Workshop *Stochastic Models And Related Topics* 2016
Jan 21–22, 2016
Dipartimento di Matematica Università di Salerno

Gradient flow to the optimal transport plan

Giovanni Pistone
giovanni.pistone@arloalberto.org
www.giannidiorestino.it

DE CASTRO STATISTICS Collegio Carlo Alberto

Jan 21, 2016

Abstract: Acknowledgments

This talk is based on work in progress joint with Luigi Malagò (Shinshu University, Japan) and Luigi Montrucchio (Università di Torino and Collegio Carlo Alberto). A research announcement has been done at an ICMS workshop, Edinburgh Sep 21–25, 2015.

Abstract: Statistical bundle

The *statistical bundle* is an extension of the *exponential manifold* introduced by G. Pistone and C. Sempi (1995). In particular, we consider a finite state space Ω and its *positive* probability functions, i.e. the points in the interior of the probability simplex, $\gamma \in \Delta^\circ(\Omega)$. For each such a γ, we consider the vector space B_γ of random variables V with $E_\gamma V = 0$. Each one dimensional regular statistical model $\theta \mapsto \gamma(\theta)$ has *Fisher score* $D\gamma(\theta) = d \log \gamma(\theta)/d\theta \in B_{\gamma(\theta)}$. It follows that B_γ is an expression of the tangent space at γ. The *statistical bundle* is the set of couples couples (γ, V), $V \in B_\gamma$, see [2]. Given a regular function $F \colon \Delta^\circ(\Omega)$, its *statistical gradient* is a section of the statistical bundle, $\gamma \mapsto F'(\gamma) \in B_\gamma$, such that $dF(\gamma(\theta))/d\theta = E_{\gamma(\theta)}(F'(\gamma(\theta))D\gamma(\theta))$. It corresponds to S-i. Amari's *natural gradient*. The *gradient flow* of F is the solution of the equation $D\gamma(\theta) = -F'(\gamma(\theta)$. The gradient flow equation is expected to go towards a minimum point of F. The most interesting cases arise when the curve $\gamma(\cdot)$ is restricted to belong to some smooth sub-model, either exponential or mixture, see eg [1].

1. Luigi Malagò and Giovanni Pistone. Natural gradient flow in the mixture geometry of a discrete exponential family.

   Entropy, 17(6):4215–4254, 2015

Abstract: Gradient flow to optimal transport

Assume $\Omega = \Omega_1 \times \Omega_2$ and denote by $\Gamma^\circ(\mu_1, \mu_2)$ the set of positive probability functions with given marginals μ_1, μ_2. Given a cost function $c \colon \Omega$, we consider minimising $F(\gamma) = E_\gamma(c)$ under the condition $\gamma \in \Gamma^\circ(\mu_1, \mu_2)$. The minimum value of this problem for $c(x, y) = |x - y|$ is the Gini's dissimilarity index, while for $c(x, y) = |x - y|^2$ its square root is the Wasserstein 2-distance. The problem does not have a minimum on positive probability function, but it does have a solution γ^* if we allow zero probabilities. In this last case it is an instance of a linear programming problem whose dual has been identified and studied by L. Kantorovich. There is a considerable body research we cannot refer to here. We want to discuss an analytic approach, consisting in computing the statistical gradient of the minimum expected cost problem and looking for a a gradient flow trajectory going from the product of the marginals $\mu_1 \otimes \mu_2$ toward the solution γ^*. To this aim we compute the sub-bundle of scores when a curve is restricted to have assigned marginals and show that such scores at each point are random variables of the interaction type, i.e. they are γ-orthogonal to constants and to simple effects.

2. Giovanni Pistone. Examples of the application of nonparametric information geometry to statistical physics.

   Entropy, 15(10):4042–4065, 2013.

   ISSN 1099-4300.

   doi: 10.3390/e15104042

Rao

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}_{p(t)}\left[U\right] &= \frac{d}{dt}\sum_{x\in\Omega} U(x)p(x;t) \\
&= \sum_{x\in\Omega} U(x)\frac{d}{dt}p(x;t) \\
&= \sum_{x\in\Omega} U(x)\frac{\frac{d}{dt}p(x;t)}{p(x;t)}\; p(x;t) \\
&= \sum_{x\in\Omega} U(x)\frac{d}{dt}\log\left(p(x;t)\right)p(x;t) \\
&= \sum_{x\in\Omega}\left(U(x)-\mathbb{E}_{p(t)}\left[U\right]\right)\frac{d}{dt}\log\left(p(x;t)\right)p(x;t) \\
&= \mathbb{E}_{p(t)}\left[\left(U-\mathbb{E}_{p(t)}\left[U\right]\right)\frac{d}{dt}\log\left(p(t)\right)\right] \\
&= \left\langle\left(U-\mathbb{E}_{p(t)}\left[U\right]\right), \frac{d}{dt}\log\left(p(t)\right)\right\rangle_{p(t)} \quad .
\end{aligned}$$

C. Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.
ISSN 0008-0659

Amari

$$
\begin{aligned}
\frac{d}{dt} f(p(t)) &= \sum_{x \in \Omega} \frac{\partial}{\partial p(x)} f(p(x;t) \colon x \in \Omega) \frac{d}{dt} p(x;t) \\
&= \sum_{x \in \Omega} \frac{\partial}{\partial p(x)} f(p(x;t) \colon x \in \Omega) \frac{\frac{d}{dt} p(x;t)}{p(x;t)} \; p(x;t) \\
&= \left\langle \nabla f(p(t)), \frac{d}{dt} \log\left(p(x;t)\right) \right\rangle_{p(t)} \\
&= \left\langle \nabla f(p(t)) - \mathbb{E}_{p(t)}\left[\nabla f(p(t))\right], \frac{d}{dt} \log\left(p(x;t)\right) \right\rangle_{p(t)} \quad ,
\end{aligned}
$$

$$\boxed{\operatorname{grad} f(p) := \nabla f(p) - \mathbb{E}_p\left[\nabla f(p)\right]}$$

Shun-Ichi Amari. Natural gradient works efficiently in learning.
Neural Computation, 10(2):251–276, feb 1998.
ISSN 0899-7667.
doi: 10.1162/089976698300017746

Statistical bundle

Definition

1.
$$B_p = \left\{ U \colon \Omega \to \mathbb{R} \middle| \mathbb{E}_p\left[U\right] = \sum_{x \in \Omega} U(x)\ p(x) = 0 \right\} \quad p \in \Delta^\circ(\Omega)$$

2.
$$\left\langle U, V \right\rangle_p = \mathbb{E}_p\left[UV\right] = \sum_{x \in \Omega} U(x)V(x)\ p(x)\ .$$

3.
$$S\Delta^\circ(\Omega) = \left\{(p, U) | p \in \Delta^\circ(\Omega), U \in B_p\right\}\ .$$

4. It is an open subset of an algebraic variety of $\mathbb{R}^\Omega \times \mathbb{R}^\Omega$

5. A *vector field* F of the statistical bundle is a *section* of the bundle i.e.,
$$F \colon \Delta^\circ(\Omega) \ni p \mapsto (p, F(p)) \in T\Delta^\circ(\Omega)$$

Affine statistical bundle

Definition

1. $$B_\eta = \left\{ U \colon \Omega \to \mathbb{R} \middle| \mathbb{E}_\eta \left[U \right] = \sum_{x \in \Omega} U(x)\ \eta(x) = 0 \right\} \quad \eta \in A_1(\Omega)$$

2. $$\left\langle U, V \right\rangle_\eta = \mathbb{E}_\eta \left[UV \right] = \sum_{\{x \in \Omega | \eta(x) \neq 0\}} U(x) V(x)\ \eta(x)\ .$$

3. The *statistical bundle* of the affine space $A_1(\Omega)$ is the linear bundle on $A_1(\Omega)$

$$SA_1(\Omega) = \left\{ (\eta, U) | \eta \in A_1(\Omega), U \in B_\eta \right\}\ .$$

4. It is a manifold isomorphic to the open subset of the Grassmanian manifold $\mathrm{Grass}(\mathbb{R}^\Omega, \#\Omega - 1)$ of sub-spaces B that do not contain constant vectors.

Why the *Statistical Bundle*?

- The notion of statistical bundle appears as a natural set up for Information Geometry IG, where the notions of score and statistical gradient do not require any parameterization.
- The setup based on the full simplex $\Delta(\Omega)$ is of interest in applications to data analysis. Methods based on the simplex lead naturally to the treatment of the infinite sample space case in cases where no natural parametric model is available.
- There are special affine atlases (frames) such that the tangent space identifies with the statistical bundle. This is a version of the Amari's dual affine connections.
- The construction extends to the affine space generated by the simplex, see the paper [1].
- In the statistical bundle there is a natural treatment of differential equations e.g. gradient flow.

[1] Nihat Ay, Jürgen Jost, Hông Vân Lê, and Lorenz Schwachhöfer. Parametrized measure models. arXiv:1510.07305 [math.DG], Oct 2015

Regular curve

Theorem

1. *Let $I \ni t \mapsto p(t)$ be a C^1 curve in $\Delta^\circ(\Omega)$.*

$$\frac{d}{dt}\mathbb{E}_{p(t)}\left[f\right] = \left\langle f - \mathbb{E}_{p(t)}\left[f\right], Dp(t)\right\rangle_{p(t)}, \quad Dp(t) = \frac{d}{dt}\log\left(p(t)\right)$$

2. *Let $I \ni t \mapsto \eta(t)$ be a C^1 curve in $A_1(\Omega)$ such that $\eta(t) \in \Delta(\Omega)$ for all t. For all $x \in \Omega$, $\eta(x;t) = 0$ implies $\frac{d}{dt}\eta(x;t) = 0$.*

$$\frac{d}{dt}\mathbb{E}_{\eta(t)}\left[f\right] = \left\langle f - \mathbb{E}_{\eta(t)}\left[f\right], D\eta(t)\right\rangle_{\eta(t)}$$

$$D\eta(x;t) = \frac{d}{dt}\log\left|\eta(x;t)\right| \quad \textit{if } \eta(x;t) \neq 0, \textit{ otherwise 0.}$$

3. *Let $I \ni t \mapsto \eta(t)$ be a C^1 curve in $A_1(\Omega)$ and assume that $\eta(x;t) = 0$ implies $\frac{d}{dt}\eta(x;t) = 0$. Hence, for each $f \colon \Delta(\Omega) \to \mathbb{R}$,*

$$\frac{d}{dt}\mathbb{E}_{\eta(t)}\left[f\right] = \left\langle f - \mathbb{E}_{\eta(t)}\left[f\right], D\eta(t)\right\rangle_{\eta(t)}$$

Statistical gradient

Definition

1. Given a function $f\colon \Delta^\circ(\Omega) \to \mathbb{R}$, its *statistical gradient* is a vector field $\Delta^\circ(\Omega) \ni p \mapsto (p, \operatorname{grad} F(p)) \in S\Delta^\circ(\Omega)$ such that for each regular curve $I \ni t \mapsto p(t)$ it holds

$$\frac{d}{dt} f(p(t)) = \left\langle \operatorname{grad} f(p(t)), Dp(t) \right\rangle_{p(t)}, \quad t \in I \ .$$

2. Given a function $f\colon A_1(\Omega) \to \mathbb{R}$, its *statistical gradient* is a vector field $A_1(\Omega) \ni \eta \mapsto (\eta, \operatorname{grad} f(\eta)) \in TA_1(\Omega)$ such that for each curve $t \mapsto \eta(t)$ with a score Dp, it holds

$$\frac{d}{dt} f(\eta(t)) = \left\langle \operatorname{grad} f(\eta(t)), D\eta(t) \right\rangle_{\eta(t)}$$

Computing grad

1. If f is a C^1 function on an open subset of $\mathbb{R}^\Omega$ containing $\Delta^\circ(\Omega)$, by writing $\nabla f(p)\colon \Omega \ni x \mapsto \frac{\partial}{\partial p(x)} f(p)$, we have the following relation between the statistical gradient and the ordinary gradient:

$$\operatorname{grad} f(p) = \nabla f(p) - \mathbb{E}_p\left[\nabla f(p)\right] \ .$$

2. If f is a C^1 function on an open subset of $\mathbb{R}^\Omega$ containing $A_1(\Omega)$, we have:

$$\operatorname{grad} f(\eta) = \nabla f(\eta) - \mathbb{E}_\eta\left[\nabla f(\eta)\right] \ .$$

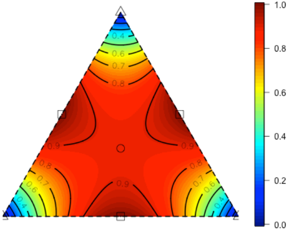
Differential equations

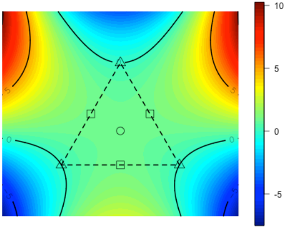
Definition (Flow)

1. Given a vector field $F\colon \Delta^\circ(\Omega)$ or $F\colon A_1(\Omega)$, the *trajectories along the vector field* are the solution of the (statistical) *differential equation*
$$\frac{D}{dt}p(t) = F(p(t)) \ .$$
2. A *flow* of the vector field F is a mapping $S\colon \Delta^\circ(\Omega) \times \mathbb{R}_{>0} \ni (p,t) \mapsto S(p,t) \in \Delta^\circ(\Omega)$, respectively $S\colon A_1(\Omega) \times \mathbb{R}_{>0} \ni (p,t) \mapsto S(p,t) \in A_1(\Omega)$, such that $S(p,0) = p$ and $t \mapsto S(p,t)$ is a trajectory along F.
3. Given $f\colon \Delta^\circ(\Omega) \to \mathbb{R}$, or $f\colon A_1(\Omega) \to \mathbb{R}$, with statistical gradient $p \mapsto (p, \operatorname{grad} f(p)) \in S\Delta^\circ(\Omega)$, respectively $\eta \mapsto (\eta, \operatorname{grad} f(p)) \in SA_1(\Omega)$, a solution of the *statistical gradient flow equation*, starting at $p_0 \in \Delta^\circ(\Omega)$, respectively $\eta_0 \in A_1(\Omega)$, at time t_0, is a trajectory of the field $-\operatorname{grad} f$ starting at p_0, respectively η_0.

Polarization measure

$$\mathrm{POL}\colon \Delta_n \ni p \mapsto 1 - 4\sum_{x=0}^{n}\left(\frac{1}{2} - p(x)\right)^2 p(x) = 4\sum_{x=0}^{n} p(x)^2(1-p(x)) \ .$$

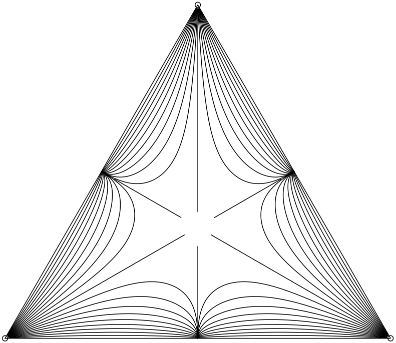


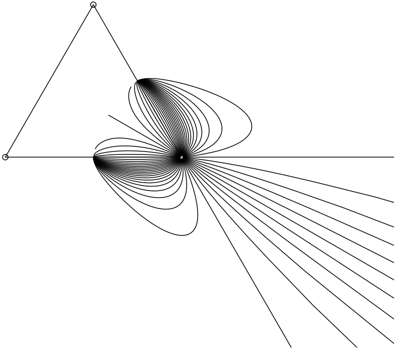


- Marta Reynal-Querol. Ethnicity, political systems and civil war. *Journal of Conflict Resolution*, 46(1):29–54, February 2002

Polarization gradient flow

$$\dot{p}(x;t) = p(x;t)\left(8p(x;t) - 12p(x;t)^2 - 8\sum_{y\in\Omega} p(y;t)^2 + 12\sum_{y\in\Omega} p(y;t)^3\right)$$





- Giovanni Pistone and Maria Piera Rogantin. The gradient flow of the polarization measure. with an appendix.

  arXiv:1502.06718, 2015

ANOVA

$\Omega = \Omega_1 \times \Omega_2$, $X \colon \Omega \to \Omega_1$, $Y \colon \Omega \to \Omega_2$, $\gamma \in \Delta^\circ(\Omega)$, $X_\# \gamma = \gamma_1$, $Y_\# \gamma = \gamma_2$

Definition (ANOVA)

$$\begin{aligned}
H_0(\gamma) &\sim \mathbb{R} && \gamma\text{-mean} \\
H_1(\gamma) &= \left\{ f \circ X \middle| f \in L^1_0(\gamma_1) \right\} && \gamma\text{-marginal effect} \\
H_2(\gamma) &= \left\{ f \circ Y \middle| f \in L^1_0(\gamma_2) \right\} && \gamma\text{-marginal effects} \\
H_{12}(\gamma) &= \left(H_0(\gamma) + H_1(\gamma) + H_2(\gamma) \right)^\perp && \gamma\text{-interactions}
\end{aligned}$$

Theorem (ANOVA)

$$L^2(\Omega) = H_0(\gamma) \oplus (H_1(\gamma) \oplus H_2(\gamma)) \oplus H_{12}(\gamma) \ ,$$

whith $f = f_0 + f_1 + f_2 + f_{1,2}$ *if, and only if,* $f_0 = \mathbb{E}_\gamma [f]$ *and*

$$\begin{aligned}
\mathbb{E}_0 [\gamma(f - f_0) | X] f_1 + \mathbb{E}_0 [\gamma(f - f_0) f_2 | X] &= 0 \ , \\
\mathbb{E}_0 [\gamma(f - f_0) f_1 | Y] + \mathbb{E}_0 [\gamma(f - f_0) | Y] f_2 &= 0 \ ,
\end{aligned}$$

ANOVA computation

$$\begin{cases} \sum_{y\in\Omega_2} \gamma(x,y)\bar{f}(x,y) & = \gamma_1(x)f_1(x) + \sum_{y\in\Omega_2} \gamma(x,y)f_2(y), \quad x\in\Omega_1 \\ \sum_{x\in\Omega_1} \gamma(x,y)\bar{f}(x,y) & = \sum_{x\in\Omega_1} \gamma(x,y)f_1(x) + \gamma_2(y)f_2(y), \quad y\in\Omega_2 \end{cases}$$

$$\begin{cases} \sum_{y\in\Omega_2} \gamma_{2|1}(y|x)\bar{f}(x,y) & = f_1(x) + \sum_{y\in\Omega_2} \gamma_{2|1}(y|x)f_2(y), \quad x\in\Omega_1 \\ \sum_{x\in\Omega_1} \gamma_{1|2}(x|y)\bar{f}(x,y) & = \sum_{x\in\Omega_1} \gamma_{1|2}(x|y)f_1(x) + f_2(y), \quad y\in\Omega_2 \end{cases}$$

$$\begin{bmatrix} I_{n_1} & \Gamma_{2|1} \\ \Gamma_{1|2}^T & I_{n_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}_{2|1} \\ \boldsymbol{f}_{1|2} \end{bmatrix}$$

$$\begin{bmatrix} I_{n_1} & \Gamma_{2|1} \\ \Gamma_{1|2}^T & I_{n_2} \end{bmatrix}^+ = \begin{bmatrix} (I_{n_1} - \Gamma_{2|1}\Gamma_{1|2}^T)^+ & -(I_{n_2} - \Gamma_{1|2}^T\Gamma_{2|1})^+\Gamma_{2|1} \\ -(I_{n_1} - \Gamma_{2|1}\Gamma_{1|2}^T)^+\Gamma_{1|2}^T & (I_{n_2} - \Gamma_{1|2}^T\Gamma_{2|1})^+ \end{bmatrix}$$

$$\boxed{P_{x_1\to x_2} = \sum_{y\in\Omega_2} \gamma_{2|1}(y|x_2)\gamma_{1|2}(x_2|y)} \qquad \boxed{Q_{y_1\to y_2} = \dots}$$

Tables with fixed marginals

Definition (Plan in $\Delta^\circ(\Omega)$)

- $\mu_1 \in \Delta^\circ(\Omega_1)$, $\mu_2 \in \Delta^\circ(\Omega_2)$
- $\Gamma^\circ(\mu_1, \mu_2) = \{\gamma \in \Delta^\circ(\Omega) | X_\# \gamma = \mu_1, Y_\# \gamma = \mu_2\}$

Bundle of $\Gamma^\circ(\mu_1, \mu_2)$

1. Let $t \mapsto \gamma(t) \in \Gamma(\mu_1, \mu_2)$ be a regular curve of $S\Delta^\circ(\Omega)$ with $\gamma(0) = \gamma$. Let $B = S_\gamma \Delta^\circ(\Omega)$ be the fiber at γ with ANOVA decomposition $B = B_1(\gamma) \oplus B_2(\gamma) \oplus B_{12}(\gamma)$. Then $D\gamma(0) \in B_{12}(\gamma)$.
2. Viceversa, given any $X \in B_{12}(\gamma)$, the curve $t \mapsto \gamma(t) = (1 + tV)\gamma$ stays in $\Gamma(\mu_1, \mu_2)$ with $X = D\gamma(0)$.

Gradient flow to the dissimilarity index

- $w \colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ is a cost, $W \colon \Delta^\circ(\Omega) \ni \gamma \mapsto \mathbb{E}_\gamma\left[w\right]$ is the expected cost. The function $W \colon \Gamma^\circ(\mu_1, \mu_2) \to \mathbb{R}$ has *statistical gradient*

$$\operatorname{grad}_{\Gamma^\circ(\mu_1,\mu_2)} W(\gamma) = (\gamma, w - \mathbb{E}_\gamma\left[w\right] - w_{1,\gamma} - w_{2,\gamma})$$

- The *gradient flow* of W is

$$D\gamma(t) = -\left(w - \mathbb{E}_{\gamma(t)}\left[w\right] - w_{1,\gamma(t)} - w_{2,\gamma(t)}\right)$$

- Should be $\lim_{t\to\infty} \gamma(t) = \gamma^* \in \Delta(\Omega)$ with $\mathbb{E}_{\gamma^*}\left[w\right] =$ the *Gini's dissimilarity* between μ_1 and μ_2.
- The extension of the gradient to $\Delta(\Omega)$ is should be zero at γ^*, namely

$$w = \mathbb{E}_{\gamma^*}\left[w\right] + w_{1,\gamma^*} + w_{2,\gamma^*} \quad \text{on } \operatorname{supp} \gamma^*$$

to be compared with the dual Kantorovich problem.

THANKS

Critical points

Theorem

1. *If* $\mathbb{R}_+ \ni t \mapsto p(t)$ *is a solution of the gradient flow of a bounded* $f \colon \Delta^\circ(\Omega) \to \mathbb{R}$*, namely,* $Dp(t) = -\operatorname{grad} f(p(t))$*,* $t > 0$*, then* $t \mapsto f(p(t))$ *is decreasing and bounded below by* $-\min f$ *.*
2. *If moreover* $t \mapsto \left\|\operatorname{grad} f(p(t))\right\|_{p(t)}^2 = \left\|Dp(t)\right\|_{p(t)}^2$ *is uniformly continuous, then* $\lim_{t\to\infty} \left\|Dp(t)\right\|_{p(t)} = 0$*.*
3. *Assume in addition that* $p \mapsto \left\|\operatorname{grad} f(p)\right\|_p$ *continuously extend to* $L \colon \Delta(\Omega)$ *and there exists a level set* $\{p \in \Delta^\circ(\Omega) | L(p) \leq a\}$ *where* L *has an unique zero* $\bar{p} \in \Delta(\Omega)$*. Hence,* $f(p(0)) \leq \alpha$ *implies* $\lim_{t\to\infty} p(t) = \bar{p}$*.*

Transports

Definition (e- and m-transport)

1. The *exponential transport*, or *e-transport*, is the family of linear mappings
$$ {}^{e}\mathbb{U}_p^q \colon B_p \ni U \mapsto U - \mathbb{E}_q\left[U\right] \in B_q \ . $$
2. The *mixture transport*, or *m-transport*, is the family of linear mappings
$$ {}^{m}\mathbb{U}_p^q \colon B_p \ni U \mapsto \frac{p}{q} U \in B_q \ . $$

1. Exponential semi-group property: ${}^{e}\mathbb{U}_q^r \, {}^{e}\mathbb{U}_p^q = {}^{e}\mathbb{U}_p^r$.
2. Mixture semi-group property: ${}^{m}\mathbb{U}_q^r \, {}^{m}\mathbb{U}_p^q = {}^{m}\mathbb{U}_p^r$.
3. Duality: $\left\langle {}^{e}\mathbb{U}_p^q U, V \right\rangle_q = \left\langle U, {}^{m}\mathbb{U}_q^p V \right\rangle_p$.
4. Conservation of the scalar product: $\left\langle {}^{e}\mathbb{U}_p^q U, {}^{m}\mathbb{U}_p^q V \right\rangle_q = \left\langle U, V \right\rangle_p$.

Hilbert transport

Definition

The *Hilbert transport*, or *h-transport*, is the family of linear mappings

$$ {}^0\mathbb{U}_p^q \colon B_p \ni U \mapsto \sqrt{\frac{p}{q}} U - \left(1 + \mathbb{E}_q\left[\sqrt{\frac{p}{q}}\right]\right)^{-1} \left(1 + \sqrt{\frac{p}{q}}\right) \mathbb{E}_q\left[\sqrt{\frac{p}{q}} U\right] \in B_q $$

1. Inverse: ${}^0\mathbb{U}_q^p \, {}^0\mathbb{U}_p^q u = u$.
2. Isometry: $\left\langle {}^0\mathbb{U}_p^q U, {}^0\mathbb{U}_p^q V \right\rangle_q = \langle U, V \rangle_p$.

- The trasports lead to a proper definition of accelleration and geodesic
- and of Hessian
- and a Taylor formula