Translations in the exponential Orlicz space with Gaussian weight

Giovanni Pistone

Nov 9, 2017
Abstract

We study the continuity of space translations on non-parametric exponential families based on the exponential Orlicz space with Gaussian reference density. I.e., \( M \) is the Gaussian standard density, \( p(x) = \exp(U(x) - \kappa) \) is a density w.r.t. \( M \). Then

\[
d_h U(x) = \lim_{t \to 0} t^{-1}(U(x + th) - U(x))
\]

is the basic computation to study differentiable exponential models.

- 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
Exponential densities on the Gaussian space

• On the Gaussian probability space

$$(\mathbb{R}^n, \mathcal{B}, M \cdot \ell),$$

$M$ being the standard Gaussian density and $\ell$ the Lebesgue measure, we consider densities of the form

$$e_M(U) = \exp (U - K_M(U)) \cdot M,$$

where $U$ belongs to the exponential Orlicz space $L^{(\cosh^{-1})}(M)$, $\mathbb{E}_M[U] = 0$, and $K_M(U)$ is a finite constant.

• $L^{(\cosh^{-1})}(M)$ is the Banach space of random variables $U$ such that

$$\mathbb{E}_M [(\cosh^{-1})(\alpha U)] < \infty \quad \text{for some } \alpha > 0,$$

with

$$\|U\|_{L^{(\cosh^{-1})}(M)} \leq 1 \iff \mathbb{E}_M [(\cosh^{-1})(\alpha U)] \leq 1.$$
Duality

• Let \((\cosh - 1)_*\) denote the convex conjugate of \((\cosh - 1)\). The mixture Orlicz space \(L^{(\cosh - 1)_*}(M)\) is defined in the same way using the conjugate convex function.

• The bilinear map

\[
L^{(\cosh - 1)}(M) \times L^{(\cosh - 1)_*}(M) \ni (U, V) \mapsto \mathbb{E}_M[UV]
\]

is a separating duality such that \((L^{(\cosh - 1)}(M))' \equiv L^{(\cosh - 1)_*}(M)\), but not the other way.

• A notable inequality holds,

\[
(\cosh - 1)_*(ay) \leq \max(1, a^2)(\cosh - 1)_*(y).
\]

• Orlicz spaces are discussed in Ch. V of J. Musielak. *Orlicz spaces and modular spaces*, volume 1034 of *Lecture Notes in Mathematics*. Springer-Verlag, 1983

• G. Pistone and C. Serni. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Statist.*,, 23(5):1543–1561, October 1995

• For a critical presentation of this approach to IG, see §3.3 of N. Ay, J. Jost, H. V. Lê, and L. Schwachhöfer. *Information Geometry*. Springer, 2017
Inclusions

- We have the following continuous inclusions:

\[ L^\infty(M) \hookrightarrow L^{(\cosh-1)}(M) \hookrightarrow L^a(M) \hookrightarrow L^{(\cosh-1)_*}(M) \hookrightarrow L^1(M), \]

- and the restrictions to the ball \( \Omega_R = \{ x \in \mathbb{R}^n | \|x\| < R \} \),

\[ L^{(\cosh-1)}(M) \rightarrow L^a(\Omega_R), \quad L^{(\cosh-1)_*}(M) \rightarrow L^1(\Omega_R), \]

are continuous.

- The exponential space \( L^{(\cosh-1)}(M) \) contains all functions \( f \in C^2(\mathbb{R}^n; \mathbb{R}) \) whose Hessian is uniformly bounded in operator’s norm. In particular, it contains all polynomials with degree up to 2, hence all functions which are bounded by such a polynomial.

- The mixture space \( L^{(\cosh-1)_*}(M) \) contains all random variables \( f: \mathbb{R}^d \rightarrow \mathbb{R} \) which are bounded by a polynomial, in particular, all polynomials.
Moment function and cumulant function

- Define the moment function $Z(U) = \mathbb{E}_M [e^U]$ and the cumulant function $K(U) = \log Z(U)$.

- The non-parametric exponential model is the set of densities $e_M(U) = \exp (U - K_M(U)) \cdot M$, where $U$ has zero $M$-expectation and belongs to the interior $S_M$ of the proper domain of the moment function.

- $Z$ and $K$ are both convex

- and the common proper domain contains the open unit ball of $L^{(\cosh - 1)}(M)$.

- $Z$ and $K$ are both Fréchet differentiable on $S_M$.

See the proof of F-differentiability in the paper.
Bounded point-wise convergence

- Let \((g_n)\) be a sequence in \(L^{(\cosh^{-1})}(M)\) and assume the bound 
  \(|g_n| \leq h\) with \(h \in L^{(\cosh^{-1})}(M)\). If \((g_n)\) point-wise converges to \(f\), 
  then \(f \in L^{(\cosh^{-1})}(M)\), but convergence in norm does not hold in general.

- Example Consider \(f(x) = x^2\) and define the approximating sequence 
  \(g_n(x) = x^2(|x| \leq n)\). Then \(g_n(x) \to f(x), n \to \infty\) and we have 
  \(0 \leq g_n(x) \leq x^2\) that is, a bound \(h(x) = x^2\) which belongs to 
  \(L^{(\cosh^{-1})}(M)\). However

\[
\int (\cosh -1)(\epsilon^{-1}(f(x) - g_n(x)))M(x) \, dx \geq \\
\frac{1}{2} \int_{|x| > n} e^{\epsilon^{-1}|x|^2} M(x) \, dx - 1 = +\infty, \quad \text{if} \ \epsilon \leq 2
\]
Application of bounded point-wise convergence

Let be given $U \in SM$ and consider the exponential family

$p(t) = \exp(tU - K(tU)) \cdot M, \ t \in ]-1,1[. \ \text{Assume there is a sequence}$

$(f_n)_n$ in $C_c^\infty (\mathbb{R}^n)$ and a bound $h \in L^{(\cosh^{-1})} (M)$ such that $f_n \to u$

point-wise and $|f_n|, |u| \leq h$. As $SM$ is open and contains 0, we have

$\alpha h \in SM$ for some $0 < \alpha < 1$. For each $t \in ]-\alpha, \alpha[$, $\exp(tf_n) \to \exp(tu)$

point-wise and $\exp(tf) \leq \exp(\alpha h)$ with $E_M[\exp(\alpha h)] < \infty$. It follows

that $K(tf_n) \to K(tu)$, so that we have the point-wise convergence of the

density $p_n(t) = \exp(tf_n - K(tf_n)) \cdot M$ to the density $p(t)$. By Scheffé’s

lemma, the convergence holds in $L^1(M)$. In particular, for each

$\phi \in C_c^\infty (\mathbb{R}^n)$, we have the convergence

$$
\int \partial_i \phi(x)p_n(x; t) \ dx \to \int \partial_i \phi(x)p(x; t) \ dx, \quad n \to \infty.
$$

for all $t$ small. By computing the derivatives, we have

$$
\int \partial_i \phi(x)p_n(x; t) \ dx = - \int \phi(x) \partial_i \left( e^{tf_n(x)} - K(tf_n) M(x) \right) \ dx =
$$

$$
\int \phi(x) (x_i - t\partial_i f_n(x)) p_n(x; t) \ dx,
$$

that is,

$$(X_i - t\partial_i f_n) p_n(t) \to -\partial_i p(t).$$
Orlicz class

The exponential class, \( C_c^{(\cosh -1)}(M) \), is the closure of \( C_c^\infty(\mathbb{R}^n) \) in the space \( L^{(\cosh -1)}(M) \).

Assume \( f \in L^{(\cosh -1)}(M) \). The following conditions are equivalent:

- The real function \( \rho \mapsto \int (\cosh -1)(\rho f(x)) M(x)dx \) is finite for all \( \rho > 0 \).
- \( f \) is the limit in \( L^{(\cosh -1)}(M) \)-norm of a sequence of bounded functions.
- \( f \in C_c^{(\cosh -1)}(M) \).
Translation by a vector

\[ \tau_h f(x) = f(x - h), \quad h \in \mathbb{R}^n \]

- For each \( h \in \mathbb{R}^n \), the mapping \( f \mapsto \tau_h f \) is linear from \( L^{(\cosh^{-1})}(M) \) to itself and \( \| \tau_h f \|_{L^{(\cosh^{-1})}(M)} \leq 2 \| f \|_{L^{(\cosh^{-1})}(M)} \) if \( |h| \leq \sqrt{\log 2} \).

- The transpose of \( \tau_h \) is defined on \( L^{(\cosh^{-1})^*}(M) \) by
  \[ \langle \tau_h f, g \rangle_M = \langle f, \tau_h^* g \rangle_M, \quad f \in L^{(\cosh^{-1})}(M), \]
  and is given by
  \[ \tau_h^* g(x) = e^{-h \cdot x + |h|^2/2} \tau_{-h} g(x). \]

- For the dual norm, the bound \( \| \tau_h^* g \|_{L^{(\cosh^{-1})^*}(M)} \leq 2 \| g \|_{L^{(\cosh^{-1})^*}(M)} \) holds if \( |h| \leq \sqrt{\log 2} \).

- If \( f \in C_c^{(\cosh^{-1})}(M) \) then \( \tau_h f \in C_c^{(\cosh^{-1})}(M) \), \( h \in \mathbb{R}^n \) and the mapping \( \mathbb{R}^n: h \mapsto \tau_h f \) is continuous in \( L^{(\cosh^{-1})}(M) \).

See proofs in the paper.
Translation by a probability

- The translation by a probability measure $\mu$ is
  $$\tau_\mu f(x) = \int f(x - y) \mu(dy).$$
- We denote by $\mathcal{P}_e$ the set of probability measures $\mu$ such that
  $$\int e^{\frac{1}{2} |h|^2} \mu(dh) < \infty$$
  e.g., $\mu$ could have a bounded support.

Let $\mu \in \mathcal{P}_e$.
- The mapping $f \mapsto \tau_\mu f$ is linear and bounded from $L^{(\cosh^{-1})}(M)$ to itself. If, moreover, $\int e^{\frac{1}{2} |h|^2} \mu(dh) \leq \sqrt{2}$, then its norm is bounded by 2.
- If $f \in C^{(\cosh^{-1})}_c(M)$ then $\tau_\mu f \in C^{(\cosh^{-1})}_c(M)$. The mapping $\mathcal{P}_e : \mu \mapsto \tau_\mu f$ is continuous at $\delta_0$ from the weak convergence to the $L^{(\cosh^{-1})}(M)$ norm.
- Mollifiers can be constructed.

See proofs in the paper.
Orlicz-Sobolev-Gauss spaces

- B. Lods and G. Pistone. Information geometry formalism for the spatially homogeneous Boltzmann equation. *Entropy*, 17(6):4323–4363, 2015
- Talk to IGAIA 2016 Conference. Submitted extended paper
- D. Brigo and G. Pistone. Projection based dimensionality reduction for measure valued evolution equations in statistical manifolds. In F. Nielsen, F. Critchley, and C. Dodson, editors, *Computational Information Geometry. For Image and Signal Processing*, Signals and Communication Technology, pages 217–265. Springer, 2017
- D. Brigo and G. Pistone. Maximum likelihood eigenfunctions of the Fokker Planck equation and Hellinger projection. arXiv:1603.04348
Poincaré inequality I

• The Gaussian Poincaré inequality in the 1-dimensional case is

\[
\int \left( f(x) - \int f(y) M(y) \, dy \right)^2 M(x) \, dx \leq \int |f'(x)|^2 M(x) \, dx ,
\]

for all \( f \in C^1_p(\mathbb{R}^n) \) i.e. \( C^1 \) and polynomially bounded.

• There exists \( \lambda > 0 \) such that for all \( f \in C^1_p(\mathbb{R}^n) \) it holds

\[
\| f - \int f(y) M(y) \, dy \|_{L^{(\cosh -1)^*}(M)} \leq \lambda^{-1} \| \nabla f \|_{L^{(\cosh -1)^*}(M)} .
\]

• If \( f \) is a density, then \( \int f(y) M(y) \, dy = 1 \) and the inequality becomes

\[
\| f - 1 \|_{L^{(\cosh -1)^*}(M)} \leq \lambda^{-1} \| \nabla f \|_{L^{(\cosh -1)^*}(M)} .
\]

i.e. a bound on the deviation from uniformity.
Poincaré inequality II

- If \( f = U_1 - U_2 \) is the difference of two \( M \)-centered random variables, for example the scores of a statistical model at two different parameters values, then

\[
\| U_1 - U_2 \|_{L^{(\cosh^{-1})^*}(M)} \leq \lambda^{-1} \| \nabla (U_1 - U_2) \|_{L^{(\cosh^{-1})^*}(M)}
\]

- In the \( L^{(\cosh^{-1})^*}(M) \)-Poincaré inequality for \((U_1 - U_2)\) difference of scores, the RHS is dominated by \( \| \nabla (U_1 - U_2) \|_{L^2(M)} \), which is a divergence of the Hyvärinen type.
Differentiable densities

Definition

The Orlicz-Sobolev (O-S) spaces with weight $M$ are

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

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

where $\partial_j$ is the derivative in the sense of distributions. The spaces $W^{1,(\cosh^{-1})}(M)$ and $W^{1,(\cosh^{-1})_*}(M)$ are both Banach spaces for the graph norms.

- $$\langle \partial_j f, \phi M \rangle = - \langle f, \partial_j \phi M - X_j \phi M \rangle = \langle f, (X_j - \partial_j)\phi M \rangle$$
- $$\langle \partial_j f, \phi \rangle_M = \langle f, \delta_j \phi \rangle_M \quad \delta_j \phi = (X_j - \partial_j)\phi$$

- Malliavin Calculus is the proper set-up
Inclusions

**Theorem**

Let $R > 0$ and let $\Omega_R$ denote the open sphere of radius $R$.

1. We have the embeddings

   \[ W^{1,((\cosh - 1)}(\mathbb{R}^n) \subset W^{1,((\cosh - 1)}(M) \subset W^{1,((\cosh - 1)}(\Omega_R) \subset W^{1,1}(\Omega_R), \quad p \geq 1. \]

2. We have the embeddings

   \[ W^{1,p}(\mathbb{R}^n) \subset W^{1,((\cosh - 1)}(\mathbb{R}^n) \subset W^{1,((\cosh - 1)}(M) \subset W^{1,1}(\Omega_R) \subset W^{1,1}(\Omega_R), \quad p > 1. \]

3. Each $u \in W^{1,((\cosh - 1)}(M)$ is Hölder-continuous of all orders on each $\overline{\Omega}_R$. 
Directional derivative

Theorem

• For each $f \in W^{1,(\cosh^{-1})}(M)$, each unit vector $h \in S^n$, and all $t \in \mathbb{R}$, it holds

$$f(x + th) - f(x) = t \int_0^1 \sum_{j=1}^n \partial_j f(x + st)h_j \, ds.$$  

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

$$\|f(x + th) - f(x)\|_{L^{(\cosh^{-1})}(M)} \leq 2t \|\nabla f\|_{L^{(\cosh^{-1})}(M)},$$

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

• For each $f \in W^{1,(\cosh^{-1})}(M)$ and each $g \in L^{(\cosh^{-1})^*}(M)$, the mapping $h \mapsto \langle \tau_h f, g \rangle_M$ is differentiable. Conversely, if $f \in L^{(\cosh^{-1})}(M)$ and $h \mapsto \tau_h f$ is weakly differentiable, then $f \in W^{1,(\cosh^{-1})}(M)$

• If $\partial_j f \in C_c^{(\cosh^{-1})}(M)$, $j = 1, \ldots, n$, then strong differentiability in $L^{(\cosh^{-1})}(M)$ holds.
Exponential family modeled on $C_c^{1,(\cosh-1)}(M)$

- Restrict the exponential family $E(M)$ to $C_c^{1,(\cosh-1)}(M)$,

$$E_1(M) = \left\{ e^{U-K_M(U)} \cdot M \mid U \in C_c^{1,(\cosh-1)}(M) \cap S_M \right\}$$

- Because of $C_c^{1,(\cosh-1)}(M) \hookrightarrow L^{\cosh-1}(M)$ the domain $C_c^{1,(\cosh-1)}(M) \cap S_M$ is open and the cumulant functional $K_M : C_c^{1,(\cosh-1)}(M) \cap S_M \to \mathbb{R}$ remains convex and differentiable.

- Every feature of the exponential manifold carries over to this case. Define $B_1(p) = B_p \cap C_c^{1,(\cosh-1)}(M)$ to be models for the tangent spaces of $E_1(M)$. The e-transport acts on these spaces

$$eU_f^g : B_1(f) \ni U \mapsto U - \mathbb{E}_g[U] \in B_1(g),$$

so that we can define the statistical bundle to be

$$T E_1(M) = \{(g, V) \mid g \in E_1(M), V \in B_1(g)\}$$

and take as charts the restrictions of the charts defined on $T E(M)$. 
For each \( f, g \in E_1(M) \) the Hyvärinen divergence is

\[
DH (g|f) = \mathbb{E}_g \left[ |\nabla \log f - \nabla \log g|^2 \right].
\]

The expression in the chart centered at \( M \) is

\[
DH_M (v|u) = DH (e_M(v)|e_M(u)) = \mathbb{E}_M \left[ |\nabla u - \nabla v|^2 e^{v-K_M(v)} \right],
\]

where \( f = e_M(u), g = e_M(v) \).

It is possible to rigorously compute the (natural) gradient of the Hyvärinen divergence.
Application to Elliptic operator

- Elliptic operator as section of the tangent bundle is

\[ A_p(x) = p(x)^{-1} \sum_{i,j=1}^{d} \frac{\partial}{\partial x_i} \left( a_{ij}(x) \frac{\partial}{\partial x_j} p(x) \right), \quad x \in \mathbb{R}^d. \]

- The expression in the statistical bundle is

\[ U \mapsto \hat{A}_M(U) = e^{U - K_M(U)} A(e^{U - K_M(U)} \cdot M) = \frac{e^{U - K_M(U)}}{e^{U - K_M(U)} \cdot M} A(e^{U - K_M(U)} \cdot M) = M^{-1} \mathcal{L}^* (e^{U - K_M(U)} \cdot M) \]

- Computation gives

\[ M^{-1} \mathcal{L}^* (e^{U - K_M(U)} \cdot M) = \]

\[ e^{U - K_M(U)} \sum_{i,j=1}^{d} \frac{\partial}{\partial x_i} \left[ a_{ij}(x) \left( \frac{\partial}{\partial x_j} U(x) - x_j \right) \right] p(x) + \]

\[ e^{U - K_M(U)} \sum_{i,j=1}^{d} a_{ij}(x) \left( \frac{\partial}{\partial x_i} U(x) - x_i \right) \left( \frac{\partial}{\partial x_j} U(x) - x_j \right) p(x). \]