A Novel Approach to Canonical Divergences within Information Geometry

Nihat Ay[1,2,3] **Shun-ichi Amari**[4]

[1] Max Planck Institute for Mathematics in the Sciences

[2] Faculty of Mathematics and Computer Science, University of Leipzig

[3] Santa Fe Institute

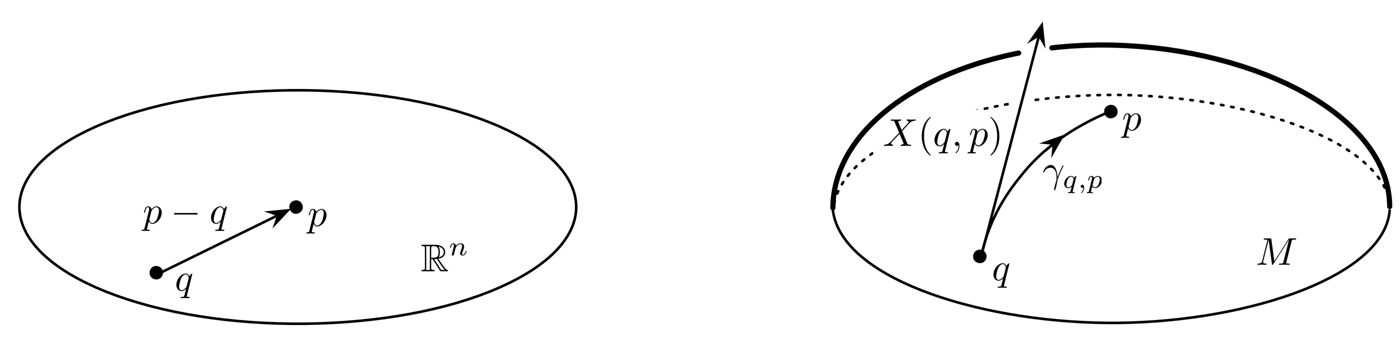
[4] RIKEN Brain Science Institute

Abstract

We propose a (locally defined) canonical divergence, associated with a Riemannian metric g and an affine connection ∇ on M. Our definition is consistent with, and naturally extends, corresponding approaches to divergence functions within information geometry.

Divergence with the Inverse Exponential Map

Main geometric idea:



(A) (B)

(A)	(B)
$q \mapsto p - q$	$q \mapsto X(q,p) := \exp_q^{-1}(p)$
$p - q = -\mathrm{grad}_q D_p$	$X(q,p) = -\mathrm{grad}_q D_p$
$q \mapsto D_p(q) = \frac{1}{2}\Vert p - q\Vert^2$	$D_p(q) = ?$

Integral representation of D_p:

$$\begin{aligned}
&\int_0^1 \langle X(\gamma(t), p), \dot\gamma(t)\rangle \, dt \\
&= -\int_0^1 \langle \mathrm{grad}_{\gamma(t)} D_p, \dot\gamma(t)\rangle \, dt = -\int_0^1 (d_{\gamma(t)} D_p)(\dot\gamma(t)) \, dt \\
&= -\int_0^1 \frac{d\, D_p \circ \gamma}{dt}(t) \, dt = D_p(\gamma(0)) - D_p(\gamma(1)) \\
&= D_p(q) - D_p(p) = D_p(q) =: D(p \,\Vert\, q)
\end{aligned}$$

Examples: KL- and α-Divergences

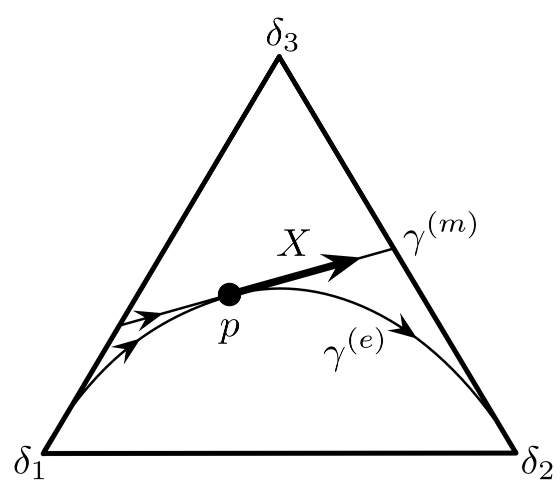
Fisher metric:

$$g_p(X, Y) = \sum_{i=1}^n \frac{1}{p_i} X_i Y_i$$

m- and e-connections:

$$\exp_p^{(m)}(X) \;=\; \sum_{i=1}^n (p_i + X_i)\,\delta_i$$

$$\exp_p^{(e)}(X) \;=\; \sum_{i=1}^n \frac{p_i \exp\left(\frac{X_i}{p_i}\right)}{\sum_{j=1}^n p_j \exp\left(\frac{X_j}{p_j}\right)}\,\delta_i$$



KL-divergence:

$$D^{(m)}(p \,\Vert\, q) \;=\; \sum_{i=1}^n p_i \log\left(\frac{p_i}{q_i}\right) \;=\; D^{(e)}(q \,\Vert\, p)$$

α-divergence for positive measures:

$$D^{(\alpha)}(p \,\Vert\, q) = \begin{cases} \frac{4}{1-\alpha^2} \sum_{i=1}^n \left(\frac{1-\alpha}{2} p_i + \frac{1+\alpha}{2} q_i - p_i^{\frac{1-\alpha}{2}} q_i^{\frac{1+\alpha}{2}} \right), & \text{if } \alpha \notin \{-1, 1\} \\ \sum_{i=1}^n \left(q_i - p_i + p_i \ln \frac{p_i}{q_i} \right), & \text{if } \alpha = -1 \\ \sum_{i=1}^n \left(p_i - q_i + q_i \ln \frac{q_i}{p_i} \right), & \text{if } \alpha = 1 \end{cases}$$

Dual Structures Induced by Divergences

Riemannian metric g:

$$g_\xi(X, Y) \;:=\; -\partial_{X_{\xi_1}} \partial_{Y_{\xi_2}} D(\xi_1 \,\Vert\, \xi_2)_{|\xi_1 = \xi,\, \xi_2 = \xi} \tag{1}$$

Affine connection ∇:

$$g_\xi(\nabla_X Y, Z) \;:=\; -\partial_{X_{\xi_1}} \partial_{Y_{\xi_1}} \partial_{Z_{\xi_2}} D(\xi_1 \,\Vert\, \xi_2)_{|\xi_1 = \xi,\, \xi_2 = \xi} \tag{2}$$

Affine connection ∇^*:

$$g_\xi(\nabla^*_X Y, Z) \;:=\; -\partial_{Z_{\xi_1}} \partial_{X_{\xi_2}} \partial_{Y_{\xi_2}} D(\xi_1 \,\Vert\, \xi_2)_{|\xi_1 = \xi,\, \xi_2 = \xi} \tag{3}$$

Theorem 1 (Eguchi, 1983)**.** *The triple (g, ∇, ∇^*) forms a torsion-free dualistic structure on M.*

Theorem 2 (Matumoto, 1993)**.** *Any torsion-free dualistic structure (g, ∇, ∇^*) on M is induced by a divergence $M \times M \to \mathbb{R}$.*

General Definition and Main Results

Definition 3. Let g be a Riemannian metric ∇ be an affine connection on M. Then, we locally define the *canonical divergence* by

$$D^{(\nabla)}(p \,\Vert\, q) \;:=\; \int_0^1 t\, \Vert\dot\gamma_{p,q}(t)\Vert^2 \, dt\,,$$

where $\gamma_{p,q} : [0,1] \to M$, is the ∇-geodesic connecting p with q, that is $\gamma_{p,q}(0) = p$ and $\gamma_{p,q}(1) = q$.

Theorem 4 (Consistency results, Ay & Amari, 2015)**.** *Let (g, ∇, ∇^*) be a torsion-free dualistic structure on M. Then:*

1. *The divergence $D^{(\nabla)}$ is consistent with (g, ∇, ∇^*) in the sense that the equations (1), (2), and (3) hold for $D = D^{(\nabla)}$.*
2. *In the special case of dual flatness, we recover the well-known canonical divergence (defined in [2])*

   $$D(p \,\Vert\, q) \;:=\; \psi(\vartheta(p)) + \varphi(\eta(q)) - \vartheta^i(p)\,\eta_i(q)\,,$$

   where ϑ and η are two dual affine coordinate systems with respect to ∇ and ∇^*, respectively, and $\eta_i = \partial_i \psi(\vartheta)$, $\varphi(\eta) = \psi(\vartheta) - \vartheta^i\,\eta_i$.
3. *In the self-dual case, that is $\nabla = \nabla^*$, we have*

   $$D^{(\nabla)}(p \,\Vert\, q) \;=\; \frac{1}{2} d(p, q)^2\,,$$

   where d denotes the Riemannian distance.

References

[1] N. Ay and S. Amari. *A Novel Approach to Canonical Divergences within Information Geometry*. Entropy **17** (2015), 8111–8129.

[2] S. Amari and H. Nagaoka. *Methods of Information Geometry*. American Mathematical Society: Providence, RI, USA; Oxford University Press: Oxford, UK, 2000.

[3] S. Egutchi. *Second order efficiency of minimum contrast estimators in a curved exponential family*. Ann. Stat. **11** (1983), 793–803.

[4] M. Henmi and R. Kobayashi. *Hooke's law in statistical manifolds and divergences*. Nagoya Math. J. **159** (2000), 1–24.

[5] T. Matumoto. *Any statistical manifold has a contrast function – On the C^3-functions taking the minimum at the diagonal of the product manifold*. Hiroshima Math. J. **23** (1993), 327–332.

Information Geometry and Its Applications IV, June 12–17, 2016, Liblice, Czech Republic.