A Brunn-Minkowski inequality for the KL-divergence

Liran Rotem

Technion – Israel Institute of Technology

Based on joint work with Gautam Aishwarya

61 Probability Encounters, In honor of Sergey Bobkov
Toulouse, May-June 2023

Convex Geometry and Information Theory

Contemporary Mathematics
Volume **545**, 2011

Fractional generalizations of Young and Brunn-Minkowski inequalities

Sergey Bobkov, Mokshay Madiman, and Liyao Wang

Abstract. A generalization of Young's inequality for convolution with sharp constant is conjectured for scenarios where more than two functions are being convolved, and it is proven for certain parameter ranges. The conjecture would provide a unified proof of recent entropy power inequalities of Barron and Madiman, as well as of a (conjectured) generalization of the Brunn-Minkowski inequality. It is shown that the generalized Brunn-Minkowski conjecture is true for convex sets; an application of this to the law of large numbers for random sets is described.

ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal of Functional Analysis 262 (2012) 3309–3339

JOURNAL OF Functional Analysis

www.elsevier.com/locate/jfa

Reverse Brunn–Minkowski and reverse entropy power inequalities for convex measures ☆

Sergey Bobkov [a], Mokshay Madiman [b,*]

[a] *School of Mathematics, University of Minnesota, 206 Church St. S.E., Minneapolis, MN 55455, USA*
[b] *Department of Statistics, Yale University, 24 Hillhouse Avenue, New Haven, CT 06511, USA*

Received 14 October 2010; accepted 18 January 2012
Available online 31 January 2012
Communicated by Cédric Villani

Abstract

We develop a reverse entropy power inequality for convex measures, which may be seen as an affine-geometric inverse of the entropy power inequality of Shannon and Stam. The specialization of this inequality to log-concave measures may be seen as a version of Milman's reverse Brunn–Minkowski inequality. The proof relies on a demonstration of new relationships between the entropy of high dimensional random vectors and the volume of convex bodies, and on a study of effective supports of convex measures, both of which are of independent interest, as well as on Milman's deep technology of M-ellipsoids and on certain information-theoretic inequalities. As a by-product, we also give a continuous analogue of some Plünnecke–Ruzsa inequalities from additive combinatorics.

Keywords: Log-concave; Convex measure; Brunn–Minkowski inequality; Entropy

Part 1: Brunn–Minkowski inequalities

The Brunn–Minkowski inequality

Given Borel sets $A, B \subseteq \mathbb{R}^n$ and $0 < t < 1$ we consider the Minkowski combination

$$(1-t)\,A + tB = \left\{(1-t)a + tb : \; a \in A, \; b \in B\right\}.$$

Theorem (Brunn–Minkowski)

We have

$$\mathrm{Vol}\left((1-t)A + tB\right)^{\frac{1}{n}} \geq (1-t)\,\mathrm{Vol}\left(A\right)^{\frac{1}{n}} + t\,\mathrm{Vol}\left(B\right)^{\frac{1}{n}}.$$

where $\mathrm{Vol}\left(\cdot\right)$ *is the usual (Lebesgue) volume.*

By homogeneity of volume this is of course the same as $\mathrm{Vol}\left(A + B\right)^{\frac{1}{n}} \geq \mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(B)^{\frac{1}{n}}$, but our formulation generalizes better.

Borell–Brascamp–Lieb

We characterize all measures that satisfy a Brunn–Minkowski inequality:

Theorem (Borell, Brascamp–Lieb)

Let μ be a Borel measure on $\mathbb{R}^n$ with density f. Fix $\kappa \in [-\infty, \frac{1}{n}]$. Then the following are equivalent:

1. *For all Borel sets $A, B \subseteq \mathbb{R}^n$ such that $\mu(A)\mu(B) > 0$ and $0 < t < 1$ we have*

$$\mu\left((1-t)A + tB\right) \geq \left((1-t)\mu(A)^{\kappa} + t\mu(B)^{\kappa}\right)^{\frac{1}{\kappa}}.$$

2. *For all $x, y \in \mathbb{R}^n$ such that $f(x)f(y) > 0$ and $0 < t < 1$ we have*

$$f\left((1-t)x + ty\right) \geq \left((1-t)f(x)^{\alpha} + tf(y)^{\alpha}\right)^{\frac{1}{\alpha}},$$

   where $\frac{1}{\alpha} + n = \frac{1}{\kappa}$.

The cases $\alpha, \kappa \in \{-\infty, 0, +\infty\}$ are interpreted in the limiting sense. This theorem can be further extended to weighted Riemannian manifolds and beyond, but we will stick to $\mathbb{R}^n$ for this talk.

Our main example: The Gaussian measure

We denote by γ the Gaussian measure on $\mathbb{R}^n$, i.e.

$$\frac{\mathrm{d}\gamma}{\mathrm{d}x} = \varphi(x) = \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2}.$$

Since $|x|^2/2$ is convex, φ is log-concave, i.e.

$$\varphi\left((1-t)x + ty\right) \geq \varphi(x)^{1-t}\varphi(t)^t$$

(i.e. $\alpha = 0$ in the Borell–Brascamp–Lieb theorem). Therefore for all Borel $A, B \subseteq \mathbb{R}^n$ we have

$$\gamma\left((1-t)A + tB\right) \geq \gamma(A)^{1-t}\gamma(B)^t.$$

(i.e. $\kappa = 0$ in the theorem). However, in general we do **not** expect an inequality of the form

$$\gamma\left((1-t)A + tB\right) \geq \left((1-t)\gamma(A)^\kappa + t\gamma(B)^\kappa\right)^{1/\kappa}$$

for any $\kappa > 0$. This can be checked directly by taking $B = \{x\}$ and letting $|x| \to \infty$.

Beyond Borell–Brascamp–Lieb

Can we do better than the inequality

$$\gamma\left((1-t)A + tB\right) \geq \gamma(A)^{1-t}\gamma(B)^{t}$$

if we assume more on the sets A and B? Taking them to be convex will not help.

Question (Gardner-Zvavitch 2010)

If $K, T \subseteq \mathbb{R}^n$ are convex and $0 \in K \cap T$, can it be true that

$$\gamma\left((1-t)K + tT\right)^{\frac{1}{n}} \geq (1-t)\gamma(K)^{\frac{1}{n}} + t\gamma(T)^{\frac{1}{n}}?$$

By taking K and T to be "very small" it is easy to see that the exponent $\frac{1}{n}$ is the best possible.

Nayar and Tkocz found a counterexample, but conjectured the inequality is true if we replace the assumption $0 \in K \cap T$ with the stronger assumption that K and T are centrally symmetric ($K = -K$ and $T = -T$, i.e. if $x \in K$ then $-x \in K$).

Infinitesimal formulation

Since we want to prove concavity of the functional

$$\rho(t) = \gamma\left((1-t)K + tT\right)^{\frac{1}{n}},$$

a natural idea is to check that $\rho''(0) \le 0$.

- Kolesnikov–Milman computed this derivative, and introduced a new idea to transform the resulting inequality from ∂K to the interior of K. They used this idea to prove a Brunn–Minkowski inequality on weighted manifolds under curvature conditions, recovering in particular Borel–Brascamp–Lieb.
- Kolesnikov-Livshyts applied these ideas to our problem, and noticed that the curvature condition of Kolesnikov-Milman only has to hold "on average" in order to have

$$\gamma\left((1-t)K + tT\right)^{\frac{1}{n}} \ge (1-t)\gamma(K)^{\frac{1}{n}} + t\gamma(T)^{\frac{1}{n}}.$$

A sufficient condition

Theorem (Kolesnikov–Livshyts, specialized to the measure γ)

Assume that for every symmetric convex body $K \subseteq \mathbb{R}^n$ and every smooth even function $u : K \to \mathbb{R}$ such that $Lu := \Delta u - \langle \nabla u, x \rangle = 1$ we have

$$\int_K \left(\left\| \nabla^2 u \right\|^2 + \left| \nabla u \right|^2 \right) \mathrm{d}\gamma \geq \kappa \cdot \gamma(K).$$

Then for all convex and symmetric $K, T \subseteq \mathbb{R}^n$ we have

$$\gamma\left((1-t)K + tT\right)^\kappa \geq (1-t)\gamma(K)^\kappa + t\gamma(T)^\kappa.$$

Theorem (Kolesnikov–Livshyts)

The inequality holds with $\kappa = \frac{1}{2n}$, even if we only assume that $0 \in K \cap T$.

A complete solution

Theorem (Eskenazis–Moschidis)

For every symmetric convex body $K \subseteq \mathbb{R}^n$ and every even smooth function $u : K \to \mathbb{R}$ with $Lu \equiv 1$ we have

$$\int_K \left(\left\| \nabla^2 u \right\|^2 + |\nabla u|^2 \right) \mathrm{d}\gamma \geq \frac{1}{n} \gamma(K).$$

Therefore, for all convex and symmetric $K, T \subseteq \mathbb{R}^n$ we have

$$\gamma\left((1-t)K + tT\right)^{\frac{1}{n}} \geq (1-t)\gamma(K)^{\frac{1}{n}} + t\gamma(T)^{\frac{1}{n}}.$$

The proof is a clever application of the Gaussian Poincaré inequality: For every smooth $f : \mathbb{R}^n \to \mathbb{R}$ we have

$$\mathrm{Var}_\gamma f \leq \int |\nabla f|^2 \, \mathrm{d}\gamma.$$

Other measures

Theorem (Cordero-Erausquin–R.)

Let $w:(0,\infty)\to\mathbb{R}$ be a non-decreasing function such that $t\mapsto w(e^t)$ is convex. Consider the measure μ with density $e^{-w(|x|)}$. Then for all convex and symmetric $K,T\subseteq\mathbb{R}^n$ we have

$$\mu\left((1-t)K+tT\right)^{\frac{1}{n}}\geq(1-t)\mu(K)^{\frac{1}{n}}+t\mu(T)^{\frac{1}{n}}.$$

Examples include $\frac{\mathrm{d}\mu}{\mathrm{d}x}=e^{-|x|^p}$ for $p>0$ and $\frac{\mathrm{d}\mu}{\mathrm{d}x}=\frac{1}{\left(1+|x|^2\right)^{\beta}}$ for $\beta>0$ (for which Borel–Brascamp–Lieb does not apply for any value of the parameters!). For this talk we will concentrate on the simplest case of the Gaussian measure γ.

Part 2: Concavity of entropy

What does this have to do with information theory?

There is a well known (yet slightly mysterious) connection between convex geometry and information theory.

Given a probability measure μ on $\mathbb{R}^n$ with density f, the (differential) entropy of μ is defined by

$$h(\mu) = -\int f \log f \, \mathrm{d}x = -\int \log f \, \mathrm{d}\mu.$$

If μ is uniform on a body K then $f = \frac{1}{|K|}\mathbb{1}_K$ and $h(\mu) = \log \mathrm{Vol}(K)$. In fact if μ is any probability measure supported on K we have by Jensen's inequality

$$h(\mu) = \int_K \log\left(\frac{1}{f}\right) f \, \mathrm{d}x \leq \log\left(\int_K \frac{1}{f} f \, \mathrm{d}x\right) = \log \mathrm{Vol}(K).$$

So we think of $e^{h(\mu)}$ as analogous to $\mathrm{Vol}(K)$.

The Entropy Power Inequality

If $e^{h(\mu)}$ is analogous to $\mathrm{Vol}(K)$, what is the analogue of the Brunn–Minkowski inequality? It is usually considered to be:

Theorem (Entropy Power Inequality, Shannon–Stam)

For all probability measures μ and ν on $\mathbb{R}^n$ we have

$$e^{\frac{2}{n}h(\mu*\nu)} \geq e^{\frac{2}{n}h(\mu)} + e^{\frac{2}{n}h(\nu)},$$

where $\mu * \nu$ denotes the convolution (corresponding to sum of independent random variables).

This looks similar to Brunn–Minkowski, but doesn't formally imply or is implied by it. By taking μ and ν to be uniform on A and B respectively we get

$$|A+B|^{\frac{2}{n}} \geq |A|^{\frac{2}{n}} + |B|^{\frac{2}{n}}$$

since $\mu * \nu$ is supported on $A+B$. But we can do better!

Optimal transport

Instead of considering $\mu * \nu$, we consider another way to interpolate between μ and ν. We do so using the standard notion of optimal transport.

Given probability measures μ and ν, we say that $T_\sharp \mu = \nu$ ("T pushes μ to ν") if for all $B \subseteq \mathbb{R}^n$ we have

$$\mu\left(T^{-1}(B)\right) = \nu(B).$$

Our goal is to find the "most efficient" transport map, i.e. to minimize

$$\int |x - Tx|^2 \, \mathrm{d}\mu(x)$$

over all transport maps T pushing μ to ν.

The Brenier map

Let $\mathcal{P}_2 = \mathcal{P}_2(\mathbb{R}^n)$ denote the class of all probability measures on $\mathbb{R}^n$ with finite second moment.

Theorem (Brenier)

If $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$ and μ is absolutely continuous, then the minimization problem

$$\min \left\{ \int |x - Tx|^2 \, \mathrm{d}\mu(x) : \; T_\sharp \mu = \nu \right\}$$

has a unique solution. Moreover this solution is characterize by being the unique transport map which is the gradient of a convex function, $T = \nabla \varphi$.

We define the 2-Wasserstein distance on $\mathcal{P}_2$ by

$$W_2^2(\mu, \nu) = \min \left\{ \int |x - Tx|^2 \, \mathrm{d}\mu(x) : \; T_\sharp \mu = \nu \right\}$$

(at least when the measures have a density)

Displacement concavity

Once we have a metric space, we can construct geodesics. Given $\mu_0, \mu_1 \in \mathcal{P}_2$ with Brenier map $T_\sharp \mu_0 = \mu_1$, define

$$\mu_t = \left((1-t) \cdot Id + tT\right)_\sharp \mu_0.$$

Since $T_t = (1-t)Id + tT$ is a Brenier map (being the gradient of a convex function), we can compute and see that $W_2\left(\mu_0, \mu_t\right) = tW_2(\mu_0, \mu_1)$ and $W_2\left(\mu_t, \mu_1\right) = (1-t)W_2\left(\mu\right)$. So $\left\{\mu_t\right\}_{t \in [0,1]}$ is the geodesic connecting μ_0 and μ_1, called the displacement interpolation.

We say that $F : \mathcal{P}_2 \to \mathbb{R}$ is displacement concave if $F(\mu_t)$ is concave in t for every such interpolation $\left\{\mu_t\right\}_{t \in [0,1]}$.

An Entropic Brunn–Minkowski, again

Theorem (Erbar–Kuwada–Sturm, 2015 (?))

$e^{\frac{1}{n}h(\mu)}$ is displacement concave on $\mathcal{P}_2(\mathbb{R}^n)$. Equivalently, for every displacement interpolation $\{\mu_t\}_{t\in[0,1]}$ we have

$$e^{\frac{1}{n}h(\mu_t)} \geq (1-t)e^{\frac{1}{n}h(\mu_0)} + t \cdot e^{\frac{1}{n}h(\mu_1)}.$$

By taking μ_0,μ_1 to be uniform on A,B we get the usual Brunn–Minkowski,

$$|(1-t)A + tB|^{\frac{1}{n}} \geq (1-t)\,|A|^{\frac{1}{n}} + t\,|B|^{\frac{1}{n}}\,.$$

Proof Sketch

$$\boxed{e^{\frac{1}{n}h(\mu_t)} \geq (1-t)e^{\frac{1}{n}h(\mu_0)} + t \cdot e^{\frac{1}{n}h(\mu_1)}}$$

Let f_t denote the density of μ_t. Since $T_t = (1-t)\,Id + tT$ pushes μ_0 to μ_t we have by change of variables:

$$\det\left(DT_t(x)\right) = \frac{f_0(x)}{f_t\left(T_t x\right)}.$$

We now compute

$$\begin{aligned}
h(\mu_t) &= -\int \log f_t(y)\mathrm{d}\mu_t(y) = -\int \log\left(f_t\left(T_t x\right)\right)\mathrm{d}\mu_0(x) \\
&= -\int \log\left(\frac{f_0(x)}{\det DT_t(x)}\right)\mathrm{d}\mu_0(x) \\
&= h(\mu_0) + \int \log\left(\det \underbrace{\left((1-t)Id + tDT\right)}_{A_t}\right)\mathrm{d}\mu_0,
\end{aligned}$$

Proof Sketch – Contd.

$$\boxed{e^{\frac{1}{n}h(\mu_t)} \geq (1-t)e^{\frac{1}{n}h(\mu_0)} + t \cdot e^{\frac{1}{n}h(\mu_1)}}$$

We saw $h(\mu_t) = C + \int \log(\det A_t)\,\mathrm{d}\mu_0$. The function $M(x,y) = \log((1-t)e^x + te^y)$ is convex, and so

$$\begin{aligned}
M\left(\frac{1}{n}h(\mu_0), \frac{1}{n}h(\mu_1)\right) &= \frac{C}{n} + M\left(\int \log(\det A_0)^{\frac{1}{n}}\,\mathrm{d}\mu_0, \int \log(\det A_1)^{\frac{1}{n}}\,\mathrm{d}\mu_0\right) \\
&\leq \frac{C}{n} + \int M\left(\log(\det A_0)^{\frac{1}{n}}, \log(\det A_1)^{\frac{1}{n}}\right)\mathrm{d}\mu_0 \\
&= \frac{C}{n} + \int \log\left((1-t)(\det A_0)^{\frac{1}{n}} + t(\det A_1)^{\frac{1}{n}}\right)\mathrm{d}\mu_0 \\
&\leq \frac{C}{n} + \int \log(\det A_t)^{\frac{1}{n}}\,\mathrm{d}\mu_0 = \frac{1}{n}h(\mu_t).
\end{aligned}$$

General measures

To discuss other measures, we generalize entropy to relative entropy (or Kullback–Leibler divergence). Given a reference measure ν we define

$$\mathsf{D}(\mu\|\nu) = \int \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\mathrm{d}\mu$$

(if μ is absolutely continuous with respect to ν). Again if $\mu = \nu_A$, i.e.

$$\mu(B) = \frac{\nu\,(B \cap A)}{\nu(A)}$$

then $-\mathsf{D}(\nu_A\|\nu) = \log \nu(A)$, and more generally if μ is supported on A then $-\mathsf{D}(\mu\|\nu) \leq \log \nu(A)$.

Therefore we think of $e^{-\mathsf{D}(\mu\|\nu)}$ as the entropic analogue of $\nu(A)$.

Borell–Brascamp–Lieb for relative entropy

Theorem (Erbar–Kuwada–Sturm)

Let ν be a Borel measure on $\mathbb{R}^n$ with density g. Fix $0 \leq \kappa \leq \frac{1}{n}$. Then the following are equivalent:

1. For every displacement interpolation $\{\mu_t\}_{t\in[0,1]}$ in $\mathcal{P}_2(\mathbb{R}^n)$ we have

$$e^{-\kappa\,\mathrm{D}(\mu_t\|\nu)} \geq (1-t)e^{-\kappa\,\mathrm{D}(\mu_0\|\nu)} + te^{-\kappa\,\mathrm{D}(\mu_1\|\nu)}$$

2. For all $x, y \in \mathbb{R}^n$ such that $g(x)g(y) > 0$ and $0 < t < 1$ we have

$$g\left((1-t)x + ty\right) \geq \left((1-t)g(x)^\alpha + tg(y)^\alpha\right)^{\frac{1}{\alpha}},$$

   where $\frac{1}{\alpha} + n = \frac{1}{\kappa}$.

This implies the classical Borell–Brascamp–Lieb. The full theorem is more general, considering weighted Riemannian manifolds satisfying a $CD(K, N)$ condition and even more general metric measure spaces.

Part 3: The obvious question

Beyond the general inequality

Consider the Gaussian γ as our reference measure. We have the general theorem for the case $\kappa = 0$, i.e.

$$-\operatorname{D}(\mu_t\|\gamma) \geq -(1-t)\operatorname{D}(\mu_0\|\gamma) - t\operatorname{D}(\mu_1\|\gamma)\,,$$

but nothing better can hold for arbitrary interpolations $\{\mu_t\}_{t\in[0,1]}$. But what if we add symmetry?

Note that if μ_t is the law of a $N(0,t)$ random variable then $\{\mu_t\}_{t\geq 0}$ is a displacement interpolation, and

$$-\operatorname{D}(\mu_t\|\gamma) = \frac{1}{2}\left(1 - t + \log t\right).$$

This is concave function on $(0,\infty)$, but $e^{-\kappa\operatorname{D}(\mu_t\|\gamma)}$ is not concave on $(0,\infty)$ for any $\kappa > 0$. **However,** $e^{-\operatorname{D}(\mu_t\|\gamma)}$ is concave for $t \in [0,1]$.

The general theorem

Theorem (Aishwarya–R.)

Let $\{\mu_t\}_{t\in[0,1]}$ *be a displacement interpolation in* $\mathcal{P}_2(\mathbb{R}^n)$ *between two even measures* μ_0 *and* μ_1*.*
Assume *every* μ_t *satisfies a Poincaré inequality with constant* 1*, i.e.*

$$\mathrm{Var}_{\mu_t}\psi \leq \int |\nabla\psi|^2 \,\mathrm{d}\mu_t. \qquad (\star)$$

Then we have

$$e^{-\frac{1}{n}\,\mathrm{D}(\mu_t\|\gamma)} \geq (1-t)e^{-\frac{1}{n}\,\mathrm{D}(\mu_0\|\gamma)} + te^{-\frac{1}{n}\,\mathrm{D}(\mu_1\|\gamma)}.$$

This explains the previous observation: If μ_t is the law of $N(0,t)$ then μ_t satisfies $(\star)$ iff $t \leq 1$. Of course in general this condition is not easy to check.

Useful corollaries – a Gaussian endpoint

Corollary

Let $\{\mu_t\}_{t\in[0,1]}$ *be a displacement interpolation in* $\mathcal{P}_2(\mathbb{R}^n)$ *between* $\mu_0 = \gamma$ *and* μ_1 *which is even and log-concave with respect to* γ *(i.e.* $\frac{\mathrm{d}\mu_1}{\mathrm{d}\gamma}$ *is log-concave). Then*

$$e^{-\frac{1}{n}\,\mathrm{D}(\mu_t\|\gamma)} \geq (1-t)e^{-\frac{1}{n}\,\mathrm{D}(\mu_0\|\gamma)} + te^{-\frac{1}{n}\,\mathrm{D}(\mu_1\|\gamma)}$$

Proof.

By Caffarelli contraction theorem the Brenier map T from μ_0 to μ_1 is a contraction. Therefore so is

$$T_t = (1-t)Id + tT.$$

Therefore, since $\mu_0 = \gamma$ satisfies Poincaré with constant 1, so does every measure $\mu_t = (T_t)_\sharp(\mu_0)$. $\square$

Useful corollaries – dimension $n = 1$

Corollary

Let $\{\mu_t\}_{t\in[0,1]}$ be a displacement interpolation in $\mathcal{P}_2\left(\mathbb{R}^1\right)$ between μ_0 and μ_1 which are both even and log-concave with respect to γ. Then

$$e^{-\mathrm{D}(\mu_t\|\gamma)} \geq (1-t)e^{-\mathrm{D}(\mu_0\|\gamma)} + te^{-\mathrm{D}(\mu_1\|\gamma)}$$

Proof.

Let T_0 and T_1 be the Brenier maps between γ and μ_0, μ_1. In dimension $n = 1$ the composition of Brenier maps is a Brenier map (since "gradient a of convex function" = "increasing"). From this one can show that

$$((1-t)T_0 + tT_1)_\sharp(\gamma) = \mu_t.$$

By Caffarelli T_0 and T_1 are contractions, hence so is $(1-t)T_0 + tT_1$, so μ_t satisfies Poincaré with constant 1. □

Proof sketch

$$\boxed{e^{-\frac{1}{n}\,\mathrm{D}(\mu_t\|\gamma)} \geq (1-t)e^{-\frac{1}{n}\,\mathrm{D}(\mu_0\|\gamma)} + te^{-\frac{1}{n}\,\mathrm{D}(\mu_1\|\gamma)}}$$

The idea is similar to the geometric case: we define $\rho(t) = e^{-\frac{1}{n}\,\mathrm{D}(\mu_t\|\gamma)}$ and prove that $\rho''(t) \leq 0$ for all $t \in [0,1]$.

Denote by f_t the density of μ_t, by T the Brenier map, and as usual $T_t(x) = (1-t)x + tTx$.
The standard conservation of mass formula states that f_t satisfies a PDE

$$\frac{\partial f_t}{\partial t} + \mathrm{div}\left(f_t \nabla\theta_t\right) = 0.$$

Here $\nabla\theta_t$ is the velocity vector field of the transport, that is

$$\frac{\mathrm{d}}{\mathrm{d}t} T_t(x) = \nabla\theta_t\left(T_t(x)\right).$$

$\nabla\theta_t$ is indeed the gradient of a potential θ_t, and moreover this potential itself satisfies the PDE

$$\frac{\partial\theta_t}{\partial t} + \frac{\left|\nabla\theta_t\right|^2}{2} = 0.$$

Proof sketch – Contd.

$$\boxed{\rho(t) = e^{-\frac{1}{n}\mathsf{D}(\mu_t\|\gamma)}}$$

Since we have formulas for all time derivatives, we can compute ρ''. Let $Lu = \Delta u - \langle \nabla u, x\rangle$ denote the Ornstein-Uhlenbeck generator. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{D}(\mu_t\|\gamma) = -\int (L\theta_t)\,\mathrm{d}\mu_t$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathsf{D}(\mu_t\|\gamma) = \int \left(\left\|\nabla^2\theta_t\right\| + |\nabla\theta_t|^2\right)\mathrm{d}\mu_t.$$

Therefore the claim $\rho''(t) \le 0$ is the same as

$$\int \left(\left\|\nabla^2\theta_t\right\| + |\nabla\theta_t|^2\right)\mathrm{d}\mu_t \ge \frac{1}{n}\left(\int (L\theta_t)\,\mathrm{d}\mu_t\right)^2.$$

But θ_t is even, and this inequality holds for all even functions! It is exactly the Kolesnikov-Livshyts criterion in the geometric case, that was proved by Eskenazis–Moschidis (when $L\theta \equiv 1$).

Proof sketch – Contd.

$$\boxed{\int \left(\left\|\nabla^2\theta\right\| + |\nabla\theta|^2\right) \mathrm{d}\mu \geq \frac{1}{n}\left(\int (L\theta)\,\mathrm{d}\mu\right)^2}$$

Define a function $u = \theta - \frac{a}{2n}|x|^2$ where $a = \int (L\theta)\,\mathrm{d}\mu$. Apply Poincaré to partial derivatives $\partial_i u$ to get

$$\int \left\|\nabla^2 u\right\|^2 \mathrm{d}\mu \geq \int |\nabla u|^2 \,\mathrm{d}\mu \geq \int \left(|\nabla\theta|^2 - \frac{2a}{n}\langle\nabla\theta, x\rangle\right) \mathrm{d}\mu.$$

On the other hand an explicit computation gives

$$\left\|\nabla^2\theta\right\|^2 = \left\|\nabla^2 u\right\|^2 + \frac{2a}{n}\Delta\theta - \frac{a^2}{n}.$$

Hence

$$\int \left(\left\|\nabla^2\theta\right\| + |\nabla\theta|^2\right) \mathrm{d}\mu \geq \int \left(2|\nabla\theta|^2 + \frac{2a}{n}\left(\Delta\theta - \langle\nabla\theta, x\rangle\right) - \frac{a^2}{n}\right) \mathrm{d}\mu$$

$$\geq \frac{2a^2}{n} - \frac{a^2}{n} = \frac{a^2}{n} = \frac{1}{n}\left(\int (L\theta)\,\mathrm{d}\mu\right)^2.$$

Congratulations Sergey!