EO357 CMStatistics 2023

B0262
Information geometry of ANOVA and transport on a finite state space

Giovanni Pistone

DE CASTRO STATISTICS

Compiled December 14, 2023

web-page: www.giannidiorestino.it (slides)
e-mail: giovanni.pistone@carloalberto.org
orcidID: 0000-0003-2841-788X
The author acknowledges the support of De Castro Statistics. He is a member of INdAM-GNAMPA and UMI-AI&ML&MAT.

Abstract

Functional ANOVA (Analysis of variance) appears in Statistics and System Theory. It is a particular orthogonal splitting of the vector space of square-integrable random variables on a product space. When the sample space is factorial, it conveniently splits the fibres of the affine bundle consisting of couples of probability functions and Fisher's scores, which we call the statistical bundle. One of the terms in the splitting is the additive model, while the other is related to the transportation model with fixed margins. This concept is known in the classical theory of contingency tables. We rephrase it and show implications to algebraic statistics, information geometry, and Kantorovich optimal transport. In this setting, the gradient flow in the transport sub-model has a limit point that solves the Kantorovich problem.

keywords: ANOVA, statistical bundle, gradient flow, additive and transportation model, Kantorovich problem.

Selected references

1968 J. Hajek, *Asymptotic Normality of Simple Linear Rank Statistics Under Alternatives*, The Annals of Mathematical Statistics, 39 (1968), pp. 325 – 346

1996 S. L. Lauritzen, *Graphical models*, The Clarendon Press Oxford University Press, 1996

1981 B. Efron and C. Stein, *The jackknife estimate of variance*, Ann. Statist., 9 (1981), pp. 586–596

2001 I. Sobol', *Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates*, Mathematics and Computers in Simulation, 55 (2001), pp. 271–280

2019 G. Peyré and M. Cuturi, *Computational optimal transport*, Foundations and Trends in Machine Learning, 11 (2019), pp. 355–607]

2021 G. Pistone, *Statistical bundle of the transport model*, in Geometric Science of Information: 5th International Conference, GSI 2021, Paris, France, July 21–23, 2021, Proceedings, Springer-Verlag, Berlin, Heidelberg, 2021, p. 752–759

2021 G. Sergeant-Perthuis, *Intersection property and interaction decomposition*, arXiv:1904.09017

2021 G. Pistone, F. Rapallo, and M. P. Rogantin, *Finite space Kantorovich problem with an MCMC of table moves*, Electron. J. Statist., 15 (2021), pp. 880–907

2022 G. Chirco and G. Pistone, *Dually affine Information Geometry modeled on a Banach space*, arXiv:2204.00917

2023 G. Pistone and M. Shoaib, *Kaniadakis' Information Geometry of Compositional Data*, Entropy, 25 (2023)

PART 1

ANOVA and Affine Statistical Bundle

ANOVA with two non-independent factors I

- Consider a product finite sample space $\Omega = \Omega_1 \times \Omega_2$. The generic probability function is denoted

$$q \colon \Omega_1 \times \Omega_2 \ni (x_1, x_2) \mapsto q(x_1, x_2) \ .$$

We denote the two margins by

$$q_1(x_1) = \sum_{y \in \Omega_2} f(x_1, y) \ , \quad q_2(x_2) = \sum_{y \in \Omega_1} q(y, x_2) \ .$$

- For each probability function q and each random variable $u \in L^2(q)$ we look for q-orthogonal decomposition of the form

$$u(x_1, x_2) = u_0 + (u_1(x_1) + u_2(x_2)) + u_{12}(x_1, x_2)$$

- Notice that we do not require $u_1 \perp u_2$ as it is done in Hajek:1968 and Sobol':2001. Cf. also Efron and Stein 1981.
- We call factors the two marginal projections of the sample space,

$$X_1 \colon (x_1, x_2) \mapsto x_1 \ , \quad X_2 \colon (x_1, x_2) \mapsto x_2 \ .$$

ANOVA with two non-independent factors II

- Cf. Lauritzen 1996 and Sergeant-Pethuis 2021.
- Consider the subsets $I \subset \{1, 2\}$, partially ordered by inclusion, that is,
$$\emptyset \prec \{1\}, \{2\} \prec \{1, 2\} \ .$$
- Each $I \neq \emptyset$ is an interaction. Let X_I be the components projection on I, $X_I = (X_j \colon j \in I)$.
- A q-effect is a random variable with zero q-mean. A q-effect of the interaction I is a q-effect of the form $f \circ X_I$ which is q-orthogonal to all $g \circ X_J$ for all $J \prec I$, that is, $J \subset I$ and $J \neq I$.
- The order of the interaction I is $\# I$. Let H_k be the vector space generated by the I-interactions of order k. H_0 contains random variables which do not depend on any X_j, $j = 1, 2$. that is, $H_0 = \mathbb{R}$.
- The space H_1 is generated by the random variables of the form $f_1 \circ X_1$ and $f_2 \circ X_2$ with
$$\mathbb{E}_q [f_1 \circ X_1] = \mathbb{E}_{q_1} [f_1] = 0 \ , \quad \mathbb{E}_q [f_2 \circ X_2] = \mathbb{E}_{q_2} [f_2] = 0 \ .$$

ANOVA with two non-independent factors III

- An element of $H_1(q)$ is of the form

$$f_1 \circ X_1 + f_2 \circ X_2 \ , \quad f_1 \in L^2_0(q_1) \ , f_2 \in L^2_0(q_2)$$

  and the representation above is unique.

- An element of $H_2(q)$ is of the form $f_{12} \circ (X_1, X_2)$ and is orthogonal to $H_\emptyset, H_{\{1\}}, H_{\{2\}}$.
- The orthogonality with respect to $H_\emptyset$ implies zero q-expectation $\mathbb{E}_q[f_{12}] = 0$.
- The orthogonality with respect to $H_\emptyset + H_{\{1\}}$ and $H_\emptyset + H_{\{2\}}$ is equivalent to zero conditional expectation with respect to each factor:

$$\mathbb{E}_q\left(f_{12} \circ (X_1, X_2) | X_1\right) = 0 \ , \quad \mathbb{E}_q\left(f_{12} \circ (X_1, X_2) | X_2\right) = 0$$

ANOVA with two non-independent factors IV

- We have a q-orthogonal decomposition of $f \in L^2(q)$ of the form

$$0 = f_0 \oplus (f_1 \circ X_1 + f_2 \circ X_2) \oplus f_{12} \circ (X_1, X_2)$$

  with $f_0 \in H_0$, $(f_1 \circ X_1 + f_2 \circ X_2) \in H_1$, and $f_{12} \circ (X_1, X_2) \in H_2$.

- Let $f \mapsto \text{Hajek}\,(q)\, f$ be the orthogonal projection of $L^2(q)$ onto H_1, the Hajek projection.
- The orthogonal decomposition of $f \in L^2(q)$ is computed as

$$f = \mathbb{E}_q\,[f] \oplus \text{Hajek}\,(q)\, f \oplus (\mathrm{I} - \mathbb{E}_q - \text{Hajek}\,(q))f \ .$$

- The computation of the Hajek projection (cf. Pistone 2001) is a least square problem in f_0, f_1, f_2 with normal equations

$$\begin{cases} \mathbb{E}_q\,[f] = f_0 \\ \qquad 0 = f_0 + f_1 \circ X_1 + \mathbb{E}_q\,(f_2 \circ X_2 | X_1) \\ \qquad 0 = f_0 + \mathbb{E}_q\,(f_1 \circ X_1 | X_2) + f_2 \circ X_2 \end{cases}$$

Affine statistical bundle I

- The affine statistical bundle is a structure that describes the joint geometry of probabilities and random variables. This justifies the adjective "statistical".
- The geometry is affine in the sense of Weyl's axioms: for each couple of points $P, Q \in \mathcal{M}$ there is vector $v = \overrightarrow{PQ}$ in such a way $Q = P + v$ and $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$.
- We consider the set of couples (q, v) such that q is a positive probability function, $q \in \mathcal{P}_{>}(\Omega)$, and v is a random variable whose q-expectation is zero, $v \in L_0^2(q)$. The vector space $L_0^2(q)$ is the fibre at q.
- We modify the original Weyl's definition to allow for vector paces depending on the base point.

Definition (Statistical bundle)

$$S\mathcal{P}_{>}(\Omega) = \left\{(q, v) \,\middle|\, q \in \mathcal{P}_{>}(\Omega), v \in L_0^2(q)\right\}$$

Exponential chart I

- We define the exponential displacement from $p \in \mathcal{P}_>(\Omega)$ to $q \in \mathcal{P}_>(\Omega)$ by

$$(p, q) \mapsto s_p(q) = \log \frac{q}{p} - \mathbb{E}_p \left[\log \frac{q}{p} \right] \in L_0^2(p) = S_p\mathcal{P}_>(\Omega) \ ,$$

  and the exponential transport between fibres by

$${}^{\mathrm{e}}\mathbb{U}_q^p \colon S_q\mathcal{P}_>(\Omega) \ni v \mapsto v - \mathbb{E}_p[v] \in S_p\mathcal{P}_>(\Omega) \ .$$

- The (generalised) parallelogram law holds true:

$$\left(\log \frac{q}{p} - \mathbb{E}_p \left[\log \frac{q}{p} \right] \right) + {}^{\mathrm{e}}\mathbb{U}_q^p \left(\log \frac{r}{q} - \mathbb{E}_q \left[\log \frac{r}{q} \right] \right) = \\ \left(\log \frac{q}{p} - \mathbb{E}_p \left[\log \frac{q}{p} \right] \right) + \left(\log \frac{r}{q} - \mathbb{E}_p \left[\log \frac{r}{q} \right] \right) = \\ \log \frac{r}{p} - \mathbb{E}_p \left[\log \frac{r}{p} \right] \ .$$

Exponential chart II

- The inverse chart (the patch) s_p^{-1} is defined on all of the fibre $S_p\mathcal{P}_>(\Omega)$ by

$$s_p^{-1}(v) = \exp\left(v - K_p(v)\right) \cdot p = e_p(v) \ , \quad K_p(v) = \log \mathbb{E}_p\left[\mathrm{e}^v\right] \ .$$

- The cumulant functional

$$K_p \colon S_p\mathcal{P}_>(\Omega) \ni v \mapsto K_p(v) = \log \mathbb{E}_p\left[\mathrm{e}^v\right]$$

has several important properties.

- It is an expression in the affine chart of the Kullback-Leibler divergence as a function of the second variable. If $s_p(q) = v$, then

$$\mathrm{D}\left(p \,\|\, q\right) = \mathbb{E}_p\left[\log \frac{p}{q}\right] = \mathbb{E}_p\left[\log \frac{p}{\exp\left(v - K_p(v)\right) \cdot p}\right] = K_p(v) \ .$$

Mixture chart

- We define the mixture displacement from $p \in \mathcal{P}_>(\Omega)$ to $q \in \mathcal{P}_>(\Omega)$ on by

$$(p, q) \mapsto \eta_p(q) = \frac{q}{p} - 1 \in L_0^2(p) = S_p\mathcal{P}_>(\Omega) \ ,$$

and the mixture transport between fibres by

$${}^{\mathrm{m}}\mathbb{U}_p^q \colon S_p\mathcal{P}_>(\Omega) \ni v \mapsto \frac{p}{q} v \in S_p\mathcal{P}_>(\Omega) \ .$$

- The (generalized) parallelogram law holds true

$$\left(\frac{q}{p} - 1\right) + \frac{q}{p}\left(\frac{r}{q} - 1\right) = \left(\frac{r}{p} - 1\right) \ .$$

- The inverse chart $\eta_p(v)$ is defined for all $v > -1$, $v \in S_p\mathcal{P}_>(\Omega)$, by

$$\eta_p^{-1}(v) = (1 + v) \cdot p \ .$$

Duality, velocity, gradient I

Duality

The exponential transport and the mixture transport are dual of each other with respect to the L_0^2 inner product, $\langle v, w \rangle_p = \mathbb{E}_p[vw]$. For all $v \in S_q\mathcal{P}_>(\Omega)$ and $w \in S_p\mathcal{P}_>(\Omega)$ it holds

$$\left\langle v, {}^{\mathrm{e}}\mathbb{U}_p^q w \right\rangle_q = \left\langle {}^{\mathrm{m}}\mathbb{U}_q^p v, w \right\rangle_p \ .$$

Affine velocity

The velocity in the chart at p of a smooth curve $t \mapsto q(t) \in \mathcal{P}_>(\Omega)$ is

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

$$\text{or} \quad \frac{d}{dt}\eta_p(q(t)) = \frac{d}{dt}\left(\frac{q(t)}{p} - 1\right) = \frac{\dot{q}(t)}{p} \ .$$

Duality, velocity, gradient II

- In the moving frame $p = q(t)$ the two representation are equal. Such an expression of the velocity,

$$\overset{\star}{q}(t) = \frac{\dot{q}(t)}{q(t)} = \frac{d}{dt} \log q(t) \ ,$$

  equals the classical Fisher's score. Notice that $\overset{\star}{q}$ is a lift to the bundle, $t \mapsto (q(t), \overset{\star}{q}(t)) \in S\mathcal{P}_{>}(\Omega)$.

- The (natural) gradient of a smooth function $\phi \colon \mathcal{P}_{>}(\Omega) \to \mathbb{R}$ is the section $\operatorname{grad} \phi$ of the statistical bundle such that for all smooth curve $t \mapsto q(t)$ it holds

$$\frac{d}{dt}\phi(q(t)) = \left\langle \operatorname{grad} \phi(q(t)), \overset{\star}{q}(y) \right\rangle_{q(t)} \ .$$

- The gradient flow of ϕ is the solution of the equation

$$\overset{\star}{q}(t) = -\operatorname{grad} \Phi(q(t)) \ .$$

PART 2

Product sample space

Transport plans

- Assume a product sample space $\Omega = \Omega_1 \times \Omega_2$ and consider the probability simplex $\mathcal{P}(\Omega_1 \times \Omega_2)$. The two marginalisation mappings are

$$\Pi_1 \colon \mathcal{P}(\Omega_1 \times \Omega_2) \ni q \mapsto \sum_{x_2 \in \Omega_2} q(\cdot, x_2) \in \mathcal{P}(\Omega_1)$$

$$\Pi_2 \colon \mathcal{P}(\Omega_1 \times \Omega_2) \ni q \mapsto \sum_{x_1 \in \Omega_1} q(x_1, \cdot) \in \mathcal{P}(\Omega_2)$$

- Each q is a transport plan from q_1 to $q_2 = q_{2|1} q_1$.
- For each given $q_1 \in \mathcal{P}(\Omega_1)$ and $q_2 \in \mathcal{P}(\Omega_2)$ define the set of transport plans as

$$\Pi(q_1, q_2) = \{q \in \mathcal{P}(\Omega_1 \times \Omega_2) \,|\, \Pi_1 q = q_1, \Pi_2 q = q_2\} \ .$$

- $\Pi(q_1, q_2)$ is non-empty, convex, and closed. Cf. the algebraic version in Pistone, Rapallo, Rogantin 2021.

Transport plans in $\mathcal{P}_>(\Omega_1 \times \Omega_2)$

- For all $q_1 \in \mathcal{P}_>(\Omega_1)$ and $q_2 \in \mathcal{P}_>(\Omega_2)$ the set of positive transport plans from q_1 to q_2 is

$$\mathring{\Pi}(q_1, q_2) = \{q \in \mathcal{P}_>(\Omega_1 \times \Omega_2) \,|\, \Pi_1 q = q_1, \Pi_2 q = q_2\}$$

- A sub-manifold of the affine statistical manifold $(\mathcal{M}, s_p, B_p, \mathbb{U}_p^q \colon p, q \in M)$ is a subset $\mathcal{N} \subset \mathcal{M}$ such that for each $q \in \mathcal{N}$ there exists a smooth splitting of the fibre at q,

$$B_q = S_q\mathcal{N} \oplus R_q\mathcal{N} \ ,$$

  and the vector space $S_q\mathcal{N}$ is the set of all velocities of curves in $\mathcal{N}$ through q.

- Basic examples of sub-manifolds of the affine statistical manifold are exponential families and mixture models. Notice that a sub-manifold of the affine statistical manifold is not forced to be an affine space.

- $\mathring{\Pi}(q_1, q_2)$ is a sub-manifold of the affine statistical manifold on $\mathcal{P}_>(\Omega_1 \times \Omega_2)$.

Velocity of a curve in $\overset{\circ}{\Pi}(q_1, q_2)$ I

- Let $t \mapsto q(t)$ be a smooth curve of $\mathcal{P}_{>}(\Omega_1 \times \Omega_2)$ with values in the set of strictly positive transport plans, $t \mapsto q(t) \in \overset{\circ}{\Pi}(q_1, q_2)$.
- Recall Fisher's score properties,

$$\overset{\star}{q}(t) = \frac{d}{dt} \log q(t) = \frac{\dot{q}(t)}{q(t)}$$

$$\frac{d}{dt}\mathbb{E}_{q(t)}[f] = \left\langle f - \mathbb{E}_{q(t)}[f], \overset{\star}{q}(t) \right\rangle_{q(t)} .$$

- For each random variable depending only on one factor

$$0 = \frac{d}{dt}\mathbb{E}_{q_j}[f_j] = \frac{d}{dt}\mathbb{E}_{q(t)}[f_j \circ X_j] = \left\langle f_j \circ X_j - \mathbb{E}_{q(t)}[f_j \circ X_j], \overset{\star}{q}(t) \right\rangle_{q(t)} = \mathbb{E}_{q(t)}[f_j \circ X_j \overset{\star}{q}(t)] .$$

  Hence $\mathbb{E}_{q(t)}(\overset{\star}{q}(t) | X_j) = 0$, $j = 1, 2$.
- That is, $\overset{\star}{q}(t)$ is a $q(t)$-interaction, $\overset{\star}{q}(t) \in H_2(q(t))$.

Velocity of a curve in $\mathring{\Pi}(q_1, q_2)$ II

- Conversely, let $q \in \mathring{\Pi}(q_1, q_2)$ and $c_{12} \in H_2(q)$. The curve $t \mapsto (1 + tc_{12}) \cdot q$ is defined for t in a neighborhood of 0, stays in $\mathring{\Pi}(q_1, q_2)$,

$$\mathbb{E}_{(1+tc_{12})\cdot q}\left[g \circ X_j\right] = \mathbb{E}_q\left[(1 + tc_{12}) g \circ X_j\right] = \mathbb{E}_{q_j}\left[g\right] \ ,$$

  and the velocity at 0 is c_{12},

$$\frac{d}{dt} \log\left((1 + tc_{12}) \cdot q\right)\Big|_{t=0} = \frac{c_{12} q}{(1 + tc_{12}) q}\Big|_{t=0} = c_{12}$$

Proposition

For all $q \in \mathring{\Pi}(q_1, q_2)$, the velocities' fibre equals the vector space of interactions,

$$S_q \mathring{\Pi}(q_1, q_2) = H_2(q)$$

Velocity of a curve in $\overset{\circ}{\Pi}(q_1, q_2)$ III

- A splitting of the statistical bundle at $q \in \overset{\circ}{\Pi}(q_1, q_2)$ is

$$S_q \mathcal{P}_> (\Omega_1 \times \Omega_2) = S_q \overset{\circ}{\Pi}(q_1, q_2) \oplus \text{Hajek}(q)\, S_q \mathcal{P}_> (\Omega_1 \times \Omega_2) \ .$$

- That is, the complement fibre $R_q \overset{\circ}{\Pi}(q_1, q_2)$ is

$$\text{Hajek}(q)\, S_q \mathcal{P}_> (\Omega_1 \times \Omega_2) = H_1(q) = \\ \left\{ f_1 \circ X_1 + f_2 \circ X_2 \,\middle|\, \mathbb{E}_{q_1}[f_1] = \mathbb{E}_{q_2}[X_2] = 0 \right\} \ ,$$

which in turn provides the exponential family of additive statistics,

$$\exp\left(f_1 \circ X_1 + f_2 \circ X_2 - K_q(f_1 \circ X_1 + f_2 \circ X_2)\right) \cdot q \ .$$

- The splitting suggests the parameterisation of each $q \in \mathcal{P}_> (\Omega_1 \times \Omega_2)$ by the margins and an interaction.

$\overset{\circ}{\Pi}(q_1, q_2)$ as an affine space

- If $q, r \in \overset{\circ}{\Pi}(q_1, q_2)$ and $c_{12} \in H_2(q) = S_q\overset{\circ}{\Pi}(q_1, q_2)$,

$$\mathbb{E}_r\left[{}^{\mathrm{m}}\mathbb{U}_q^r c_{12} g_i \circ X_i\right] = \mathbb{E}_r\left[\left(\frac{q}{r} c_{12}\right) g_i \circ X_i\right] = \mathbb{E}_q\left[c_{12}\, g_i \circ X_i\right] = 0$$

  that is, $\frac{q}{r} c \in H_2(r) = S_r\overset{\circ}{\Pi}(q_1, q_2)$.

- We have defined a co-cycle of transports

$$S\overset{\circ}{\Pi}(q_1, q_2) = \left\{(q, c) \middle| q \in \overset{\circ}{\Pi}(q_1, q_2), c \in H_2(q)\right\}$$

- The dual transport is computed as follows. If $q, r \in \overset{\circ}{\Pi}(q_1, q_2)$, $c_{12} \in S_q\overset{\circ}{\Pi}(q_1, q_2) = H_2(q)$, and $d_{12} \in S_r\overset{\circ}{\Pi}(q_1, q_2) = H_2(r)$, then

$$\left\langle {}^{\mathrm{m}}\mathbb{U}_q^r c, d \right\rangle_r = \mathbb{E}_q\left[cd\right] = \left\langle c, d - \mathrm{Hajek}\left(q\right) d \right\rangle_q = \left\langle c, \left({}^{\mathrm{m}}\mathbb{U}_q^r\right)^T d \right\rangle_q$$

$$\text{that is,} \quad \left({}^{\mathrm{m}}\mathbb{U}_q^r\right)^T = \left(I - \mathrm{Hajek}\left(q\right)\right) .$$

Geodesics

- Let us compute the mixture geodesic. If $(q, c) \in S\mathring{\Pi}(q_1, q_2)$, an m-geodesic is a curve in $t \mapsto q(t) \in \mathring{\Pi}(q_1, q_2)$ with "constant" velocity.
- Let $(q(0), \overset{\star}{q}(0)) = (q, c)$ and $\overset{\star}{q}(t) = {}^{\mathrm{m}}\mathbb{U}_q^{q(t)} c$. It follows

$$\frac{\dot{q}(t)}{q(t)} = \frac{q}{q(t)} c \quad \text{then} \quad q(t) = (1 + tc) \cdot q \ .$$

The m-geodesic from q in the direction c is $t \mapsto (1 + tc) \cdot q$.

- The affine chart is the geodesic at $t = 1$:

$$\mathring{\Pi}(q_1, q_2) \times \mathring{\Pi}(q_1, q_2) \ni (q, r) \mapsto \frac{r}{q} - 1$$

- The e-geodesic from q in the direction c is the solution of

$$\overset{\star}{q}(t) = (I - \operatorname{Hajek}(q(t)))c \ .$$

- A solution of this equation requires the computation of the Hajek projection.

Gradient of the expected cost

Let us discuss the Optimal Transport OT problem in the framework of the affine statistical bundle.

- $c \colon \Omega_1 \times \Omega_2 \to \mathbb{R}_{\geq}$ is the cost function and the expected cost function is
$$C \colon \mathcal{P}_{>}(\Omega_1 \times \Omega_2) \ni q \mapsto \mathbb{E}_q[c] \ .$$
- The function $q \mapsto C(q)$ restricted to the open transport model $q \in \overset{\circ}{\Pi}(q_1, q_2)$ has gradient in $S\overset{\circ}{\Pi}(q_1, q_2)$ given by

$$\frac{d}{dt} C(q(t)) = \frac{d}{dt} \mathbb{E}_{q(t)}[c] = \left\langle c - \mathbb{E}_{q(t)}[c], \overset{\star}{q}(t) \right\rangle_{q(t)} = \left\langle \left(c - \mathbb{E}_{q(t)}[c]\right) - \operatorname{Hajek}(q(t)) \left(c - \mathbb{E}_{q(t)}[c]\right), \overset{\star}{q}(t) \right\rangle_{q(t)} \ ,$$

that is,

$$\operatorname{grad} C(q) = (c - C(q)) - \operatorname{Hajek}(q) (c - C(q))$$

Gradient flow of the OT cost

- The equation of the gradient flow of C is

$$\overset{\star}{q}(t) = -\left(c - C(q(t)) - \text{Hajek}\left(q(t)\right)\left(c - C(q(t))\right)\right) \ .$$

- Notice that the gradient above is the projection onto the space orthogonal to the space of simple effects. Hence, it is actually well defined for all $q \in \mathcal{P}\left(\Omega_1 \times \Omega_2\right)$. If $\hat{q}$ is a zero of this extended map, then c equals the sum of two functions in one variable on the support of $\hat{q}$.
- If a solution $t \mapsto q(t)$ of the gradient flow equation converges to a transport plan $\bar{q} = \lim_{t\to\infty} q(t) \in \Pi(q_1, q_2)$, then $\mathbb{E}_{\bar{q}}\left[c\right]$ is the value of the Kantorovich optimal transport problem.
- The form of the gradient is compatible with the classical result in OT: if $\bar{q}$ is an optimal plan, that the cost is equal to the sum of two univariate potentials. Cf., e.g., Peyré and Cuturi 2019.