Imperial College London

SAGEST 2018

Exponential manifold on the Gaussian space: Orlicz-Sobolev spaces

Giovanni Pistone

www.giannidiorestino.it

DE CASTRO STATISTICS Collegio Carlo Alberto

June 21, 2018

Abstract

Given a Gaussian space $(\mathbb{R}^n, \gamma)$, consider the set $\mathcal{M}$ of positive densities p which are connected to the unit density by an open Hellinger arc. The elements of $\mathcal{M}$ are precisely the densities of the form $e^{u-K(u)}$ where $E_\gamma(u) = 0$, $K(u)$ is a normalising constant, and u belongs to the exponential Orlicz space with weight γ. $\mathcal{M}$ is a manifold for an affine atlas of charts. The Gaussian assumption provides the exponential manifold with special features. Applications include the study of Boltzmann equation and the study the a gradient flow to the distribution with minimal Wasserstein distance.

Here, we discuss the manifold of smooth densities by taking as model space for the exponential manifold the Orlicz-Sobolev space with Gaussian weight γ. Statistical applications involving smooth densities are: Hyvärinen divergence and the finite-dimensional projection of solution of evolution equations for densities.

- G. Pistone. Nonparametric information geometry. In F. Nielsen and F. Barbaresco, editors, *Geometric science of information*, volume 8085 of *Lecture Notes in Comput. Sci.*, pages 5–36. Springer, Heidelberg, 2013. First International Conference, GSI 2013 Paris, France, August 28-30, 2013 Proceedings
- B. Lods and G. Pistone. Information geometry formalism for the spatially homogeneous Boltzmann equation. *Entropy*, 17(6):4323–4363, 2015
- G. Pistone. Information geometry of the Gaussian space. arXiv:1803.08135, 2018

Example: scoring rule

- On the statistical model $\mathcal{M}$ of positive densities on a measure space $(X, \mathcal{X}, \mu)$, a local scoring rule is a mapping $S \colon \mathcal{M} \ni q \mapsto S(\cdot, q)$ with values in random variables. The qualification "local" means that the scoring rule depends on the sample point x only.
- The risk under a positive density $p \in \mathcal{P}$ is $d(p, q) = \mathbb{E}_p [S(q)]$. Notice that we assume that the expected value is defined for each couple p, q $\mathcal{M}$.
- The scoring rule is proper is $q \mapsto d(p, q)$ is minimized at $q = p$ only that is, $d(p, q) \geq d(p, p)$ and $d(p, q) = d(p, p)$ implies $q = p$.
- There is a sampling version of the objective function, $\hat{d}(q) = \sum_{j=1}^{N} S(X_j, Q)$ with (X_j) IID p, and $\hat{q} = \operatorname{argmin} \hat{d}(q)$ is an estimator of p e.g., $S(x, q) = -\log q(x)$.
- The divergence associate to S is $D(p, q) = d(p, q) - d(p, p)$ and minimization of $q \mapsto D(p, q)$ is equivalent to the minimization of $q \mapsto d(p, q)$. However, $D(p, q)$ has no sampling version.

- A. Hyvärinen. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.*, 6:695–709, 2005
- M. Parry, A. P. Dawid, and S. Lauritzen. Proper local scoring rules. *Ann. Statist.*, 40(1):561–592, 2012

Example: Hyvärinen divergence I

- Let us assume now that the sample space is the n-dimensional real space and each density q in $\mathcal{M}$ is strictly positive and such that $\partial_j \log q = \partial_j q / q$ is square integrable for each $p \in \mathcal{M}$.
- The Hyvärinen divergence is

$$\mathrm{DH}\,(p|q) = \frac{1}{2} \int \left| \nabla \log p(x) - \nabla \log q(x) \right|^2 p(x)\ dx < \infty$$

- By expanding the squared norm of the difference, we obtain

$$\frac{1}{2} \int \left| \nabla \log p(x) \right|^2 p(x)\ dx + \frac{1}{2} \int \left| \nabla \log q(x) \right|^2 p(x)\ dx - \int \nabla \log p(x) \cdot \nabla \log q(x)\ p(x)\ dx\ ,$$

where the first term does not depend on q.

Example: Hyvärinen divergence II

- If $\nabla \log p = \nabla p / p$ and the border terms in the integration by parts are zero

$$-\int \nabla \log p(x) \cdot \nabla \log q(x) \; p(x) \; dx = \\ -\int \nabla p(x) \cdot \nabla \log q(x) \; dx = \int \Delta \log q(x) \; p(x) \; dx$$

- The Hyvärinen score is

$$S_{\mathrm{H}}(q) = \Delta \log q(x) + \frac{1}{2} \left| \nabla \log q(x) \right|^2 \; .$$

- Minimization of the expected Hyvärinen score is the same as minimization of the Hyvärinen divergence.
- All assumptions made are satisfied if $\mathcal{M}$ is the multivariate Gaussian model. This provides an example where a statistical method requires a detailed discussion of the properties of the spatial derivatives of the statistical model.

Example: Hyvärinen divergence III

- On the Gaussian space $(\mathbb{R}^n, \gamma)$, $\gamma(x) = (2\pi)^{-n/2} \mathrm{e}^{-|x|^2/2}$, consider the densities of exponential form $p = \mathrm{e}^{u-K(u)} \cdot \gamma$. Then, at least formally,

$$\mathrm{DH}\,(p|q) = \frac{1}{2} \int |\nabla u - \nabla v|^2 \, \mathrm{e}^{u-K(u)} \gamma(x) \; dx$$

  In this case, the ∇ operator could be taken in the sense of the analysis of the Gaussian space. But, $\mathrm{DH}\,(p|q) < \infty$?

- A variation where the integrability issue does not appear consistes of the substitution the log function with the Nigel Newton balanced chart $\log_A(t) = \int_1^y ds/A(s)$, with $A(t) = s/(1+s)$. A possible definition in then

$$\frac{1}{2} \int |\nabla \log_A p(x) - \nabla \log_A q(x)|^2 \, A(p(x)) \; dx$$

  where the cancellation holds and $A \circ p$ is bounded.

- P. Malliavin. *Integration and probability*, volume 157 of *Graduate Texts in Mathematics*. Springer-Verlag, 1995. With the collaboration of Hlne Airault, Leslie Kay and Grard Letac, Edited and translated from the French by Kay, With a foreword by Mark Pinsky
- N. J. Newton. An infinite-dimensional statistical manifold modelled on Hilbert space. *J. Funct. Anal.*, 263(6):1661–1681, 2012

IG is the geometry of the statistical bundle

- In a typical statistical set up, we have a set of positive densities $\mathcal{M}$ and a set of random variables B. We need the smoothness of a given map $\mathcal{M} \times B \ni (q, S) \mapsto F(q, S) \in \mathbb{R}$.
- A natural structure consists of endowing the model $\mathcal{M}$ with a differentiable atlas of charts and take as B a set of linear fibers on the manifold.
- Let be given an atlas on $\mathcal{M}$. A statistical bundle on $\mathcal{M}$ is

$$T\mathcal{M} = \{(p, u) | p \in \mathcal{M}, u \in B_p, \mathbb{E}_p [u] = 0\}$$

- Moreover, each fiber B_p is to be an expression in the atlas of the tangent space at p, $T_p\mathcal{M} \equiv B_p$. This last requirement is not trivial. For example, in general $L_0^2(p) \neq L_0^2(q)$.

- P. Gibilisco and G. Pistone. Connections on non-parametric statistical manifolds by Orlicz space geometry. *IDAQP*, 1(2):325–347, 1998

Orlicz model space

- If $\phi(y) = \cosh y - 1$, the Orlicz Φ-space $L^{\Phi}(p)$ is the vector space of all random variables u such that $\mathbb{E}_p[\Phi(\alpha U)]$ is finite for some $\alpha > 0$.
- $u \in L^{(\cosh -1)}(p)$ if, and only if, the moment generating function $\alpha \mapsto \mathbb{E}_p[\mathrm{e}^{\alpha u}]$ is finite in around 0 that is, $L^{(\cosh -1)}(p)$ is the space of sufficient statistics u in the exponential family $\theta \mapsto p_\theta \propto \mathrm{e}^{\theta u} \cdot p$.
- $L^{(\cosh -1)}(p)$ is a Banach space. The set

$$\left\{ u \in L^{(\cosh -1)}(p) \middle| \mathbb{E}_p[(\cosh -1)(u)] \leq 1 \right\}$$

  is the closed unit ball.
- If $(\cosh -1)_*$ is the convex conjugate of $(\cosh -1)$ we can define the Orlicz space $L^{(\cosh -1)_*}(p)$. The exponential space $L^{(\cosh -1)}(p)$ is the dual of the mixture space $L^{(\cosh -1)_*}(p)$ in the duality $(u, f) \mapsto \mathbb{E}_p[uf]$.

Maximal exponential family

- We define $B_p = \left\{ u \in L^{(\cosh -1)}(p) \middle| \mathbb{E}_p[u] = 0 \right\}$
- For each $p \in \mathcal{P}_>$, the moment generating functional is the positive lower-semi-continuous convex function $G_p \colon B_p \ni u \mapsto \mathbb{E}_p\left[\mathrm{e}^U\right]$.
- The cumulant generating functional is the non-negative lower-semi-continuous convex function $K_p = \log G_p$.
- The interior of the common proper domain $\{u | G_p(u) < +\infty\}^\circ = \{u | K_p(u) < \infty\}^\circ$ is an open convex set $\mathcal{S}_p$ containing the open unit ball (for the norm of the Orlicz space B_p).
- For each $p \in \mathcal{P}_>$, the maximal exponential family at p is

$$\mathcal{E}(p) = \left\{ \mathrm{e}^{u - K_p(u)} \cdot p \middle| u \in \mathcal{S}_p \right\}.$$

Portmanteau theorem

If $p, q \in \mathcal{P}_>$ we write $p \smile q$ if p and q are connected by an open exponential arc. It is an equivalence relation.
The following statements are equivalent:

1. $q \in \mathcal{E}\left(p\right)$;
2. $p \smile q$;
3. $\mathcal{E}\left(p\right) = \mathcal{E}\left(q\right)$;
4. $L^{(\cosh -1)}\left(p\right) = L^{(\cosh -1)}\left(q\right)$;
5. $\log\left(\frac{q}{p}\right) \in L^{(\cosh -1)}\left(p\right) \cap L^{(\cosh -1)}\left(q\right)$.
6. $\frac{q}{p} \in L^{1+\epsilon}(p)$ and $\frac{p}{q} \in L^{1+\epsilon}(q)$ for some $\epsilon > 0$.

Because of Item 4, all B_q, $q \in \mathcal{E}\left(p\right)$, are isomorphic under the mapping ${}^{\mathrm{e}}\mathbb{U}_p^q u = u - \mathbb{E}_q\left[u\right]$.

- A. Cena and G. Pistone. Exponential statistical manifold. *Ann. Inst. Statist. Math.*, 59(1):27–56, 2007
- M. Santacroce, P. Siri, and B. Trivellato. New results on mixture and exponential models by Orlicz spaces. *Bernoulli*, 22(3):1431–1447, 2016

e-charts

- For each $p \in \mathcal{E}$, consider the chart $s_p \colon \mathcal{E} \to B_p$

$$s_p(q) = \log\left(\frac{q}{p}\right) - \mathbb{E}_p\left[\log\left(\frac{q}{p}\right)\right]$$

- The inverse of each chart e_p is

$$e_p = s_p^{-1} \colon \mathcal{S}_p \ni u \mapsto \mathrm{e}^{u - K_p(u)} \cdot p$$

- $\{s_p | p \in \mathcal{P}_>\}$ is an affine atlas on $\mathcal{P}_>$ that defines the exponential manifold.
- Each equivalent class of connected densities $\mathcal{E}$ is a connected component of the exponential manifold.
- The information closure of any $\mathcal{E}\,(p)$ is $\mathcal{P}_\geq$. The reverse information closure of any $\mathcal{E}\,(p)$ is $\mathcal{P}_>$.

- D. Imparato and B. Trivellato. Geometry of extended exponential models. In *Algebraic and geometric methods in statistics*, pages 307–326. Cambridge Univ. Press, 2010

Summary

$$
\boxed{p \smile q} \implies
\boxed{
\begin{array}{ccccccc}
\mathcal{E}(p) & \xrightarrow{s_p} & \mathcal{S}p & \xrightarrow{I} & B_p & \xrightarrow{I} & L^{(\cosh - 1)}(p) \\
\| & & \downarrow{\scriptstyle s_q \circ s_p^{-1}} & & \downarrow{\scriptstyle d(s_q \circ s_p^{-1})} & & \| \\
\mathcal{E}(q) & \xrightarrow[s_q]{} & \mathcal{S}q & \xrightarrow[I]{} & B_q & \xrightarrow[I]{} & L^{(\cosh - 1)}(q)
\end{array}
}
$$

- If $p \smile q$, then $\mathcal{E}(p) = \mathcal{E}(q)$ and $L^{(\cosh - 1)}(p) = L^{(\cosh - 1)}(q)$.
- $B_p = \left\{ u \in L^{(\cosh - 1)}(p) \middle| \mathbb{E}_p[u] = 0 \right\}$
- $\mathcal{S}p \neq \mathcal{S}q$ and $s_q \circ s_p^{-1} \colon \mathcal{S}p \to \mathcal{S}q$ is affine

$$
s_q \circ s_p^{-1}(u) = u - \mathbb{E}_q[u] + \log\left(\frac{p}{q}\right) - \mathbb{E}_q\left[\log\left(\frac{p}{q}\right)\right]
$$

- The tangent application is $d(s_q \circ s_p^{-1})(u)[v] = v - \mathbb{E}_{e_p(u)}[v] = {}^{\mathrm{e}}\mathbb{U}_p^{e_p(u)} v$ (does not depend on p).

Gaussian space

- The Gaussian maximal exponential manifold is $\mathcal{E}(\gamma)$ with

$$\gamma(x) = (2\pi)^{-n/2} \exp\left(-|x|^2/2\right), \qquad x \in \mathbb{R}^n$$

- The relevant Orlicz spaces are the exponential space $L^{(\cosh -1)}(\gamma)$ and the mixture space $L^{(\cosh -1)_*}(\gamma)$.
- The mixture space $L^{(\cosh -1)_*}(\gamma)$ is separable; its dual is the exponential space $L^{(\cosh -1)}(\gamma)$.
- A positive density $f \in \mathcal{P}_>$ has finite entropy if, and only if, f belongs to the mixture space

$$-\int f(x) \log f(x) \gamma(x)\, dx < +\infty \quad \Leftrightarrow \quad f \in L^{(\cosh -1)_*}(\gamma) \ .$$

- $L^\infty(\gamma) \hookrightarrow L^{(\cosh -1)}(\gamma) \hookrightarrow L^a(\gamma) \hookrightarrow L^{(\cosh -1)_*}(\gamma) \hookrightarrow L^1(\gamma),\ a > 1$
- Restriction to the ball Ω_R: $L^{(\cosh -1)}(\gamma) \to L^a(\Omega_R)$, $a \geq 1$, and $L^{(\cosh -1)_*}(\gamma) \to L^1(\Omega_R)$.

- G. Pistone. Information geometry of the Gaussian space. arXiv:1803.08135, 2018

Notable elements in $L^{(\cosh -1)}(\gamma)$ and $L^{(\cosh -1)_*}(\gamma)$ I

- The exponential space $L^{(\cosh -1)}(\gamma)$ contains all polynomials with degree up to 2 and all functions which are bounded by such a polynomial.
- The mixture space $L^{(\cosh -1)_*}(\gamma)$ contains all $f \colon \mathbb{R}^d \to \mathbb{R}$ which are bounded by a polynomial, in particular, all polynomials.
- Poincaré inequality If $u \in \operatorname{Dom}(\nabla)$ in the sense of the Gaussian space i.e., $u, \partial_j u \in L^2(\gamma)$ then

$$\int \left| u(x) - \int u(y)\gamma(y)\ dy \right|^2 \gamma(x)\ dx \ \leq \int \left\| \nabla u(x) \right\|^2 \gamma(x)\ dx$$

- $f \in C^1_{\mathrm{p}}(\mathbb{R}^n)$

$$\left\| f - \int f(y)\gamma(y)\ dy \right\|_{L^{(\cosh -1)_*}(\gamma)} \leq \text{const} \left\| |\nabla f| \right\|_{L^{(\cosh -1)_*}(\gamma)}$$

In particular, if f is a density of the Gaussian space,

Notable elements in $L^{(\cosh -1)}(\gamma)$ and $L^{(\cosh -1)_*}(\gamma)$ II

- $f \in C_p^1(\mathbb{R}^n)$

$$\left\| f - 1 \right\|_{L^{(\cosh -1)_*}(\gamma)} \leq \text{const} \left\| \left| \nabla f \right| \right\|_{L^{(\cosh -1)_*}(\gamma)}$$

  This inequality is similar to an bound on the entropy.

- If $f, g \in C_p^2(\mathbb{R}^n)$ and $|x \cdot y| \leq |x|_1 \, |y|_2$ then

$$\left| \operatorname{Cov}_\gamma(f, g) \right| \leq \left| \left\| \nabla f \right\|_{L^{(\cosh -1)_*}(\gamma)} \right|_1 \left| \left\| \nabla g \right\|_{(L^{(\cosh -1)_*}(\gamma))^*} \right|_2 \quad .$$

  If f is a density of the Gaussian space,

$$\operatorname{Cov}_\gamma(f, g) = \int g(x) f(x) \gamma(x) \, dx \; - \int g(x) \gamma(x) \, dx \quad .$$

- I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012. From Stein's method to universality
- G. Pistone. Information geometry of the Gaussian space. arXiv:1803.08135, 2018

Velocity and scores

- $\mathcal{E}(\gamma) = \{q = u \in \mathcal{S}_\gamma | \subset\} L^{(\cosh -1)_*}(\gamma)$
- The inverse of the chart $s_\gamma : q \mapsto u$ is the exponential mapping $e_\gamma = s_\gamma^{-1} \colon \mathcal{S}_\gamma \to \mathcal{E}(\gamma)$, that is, $e_\gamma(u) = q$.
- The exponential mapping e_γ is defined on an open set of $L^{(\cosh -1)}(\gamma)$ and has values in $L^{(\cosh -1)_*}(\gamma)$. The chart mapping s_γ is not smooth and induces on $\mathcal{E}(\gamma)$ a special topology.
- The mapping $e_\gamma \colon \mathcal{S}_\gamma \ni u \mapsto \mathrm{e}^{u-K(u)} L^{(\cosh -1)_*}(\gamma)$ is continuously differentiable, with derivative in the direction $h \in L^{(\cosh -1)}(\gamma)$

$$d_h e_\gamma(u) = e_\gamma(u)(h - \mathbb{E}_{e_\gamma(u)}[h])$$

- If $\theta \mapsto u(\theta)$ is a smooth curve in $L^{(\cosh -1)}(\gamma)$, then $\theta \mapsto p(\theta) = e_\gamma(u(\theta))$ is a smooth curve in $L^{(\cosh -1)_*}(\gamma)$ and $\dot{p}(\theta) = p(\theta)(\dot{u}(\theta) - \mathbb{E}_{p(\theta)}[\dot{u}(\theta)])$, that is the expression of the velocity in the statistical bundle is

$$Su(\theta) = \dot{u}(\theta) - \mathbb{E}_{p(\theta)}[\dot{u}(\theta)] = \frac{\dot{u}(\theta)}{u(\theta)} = \frac{d}{d\theta} \log p(\theta) \in B_{p(\theta)}$$

- G. Pistone. Information geometry of the Gaussian space. arXiv:1803.08135, 2018

Natural gradient

- Given a scalar field $\Phi \colon \mathcal{E} \to \mathbb{R}$ the natural gradient is the section of the statistical bundle $\operatorname{grad} \Phi$ such that for all smooth curve $\theta \mapsto p(\theta) = \mathrm{e}^{u(\theta) - K(u(\theta))}$ it holds

$$\frac{d}{d\theta} \Phi(p(\theta) = \left\langle \operatorname{grad} \Phi(p(\theta)), Sp(\theta) \right\rangle_{p(\theta)}$$

where $\left\langle f, g \right\rangle_p = \mathbb{E}_p \left[fg \right]$, $f \in L^{(\cosh -1)_*}(p)$ and $g \in L^{(\cosh -1)}(p)$.

- For example, the natural gradient of the entropy $H(p) = -\int p(x) \log p(x) \gamma(x) \; dx$ is $\operatorname{grad} H(p) = -\log p - H(p)$.
- The gradient flow of Φ is the solution of the equation $Sp(\theta) = \operatorname{grad} \Phi(p(\theta)$. For example, the gradient flow of the entropy is

$$\frac{d}{d\theta} \log p(\theta) = -\log p(\theta) + \int p(x) \log p(x) \gamma(x) \; dx$$

- S.-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, feb 1998

Transport plan, Wasserstein

- Consider the product Gaussian space $(\mathbb{R}^{2n}, \gamma \otimes \gamma)$ with projection X and Y. The marginalization mapping

$$\mathcal{E}(\gamma \otimes \gamma) \ni p \mapsto (X_{\#}p, Y_{\#}p) \in \mathcal{E}(\gamma) \times \mathcal{E}(\gamma)$$

  has fibers

$$\Gamma(p_1, p_2) = \{p \in \mathcal{E}(\gamma \otimes \gamma) | X_{\#}p = p_1, Y_{\#}p = p_2\}$$

  which is a sub-manifolds of $\mathcal{E}(\gamma \otimes \gamma)$ called transport plan.

- The 2-Wasserstein functional $W(p) = \mathbb{E}_p\left[|X - Y|^2\right]$ has natural gradient

$$\operatorname{grad} W(p) = |X - Y|^2 - \mathbb{E}_p\left[|X - Y|^2\right]$$

- The restriction and projection of grad W on the statistical bundle of the sub-manifold of the transport plan $\Gamma(p_1, p_2)$ gives an equation for the gradient flow. The value of the 2-Wasserstein functional along the flow converges to the 2-Wasserstein distance between p_1 and p_2.

- L. Malago, G. Pistone, Talk at ICMS, Edinburgh 2015

Orlicz-Sobolev with Gaussian weight (GOS)

- The GOS spaces with weight M are the vector spaces

$$W^{1,(\cosh -1)}(\gamma) = \left\{ f \in L^{(\cosh -1)}(\gamma) \middle| \partial_j f \in L^{(\cosh -1)}(\gamma), j = 1, \dots, n \right\}$$

$$W^{1,(\cosh -1)_*}(\gamma) = \left\{ f \in L^{(\cosh -1)_*}(\gamma) \middle| \partial_j f \in L^{(\cosh -1)_*}(\gamma), j = 1, \dots, n \right\}$$

where ∂_j is the derivative in the sense of distributions.

- Both are Banach spaces with the norm of the graph

$$\|f\|_{W^{1,(\cosh -1)}(\gamma)} = \|f\|_{L^{(\cosh -1)}(\gamma)} + \sum_{j=1}^{n} \|\partial_j f\|_{L^{(\cosh -1)}(\gamma)}$$

$$\|f\|_{W^{1,(\cosh -1)_*}(\gamma)} = \|f\|_{L^{(\cosh -1)}(\gamma)} + \sum_{j=1}^{n} \|\partial_j f\|_{L^{(\cosh -1)}(\gamma)}$$

- J. Musielak. *Orlicz spaces and modular spaces*, volume 1034 of *Lecture Notes in Mathematics*. Springer-Verlag, 1983
- B. Lods and G. Pistone. Information geometry formalism for the spatially homogeneous Boltzmann equation. *Entropy*, 17(6):4323–4363, 2015

Remarks

- As $\phi \in C_0^\infty(\mathbb{R}^n)$ implies $\phi\gamma \in C_0^\infty(\mathbb{R}^n)$, for each $f \in W^{1,(\cosh -1)_*}(\gamma)$ we have

$$\langle \partial_j f, \phi \rangle_\gamma = \langle \partial_j f, \phi\gamma \rangle = -\langle f, \gamma\partial_j\phi - X_j\gamma\phi \rangle = \langle f, (X_j - \partial_j)\phi \rangle_\gamma$$

- The extension of the Stein operator $\delta_j = X_j - \partial_j$ to both $W^{1,(\cosh -1)}(\gamma)$ and $W^{1,(\cosh -1)_*}(\gamma)$ is of interest.
- Assume $f \in W^{1,a}(\gamma)$ for all $a \geq 1$. Then $\partial_j f \in L^a(\gamma)$ and

$$\int |x_j f(x)|^a \gamma(x)\, dx \;\leq$$
$$\left(\int |x_j|^{2a} \gamma(x)\, dx\right)^{1/2} \left(\int |u(x)|^{2a} \gamma(x)\, dx\right)^{1/2}$$

so that $\delta_j \colon \cap_{a \geq 1} W^{1,a}(\gamma) \to \cap_{a \geq 1} W^{1,a}(\gamma)$. In particular, $\delta_j \colon W^{1,(\cosh -1)}(\gamma) \to \cap_{a \geq 1} W^{1,a}(\gamma)$.

- P. Malliavin. *Stochastic analysis*, volume 313 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, 1997

Smoothness of GOS spaces I

- Every $u \in W^{1,(\cosh -1)}(\gamma)$ when restricted to an open ball of radius $R > 0$ belongs to the Sobolev space $W^{1,a}(\Omega_R)$ for all $a \geq 1$ i.e. $u_R \in \cap_{a \geq 1} W^{1,a}(\Omega_R)$.
- Every $f \in W^{1,(\cosh -1)_*}(\gamma)$ when restricted to an open ball of radius $R > 0$ belongs to the dual of the space $\cap_{a \geq 1} W^{1,a}(\Omega_R)$, in particular to $W^{1,1}(\Omega_R)$.
- Sobolev Each $u \in W^{1,(\cosh -1)}(\gamma)$ is a.s. continuous and Hölder of all orders on each $\overline{\Omega}_R$.
- If $u \in W^{1,(\cosh -1)}(\gamma)$, then $u, \partial_j u \in L^a(\gamma)$ for all $a \geq 1$ i.e.,

$$\mathrm{e}^{-\frac{1}{2a}|X|^2} u, \mathrm{e}^{-\frac{1}{2a}|X|^2} \partial_j u \in L^a(\mathbb{R}^n)$$

As

$$\partial_j \mathrm{e}^{-\frac{1}{2a}|X|^2} u = -\frac{1}{a} x_j \mathrm{e}^{-\frac{1}{2a}|X|^2} u + \mathrm{e}^{-\frac{1}{2a}|X|^2} \partial_j u$$

it follows

$$\left(\mathrm{e}^{-\frac{1}{2a}|X|^2} u\right) \in W^{1,a}(\mathbb{R}^n) \quad a \geq 1$$

Smoothness of GOS spaces II

- Morrey Because of

$$W^{1,(\cosh -1)}(\gamma) \ni u \mapsto \left(\mathrm{e}^{-\frac{1}{2a}|X|^2} u\right) \in W^{1,a}(\mathbb{R}^n) \quad a \geq 1$$

  it holds for each $a > n$ the uniform bound

$$u \in W^{1,(\cosh -1)}(\gamma) \quad \Rightarrow \\ \mathrm{e}^{-\frac{1}{2a}|x|^2} |u(x)| \leq C(n,a) \left\| \mathrm{e}^{-\frac{1}{2a}|x|^2} u \right\|_{W^{1,a}(\mathbb{R}^n)} \quad \text{a.s.}$$

  and the RHS is dominated by $\|u\|_{W^{1,(\cosh -1)}(\gamma)}$.

- The same assumption implies the global Hölder inequality

$$\mathrm{e}^{-\frac{1}{2a}|x|^2} u(x) - \mathrm{e}^{-\frac{1}{2a}|y|^2} u(y) \leq C(n.a) |x-y|^{1-n/a} \left\| \mathrm{e}^{\frac{1}{2a}|X|^2} u \right\|_{L^a(\mathbb{R}^n)} \leq \\ C(n.a) |x-y|^{1-n/a} \|u\|_{W^{1,(\cosh -1)}(\gamma)}$$

- The previous inequalities are not optimal!

Smoothness of GOS spaces III

- **Remark** We expect the space $W^{\infty,\cosh -1}(\gamma)$ of functions whose derivatives of all order belong to $L^{(\cosh -1)}(\gamma)$ to have infinitely differentiable elements. This provides an interesting class of random variables on the Gaussian space defined only by the differentiability and the integrability condition.
- If Φ is a diffeomorphism of $\mathbb{R}^n$, then

$$\Phi_*\gamma(x) = \exp\left(-\frac{1}{2}\left(\left|\Phi^{-1}(x)\right|^2 - |x|^2\right)\right)\left|\det\left(J\Phi^{-1}(x)\right)\right|\gamma(x)$$

and it would be interesting to have

$$-\frac{1}{2}\left(\left|\Phi^{-1}\right|^2 - |X|^2\right) - \log\left|\det\left(J\Phi^{-1}\right)\right| \in W^{\infty,\cosh -1}(\gamma)$$

in order to connect with the literature on the geometry of densities induced by the geometry of the group of diffeomorphisms.

- R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003
- H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011

Exponential family modeled on $W^{1,(\cosh -1)}(\gamma)$

- If we restrict the exponential family $\mathcal{E}(\gamma)$ to $W^{1,(\cosh -1)}(\gamma)$,

$$W_\gamma = W^{1,(\cosh -1)}(\gamma) \cap B_\gamma = \left\{ u \in W^{1,(\cosh -1)}(\gamma) \middle| \mathbb{E}_\gamma[u] = 0 \right\}$$

  we obtain the non-parametric exponential family

$$\mathcal{E}_1(\gamma) = \left\{ \mathrm{e}^{u-K(u)} \cdot \gamma \middle| u \in W^{1,(\cosh -1)}(\gamma) \cap \mathcal{S}_\gamma \right\}$$

- Because of $W^{1,(\cosh -1)}(\gamma) \hookrightarrow L^{(\cosh -1)}(\gamma)$ the set $W^{1,(\cosh -1)}(\gamma) \cap \mathcal{S}\gamma$ is open in W_γ and the cumulant functional $K : W^{1,(\cosh -1)}(\gamma) \cap \mathcal{S}\gamma \to \mathbb{R}$ is convex and differentiable.
- Many features of the exponential manifold carry over to this case. In particular, we can define for each $f \in \mathcal{E}_1(\gamma)$ the space

$$W_f = W^{1,(\cosh -1)}(\gamma) \cap B_\gamma = \left\{ u \in W^{1,(\cosh -1)}(\gamma) \middle| \mathbb{E}_f[u] = 0 \right\}$$

  to be models for the tangent spaces of $\mathcal{E}_1(\gamma)$. The e-transport acts on these spaces, ${}^{\mathrm{e}}\mathbb{U}_f^g : W_f \ni u \mapsto u - \mathbb{E}_g[u] \in W_g$, so that we can define the statistical bundle to be

$$S\,\mathcal{E}_1(\gamma) = \{(g, v) | g \in \mathcal{E}_1(\gamma), v \in W_f\}$$

  and take as charts the restrictions of the charts defined on $S\,\mathcal{E}(\gamma)$.

Calculus on $W^{1,(\cosh -1)}(\gamma)$ — I

- The exponential class, $C_0^{(\cosh -1)}(\gamma)$, is the closure of $C_0(\mathbb{R}^n)$ in the exponential space $L^{(\cosh -1)}(\gamma)$. The space $C_0^\infty(\mathbb{R}^n)$ is dense in $C_0^{(\cosh -1)}(\gamma)$.
- Assume $f \in L^{(\cosh -1)}(\gamma)$ and write $f_R(x) = f(x)(|x| > R)$. The following conditions are equivalent:
  1. The real function $\rho \mapsto \int (\cosh -1)(\rho f(x))\gamma(x)\, dx$ is finite for all $\rho > 0$;
  2. $f \in C_0^{(\cosh -1)}(\gamma)$;
  3. $\lim_{R\to\infty} \|f_R\|_{L^{(\cosh -1)}(\gamma)} = 0$.
- Translation by a vector
  1. For each $h \in \mathbb{R}^n$, the translation mapping $L^{(\cosh -1)}(\gamma) \ni f \mapsto \tau_h f$ is linear and bounded from $L^{(\cosh -1)}(\gamma)$ to itself. In particular,

$$\|\tau_h f\|_{L^{(\cosh -1)}(\gamma)} \leq 2\,\|f\|_{L^{(\cosh -1)}(\gamma)} \quad \text{if} \quad |h| \leq \sqrt{\log 2}\ .$$

Calculus on $W^{1,(\cosh -1)}(\gamma)$ — II

2. For all $g \in L^{(\cosh -1)_*}(\gamma)$ we have

$$\langle \tau_h f, g \rangle_\gamma = \langle f, \tau_h^* g \rangle_\gamma , \quad \tau_h^* g(x) = \mathrm{e}^{-h\cdot x - \frac{1}{2}|h|^2} \tau_{-h} g(x) ,$$

   and $|h| \leq \sqrt{\log 2}$ implies $\left\| \tau_h^* g \right\|_{L^{(\cosh -1)}(\gamma))^*} \leq 2 \left\| g \right\|_{L^{(\cosh -1)}(\gamma))^*}$. The translation mapping $h \mapsto \tau_h^* g$ is continuous in $L^{(\cosh -1)_*}(\gamma)$.

3. If $f \in C_0^{(\cosh -1)}(\gamma)$ then $\tau_h f \in C_0^{(\cosh -1)}(\gamma)$, $h \in \mathbb{R}^n$, and the mapping $\mathbb{R}^n \colon h \mapsto \tau_h f$ is continuous in $L^{(\cosh -1)}(\gamma)$.

- Continuity and directional derivative

   1. For each $v \in W^{1,(\cosh -1)}(\gamma)$, each unit vector h, and all $t \in \mathbb{R}$, it holds

$$v(x+th) - v(x) = t \int_0^1 \nabla v(x+sth) \cdot h \, ds .$$

   Moreover, $|t| \leq \sqrt{2}$ implies

$$\left\| v(x+th) - v(x) \right\|_{L^{(\cosh -1)}(\gamma)} \leq 2t \left\| \nabla v \right\|_{L^{(\cosh -1)}(\gamma)} ,$$

Calculus on $W^{1,(\cosh -1)}(\gamma)$ — III

especially, $\lim_{t\to 0} \|v(x+th) - v(x)\|_{L^{(\cosh -1)}(\gamma)} = 0$ uniformly in h.

2. For each $v \in W^{1,(\cosh -1)}(\gamma)$ the mapping $h \mapsto \tau_h v$ is differentiable from $\mathbb{R}^n$ to $L^{\infty-0}(M)$ with gradient ∇v at $h = 0$.
3. For each $v \in W^{1,(\cosh -1)}(\gamma)$ and each $f \in L^{(\cosh -1)_*}(\gamma)$, the mapping $h \mapsto \langle \tau_h v, f \rangle_\gamma$ is differentiable with derivative $\langle \tau_h \nabla v \cdot h, f \rangle_\gamma$.
4. If $\partial_j v \in C_0^{(\cosh -1)}(\gamma)$, $j = 1, \dots, n$, then strong differentiability in $L^{(\cosh -1)}(\gamma)$ holds.

- Calculus in $C_0^{1,(\cosh -1)}(\gamma)$

1. For each $f \in C_0^{1,(\cosh -1)}(\gamma)$ the sequence $f * \omega_n$, $n \in \mathbb{N}$, belongs to $C^\infty(\mathbb{R}^n) \cap W^{1,(\cosh -1)}(\gamma)$. Precisely, for each n and $j = 1, \dots, n$, we have the equality $\partial_j(f * \omega_n) = (\partial_j f) * \omega_n$; the sequences $f * \omega_n$, respectively $\partial_j f * \omega_n$, $j = 1, \dots, n$, converge to f, respectively $\partial_j f$, $j = 1, \dots, n$, strongly in $L^{(\cosh -1)}(\gamma)$.
2. Same statement is true if $f \in W^{1,(\cosh -1)_*}(\gamma)$.

Calculus on $W^{1,(\cosh -1)}(\gamma)$ — IV

3. Let be given $f \in C_0^{1,(\cosh -1)}(\gamma)$ and $g \in W^{1,(\cosh -1)_*}(\gamma)$. Then $fg \in W^{1,1}(\gamma)$ and $\partial_j(fg) = \partial_j fg + f\partial_j g$.
4. Let be given $F \in C^1(\mathbb{R})$ with $\|F'\|_\infty < \infty$. For each $u \in C_0^{1,(\cosh -1)}(\gamma)$, we have $F \circ u, F' \circ u\ \partial_j u \in C_0^{(\cosh -1)}(\gamma)$ and $\partial_j F \circ u = F' \circ u\ \partial_j u$, in particular $F(u) \in C_0^{1,(\cosh -1)}(\gamma)$.

- Product
  1. If $u \in \mathcal{S}_\gamma$ and $f_1, \dots, f_m \in L^{(\cosh -1)}(\gamma)$, then $f_1 \cdots f_m \mathrm{e}^{u-K(u)} \in L^\alpha(\gamma)$ for some $\alpha > 1$, hence it is in $L^{(\cosh -1)_*}(\gamma)$.
  2. If $u \in \mathcal{S}_\gamma \cap C_0^{1,(\cosh -1)}(\gamma)$ and $f \in C_0^{1,(\cosh -1)}(\gamma)$, then

$$f\mathrm{e}^{u-K(u)} \in W^{1,(\cosh -1)_*}(\gamma) \cap C(\mathbb{R}^n) ,$$

and its distributional partial derivatives are $(\partial_j f + f\partial_j u)\mathrm{e}^{u-K(u)}$

- M. R. Grasselli. Dual connections in nonparametric classical information geometry. *Ann. Inst. Statist. Math.*, 62(5):873–896, 2010
- B. Lods and G. Pistone. Information geometry formalism for the spatially homogeneous Boltzmann equation. *Entropy*, 17(6):4323–4363, 2015
- G. Pistone. Information geometry of the Gaussian space. arXiv:1803.08135, 2018