UNIQUENESS OF THE FISHER–RAO METRIC ON THE SPACE OF SMOOTH DENSITIES

MARTIN BAUER, MARTINS BRUVERIS, PETER W. MICHOR

Abstract. On a closed manifold of dimension greater than one, every smooth weak Riemannian metric on the space of smooth positive probability densities, that is invariant under the action of the diffeomorphism group, is a multiple of the Fisher–Rao metric.

Introduction. The Fisher–Rao metric on the space $\mathrm{Prob}(M)$ of probability densities is of importance in the field of information geometry. Restricted to finite-dimensional submanifolds of $\mathrm{Prob}(M)$, so-called statistical manifolds, it is called Fisher's information metric [1]. The Fisher–Rao metric has the property that it is invariant under the action of the diffeomorphism group. The interesting question is whether it is the unique metric possessing this invariance property. A uniqueness result was established [4, p. 156] for Fisher's information metric on finite sample spaces and [2] extended it to infinite sample spaces.

The Fisher–Rao metric on the infinite-dimensional manifold of all positive probability densities was studied in [5], including the computation of its curvature. A consequence of our main theorem in this article is the infinite-dimensional analogue of the result in [4]:

Theorem. *Let M be a compact manifold without boundary of dimension ≥ 2. Then any smooth weak Riemannian metric on the space* $\mathrm{Prob}(M)$ *of smooth positive probability densities, that is invariant under the action of the diffeomorphism group of M, is a multiple of the Fisher–Rao metric.*

The situation for a 1-dimensional manifold is described at the end of the paper. Our result holds for smooth positive probability densities on a compact manifold. However, the proof can be adapted to a suitable (and there are many choices) space of densities on a non-compact manifold.

Acknowledgments. This question was brought to our attention during a workshop at Öli-Hütte above Bad Gastein in Austria, July 14–20, 2014. We thank all the participants of the workshop for the friendly atmosphere and helpful discussions.

Date: May 10, 2016 .

2010 *Mathematics Subject Classification.* Primary 58B20, 58D15.

Key words and phrases. Fisher–Rao Metric; Information Geometry; Invariant Metrics; Space of Densities; Groups of Diffeomorphisms.

MB was supported by 'Fonds zur Förderung der wissenschaftlichen Forschung, Projekt P 24625'.

The space of densities. Let M^m be a smooth manifold without boundary. Let (U_α, u_α) be a smooth atlas for it. The *volume bundle* $(\mathrm{Vol}(M), \pi_M, M)$ of M is the 1-dimensional vector bundle (line bundle) which is given by the following cocycle of transition functions:

$$\psi_{\alpha\beta} : U_{\alpha\beta} = U_\alpha \cap U_\beta \to \mathbb{R} \setminus \{0\} = GL(1,\mathbb{R}),$$
$$\psi_{\alpha\beta}(x) = |\det d(u_\beta \circ u_\alpha^{-1})(u_\alpha(x))| = \frac{1}{|\det d(u_\alpha \circ u_\beta^{-1})(u_\beta(x))|}.$$

Vol(M) is a trivial line bundle over M, but there is no natural trivialization. There is a natural order on each fiber. Since $\mathrm{Vol}(M)$ is a natural bundle of order 1 on M, there is a natural action of the group $\mathrm{Diff}(M)$ on $\mathrm{Vol}(M)$, given by

$$\begin{tikzcd} \mathrm{Vol(M)} \arrow[r, "|\det(T\varphi^{-1})|\circ\varphi"] \arrow[d] & \mathrm{Vol(M)} \arrow[d] \\ M \arrow[r, "\varphi"] & M \end{tikzcd}.$$

If M is orientable, then $\mathrm{Vol}(M) = \Lambda^m T^*M$. If M is not orientable, let $\tilde{M}$ be the orientable double cover of M with its deck-transformation $\tau : \tilde{M} \to \tilde{M}$. Then $\Gamma(\mathrm{Vol}(M))$ is isomorphic to the space $\{\omega \in \Omega^m(\tilde{M}) : \tau^*\omega = -\omega\}$. These are the 'formes impaires' of de Rham. See [10, 13.1] for this.

Sections of the line bundle $\mathrm{Vol}(M)$ are called densities. The space $\Gamma(\mathrm{Vol}(M))$ of all smooth sections is a Fréchet space in its natural topology; see [9]. For each section α of $\mathrm{Vol}(M)$ of compact support the integral $\int_M \alpha$ is invariantly defined as follows: Let (U_α, u_α) be an atlas on M with associated trivialization $\psi_\alpha : \mathrm{Vol}(M)|_{U_\alpha} \to \mathbb{R}$, and let f_α be a partition of unity with $\mathrm{supp}(f_\alpha) \subset U_\alpha$. Then we put

$$\int_M \mu = \sum_\alpha \int_{U_\alpha} f_\alpha \mu := \sum_\alpha \int_{u_\alpha(U_\alpha)} f_\alpha(u_\alpha^{-1}(y)).\psi_\alpha(\mu(u_\alpha^{-1}(y)))\, dy.$$

The integral is independent of the choice of the atlas and the partition of unity.

The Fisher–Rao metric. Let M^m be a smooth compact manifold without boundary. We denote by $\mathrm{Dens}_+(M)$ the space of smooth positive densities on M, i.e. $\mathrm{Dens}_+(M) = \{\mu \in \Gamma(\mathrm{Vol}(M)) : \mu(x) > 0\ \forall x \in M\}$. Let $\mathrm{Prob}(M)$ be the subspace of positive densities with integral 1 on M. Both spaces are smooth Fréchet manifolds, in particular they are open subsets of the affine spaces of all densities and densities of integral 1 respectively. For $\mu \in \mathrm{Dens}_+(M)$ we have $T_\mu \mathrm{Dens}_+(M) = \Gamma(\mathrm{Vol}(M))$ and for $\mu \in \mathrm{Prob}(M)$ we have

$$T_\mu \mathrm{Prob}(M) = \{\alpha \in \Gamma(\mathrm{Vol}(M)) : \int_M \alpha = 0\}.$$

The Fisher–Rao metric is a Riemannian metric on $\mathrm{Prob}(M)$ and is defined as follows:

$$G^{\mathrm{FR}}_\mu(\alpha, \beta) = \int_M \frac{\alpha}{\mu}\frac{\beta}{\mu}\mu.$$

This metric is invariant under the associated action of $\mathrm{Diff}(M)$ on $\mathrm{Prob}(M)$, since

$$\Big((\varphi^*)^* G^{\mathrm{FR}}\Big)_\mu(\alpha,\beta) = G^{\mathrm{FR}}_{\varphi^*\mu}(\varphi^*\alpha,\varphi^*\beta) = \int_M \Big(\frac{\alpha}{\mu}\circ\varphi\Big)\Big(\frac{\beta}{\mu}\circ\varphi\Big)\varphi^*\mu = \int_M \frac{\alpha}{\mu}\frac{\beta}{\mu}\mu\,.$$

The uniqueness result for the Fisher–Rao metric follows from the following classification of $\mathrm{Diff}(M)$-invariant bilinear forms on $\mathrm{Dens}_+(M)$.

Main Theorem. *Let M be a compact manifold without boundary of dimension ≥ 2. Let G be a smooth (equivalently, bounded) bilinear form on $\mathrm{Dens}_+(M)$ which is invariant under the action of $\mathrm{Diff}(M)$. Then*

$$G_\mu(\alpha,\beta) = C_1(\mu(M))\int_M \frac{\alpha}{\mu}\frac{\beta}{\mu}\,\mu + C_2(\mu(M))\int_M \alpha\cdot\int_M \beta$$

for some functions C_1, C_2 of the total volume $\mu(M)$.

To see that this theorem implies the uniqueness of the Fisher–Rao metric, note that if G is a $\mathrm{Diff}(M)$-invariant Riemannian metric on $\mathrm{Prob}(M)$, then we can equivariantly extend it to $\mathrm{Dens}_+(M)$ via

$$G_\mu(\alpha,\beta) = G_{\frac{\mu}{\mu(M)}}\left(\alpha - \Big(\int_M\alpha\Big)\frac{\mu}{\mu(M)}, \beta - \Big(\int_M\beta\Big)\frac{\mu}{\mu(M)}\right)\,.$$

Relations to right-invariant metrics on diffeomorphism groups. Let $\mu_0 \in \mathrm{Prob}(M)$ be a fixed smooth positive probability density. In [7] it has been shown, that the degenerate, $\dot H^1$-metric $\frac{1}{2}\int_M \mathrm{div}^{\mu_0}(X).\,\mathrm{div}^{\mu_0}(X).\mu_0$ on $\mathfrak{X}(M)$ is invariant under the adjoint action of $\mathrm{Diff}(M,\mu_0)$. Thus the induced degenerate right invariant metric on $\mathrm{Diff}(M)$ descends to a metric on

$$\mathrm{Prob}(M) \cong \mathrm{Diff}(M,\mu_0)\backslash\mathrm{Diff}(M) \quad \text{via} \quad \mathrm{Diff}(M)\ni\varphi\mapsto\varphi^*\mu_0\in\mathrm{Prob}(M)$$

which is invariant under the right action of $\mathrm{Diff}(M)$. This metric turns out to be the Fisher–Rao metric on $\mathrm{Prob}(M)$. In [11], the $\dot H^1$-metric was extended to a nondegenerate metric on $\mathrm{Diff}(M)$, that also descends to the Fisher–Rao metric. A consequence of our uniqueness result is the following:

Corollary. *Let $\dim(M)\geq 2$. If a weak right-invariant (possibly degenerate) Riemannian metric $\tilde G$ on $\mathrm{Diff}(M)$ descends to a metric G on $\mathrm{Prob}(M)$ via the right action, i.e., the mapping $\varphi\mapsto\varphi^*\mu_0$ from $(\mathrm{Diff}(M),\tilde G)$ to $(\mathrm{Prob}(M),G)$ is a Riemannian submersion, then G has to be a multiple of the Fisher–Rao metric.*

Note that any right-invariant metric $\tilde G$ on $\mathrm{Diff}(M)$ descends to a metric on $\mathrm{Prob}(M)$ via the left action $\varphi\mapsto\varphi_*\mu_0$; but this is not $\mathrm{Diff}(M)$-invariant in general.

For $M = S^1$ the descending property is much less restrictive, since in this case the group of volume preserving diffeomorphism is generated by constant vector fields only. Thus any right-invariant metric on the homogenous space $\mathrm{Diff}(S^1)/S^1$ descends to a $\mathrm{Diff}(S^1)$ invariant metric on $\mathrm{Prob}(S^1)$, e.g., the homogenous Sobolev metric of order $n\geq 1$:

$$G_{\mathrm{Id}}(X,Y) = \sum_{k=1}^n \int_{S^1} \partial_\theta^k X.\partial_\theta^k Y\,d\theta\,.$$

For $n = 1$ the metric descends to the Fisher–Rao metric and for $n = 2$ we obtain a higher order metric. For the one-dimensional situation see also the last section of this article, where relations between metrics on $\mathrm{Dens}_+(S^1)$ and $\mathrm{Met}(S^1)$ are discussed.

Related Work. In [2] the authors prove a related result about the uniqueness of an invariant 2-tensor field on the space of probability densities. However they assume that the tensor is defined on all integrable densities, non-smooth as well as singular, and is invariant not only under smooth diffeomorphisms, but under all sufficient statistics. This is a significantly more restrictive definition and a stronger invariance assumption, allowing the authors to consider probability densities that are step functions, thus reducing the problem to the finite-dimensional case of [4]. Instead, we only require the Riemannian metric to be defined on all smooth, positive densities and to be invariant under smooth diffeomorphisms. One can see the space $\mathrm{Prob}(M)$ as the smooth, regular core of the space of all probability densities.

Proof of the Main Theorem. Let us fix a basic probability density μ_0. By the Moser trick [12], see [10, 31.13] or the proof of [9, 43.7] for proofs in the notation used here which can be lifted to the orientable double cover, there exists for each $\mu \in \mathrm{Dens}_+(M)$ a diffeomorphism $\varphi_\mu \in \mathrm{Diff}(M)$ with $\varphi_\mu^*\mu = \mu(M)\mu_0 =: c.\mu_0$ where $c = \mu(M) = \int_M \mu > 0$. Then

$$\big((\varphi_\mu^*)^* G\big)_\mu(\alpha,\beta) = G_{\varphi_\mu^*\mu}(\varphi_\mu^*\alpha, \varphi_\mu^*\beta) = G_{c.\mu_0}(\varphi_\mu^*\alpha, \varphi_\mu^*\beta)\,.$$

Thus it suffices to show that for any $c > 0$ we have

$$G_{c\mu_0}(\alpha,\beta) = C_1(c).\int_M \frac{\alpha}{\mu_0}\frac{\beta}{\mu_0}\mu_0 + C_2(c)\int_M \alpha \cdot \int_M \beta$$

for some functions C_1, C_2 of the total volume $c = \mu(M)$. Since $c \mapsto c.\mu_0$ is a smooth curve in $\mathrm{Dens}_+(M)$, the functions C_1 and C_2 are then smooth in c. Both bilinear forms are still invariant under the action of the group $\mathrm{Diff}(M, c\mu_0) = \mathrm{Diff}(M,\mu_0) = \{\psi \in \mathrm{Diff}(M) : \psi^*\mu_0 = \mu_0\}$. The bilinear form

$$T_{\mu_0}\mathrm{Dens}_+(M) \times T_{\mu_0}(M)\,\mathrm{Dens}_+(M) \ni (\alpha,\beta) \mapsto G_{c\mu_0}\Big(\frac{\alpha}{\mu_0}\mu_0, \frac{\beta}{\mu_0}\mu_0\Big)$$

can be viewed as a bilinear form

$$C^\infty(M) \times C^\infty(M) \ni (f,g) \mapsto G_c(f,g)\,.$$

We will consider now the associated bounded mapping

$$\check G_c : C^\infty(M) \to C^\infty(M)' = \mathcal{D}'(M)\,.$$

(1) The Lie algebra $\mathfrak{X}(M,\mu_0)$ of $\mathrm{Diff}(M,\mu_0)$ consists of vector fields X with

$$0 = \mathrm{div}^{\mu_0}(X) := \frac{\mathcal{L}_X\mu_0}{\mu_0}.$$

On an oriented open subset $U \subset M$, each density is an m-form, $m = \dim(M)$, and $\mathrm{div}^{\mu_0}(X) = di_X\mu_0/\mu_0$.

The mapping $\hat{\imath}_{\mu_0} : \mathfrak{X}(U) \to \Omega^{m-1}(U)$ given by $X \mapsto i_X\mu_0$ is an isomorphism. The Lie subalgebra $\mathfrak{X}(U,\mu_0)$ of divergence free vector fields corresponds to the space of closed $(m-1)$-forms.

Denote by $\mathfrak{X}_{\text{exact}}(M,\mu_0)$ the set (not a vector space) of 'exact' divergence free vector fields $X = \hat{\imath}_{\mu_0}^{-1}(d\omega)$, where $\omega \in \Omega_c^{m-2}(U)$ for an oriented open subset $U \subset M$.

(2) *If for $f \in C^\infty(M)$ and a connected open set $U \subseteq M$ we have $(\mathcal{L}_X f)|U = 0$ for all $X \in \mathfrak{X}_{exact}(M,\mu_0)$, then $f|U$ is constant.*

Since we shall need some of the notation later on, we give details. Let $x \in U$. For every tangent vector $X_x \in T_xM$ we can find a vector field $X \in \mathfrak{X}_{\text{exact}}(M,\mu_0)$ such that $X(x) = X_x$; to see this, choose a chart (U_x, u) near x such that $\mu_0|U_x = du^1 \wedge \cdots \wedge du^m$, and choose $g \in C_c^\infty(U_x)$, such that $g = 1$ near x. Then $X := \hat{\imath}_{\mu_0}^{-1} d(g.u^2.du^3 \wedge \cdots \wedge du^m) \in \mathfrak{X}_{\text{exact}}(M,\mu_0)$ and $X = \partial_{u^1}$ near x. So we can produce a basis for T_xM and even a local frame near x. Thus $\mathcal{L}_X f|U = 0$ for all $X \in \mathfrak{X}_{\text{exact}}(M,\mu_0)$ implies $df = 0$ and hence f is constant.

(3) *If for a distribution $A \in \mathcal{D}'(M)$ and a connected open set $U \subseteq M$ we have $\mathcal{L}_X A|U = 0$ for all $X \in \mathfrak{X}_{exact}(M,\mu_0)$, then $A|U = C\mu_0|U$ for some constant C, meaning $\langle A, f\rangle = C\int_M f\mu_0$ for all $f \in C_c^\infty(U)$.*

Because $\langle \mathcal{L}_X A, f\rangle = -\langle A, \mathcal{L}_X f\rangle$, the invariance property $\mathcal{L}_X A|U = 0$ implies $\langle A, \mathcal{L}_X f\rangle = 0$ for all $f \in C_c^\infty(U)$. Clearly, $\int_M (\mathcal{L}_X f)\mu_0 = 0$. For each $x \in U$ let $U_x \subset U$ be an open oriented chart which is diffeomorphic to $\mathbb{R}^m$. Let $g \in C_c^\infty(U_x)$ satisfy $\int_M g\mu_0 = 0$; we will show that $\langle A, g\rangle = 0$. Because the integral over $g\mu_0$ is zero, the compact cohomology class $[g\mu_0] \in H_c^m(U_x) \cong \mathbb{R}$ vanishes; thus there exists $\alpha \in \Omega_c^{m-1}(U_x) \subset \Omega^{m-1}(M)$ with $d\alpha = g\mu_0$. Since U_x is diffeomorphic to $\mathbb{R}^m$, we can write $\alpha = \sum_j f_j d\beta_j$ with $\beta_j \in \Omega^{m-2}(U_x)$ and $f_j \in C_c^\infty(U_x)$. Choose $h \in C_c^\infty(U_x)$ with $h = 1$ on $\bigcup_j \operatorname{supp}(f_j)$, so that $\alpha = \sum_j f_j d(h\beta_j)$ and $h\beta_j \in \Omega^{m-2}(M)$. In particular the vector fields $X_j = \hat{\imath}_{\mu_0}^{-1} d(h\beta_j)$ lie in $\mathfrak{X}_{\text{exact}}(M,\mu_0)$ and we have the identity $\sum_j f_j.i_{X_j}\mu_0 = \alpha$. This means

$$\sum_j (\mathcal{L}_{X_j} f_j)\mu_0 = \sum_j \mathcal{L}_{X_j}(f_j\mu_0) = \sum_j di_{X_j}(f_j\mu_0) = d\Big(\sum_j f_j.i_{X_j}\mu_0\Big) = d\alpha = g\mu_0$$

$$\sum_j \mathcal{L}_{X_j} f_j = g\,,$$

leading to

$$\langle A, g\rangle = \sum_j \langle A, \mathcal{L}_{X_j} f_j\rangle = -\sum_j \langle \mathcal{L}_{X_j} A, f_j\rangle = 0\,.$$

So $\langle A, g\rangle = 0$ for all $g \in C_c^\infty(U_x)$ with $\int_M g\mu_0 = 0$. Finally, choose a function φ with support in U_x and $\int_M \varphi\mu_0 = 1$. Then for any $f \in C_c^\infty(U_x)$, the function defined by $g = f - (\int_M f\mu_0).\varphi$ in $C^\infty(M)$ satisfies $\int_M g\mu_0 = 0$ and so

$$\langle A, f\rangle = \langle A, g\rangle + \langle A, \varphi\rangle \int_M f\mu_0 = C\int_M f\mu_0\,,$$

with $C_x = \langle A, \varphi\rangle$. Thus $A|U_x = C_x\mu_0|U_x$. Since U is connected, the constants C_x are all equal: Choose $\varphi \in C_c^\infty(U_x \cap U_y)$ with $\int \varphi\mu_0 = 1$. Thus (3) is proved.

(4) *The operator $\check{G}_c : C^\infty(M) \to \mathcal{D}'(M)$ has the following property: If for $f \in C^\infty(M)$ and a connected open $U \subseteq M$ the restriction $f|U$ is constant, then we have $\check{G}(f)|U = C_U(f)\mu_0|U$ for some constant $C_U(f)$.*

To see (4), for $x \in U$, choose $g \in C^\infty(M)$ with $g = 1$ near $M \setminus U$ and $g = 0$ on a neighborhood V of x. Then for any $X \in \mathfrak{X}_{\text{exact}}(M, \mu_0)$, that is $X = \hat{\iota}_{\mu_0}^{-1}(d\omega)$ for some $\omega \in \Omega_c^{m-2}(W)$ where $W \subset M$ is an oriented open set, let $Y = \hat{\iota}_{\mu_0}^{-1}(d(g\omega))$. The vector field $Y \in \mathfrak{X}_{\text{exact}}(M, \mu_0)$ equals X near $M \setminus U$ and vanishes on V. Since f is constant on U, $\mathcal{L}_X f = \mathcal{L}_Y f$. For all $h \in C^\infty(M)$ we have

$$\left\langle \mathcal{L}_X \check{G}_c(f), h \right\rangle = \left\langle \check{G}_c(f), -\mathcal{L}_X h \right\rangle = -G_c(f, \mathcal{L}_X h) = G_c(\mathcal{L}_X f, h) = \left\langle \check{G}_c(\mathcal{L}_X f), h \right\rangle,$$

since G_c is invariant. Thus also

$$\mathcal{L}_X \check{G}_c(f) = \check{G}_c(\mathcal{L}_X f) = \check{G}_c(\mathcal{L}_Y f) = \mathcal{L}_Y \check{G}_c(f).$$

Now Y vanishes on V and therefore so does $\mathcal{L}_X \check{G}_c(f)$. By (3) we have $\check{G}_c(f)|V = C_V(f)\mu_0|V$ for some $C_V(f) \in \mathbb{R}$. Since U is connected, all the constants $C_V(f)$ have to agree, giving a constant $C_U(f)$, depending only on U and f. Thus (4) follows.

By the Schwartz kernel theorem [6, Theorem 5.2.1], $\check{G}_c$ has a kernel $\hat{G}_c$, which is a distribution (generalized function) in

$$\mathcal{D}'(M \times M) \cong \mathcal{D}'(M) \mathbin{\bar{\otimes}} \mathcal{D}'(M) = (C^\infty(M) \mathbin{\bar{\otimes}} C^\infty(M))' \cong L(C^\infty(M), \mathcal{D}'(M)).$$

Note the defining relations

$$G_c(f, g) = \langle \check{G}_c(f), g \rangle = \langle \hat{G}_c, f \otimes g \rangle.$$

Moreover, $\hat{G}_c$ is invariant under the diagonal action of $\operatorname{Diff}(M, \mu_0)$ on $M \times M$. In view of the tensor product in the defining relations, the infinitesimal version of this invariance is: $\mathcal{L}_{X \times 0 + 0 \times X} \hat{G}_c = 0$ for all $X \in \mathfrak{X}(M, \mu_0)$.

(5) *There exists a constant $C_2 = C_2(c)$ such that the distribution $\hat{G}_c - C_2 \mu_0 \otimes \mu_0$ is supported on the diagonal of $M \times M$.*

Namely, if $(x, y) \in M \times M$ is not on the diagonal, then there exist open neighborhoods U_x of x and U_y of y in M such that $\overline{U_x} \times \overline{U_y}$ is disjoint to the diagonal, or $\overline{U_x} \cap \overline{U_y} = \emptyset$. Choose any functions $f, g \in C^\infty(M)$ with $\operatorname{supp}(f) \subset U_x$ and $\operatorname{supp}(g) \subset U_y$. Then $f|(M \setminus \overline{U_x}) = 0$, so by (4), $\check{G}_c(f)|(M \setminus \overline{U_x}) = C_{M \setminus \overline{U_x}}(f).\mu_0$. Therefore,

$$\begin{aligned} G_c(f, g) &= \langle \hat{G}_c, f \otimes g \rangle = \langle \check{G}_c(f), g \rangle \\ &= \langle \check{G}_c(f)|(M \setminus \overline{U_x}), g|(M \setminus \overline{U_x}) \rangle, \qquad \text{since } \operatorname{supp}(g) \subset U_y \subset M \setminus \overline{U_x}, \\ &= C_{M \setminus \overline{U_x}}(f) \cdot \int_M g\mu_0 \end{aligned}$$

By applying the argument for the transposed bilinear form $G_c^T(g, f) = G_c(f, g)$, which is also $\operatorname{Diff}(M, \mu_0)$-invariant, we arrive at

$$G_c(f, g) = G_c^T(g, f) = C'_{M \setminus \overline{U_y}}(g) \cdot \int_M f\mu_0.$$

Fix two functions f_0, g_0 with the same properties as f, g and additionally $\int_M f_0\mu_0 = 1$ and $\int_M g_0\mu_0 = 1$. Then we get $C_{M\setminus\overline{U_x}}(f) = C'_{M\setminus\overline{U_y}}(g_0)\int_M f\mu_0$, and so

$$\begin{aligned} G_c(f,g) &= C'_{M\setminus\overline{U_y}}(g_0)\int_M f\mu_0 \cdot \int_M g\mu_0 \\ &= C_{M\setminus\overline{U_x}}(f_0)\int_M f\mu_0 \cdot \int_M g\mu_0 \,. \end{aligned}$$

Since $\dim(M) \geq 2$ and M is connected, the complement of the diagonal in $M \times M$ is also connected, and thus the constants $C_{M\setminus\overline{U_x}}(f_0)$ and $C'_{M\setminus\overline{U_y}}(g_0)$ cannot depend on the functions f_0, g_0 or the open sets U_x and U_y as long as the latter are disjoint. Thus there exists a constant $C_2(c)$ such that for all $f, g \in C^\infty(M)$ with disjoint supports we have

$$G_c(f,g) = C_2(c)\int_M f\mu_0 \cdot \int_M g\mu_0$$

Since $C_c^\infty(U_x \times U_y) = C_c^\infty(U_x)\,\bar{\otimes}\,C_c^\infty(U_y)$, this implies claim (5).

Now we can finish the proof. We replace $\hat{G}_c \in \mathcal{D}'(M \times M)$ by $\hat{G}_c - C_2\mu_0 \otimes \mu_0$ and thus assume without loss that the constant C_2 in (5) is 0. Let (U, u) be an oriented chart on M such that $\mu_0|U = du^1 \wedge \cdots \wedge du^m$. The distribution $\hat{G}_c|U \times U \in \mathcal{D}'(U \times U)$ has support contained in the diagonal and is of finite order k, since M is compact. By [6, Theorem 5.2.3], the corresponding operator $\check{G}_c : C_c^\infty(U) \to \mathcal{D}'(U)$ is of the form $\hat{G}_c(f) = \sum_{|\alpha|\leq k} A_\alpha.\partial^\alpha f$ for $A_\alpha \in \mathcal{D}'(U)$, so that $G(f,g) = \langle \check{G}_c(f), g\rangle = \sum_\alpha \langle A_\alpha, (\partial^\alpha f).g\rangle$. Moreover, the A_α in this representation are uniquely given, as is seen by a look at [6, Theorem 2.3.5].

For $x \in U$ choose an open set U_x with $x \in U_x \subset \overline{U_x} \subset U$, and choose $X \in \mathfrak{X}_{\text{exact}}(M, \mu_0)$ with $X|U_x = \partial_{u^i}$, as in the proof of (2). For functions $f, g \in C_c^\infty(U_x)$ we then have, by the invariance of G_c,

$$\begin{aligned} 0 &= G_c(\mathcal{L}_X f, g) + G_c(f, \mathcal{L}_X g) = \langle \hat{G}_c|U \times U, \mathcal{L}_X f \otimes g + f \otimes \mathcal{L}_X g\rangle \\ &= \sum_\alpha \langle A_\alpha, (\partial^\alpha\partial_{u^i} f).g + (\partial^\alpha f)(\partial_{u^i} g)\rangle \\ &= \sum_\alpha \langle A_\alpha, \partial_{u^i}((\partial^\alpha f).g)\rangle = \sum_\alpha \langle -\partial_{u^i} A_\alpha, (\partial^\alpha f).g\rangle \,. \end{aligned}$$

Since the corresponding operator has again a kernel distribution which is supported on the diagonal, and since the distributions in the representation are unique, we can conclude that $\partial_{u^i} A_\alpha|U_x = 0$ for each α, and each i.

To see that this implies that $A_\alpha|U_x = C_\alpha\mu_0|U_x$, let $f \in C_c^\infty(U_x)$ with $\int_M f\mu_0 = 0$. Then, as in (3), there exists $\omega \in \Omega_c^{m-1}(U_x)$ with $d\omega = f\mu_0$. In coordinates we have $\omega = \sum_i \omega_i.du^1 \wedge \ldots \widehat{du^i} \cdots \wedge du^m$, and so $f = \sum_i (-1)^{i+1}\partial_{u^i}\omega_i$ with $\omega_i \in C_c^\infty(U_x)$. Thus

$$\langle A_\alpha, f\rangle = \sum_i (-1)^{i+1}\langle A_\alpha, \partial_{u^i}\omega_i\rangle = \sum_i (-1)^i \langle \partial_{u^i} A_\alpha, \omega_i\rangle = 0 \,.$$

Hence $\langle A_\alpha, f\rangle = 0$ for all $f \in C_c^\infty(U_x)$ with zero integral and as in the proof of (3) we can conclude that $A_\alpha|U_x = C_\alpha\mu_0|U_x$.

But then $G_c(f,g) = \int_{U_x} (Lf).g\mu_0$ for the differential operator $L = \sum_{|\alpha|\le k} C_\alpha \partial^\alpha$ with constant coefficients on U_x. Now we choose $g \in C_c^\infty(U_x)$ such that $g = 1$ on the support of f. By the invariance of G_c we have again

$$0 = G_c(\mathcal{L}_X f, g) + G_c(f, \mathcal{L}_X g) = \int_{U_x} L(\mathcal{L}_X f).1.\mu_0 + 0$$

for each $X \in \mathfrak{X}(M,\mu_0)$. Thus the distribution $f \mapsto \int_{U_x} L(f)\mu_0$ vanishes on all functions of the form $\mathcal{L}_X f$, and by (3) we conclude that $L(\quad).\mu_0 = C_x.\mu_0$ in $\mathcal{D}'(U_x)$, or $L = C_x \operatorname{Id}$. By covering M with open sets U_x, we see that all the constants C_x are the same. This concludes the proof of the Main Theorem. $\square$

Invariant metrics on $\operatorname{Dens}_+(S^1)$**.** It is interesting to consider the case $M = S^1$, which is not covered by the theorem. In the following let $M = S^1$. Then positive densities can be represented by positive one-forms. The space of all positive densities is isomorphic to the space of all Riemannian metrics on S^1 via the $\operatorname{Diff}(S^1)$-equivariant mapping

$$\Phi = (\quad)^2 : \operatorname{Dens}_+(S^1) \to \operatorname{Met}(S^1), \qquad \Phi(f d\theta) = f^2 d\theta^2 .$$

On $\operatorname{Met}(S^1)$ there exists a variety of $\operatorname{Diff}(S^1)$-invariant metrics; see [3]. We can take for example the family of Sobolev-type metrics. Write $g \in \operatorname{Met}(S^1)$ in the form $g = \tilde{g} d\theta^2$ and $h = \tilde{h} d\theta^2$, $k = \tilde{k} d\theta^2$ with $\tilde{g}, \tilde{h}, \tilde{k} \in C^\infty(S^1)$. Then for any integer n, the following metrics are $\operatorname{Diff}(S^1)$-invariant,

$$G_g^l(h,k) = \int_{S^1} \frac{\tilde{h}}{\tilde{g}} . (1 + \Delta^g)^n \left(\frac{\tilde{k}}{\tilde{g}} \right) \sqrt{\tilde{g}}\, d\theta ;$$

here Δ^g denotes the Laplacian on S^1 with respect to the metric g. Due to the equivariance of Φ, the pullback via Φ of any of these metrics yields a $\operatorname{Diff}(S^1)$-invariant metric on $\operatorname{Dens}_+(M)$, given by

$$G_\mu(\alpha,\beta) = 4 \int_{S^1} \frac{\alpha}{\mu} . \left(1 + \Delta^{\Phi(\mu)}\right)^n \left(\frac{\beta}{\mu} \right) \mu .$$

For $n = 0$ we obtain 4 times the Fisher–Rao metric. For $n \ge 1$ we see by the number of derivatives involved in the expression for $G_\mu(\alpha,\beta)$, that we obtain different $\operatorname{Diff}(S^1)$-invariant metrics on $\operatorname{Dens}_+(M)$ as well as on $\operatorname{Prob}(S^1)$.

References

[1] S.-I. Amari. *Differential-geometrical methods in statistics*, volume 28 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1985.

[2] N. Ay, J. Jost, H. V. Le, and L. Schwachhöfer. Information geometry and sufficient statistics. *The annals of statistics*, 2014.

[3] M. Bauer, P. Harms, and P. W. Michor. Sobolev metrics on the manifold of all Riemannian metrics. *J. Differential Geom.*, 94(2):187–208, 2013.

[4] N. N. Čencov. *Statistical decision rules and optimal inference*, volume 53 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, R.I., 1982. Translation from the Russian edited by Lev J. Leifman.

[5] T. Friedrich. Die Fisher-Information und symplektische Strukturen. *Math. Nachr.*, 153:273–296, 1991.

[6] L. Hörmander. *The analysis of linear partial differential operators. I*, volume 256 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1983. Distribution theory and Fourier analysis.

[7] B. Khesin, J. Lenells, G. Misiołek, and S. C. Preston. Geometry of diffeomorphism groups, complete integrability and geometric statistics. *Geom. Funct. Anal.*, 23(1):334–366, 2013.

[8] I. Kolář, P. W. Michor, and J. Slovák. *Natural operations in differential geometry.* Springer-Verlag, Berlin, 1993.

[9] A. Kriegl and P. W. Michor. *The convenient setting of global analysis*, volume 53 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 1997.

[10] P. W. Michor. *Topics in differential geometry*, volume 93 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2008.

[11] K. Modin. Generalized Hunter–Saxton equations, optimal information transport, and factorisation of diffeomorphisms. *J. Geom. Anal.*, 2014.

[12] J. Moser. On the volume elements on a manifold. *Trans. Amer. Math. Soc.*, 120:286–294, 1965.

Martin Bauer, Peter W. Michor: Fakultät für Mathematik, Universität Wien, Oskar-Morgenstern-Platz 1, A-1090 Wien, Austria.

Martins Bruveris: Department of Mathematics, Brunel University London, Uxbridge, UB8 3PH, United Kingdom

E-mail address: `bauer.martin@univie.ac.at`

E-mail address: `martins.bruveris@brunel.ac.uk`

E-mail address: `peter.michor@univie.ac.at`