f_h$. To see this, define the tensor $d^* R$ by $d^* R(u, v) = -\text{tr} D. R(\cdot, u)v$. Then $d^* R$ satisfies d^*

$R(v_1, v_2), v_3 = \diamond(D_{v_3}\mathrm{Ric}^!)(v_1), v_2 \diamond - \diamond(D_{v_2}\mathrm{Ric}^!)(v_3), v_1$. Noting that $R(\nabla\varphi)(u, v) = R(\nabla\varphi,u)v$, to simplify notation, we also define $R^!_{\varphi\varphi} = d^*R + D\mathrm{Ric}^! + R(\nabla\varphi)$. (28)

The bound on Df_h requires restriction to the smaller and smoother class H_1; the same is required when bounding $D^2 f_h$. Let

$$\mathsf{H}_2 := \{h \in \mathsf{C}_0(M) \cap \mathsf{C}^2(M)\mid C_0(h) < \infty, C_1(h) < \infty, C_2(h) < \infty\}. \text{ (29)}$$

assume that the conditions of Theorem 3 hold and that $\chi_1 = \sup l_{\mathrm{op}}(x)$ and $\chi_2 = m \sup_{\mathrm{op}}(x)$

$$|R^!_\varphi| \quad |R|^2$$

$$x \in M \quad x \in M$$

are both finite, where $R^!$ is defined by (28). Further, assume that $\mathrm{Ric}^!_\varphi, R^!_{\varphi\varphi}$ and R are all Lipschitz with finite Lipschitz constants $L(\mathrm{Ric}^!_\varphi), L(R^!_{\varphi\varphi})$ and $L(R)$ respectively. For every $h \in \mathsf{H}_2$ with f_h defined in (18):

(i) *If $\chi_2 = 0, D^2 f_h$ exists and is Lipschitz with constant*

$$13\ 13$$

$$2$$

$$C_2(f_h) \leq C_2(h) + C_1(h)C_2(\varphi) + C_0(h)\, C_3(\varphi) + C_2(\varphi).$$

$$3\kappa\ 4\kappa^2 4\kappa^2 4\kappa^3$$

(ii) *If $\chi>0$ and $\kappa>1/2$, then $D^2 f_h$ exists and is Lipschitz with constant*

$$2$$

$$\left\{ \left(\left(\left| \begin{array}{c} \blacklozenge \\ \end{array} \right| 1/2 \right) \right) \right\}$$

$$L(\mathrm{Ric}^{!})$$

φ

$$2\kappa 2$$
$$+$$
$$4$$

$\chi 2$

$$_1 + 2\,\chi$$

$\underline{2}$

$$1$$
$$C_2(f_h) \le C_2(h)$$
$$+ C_1(h)$$

with

$$3\kappa\ 8\kappa-14\kappa+1$$
$$\left| \left| \left(\left| \right| \right) \right) \right.$$
$$2\tilde{\beta}$$
$$+ C_0(h)$$
$$2\kappa-1$$

where

$$123$$
$$\tilde{\beta}^2 = \frac{\beta}{} + \frac{\beta}{} + \frac{\beta}{},$$

$$\chi 2$$

$$1_{\underline{1}} + 6\chi$$

$$\underline{2}$$

$$\beta_1 = 2mL(R)^2 + L(R^{!})^2 + L(\mathrm{Ric}^{!})^2,$$

$$2\ ^\circ 4\kappa+1\ ^\varphi$$

χ

$\underline{1}$

$$\beta_2 = L(\mathrm{Ric}^{!})L(R^{!}),$$

$\varphi\varphi$

κ

$$\begin{array}{c} 2 \\ \blacklozenge \ \chi \\ 1 \\ \kappa^2 2 \end{array}$$
$$1 + 2\,\chi L(\mathrm{Ric}^{!})^2$$

$2\ \varphi$

$$\beta=$$

$$3$$
$$.$$

Remark 6. Note that, $\chi_2 = 0$ corresponds to M being a flat manifold, such as a Euclidean space, a cylinder or a flat torus. Consequently, $\chi = L(\mathrm{Ric}^{!}) = C_2(\varphi)$ and $L(R^{!}) = C_3(\varphi)$. Our result thus

recovers the corresponding bound given in Mackey and Gorham (2016)for $\mathbb{R}^m$, where L_i, $M_i(h)$ and k in Mackey and Gorham (2016) correspond respectively to $C_{i-1}(\varphi)$, $C_{i-1}(h)$ and 2κhere. Our result establishes that their upper bound also holds for general complete and connected flat manifolds.

On the other hand, if M is locally symmetric, we have $DR = 0$. Then, it follows from (21) and (22) that $L(\mathrm{Ric}^!_\varphi)= L(D_u(\nabla\varphi))= C_2(\varphi)$ and $L(R^!_\varphi)= L(D(\mathrm{Ric}^!_\varphi))= C_3(\varphi)$. As symmetric manifolds are locally symmetric, this will hold for a class of familiar manifolds, such as spheres, hyperbolic spaces, projective spaces and the space of positive definite symmetric matrices. Pertinently, the upper bound for $C_2(f_h)$in Proposition 4 when $\chi_2 = 0$ is not the limit, as $\chi_2 \to 0$, of that for $\chi_2 > 0$. In addition, we need an extra requirement for $\kappa$when $\chi_2 > 0$.

Proof. The proof uses Lemmas 4, 5, 6 and 7 given in Appendix B of Supplementary Material. Consider the vector field V^x_t along the path $X_{x,t}$ which satisfies the stochastic covariant Itô equation

$$DV^x_t = R(\Xi \, dB_t, u^x_t)v^x_t - R^!_{\varphi t}(u^x_t, v^x_t)+ \tfrac{1}{2}\mathrm{Ric}^!_{\varphi t}(V^x_t) \, dt \quad (30)$$

with $V^x_0 = 0$, where Ξ is defined in (4), $R^!_\varphi$ and $\mathrm{Ric}^!_\varphi$ are defined by (28) and (20) respectively, and

where u^x_t and v_t are the solutions of (24) both with the underlying path $X_{x,t}$ and with the initial conditions $u^x_0 = u$ and $v_0 = v$ respectively. It is known that, for h satisfying the given conditions, $N_s =$

$D^2 \mathrm{E}_s h(X_{X_{x,s}},t-s)(u^x_s,v)+ D \, \mathrm{E}_s h(X_{X_{x,s}},t-s)(V^x_s)$is a local martingale for $0 \leq s \leq t$, for any fixed $t > 0$(Thompson, 2020, Lemma 11.3). Since

$$|N_s|\leq |\mathrm{E}D^2 h(X_{X_{x,s}},t-s)|_{\mathrm{op}}|u^x_s||v_s|+ |\mathrm{E}\, Dh(X_{X_{x,s}},t-s)|_{\mathrm{op}}|V^x_s|$$

$$\leq C_1(h)|u||v|+ C_0(h)|V^x_s|,$$

it follows from Lemmas 4 and 6 (Appendix B of Supplementary Material) that $E\,|N| < \infty$ so that

N is in fact a martingale for $0 \le s \le t$. Thus $E\,N_0 = E\,N_s$, which implies that, for any fixed $t > 0$ and $u, v \in T_x(M)$,

$$D^2 E\,h(X_{x,t})(u, v) = E\,D^2 h(X_{x,t})(u_t, v) + E\,Dh(X_{x,t})(V^x_{tt}).$$

Then, the definition of f_h, the Dominated Convergence Theorem and Theorem 2 together ensure that $D^2 f_h$ exists and that, for any $u, v \in T_x(M)$,

$$D^2 f_h(x)(u,v) = \int_0^\infty E\,D^2 h(X_{x,t})(u_t, v) + E\,Dh(X_{x,t})(V^x_{tt})\,dt. \tag{31}$$

Now, we construct a pair of diffusions $(X_{x,t}, Y_{y,t})$, starting from (x, y), as in Theorem 2. Since we need to apply Lemmas 5 and 7 (Appendix B of Supplementary Material) to the processes related to $(X_{x,t}, Y_{y,t})$ in the following proof, it is necessary to take the parameter e in the construction of $(X_{x,t}, Y_{y,t})$ to be 6. As in the proof of Proposition 3, write u^{yy}_{tt} and v for the solutions of (24) with the underlying path $X_{x,t}$ replaced by $Y_{y,t}$ and with the respective initial conditions $u^{yy}_{00} = \Pi_{\gamma_{x,y}}(u)$ and $v = \Pi_{\gamma_{x,y}}(v)$.

Also, let $\tilde{u}^x$ denote $\Pi_\gamma(u^x)$, and similarly for $\tilde{v}$ and $V^{\tilde{y}}$. Then,

$$\left|(D^2 f_h(x) - \Pi_{\gamma_{x,y}} D^2 f_h(y))(u,v)\right|$$

$$\le \int_0^\infty E\,D^2 h(X_{x,t})(u^{xx}_t, v_t) - D^2 h(Y_{y,t})(u^{yy}_t, v_t)\,dt \tag{32}$$

$$+ \int_0^\infty E\,Dh(X_{x,t})(V^x_{tt}) - Dh(Y_{y,t})(V^{\tilde{y}})\,dt.$$

Under the given conditions on h, the first term on the right hand side of (32) can be estimated as

$$\int_0^\infty$$

$$\int_0^\infty E \left| D^2h(X_{x,t})(u_t, v) - D^2h(Y_{y,t})(u_t, v) \right| dt$$

$$\leq \int_0^\infty \int_0^\infty E \left| D^2h(X_{x,t}) - \Pi_\gamma D^2h(Y_{y,t}) \right|_{X_{x,t},Y_{y,t}} (u_t, v) \, dt$$

$$+ \int_0^\infty \int_0^\infty E \left| D^2h(Y_{y,t})(\tilde{u}_t - u_t, \tilde{v}) \right| dt + E \left| D^2h(Y_{y,t})(u_t, \tilde{v} - v) \right| dt$$

$$\leq \int_0^\infty \int_0^\infty C_2(h) \, E\, \rho(X_{x,t}, Y_{y,t}) |u_t||v| \, dt + C_1(h)\, E \left[|\tilde{u}_t - u_t||\tilde{v}| + |u_t||\tilde{v} - v| \right] dt.$$

Similarly, for the second term on the right hand side of (32), we have that

$$\int_0^\infty E \left| Dh(X_{x,t})(V^x_t) - Dh(Y_{y,t})(V^y_t) \right| dt$$

$$\int_0^\infty \int_0^\infty \leq E \left| Dh(X_{x,t}) - \Pi_\gamma Dh(Y_{y,t}) \right|_{X_{x,t},Y_{y,t}} (V^x_t) \, dt + E \left| Dh(Y_{y,t})\, \tilde{V}^x_t - V^y_t \right| dt$$

$$\int_0^\infty \int_0^\infty \leq C_1(h)\, E\, \rho(X_{x,t}, Y_{y,t}) |V^x_t| \, dt + C_0(h)\, E \left| \tilde{V}^x_t - V^y_t \right| dt.$$

By the Hölder inequality, Theorem 2 and Lemmas 4, 5, 6 and 7 (Appendix B of Supplementary Material), it follows from the above estimations and from (32) that, if $\chi_2 > 0$,

$$\frac{1}{(D^2 f_h(x) - \Pi_{\gamma_{x,y}} D^2 f_h(y))(u,v)\rho(x,y)|u||v|} \left(\begin{array}{c} \left[\left[\left[\left[\\ L(\mathrm{Ric}) \ 2 \ \chi + \chi_2^{1/2} \end{array}\right.\right.\right.\right.$$

$$\frac{\varphi}{+ \frac{21}{2\kappa^2}}$$

$$\left. \begin{array}{c} \left] \ | \ | \ | \ \right\} \\ 2\beta^{-1} \ 4 \end{array}\right.$$

$$\leq C_2(h)$$

$$+ \ C_1(h) \ + \ C_0(h)$$

$$\frac{}{3\kappa \ 4\kappa + \ 18\kappa - 1 \ 2\kappa - 1} \ | \ | \ \lfloor \ | \ | \ \rfloor$$

when $\kappa > 1/2$, as required.

If $\chi_2 = 0$, we need to modify the above application of Lemmas 6 and 7 (Appendix B of Supplementary Material). This results in

$$\frac{1}{(D^2 f_h(x) - \Pi_{\gamma_{x,y}} D^2 f_h(y))(u,v)}$$

$$\rho(x,y)|u||v|$$

$$13 \ 13$$

$$2$$

$$\leq \ C_2(h) + \ C_1(h)C_2(\varphi) + \ C_0(h) \ C_3(\varphi) + \ C_2(\varphi).$$

$$3\kappa \ 4\kappa^2 4\kappa^2 4\kappa^3$$

This shows that Ddf_h is Lipschitz with the required constant.

5. Application to bounding integral (semi-)metrics

A key application of Stein's method is in obtaining upper bounds on an integral (semi-)metric $d_H(X,Z)$, with respect to some function class H, for an arbitrary random variable $Z \sim \nu$. Exploiting the characterising property of the operator L_φ,

$$E[h(Z)] - E[h(X)] = E L_\varphi f_h(Z) , \ \forall h \in H,$$

the task then reduces to obtaining a uniform upper bound on $E L_\varphi f_h(Z)$ over functions f_h using the Stein factors. The quantity d_H is clearly a semi-metric and is a metric only if H separates points in the set of signed measures on M.

5.1. Wasserstein distance between μ_φ and μ_ψ

The result of Theorem 3 in conjunction with the first-order bound in Proposition 2 can be used to obtain an

upper bound on the 1-Wasserstein distance between certain types of random variables. For this we consider the function class

$$H^1_{\leq 1} := \{h \in C(M) \mid h \text{ is Lipschitz with } C_0(h) \leq 1\},$$

under which d_H is a bonafide metric. The 1-Wasserstein distance between two random variables Z_1 and Z_2 on M is then defined as

$$d_W(Z_1, Z_2) := \sup_{h \in H^1_{\leq 1}} |E[h(Z_1)] - E[h(Z_2)]|.$$

Theorem 4. *Assume that the conditions of Theorem 3 hold. Let $Z \sim \mu_\psi$ such that $E[\rho(Z,x)] < \infty$ for some $x \in M$, where ψ satisfies (3) with some constant $\kappa' > 0$. Then*

$$d_W(Z,X) \leq \frac{1}{2\kappa} E[|\nabla(\psi - \varphi)(Z)|].$$

Proof. The proof pursues a similar argument to that of Proposition 4.1 of Mijoule, Reinert and Swan (2019). Note first that

$$\sup_{h \in H^1_{\leq 1}} |E[h(Z)] - E[h(X)]| = \sup_{h \in H^1_{\leq 1} \cap C_0(M)} |E[h(Z)] - E[h(X)]|.$$

have by Theorem 3 that $E[h(Z)] - E[h(X)] = E L_\varphi f_h(Z)$. On the other hand, the given assumption that $Z \sim \mu_\psi$, where ψ satisfies (3), also implies that $E L_\psi f_h(Z) = 0$ for $h \in H^1_{\leq 1} \cap C_0(M)$. Noting that

$$L_\varphi f_h(x) = L_\psi f_h(x) + \langle \nabla\psi(x) - \nabla\varphi(x), f_h(x) \rangle,$$

we obtain

$$E[h(Z)] - E[h(X)] = E[\langle \nabla\psi(Z) - \nabla\varphi(Z), \nabla f_h(Z)\rangle],$$

so that the result follows from Proposition 2. **Example 2.**

Assume that $M = S^m$ and that all probability measures μ_ϕ involved satisfy the condition $\text{Hess}^\phi \geq (2\kappa - (m-1))g$, for some $\kappa > 0$.

(i) The functions φ and ψ corresponding to von Mises-Fisher distributions $M(x_1,c_1)$ and $M(x_2,c_2)$ are respectively $-c_1 \cos\rho(x_1,x)$ and $-c_2 \cos\rho(x_2,x)$. Then,

$$|\nabla(\psi - \varphi)(x)| = c^* |\sin\rho(x^*,x)| \leq c^* \rho(x^*,x) \leq c^* \{\rho(x^*,x_2) + \rho(x_2,x)\},$$

where $c^* = |c_2 x_2 - c_1 x_1|$ and $x^* = (c_2 x_2 - c_1 x_1)/c^*$. From the symmetry between φ and ψ, it follows that the Wasserstein-1 distance d_W between $M(x_1,c_1)$ and $M(x_2,c_2)$ is bounded:

$$d_W(X_1,X_2) \leq \frac{|c_2 x_2 - c_1 x_1|}{4\kappa} \sum_{i=1}^{2} (\rho(x^*,x_i) + E[\rho(x_i,X_i)]),$$

where $X_i \sim M(c_i,x_i)$.

(ii) The function ψ corresponding to the Fisher-Watson distribution

$$W(x_1,x_2,c_1,c_2) \propto e^{c_1\langle x_1,x\rangle + c_2\langle x_2,x\rangle^2} dvol(x),$$

where $x_1, x_2 = 0$, is $-c_1 \cos \rho(x_1,x) - c_2 \cos^2 \rho(x_2,x)$. If μ_φ is the von Mises-Fisher distribution $M(x_1,c_1)$, then

$$|\nabla(\psi - \varphi)(x)| = c_2 |\sin(2\rho(x_2,x))|.$$

Hence, for $X \sim M(x_1,c_1)$ and $Z \sim W(x_1,x_2,c_1,c_2)$,

$$d_W(X,Z) \leq \frac{c_2}{2\kappa} E\left[|\sin(2\rho(x_2,Z))|\right].$$

(iii) Let $m > 2$ and $M = SO(m)$ with the bi-invariant metric determined by $g(E_1,E_2) = -\frac{1}{2}\operatorname{tr}(E_1 E_2)$ for skew-symmetric E_1, E_2. Assume that, for $S \in M$, $\varphi(S) = -c\operatorname{tr}(S_0 S)$ with $S_0 \in SO(m)$ and that constant $c > 0$. Then, μ_φ is a von Mises-Fisher distribution on $SO(m)$. Since for any skew-symmetric matrix E

$$\lim_{t \to 0} \frac{\varphi(Se^{tE}) - \varphi(S)}{t} = -c\operatorname{tr}(S_0 SE),$$

we have that $\nabla \varphi(S) = \frac{c}{2} S\{(S_0 S)^\top - S_0 S\}$. This implies that

$$|\nabla \varphi(S)|^2 = \frac{c^2}{2}\langle S\{(S_0 S)^\top - S_0 S\}, S((S_0 S)^\top - S_0 S)\rangle_S$$

$$c_2 = \frac{c^2}{2}\langle (S_0 S)^\top - S_0 S, (S_0 S)^\top - S_0 S\rangle_I$$

$$= -\frac{c^2}{2}\operatorname{tr}\left((S_0 S)^\top - S_0 S\right)^2 = c^2(m - \operatorname{tr}((S_0 S))).$$

If $c \in (0,(m-2)/2)$, there is a $\kappa > 0$ such that the Bakry-Emery curvature criterion (A1) holds, as seen in Example 1(iv). Then, if Z is a uniform random variable on $SO(m)$, $S_0 Z$ is also a uniform random variable and so

$$d_W(Z,X) \leq \frac{c}{2\kappa} E\sqrt{m - \operatorname{tr}(Z^2)}.$$

5.2. Integral semi-metrics for general distributions

If $h \in H_2$, the result of Theorem 3, together with Propositions 3 and 4, enable us to bound $E[h(Z)] - E[h(X)]$ for a more general random variable Z on M as follows, where H_2 is as defined in (29).

Corollary 2. *Assume that the conditions of Proposition 4 hold. Assume further that φ is Lipschitz with Lipschitz constants $C_i(\varphi), i = 0,1$. Then for every $h \in H_2$*

$$|E[h(Z)] - E[h(X)]| \leq \eta E[\rho(Z,X)],$$

where

$$\eta = mC_2(f_h) + C_0(\varphi)C_1(f_h) + C_1(\varphi)C_0(f_h)$$

and where $C_i(f_h)$ are bounded as in Propositions 3 and 4.

Proof. It follows from a direct estimation of $|\mathrm{E}\, L_\varphi(f_h)(Z)|$ that

$$|\mathrm{E}\, L_\varphi(f_h)(Z)| = |\mathrm{E}\, L_\varphi(f_h)(Z) - L_\varphi(f_h)(X)|$$

$$\leq |\mathrm{E}\,[(\Delta(f_h)(Z) - \Delta(f_h)(X))]|$$

$$+ |\mathrm{E}\,[\langle\nabla\varphi(Z),\nabla f_h(Z)\rangle - \langle\nabla\varphi(X),\nabla f_h(X)\rangle]|$$

$$\leq mC_2(f_h)\mathrm{E}\,[\rho(Z,X)]$$

$$+ \mathrm{E}\,\langle\nabla\varphi(Z),\nabla f_h(Z)\rangle - \Pi_\gamma_{X,Z}\langle\nabla\varphi(X),\nabla f_h(Z)\rangle$$

$$+ \mathrm{E}\,\langle\nabla\varphi(X),\Pi_\gamma_{Z,X}\nabla f_h(Z)\rangle - \langle\nabla\varphi(X),\nabla f_h(X)\rangle$$

$$\leq \{mC_2(f_h) + C_0(f_h)C_1(\varphi) + C_1(f_h)C_0(\varphi)\}\mathrm{E}\,[\rho(Z,X)]$$

as required.

A further simplification occurs when M is compact.

Corollary 3. *If* M *is compact then, for any Lipschitz function on* M *with* $C_0(h) \leq 1$, *any fixed* $\mathrm{E} > 0$ *and* $s > 0$, *there exists a* $g \in C^2(M)$ *with Lipschitz constants* $C_i(g), i = 0,1,2$, *such that* $C_0(g) \leq 1 + s$ *and*

$$|\mathrm{E}\,[h(Z)] - \mathrm{E}\,[h(X)]| \leq 2\mathrm{E} + mC_2(f_g) + C_0(f_g)C_1(\varphi) + C_1(f_g)C_0(\varphi)\,\mathrm{E}\,[\rho(Z,X)].$$

Proof. Since M is compact, any $g \in C^\infty(M)$ has bounded derivatives, and thus possesses finite Lipschitz constant $C_i(g), i = 0,1,2,\ldots,k$ for every k. This ensures that Lipschitz constants $C_i(f_g), i = 0,1,2$ of the Stein equation solution f_g are finite.

The existence of the requisite $g \in C^2(M)$ is guaranteed by the result in Azagra et al. (2007) on existence of a C^∞ Lipschitz approximation of a Lipschitz function. By Theorem 1 in Azagra et al. (2007), for every Lipschitz function h on M with Lipschitz constant 1 and for every $\mathrm{E}, s > 0$, there exists a $g \in C^\infty(M)$ such that $\sup_{x \in M}|g(x) - h(x)| < \mathrm{E}$ with $C_0(g) \leq 1 + s$. Thus, by applying Corollary 2 to g, we have

$$|\mathrm{E}\,[h(Z)] - \mathrm{E}\,[h(X)]|$$

$$\leq |\mathrm{E}\,[h(Z)] - \mathrm{E}\,[g(Z)]| + |\mathrm{E}\,[g(X)] - \mathrm{E}\,[h(X)]| + |\mathrm{E}\,[g(Z)] - \mathrm{E}\,[g(X)]|$$

$$\leq 2\mathrm{E} + |\mathrm{E}\,[g(Z)] - \mathrm{E}\,[g(X)]|$$

$$\leq 2\mathrm{E} + mC_2(f_g) + C_0(f_g)C_1(\varphi) + C_1(f_g)C_0(\varphi)\,\mathrm{E}\,[\rho(Z,X)],$$

Consider the function class

as required.

$$H_{\leq 1}^2 = \{h \in C^2(M)\,|\,h \text{ is Lipschitz with } C_0(h) \leq 1, C_1(h) \leq 1, C_2(h) \leq 1\}.$$

Since

$$\sup_{h \in H^2} |\mathrm{E}\,[h(Z)] - \mathrm{E}\,[h(X)]| = \sup_{h \in H_{\leq 1}^2 \cap C_0(M)_{\leq 1}} |\mathrm{E}\,[h(Z)] - \mathrm{E}\,[h(X)]|,$$

from Propositions 1, 3 and 4, as well as Corollary 2, the following result on the bound for the integral (semi-)metric

$$d_1(Z, X) := \sup_{h \in H_{s,1}^2} |E[h(Z)] - E[h(X)]|,$$

is immediate.

Theorem 5. *Assume that the conditions of Proposition 4 hold, and that φ is Lipschitz with Lipschitz constants $C_i(\varphi), i = 0,1$. Then, for any random variable Z on M,*

$$d_1(Z, X) \leq \eta^* E[\rho(Z, X)],$$

where, if $\chi_2 = 0$,

11 3

$$\eta^*$$

$$= m + (3C_2(\varphi) + C_3(\varphi)) + C_2(\varphi)^2$$

$3\kappa\ 4\kappa^2 4\kappa^3\ 1\ \underline{C_2(\varphi)}\ 1$

w
hi
le,
if
χ_2
>
0,

+

C

0

(

φ

)

+

+

C

1

(

φ

)

$2\kappa\ 2\kappa^2\kappa$

$$1\ \frac{L(\mathrm{Ric}^!)}{42\ \underline{\chi_2 + \chi}}^{2^{1/2}}\ 2\tilde{\beta}$$

$\underline{1}$

$$+\ ^\varphi\ +\ l\ +$$

$2\kappa 2$

$$\left(\left|\left|\right\}\right)\left|\left|\right\}$$

$$\eta^*$$

$$= m$$

$$\left\{ \frac{3\kappa}{} \left(\frac{8\kappa-14\kappa+1}{} \right) \left(\frac{2\kappa-1}{} \right) \right\}$$

$$L(\mathrm{Ric}^!)1_{\varphi}$$

$$+ \frac{2\kappa2}{1}$$

$$+ C_0(\varphi) + C_1(\varphi)$$

$$2\kappa'$$

$$\left| \left| \left(^{\kappa} \right) \right| \right|$$

and where the constants χ_1, χ_2 and $\tilde{\beta}$ are as in Proposition 4.

Acknowledgements

Supplementary Material

Auxiliary results (DOI: 10.3150/23-BEJ1625SUPP; .pdf). Appendix A contains a result on behaviour of the distance function around the first non-conjugate point of the cut locus, which is used in the proofs in Section 3. Appendix B is devoted to estimation of certain stochastic vector fields used in the proofs in Section 4.2.

References

Azagra, D., Ferrera, J., López-Mesas, F. and Rangel, Y. (2007). Smooth approximation of Lipschitz functions on Riemannian manifolds. *J. Math. Anal. Appl.* 326 1370–1378. MR2280987 https://doi.org/10.1016/j.jmaa.2006.03.088

Bakry, D. (1986). Un critère de non-explosion pour certaines diffusions sur une variété riemannienne complète. *C. R. Acad. Sci. Paris Sér. I Math.* 303 23–26. MR0849620 Barbour, A.D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* 25 175–184. MR0974580 https://doi.org/10.1017/s0021900200040341

Barbour, A.D. and Chen, L.H.Y. (2014). Stein's (magic) method. Preprint. Available at arXiv:1411.1179. Barden, D. and Le, H. (1997). Some consequences of the nature of the distance function on the cut locus in a Riemannian manifold. *J. Lond. Math. Soc. (2)* 56 369–383. MR1489143 https://doi.org/10.1112/S002461079700553X

Chatterjee, S. and Meckes, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* 4 257–283. MR2453473

Cheeger, J. and Ebin, D.G. (1975). *Comparison Theorems in Riemannian Geometry*. North-Holland Mathematical Library 9. Amsterdam: North-Holland. MR0458335

Chen, P., Nourdin, I., Xu, L. and Yang, X. (2019). Multivariate stable approximation in Wasserstein distance by Stein's method. Preprint. Available at arXiv:1911.12917.

Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0838085 https://doi.org/10.1002/9780470316658

Fang, X., Shao, Q.-M. and Xu, L. (2019). Multivariate approximations in Wasserstein distance by Stein's method and Bismut's formula. *Probab. Theory Related Fields* 174 945–979. MR3980309 https://doi.org/10.1007/s00440-018-0874-5

Greene, R.E. and Wu, H. (1979). *Function Theory on Manifolds Which Possess a Pole*. *Lecture Notes in Math.* 699. Berlin: Springer. MR0521983 Kendall, W.S. (1986a). Stochastic differential geometry, a coupling property, and harmonic maps. *J. Lond. Math. Soc. (2)* 33 554–566. MR0850971 https://doi.org/10.1112/jlms/s2-33.3.554 Kendall, W.S. (1986b). Nonnegative Ricci curvature and the Brownian coupling property. *Stochastics* 19 111–129. MR0864339 https://doi.org/10.1080/17442508608833419

Kent, J.T. (1994). The complex Bingham distribution and shape analysis. *J. Roy. Statist. Soc. Ser. B* 56 285–299. MR1281934

Kobayashi, S. and Nomizu, K. (1963). *Foundations of Differential Geometry. Vol I*. New York: Wiley. MR0152974

Kuwada, K. (2010). Duality on gradient estimates and Wasserstein controls. *J. Funct. Anal.* 258 3758–3774. MR2606871 https://doi.org/10.1016/j.jfa.2010.01.010

Le, H.L. and Barden, D. (1995). Itô correction terms for the radial parts of semimartingales on manifolds. *Probab. Theory Related Fields* 101 133–146. MR1314177 https://doi.org/10.1007/BF01192198 Le, H., Lewis, A., Bharath, K. and Fallaize, C. (2024). Supplement to "A diffusion approach to Stein's method on Riemannian manifolds." https://doi.org/10.3150/23-BEJ1625SUPP

Lewis, A. (2021). Stein's Method for Probability Distributions on S^1. Preprint. Available at arXiv:2105.13199.

Mackey, L. and Gorham, J. (2016). Multivariate Stein factors for a class of strongly log-concave distributions. *Electron. Commun. Probab.* 21 Paper No. 56, 14 pp. MR3548768 https://doi.org/10.1214/16-ecp15 Meckes, E. (2009). On Stein's method for multivariate normal approximation. In *High Dimensional Probability V: The Luminy Volume*. *Inst. Math. Stat. (IMS) Collect.* 5 153–178. Beachwood, OH: IMS. MR2797946 https://doi.org/10.1214/09-IMSCOLL511

Mijoule, G., Reinert, G. and Swan, Y. (2019). Stein operators, kernels and discrepancies for multivariate continuous distributions. Preprint. Available at arXiv:1806.03478.

O'Neill, B. (1983). *Semi-Riemannian Geometry: With Applications to Relativity*. *Pure and Applied Mathematics* 103. New York: Academic Press. MR0719023

Röllin, A. (2012). On the optimality of Stein factors. In *Probability Approximations and Beyond*. *Lect. Notes Stat.* 205 61–72. New York: Springer. MR3289376 https://doi.org/10.1007/978-1-4614-1966-2_5

Ross, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* 8 210–293. MR2861132 https://doi.org/10.1214/11-PS182

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory* 583–602. Berkeley, CaA: Univ. California Press. MR0402873

Sturm, K.-T. (2006a). On the geometry of metric measure spaces. I. *Acta Math.* 196 65–131. MR2237206 https://doi.org/10.1007/s11511-006-0002-8

Sturm, K.-T. (2006b). On the geometry of metric measure spaces. II. *Acta Math.* 196 133–177. MR2237207 https://doi.org/10.1007/s11511-006-0003-7

Thalmaier, A. (1997). On the differentiation of heat semigroups and Poisson integrals. *Stoch. Stoch. Rep.* 61 297–321. MR1488139 https://doi.org/10.1080/17442509708834123 Thompson, J. (2020). Approximation of Riemannian measures by Stein's method. Preprint. Available at arXiv: 2001.009. von Renesse, M.-K. and Sturm, K.-T. (2005). Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.* 58 923–940. MR2142879 https://doi.org/10.1002/cpa.20060