Comparison of the Kim-Milman and Brenier maps

Anastasiya Tanana
E-mail: anastasiya.tanana@gmail.com

Abstract
It is shown that the non-expansive map pushing a Gaussian measure \( \mu \) onto a probability measure log-concave with respect to \( \mu \) obtained by Kim and Milman in [4] is in general different from the Brenier map. The argument continues Example 6.1 in [4].

1 Introduction

In this section we introduce the two maps mentioned in the title of this note. For two Borel probability measures \( \mu \) and \( \nu \) on \( \mathbb{R}^n \) a Borel map \( T : \mathbb{R}^n \to \mathbb{R}^n \) is said to push \( \mu \) forward to \( \nu \) (or transport \( \mu \) onto \( \nu \)), denoted by \( T_\# \mu = \nu \), if \( \mu(T^{-1}(\Omega)) = \nu(\Omega) \) for every Borel set \( \Omega \subset \mathbb{R}^n \), or equivalently, if for every bounded Borel function \( \zeta : \mathbb{R}^n \to \mathbb{R} \)

\[
\int_{\mathbb{R}^n} (\zeta \circ T)(x)d\mu(x) = \int_{\mathbb{R}^n} \zeta(y)d\nu(y).
\]

We consider a Gaussian measure \( \mu \) with density

\[
d\mu(x) = \frac{\sqrt{\det(A)}}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}x^TAx\right),
\]

where \( A \) is a symmetric positive definite matrix, and a Borel probability measure \( \nu \) log-concave with respect to \( \mu \), that is, \( d\nu = \exp(-F)d\mu \) for a convex function \( F : \mathbb{R}^n \to \mathbb{R} \).

1.1 The Brenier map

This map comes from the Monge-Kantorovich optimal transport problem with quadratic cost, that is, the problem of finding a minimizer of the functional

\[
\int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|^2 d\pi(x, y)
\]

over all couplings \( \pi \) of \( \mu \) and \( \nu \) (a Borel probability measure \( \pi \) on \( \mathbb{R}^n \times \mathbb{R}^n \) is said to be a coupling of \( \mu \) and \( \nu \) if for every Borel set \( \Omega \subset \mathbb{R}^n \), \( \pi(\Omega \times \mathbb{R}^n) = \mu(\Omega) \) and \( \pi(\mathbb{R}^n \times \Omega) = \nu(\Omega) \)).

A result stated by Brenier [1] and refined by McCann [5] can be formulated as follows:

**Theorem 1.1.** Let \( \mu, \nu \) be Borel probability measures on \( \mathbb{R}^n \) and assume that \( \mu \) is absolutely continuous with respect to the Lebesgue measure. Then there exists a unique, up to a \( \mu \)-nullset, measurable map \( T \) such that \( T_\# \mu = \nu \) and \( T = \nabla \varphi \) for some convex function \( \varphi \). If in addition \( \mu \) and \( \nu \) have finite second order moments, then \( (\text{Id} \times \nabla \varphi)_\# \mu \) is the unique solution of the Monge-Kantorovich optimal transport problem with quadratic cost.

The map \( \nabla \varphi \), defined up to a \( \mu \)-nullset, is referred to as the Brenier map.

It was observed by Caffarelli in [2] that the Brenier map transporting a Gaussian measure \( \mu \) onto a probability measure \( \nu \) log-concave with respect to \( \mu \) is non-expansive (i.e., 1-Lipschitz).
1 Introduction

1.2 The Kim-Milman map

Kim and Milman’s construction in [4] produces a non-expansive map transporting Gaussian $\mu$ onto a probability measure $\nu$ log-concave with respect to $\mu$ via interpolation along the heat flow. We sketch the construction here.

They consider the second-order differential operator

$$L = \exp\left(\frac{1}{2}x^\top Ax\right) \nabla \cdot \left(\exp\left(-\frac{1}{2}x^\top Ax\right) \nabla \right) = \Delta - Ax \cdot \nabla,$$

and the corresponding Fokker-Planck equation

$$\frac{d}{dt} P^A_t(f) = LP^A_t(f),$$

$$P^A_0(f) = f.$$(1)

It is known that for a sufficiently regular function $f$ the solution to (1) is given by the Mehler formula [3]

$$P^A_t(f)(x) = \int_{\mathbb{R}^n} f \left(\exp(-tA)x + \sqrt{\text{Id} - \exp(-2tA)}y\right) d\mu(y).$$

Then they introduce the flow of probability measures $\nu_t$ defined by $d\nu_t = P^A_t(\exp(-F)) d\mu$ so that $\nu_0 = \nu$ and $\nu_t \to \mu$ as $t \to \infty$ (for example, in $L^1$). The equation (1) and the definition of $L$ immediately imply that the densities of $\nu_t$ solve the linear transport equation

$$\frac{d}{dt} \left( \frac{d\nu_t}{dx} \right) - \nabla \cdot \left( \frac{d\nu_t}{dx} \nabla \log P^A_t(\exp(-F)) \right) = 0.$$

Theory for this equation gives (for example, see [8, Theorem 5.34]) that if there exists a locally Lipschitz family of diffeomorphisms $\{S_t\}_{t \in [0,\infty)}$ solving the initial value problem

$$\frac{d}{dt}S_t(x) = w_t(S_t(x)), \quad S_0(x) = x,$$(2)

for the velocity field $w_t(x) = -\nabla \log P^A_t(\exp(-F))(x)$, then $S_t \# \nu = \nu_t$. The maps $S_t$ are globally well-defined under some conditions on $F$ (for general $F$ the final non-expansive map $T_\# \mu = \nu$ is obtained by approximation) and the equation (2) implies by differentiation that

$$\frac{d}{dt} DS_t(x) = Dw_t\big|_{S_t(x)} DS_t(x), \quad DS_0 \equiv \text{Id}.$$

By the Prékopa-Leindler inequality (see Theorems 3 and 6 in [2]) $\log P^A_t(\exp(-F))$ is a concave function and thus $Dw_t = -D^2 \log P^A_t(\exp(-F))$ is positive semidefinite at each point. This implies that

$$\frac{d}{dt}(DS_t)^\top(x)DS_t(x) = (DS_t)^\top(x) \left[ (Dw_t)^\top\big|_{S_t(x)} + Dw_t\big|_{S_t(x)} \right] (DS_t)(x) \geq 0,$$

and therefore $S_t$ are expansions for all $t \geq 0$. Their inverses $T_t = S_t^{-1}$ are then non-expansive and can be shown to converge (uniformly on compact sets, up to a subsequence) to a non-expansive map $T$. Since $T_t \# \nu_t = \nu$, in the limit $T_\# \mu = \nu$. 
2 Comparison

In the last chapter of [4], Kim and Milman compare their map $T$ with the Brenier map. They give a sufficient condition for the two maps to be the same (in particular, when $n = 1$, or when $\mu$ and $\nu$ are both radially symmetric, the maps do coincide), but are unable to show that in some case they are different. Continuing Example 6.1 in [4], we show that there exist Gaussian measures $\mu$ and $\nu$ such that the construction does not give the Brenier map between them.

Example 2.1. We consider the special case $\frac{d\nu_t(x)}{dt} = c_t \cdot \exp \left( -\frac{1}{2} x^T A_t x \right), \quad \frac{d\nu_0}{dt} = c_0 \cdot \exp \left( -\frac{1}{2} x^T B_0 x \right)$, where $A$ and $B$ are symmetric positive definite matrices, and achieve a contradiction assuming that for all such $A$ and $B$ the Kim-Milman map between $\mu$ and $\nu$ coincides with the Brenier map.

Fix a pair of symmetric positive definite non-commuting matrices $A$ and $B$ and the corresponding pair of measures $\mu$ and $\nu$. The Mehler formula can be used to obtain that

$$P_t^A \left( c_0 \exp \left( -\frac{1}{2} x^T B_0 \right) \right) (x) = c_t \exp \left( -\frac{1}{2} x^T B_t x \right)$$

for some constants $c_t$ and constant in space symmetric matrices $B_t$ (with $B_0 = B$), which are positive semidefinite by the Prékopa-Leindler inequality and decaying to 0 as $t \to \infty$. The vector field $w_t(x) = -\nabla \log P_t^A(\exp(-\frac{1}{2} x^T B_0))(x) = B_t x$ is continuous in $t$ and globally Lipschitz in $x$ on $[0, T] \times \mathbb{R}^n$ for every $T \geq 0$. These conditions imply that $w_t$ uniquely defines for each $x \in \mathbb{R}^n$ the curve $S_t(x)$ solving (2) on the time interval $[0, \infty)$. In particular, it uniquely defines the flow of diffeomorphisms $S_t$ solving (2).

The Mehler formula also gives the explicit expression for $\nu_t$:

$$d\nu_t = d_t \exp \left( -\frac{1}{2} x^T (A + B_t) x \right) dx,$$

where $d_t = \frac{\sqrt{\det(A+B_t)}}{(2\pi)^{\frac{n}{2}}}$ are the normalizing constants. Hence, $\nu_t$ are also Gaussian and log-concave with respect to $\mu$. Fix $t \geq 0$ and consider Kim and Milman’s construction for measures $\mu$ and $\nu$. Notice that the flow of measures interpolating between $\nu$ and $\mu$ is the time-shifted initial flow $\nu_t$:

$$d\tilde{\nu}_s = P_s^A \left( P_t^A \left( c_0 \exp \left( -\frac{1}{2} x^T B_0 \right) \right) \right) d\mu = P_s^A \left( c_0 \exp \left( -\frac{1}{2} x^T B_0 \right) \right) d\mu = d\nu_{t+s}, \quad \forall s \geq 0.$$

This is a consequence of the semigroup property for $P^A$: $P^A_s \circ P^A_t = P^A_{s+t}$ for all $s, t \geq 0$, which follows, for example, from the Mehler formula. For the same reason, the corresponding velocity field $\tilde{w}_s = -\nabla \log P_s^A(P_t^A(c_0 \exp(-\frac{1}{2} x^T B_0)))$ is the time-shift of the initial one: $\tilde{w}_s = w_{t+s}$. This implies that the flow of diffeomorphisms $S_t$ along $\nu$ and the flow of diffeomorphisms $\tilde{S}_s$ along $\tilde{w}_s$ ($\tilde{S}_s \# \tilde{\nu} = \tilde{\nu}_s$) satisfy

$$S_{t+s} = \tilde{S}_s \circ S_t, \quad \forall s \geq 0.$$

Then the inverse diffeomorphisms $T_s = S_s^{-1}$ and $\tilde{T}_s = \tilde{S}_s^{-1}$ satisfy the relation

$$\tilde{T}_s = S_t \circ T_{t+s}, \quad \forall s \geq 0. \quad (3)$$

Denote by $T_{0, opt}$ the Brenier map between $\mu$ and $\nu$, and by $T_{t, opt}$ the Brenier map between $\mu$ and $\tilde{\nu} = \nu_t$. By our assumption, $T_{t+s} \to T_{0, opt}$ and $\tilde{T}_s \to T_{t, opt}$ as $s \to \infty$. In particular, taking the limit as $s \to \infty$ in (3) gives

$$T_{t, opt} = S_t \circ T_{0, opt}, \quad \forall t \geq 0. \quad (4)$$
Since $\nu_t$ and $\mu$ are Gaussian, the Brenier map between $\nu_t$ and $\mu$ is given explicitly (see Example 1.7 in \cite{[6]}) by multiplication by the symmetric positive definite matrix
\[
A^{1/2}(A^{1/2}(A + B_t)A^{1/2})^{-1/2}A^{1/2}.
\]
Therefore, the Brenier map $T_{t,opt}$ between $\mu$ and $\nu_t$, being the unique map pushing $\mu$ forward to $\nu_t$ which is a gradient of a convex function, should be given by multiplication by the inverse of this matrix, i.e.,
\[
DT_{t,opt}(x) = A^{-1/2}(A^{1/2}(A + B_t)A^{1/2})^{1/2}A^{-1/2}, \quad \forall x \in \mathbb{R}^n.
\]
Recall that the flow $S_t$ satisfies $\frac{d}{dt}DS_t(x) = Dw_t|_{S_t(x)}DS_t(x)$, $DS_0 \equiv \text{Id}$. Since $Dw_t \equiv B_t$, $S_t$ are given by multiplication by constant in space matrices $DS_t$ satisfying the ODE
\[
\frac{d}{dt}DS_t = B_t(DS_t).
\]
Multiplying this ODE from the right by the matrix $DT_{0,opt}$, from \cite{[11]} we obtain that $DT_{t,opt}$ satisfy the ODE $\frac{d}{dt}DT_{t,opt} = B_t(DT_{t,opt})$ as well. In particular, since $DT_{t,opt}$ are symmetric, $B_t(DT_{t,opt})$ should be symmetric for all $t$. Consider $t = 0$: $B_0(DT_{0,opt}) = BA^{-1/2}(A^{1/2}(A + B)A^{1/2})^{1/2}A^{-1/2}$. This matrix is symmetric if and only if $A^{1/2}BA^{-1/2}(A^{1/2}(A + B)A^{1/2})^{1/2}$ is symmetric. But then $A^{1/2}BA^{-1/2}(A^{1/2}(A + B)A^{1/2}) = A^{1/2}B(A + B)A^{1/2}$ is symmetric as well (here we have used that for symmetric positive definite matrices $C, D$ we have: $CD^{1/2}$ is symmetric $\Rightarrow CD = CD^{1/2}D^{1/2} = D^{1/2}CD^{1/2} = DC \Rightarrow CD$ is symmetric), implying that $B(A + B)$ is symmetric, i.e., $A$ and $B$ commute. Since $A$ and $B$ were non-commuting, the assumption was not correct, meaning that the Kim-Milman map does not generically coincide with the Brenier map.

Remark. It is still of interest whether by any chance the construction always gives the Brenier map in the special case when $\mu$ is the standard Gaussian measure. When $\mu$ is standard Gaussian and $\nu$ is Gaussian, it does give the Brenier map: it can be shown that the matrices $B_t$ obtained by the Mehler formula all commute and hence the solution of $\frac{d}{dt}DS_t(x) = Dw_t|_{S_t(x)}DS_t(x)$, $DS_0 \equiv \text{Id}$, is well-defined by the formula
\[
DS_t(x) = \exp \left( \int_0^t B_s ds \right), \quad \forall x \in \mathbb{R}^n.
\]
Thus, $DS_t$ are symmetric positive definite, constant in space matrices. The same holds for their inverses $DT_t$, i.e., $T_t$ are affine functions and are the gradients of convex quadratic functions. A (pointwise up to a subsequence) limit $T$ of $T_t$ as $t \to \infty$ will preserve these properties and, therefore, will be the gradient of a convex function transporting $\mu$ onto $\nu$, that is, the Brenier map.
References

[1] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

[2] L. A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Communications in Mathematical Physics*, 214(3):547–563, 2000.

[3] G. Hargé. A particular case of correlation inequality for the Gaussian measure. *The Annals of Probability*, 27(4), 1999.

[4] Y.-H. Kim and E. Milman. A generalization of Caffarelli’s contraction theorem via (reverse) heat flow. *Mathematische Annalen*, 354(3):827–862, 2012.

[5] R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–323, 1995.

[6] R. J. McCann. A convexity principle for interacting gases. *Advances in Mathematics*, 128:153–179, 1997.

[7] A. Prékopa. On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum*, 34:335–343, 1973.

[8] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.