The optimal transport problem

Luigi Ambrosio

Scuola Normale Superiore, Pisa
luigi.ambrosio@sns.it
http://cvgmt.sns.it

Outline

1. The problem of optimal transportation
2. Structure of optimal transport maps
3. The metric side of optimal transportation
4. Some applications
5. The differential side of optimal transportation
6. Gradient flows, optimal transportation and (nonlinear, diffusion) PDE's

A few references

[1] L.AMBROSIO. *Lecture notes on optimal transport problems.* In: Lecture Notes in Mathematics **1812**, Springer, 2003.

[2] L.AMBROSIO, N.GIGLI, G.SAVARÉ. *Gradient flows in metric spaces and in spaces of probability measures.* Lectures in Mathematics ETH, Birkhäuser, 2005.

[3] L.C. EVANS. *Partial differential equations and Monge-Kantorovich mass transfer.* Current developments in Mathematics, International Press, 1999.

[4] F.OTTO. *The geometry of dissipative evolution equations: the porous medium equation.* Comm. PDE, **26** (2001), 101–174.

[5] S.T. RACHEV, L. RÜSCHENDORF. *Mass transportation problems.* Voll. I, II, Springer, 1989.

[6] C. VILLANI. *Topics in optimal transportation.* American Mathematical Society, 2003.

[7] C. VILLANI. *Optimal Transport, Old and New.* Springer, 2009.

1. The Monge problem of optimal transportation

The problem can be informally described as follows: given $X,\ Y \subset \mathbb{R}^n$, we have two distributions of mass $\rho(x)$ in X and $\rho'(y)$ in Y satisfying the *mass balance condition*

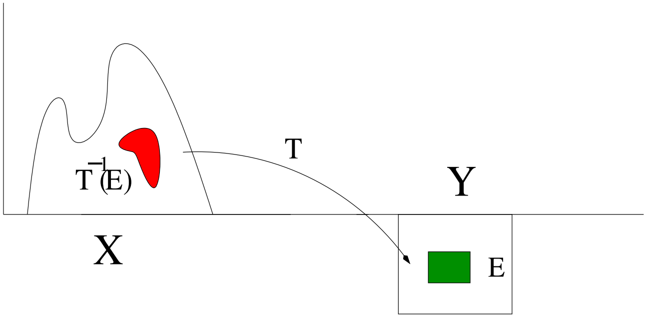
$$\int_X \rho(x)\,dx = \int_Y \rho'(y)\,dy$$

and we want to move ρ into ρ' in such a way that the work done is minimal.

The admissible movements are described by a *transport map* $T : X \to Y$ such that the *local mass balance condition* holds:

$$\int_{T^{-1}(E)} \rho(x)\,dx = \int_E \rho'(y)\,dy \qquad \forall E \subset Y.$$

The local mass balance condition



Since *work=mass*$\times$*displacement*, we have to minimize

$$\mathcal{E}(T) := \int_X |T(x) - x|\rho(x)\,dx$$

among all admissible transport maps T.

Example (Book shifting)

We let $n = 1$, M integer, $X = [0, M]$, $Y = [1, M + 1]$. Then, the map $x \mapsto x + 1$ is optimal, but the map

$$T(x) = \begin{cases} x + M & \text{if } x \in [0, 1); \\ x & \text{if } x \in [1, M] \end{cases}$$

is optimal as well.

Despite its very classical and “natural” structure, this variational problem was not considered so much, in contrast with the variational problems, for instance, arising from Mechanics:

$$\mathcal{A}(x) := \int_0^1 L\left(t, x(t), \dot{x}(t)\right)\, dt.$$

As a matter of fact, even some basic issues, as the analogue of the Euler-Lagrange equations

$$\frac{d}{dt} L_{\dot{x}}\left(t, x(t), \dot{x}(t)\right) = L_x\left(t, x(t), \dot{x}(t)\right)$$

were not understood, until much more recent times.
Indeed, the problem could be attacked successfully only with the modern tools of Measure Theory and Functional Analysis, with the seminal work of Kantorovich, in 1940.

In even more recent times (last 15-20 years) many more connections are emerging between this theory and many other fields: Shape Optimization, Geometric and Functional inequalities, Nonlinear diffusion, Partial Differential Equations, Riemannian Geometry.

A (surely non exhaustive) list includes: Barthe, Bernard, Brenier, Buttazzo, Mc Cann, Cavalletti, Caffarelli, A., Carrillo, Gangbo, Gigli, De Philippis, Fathi, Figalli, Cordero Erausqin, Evans, Kinderlehrer, Savaré, Pratelli, Bouchittè, Feldman, Lott, Mondino, Naber, Otto, Rachev, Rüschendorf, Sturm, Toscani, Villani, Von Renesse.

A modern formulation of the optimal transport problem

We consider:
- a probability measure μ in X;
- a probability measure ν in Y;
- a function $c : X \times Y \to [0, +\infty]$.

Then, we minimize the energy

$$\mathcal{E}(T) := \int_X c\,(x, T(x))\; d\mu(x)$$

among all maps T satisfying

$$\mu(T^{-1}(E)) = \nu(E) \qquad \forall E \subset Y$$

(in short, we will write $T_{\#}\mu = \nu$).

An even more general formulation, allowing *transport plans* instead of transport maps, was considered by Kantorovich, and is very popular and studied in Probability: *find a law in $X \times Y$ whose marginals are μ and ν, and such that the expectation of c is minimal.*

2. Structure of optimal transport maps

One of the basic tools to show existence of optimal transport maps is a *duality* formula: the minimum in Monge's problem is the supremum (and in lucky cases the maximum) of

$$\int_X \varphi \, d\mu(x) + \int_Y \psi \, d\nu(y)$$

among all pairs $\varphi : X \to \mathbb{R}$, $\psi : Y \to \mathbb{R}$ satifying

$$\varphi(x) + \psi(y) \leq c(x, y) \qquad \forall (x, y) \in X \times Y.$$

As a consequence of this fact, for many cost functions c, strong restrictions arise on the possible places y where mass initially at x could be sent (in an optimal way!), namely

$$\varphi(x) + \psi(y) = c(x, y).$$

Cost=distance

Assume for instance that $X = Y = \mathbb{R}^n$, and that $c(x, y) = |x - y|$. Then, by the minimality of

$$x' \mapsto |x' - y| - \varphi(x') - \psi(y)$$

at $x' = x$, *if φ is differentiable at x*, we get

$$\frac{x - y}{|x - y|} = \nabla\varphi(x).$$

Therefore the direction of transportation is given by $-\nabla\varphi(x)$ and only a (unavoidable) 1-dimensional degree of freedom is left.

Theorem (Evans–Gangbo, '95)

Assume that $X = Y = \mathbb{R}^n$, $c(x, y) = |x - y|$, and $\mu \ll \mathscr{L}^n$. Then there exists an optimal transport map.

Cost=distance^2

Assume for instance that $X = Y = \mathbb{R}^n$, and that $c(x,y) = \frac{1}{2}|x-y|^2$. Then, by the minimality of

$$x' \mapsto \frac{1}{2}|x'-y|^2 - \varphi(x') - \psi(y)$$

at $x' = x$, *if φ is differentiable at x*, we get

$$y = x - \nabla\varphi(x) = \nabla\left(\frac{1}{2}|x|^2 - \varphi(x)\right).$$

Theorem (Brenier, Knott–Smith, '80)

Assume that $X = Y = \mathbb{R}^n$, $c(x,y) = \frac{1}{2}|x-y|^2$, and $\mu \ll \mathscr{L}^n$. Then there exists a unique optimal transport map. Furthermore, this map is the gradient of a convex function.

Cost=distance^2 on Riemannian manifolds

Assume for instance that $X = Y = M$, a compact Riemannian manifold, and that $c(x,y) = \frac{1}{2} d_M^2(x,y)$. Then, nowithstanding the lack of differentiability of $d_M^2(\cdot,y)$ in the large, we have:

Theorem (McCann, '01)

Assume that $X = Y = M$, $c(x,y) = \frac{1}{2} d_M^2(x,y)$, and $\mu \ll \mathrm{Vol}_M$. Then there exists a unique optimal transport map, representable by

$$T(x) = \exp_x\left(-\nabla\varphi(x)\right) \qquad \text{for } \mu\text{-a.e. } x.$$

Furthermore, T never goes "beyond" the cut locus, namely $t \mapsto \exp_x\left(-t\nabla\varphi(x)\right)$ is a minimizing geodesic in $[0,s]$ for all $s < 1$.

3. The metric side of optimal transportation

The minimum value in Monge's (or Kantorovich's) problem can be used to define a distance, called Wasserstein distance, between probability measures in X. In the case cost=distance, we set

$$W_1(\mu,\nu) := \inf\left\{\int_X d(x,T(x))\,d\mu(x) :\ T_\#\mu = \nu\right\}.$$

In the case cost=distance^2, instead, we set

$$W_2(\mu,\nu) := \inf\left\{\sqrt{\int_X d^2(x,T(x))\,d\mu(x)} :\ T_\#\mu = \nu\right\}.$$

The "manifold" $\mathscr{P}_2(X)$ of probability measures on X with finite quadratic moments becomes in this way a metric space, which inherits many properties of X (e.g., compact if X is compact, complete if X is complete, PC if X is PC, non-branching if X is non-branching,....).

W_2 metrizes weak convergence plus convergence of moments

Theorem

For $\mu_n,\ \mu \in \mathscr{P}_2(X)$, one has that $W_2(\mu_n, \mu) \to 0$ if and only if

$$\lim_{n\to\infty} \int_X \phi \, d\mu_n = \int_X \phi \, d\mu \qquad \forall \phi \in C_b(X)$$

and

$$\lim_{n\to\infty} \int_X d^2(x, \bar{x}) \, d\mu_n(x) = \int_X d^2(x, \bar{x}) \, d\mu(x)$$

for at least one (and thus for all) $\bar{x} \in X$.

Geodesics in the Wasserstein space

Having put a metric structure on $\mathscr{P}_2(X)$, it is natural to study geodesics $\{\mu_t\}_{t\in[0,1]}$ (i.e. length minimizing curves) in this space. Up to a reparameterization, they are characterized by

$$W_2(\mu_s,\mu_t) = |t-s| W_2(\mu_0,\mu_1) \qquad s,\, t\in[0,1].$$

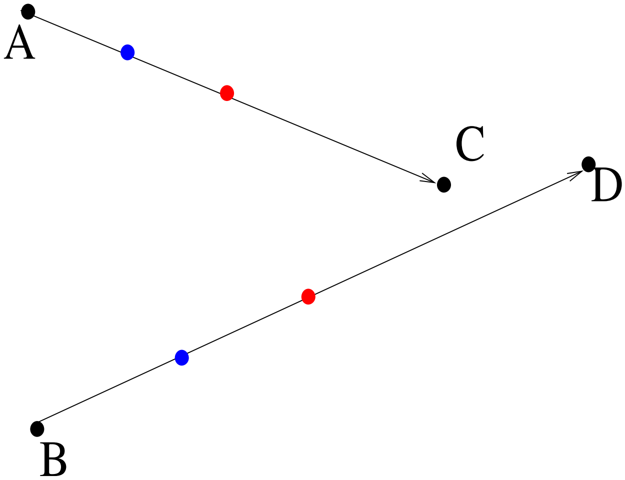
For instance, in the case $X = Y = M$, compact Riemannian manifold, we have a complete characterization:

Theorem

Assume that $\mu \ll \mathrm{Vol}_M$ and let $T(x) = \exp_x(-\nabla\varphi(x))$ be the optimal transport map between μ and ν. Then

$$\mu_t := (T_t)_\#\mu \quad \textit{with} \quad T_t(x) := \exp_x(-t\nabla\varphi(x)) \qquad t\in[0,1]$$

is the unique constant speed geodesic between μ and ν.



Interpolation between $(\delta_A + \delta_B)/2$ and $(\delta_C + \delta_D)/2$

4. Some applications

The theory of optimal transportation provides a new "nonlinear" perspective on $\mathscr{P}(X)$ that is very useful and suggestive in many applications.
Let us consider for instance the problem of interpolating between two probability densities $\rho,\ \rho'$ in $\mathbb{R}^n$. The linear, canonical way:

$$\rho_t := (1-t)\rho + t\rho' \qquad t \in [0,1].$$

The "Wasserstein" way (I=identity map): whenever $T_\#\rho = \rho'$, one defines the interpolating curve

$$\rho_t := ((1-t)I + tT)_\#\,\rho \qquad t \in [0,1].$$

This is still linear, but at the level of transport maps, and nothing but the geodesic interpolation, if T is an optimal transport map.

We may for instance consider a model in which the free energy is given by

$$\mathcal{E}(\rho) := \int \rho \ln \rho \, dx + \int V(x)\rho(x)\, dx + \int\int W(x-y)\rho(x)\rho(y)\, dxdy.$$

In general, because of the interaction potential term, this functional *is not* convex with respect to the linear interpolation, while *it is* convex with respect to the Wasserstein one, if V and W are convex.
This interpolation argument has been used to show uniqueness of ground states.
Finally, notice that the potential energy term $\rho \mapsto \int V\rho$ is linear w.r.t. the standard linear structure of $\mathscr{P}(\mathbb{R}^n)$, but *nonlinear* w.r.t. the Wasserstein one:

$$\int V\rho_t \, dx = \int V\left((1-t)x + tT(x)\right)\rho_0(x)\, dx \qquad t \in [0,1].$$

Indeed, we will see that the "Wasserstein gradient" is ∇V and it is not constant!

The Brunn-Minkowski inequality

Given A, $B \subset \mathbb{R}^n$ compact, this inequality says that

$$\mathrm{Vol}^{1/n}(A+B) \geq \mathrm{Vol}^{1/n}(A) + \mathrm{Vol}^{1/n}(B),$$

where $A+B$ is the Minkowski sum of A and B:

$$A+B := \{a+b: \ a \in A, \ b \in B\}.$$

This inequality can be used, for instance, to prove another important inequality (with sharp constant $C(n)$), the *isoperimetric* one:

$$\mathrm{Vol}^{1/n}(A) \leq C(n)\mathrm{Area}^{1/(n-1)}(\partial A) \qquad n > 1.$$

Proof of BM via optimal transportation

McCann pointed out in his PhD thesis that a direct proof via optimal transportation of the Brunn-Minkowski inequality, in the scaled version

$$\mathrm{Vol}^{1/n}\left(\frac{A+B}{2}\right) \geq \frac{1}{2}\mathrm{Vol}^{1/n}(A) + \frac{1}{2}\mathrm{Vol}^{1/n}(B)$$

can be achieved as follows. First, define the energy

$$\mathcal{E}(\rho) := \int \rho^{1-\frac{1}{n}}\, dx$$

and show that $\mathcal{E}$ is *concave* along Wasserstein geodesics. Then, set

$$\rho_A(x) := \begin{cases} \frac{1}{\mathrm{Vol}(A)} & \text{if } x \in A \\ 0 & \text{if } x \notin A, \end{cases} \qquad \rho_B(x) := \begin{cases} \frac{1}{\mathrm{Vol}(B)} & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$$

and denote by $\{\rho_t\}_{t\in[0,1]}$ the constant speed geodesic between ρ_A and ρ_B. Then, the conclusion follows by

$$\mathcal{E}(\rho_0) = \mathrm{Vol}^{1/n}(A), \quad \mathcal{E}(\rho_1) = \mathrm{Vol}^{1/n}(B), \quad \mathcal{E}(\rho_{1/2}) \leq \mathrm{Vol}^{1/n}\left(\frac{A+B}{2}\right).$$

The differential side of optimal transportation

Having a metric structure on $\mathscr{P}(X)$, we may ask ourselves whether a deeper structure (differential, Riemannian) exists, compatible with the Wasserstein distance, when X has a differentiable structure.
Let $X = \mathbb{R}^n$. The basic ingredient is the *continuity equation*

$$\frac{d}{dt}\mu_t + \mathrm{div}\,(\boldsymbol{v}_t\mu_t) = 0$$

describing the evolution of a time-dependent mass distribution μ_t under the action of a velocity field $\boldsymbol{v}_t(x)$. According to this equation, infinitesimal variations $s = \delta\mu \in T_\mu\mathscr{P}_2(\mathbb{R}^d)$ of μ are coupled to the velocity $\boldsymbol{v}$ by

$$\delta\mu + \mathrm{div}\,(\boldsymbol{v}\mu) = 0.$$

Otto calculus

Looking for gradient vector fields, one is led to the coupling $-\mathrm{div}(\mu\nabla\phi) = s$ linking potential functions ϕ to tangent vectors s, and to the metric (Otto)

$$\mathbf{g}_\mu(s,s') := \int \nabla\phi\cdot\nabla\phi'\,d\mu \quad \text{with } -\mathrm{div}(\mu\nabla\phi) = s,\ -\mathrm{div}(\mu\nabla\phi') = s'.$$

Having defined a tangent bundle and a metric on it, the "Riemannian" distance $d(\nu,\nu')$ induced by this metric is

$$\inf\left\{\sqrt{\int_0^1\int |\boldsymbol{v}_t|^2\,d\mu_t\,dt}:\ \frac{d}{dt}\mu_t + \mathrm{div}\,(\boldsymbol{v}_t\mu_t) = 0,\ \mu_0 = \nu,\ \mu_1 = \nu'\right\}.$$

It turns out that (Benamou-Brenier) this distance is precisely the Wasserstein distance W_2. So, $\mathscr{P}(\mathbb{R}^n)$ is a kind of infinite-dimensional Riemannian manifold. This has been the object of several investigations in the last 10-15 years, and by now a complete and rigorous theory is available.

The heat flow is the W_2-gradient flow of the Entropy!

Let's see how with this calculus we can (at least formally) recover the heat equation as gradient flow of the entropy functional

$$\mathrm{Ent}(\rho) := \int_{\mathbb{R}^n} \rho \log \rho \, dx$$

in $\mathscr{P}_2(\mathbb{R}^n)$. Indeed,

$$d_\rho \mathrm{Ent}(s) = \int_{\mathbb{R}^n} (1 + \log \rho) s \, dx = \int_{\mathbb{R}^n} s \log \rho \, dx$$

and, if we represent $s = -\mathrm{div}\,(\boldsymbol{w}\rho)$, we get

$$d_\rho \mathrm{Ent}(s) = \int_{\mathbb{R}^n} \langle \nabla \log \rho, \boldsymbol{w} \rangle \rho \, dx$$

which tells us, remembering our metric tensor $\mathbf{g}_\rho$, that the "Wasserstein gradient" $\nabla^W \mathrm{Ent}$ of the Entropy at ρ is $\nabla \log \rho$.

The heat flow is the W_2-gradient flow of the Entropy!

In an analogous way, it can be seen that

$$\nabla^W \int_{\mathbb{R}^n} V \, d\mu = \nabla V \quad \text{viewed as an element of } L^2(\mu; \mathbb{R}^n)!$$

and that

$$\nabla^W \int_{\mathbb{R}^n} K(x-y) \, d\mu \times \mu = \nabla (K * \mu) \quad \text{viewed as an element of } L^2(\mu; \mathbb{R}^n).$$

Now, coming back to Ent, writing

$$\frac{d}{dt}\rho_t + \mathrm{div}\left(-\frac{\nabla \rho_t}{\rho_t}\rho_t\right) = \frac{d}{dt}\rho_t - \Delta \rho_t = 0$$

we realize that the velocity field $\boldsymbol{v}_t$ is $-\nabla \log \rho_t = -\nabla^W \mathrm{Ent}(\rho_t)$.

Another key identification

When we look at the heat flow from the W_2 point of view, the rate of energy dissipation $\frac{d}{dt}\mathrm{Ent}(\rho_t)$ is

$$-|\nabla^-\mathrm{Ent}|^2(\rho_t)$$

where $|\nabla^-\mathrm{Ent}|$ is a one-sided gradient, the so-called descending slope:

$$|\nabla^-\mathrm{Ent}|(\rho) := \limsup_{\sigma\to\rho} \frac{[\mathrm{Ent}(\rho)-\mathrm{Ent}(\sigma)]^+}{W_2(\rho,\sigma)}$$

is the so-called descending slope.

On the other hand, a direct "conventional" computation gives

$$\frac{d}{dt}\int \rho_t \log\rho_t\,dx = \int (1+\log\rho_t)\Delta\rho_t\,dx = -\int \frac{|\nabla\rho_t|^2}{\rho_t}\,dx.$$

Hence (at least formally, taking limits as $t\to 0$)

$$|\nabla^-\mathrm{Ent}|^2(\rho) = \int \frac{|\nabla\rho|^2}{\rho}\,dx = 4\int |\nabla\sqrt{\rho}|^2\,dx \quad !$$

Gradient flows, optimal transportation and (nonlinear, diffusion) PDE's

Let $\mathcal{M}$ be a smooth manifold, and $F : \mathcal{M} \to \mathbb{R}$. The gradient flow of F starting from $\bar{x}$ is the solution to the ODE

$$\begin{cases} \dot{x}(t) = -\nabla F(x(t)) \\ x(0) = \bar{x}, \end{cases} \qquad x : [0, T] \to \mathcal{M}$$

Remarks. (1) A metric on $\mathcal{M}$ is needed, to identify $dF(x)$ (a covector) with $\nabla F(x)$ (a vector).
(2) The energy dissipation identity holds:

$$\frac{d}{dt} F(x(t)) = dF_x(\dot{x}(t)) = -|\nabla F(x(t))|^2 .$$

A model case: uniformly convex functions in $\mathbb{R}^n$

Assume that $\mathcal{M} = \mathbb{R}^n$, and that the following uniform convexity condition holds ($\lambda > 0$):

$$\sum_{i,j=1}^{n} \frac{\partial^2 F}{\partial x_i \partial x_j}(x)\xi_i\xi_j \geq \lambda|\xi|^2 \qquad \forall \xi \in \mathbb{R}^n.$$

In this case:

(1) solutions to the gradient flow converge exponentially fast to the unique minimum $x_{\min}$ of F;

(2) the semigroup induced by the gradient flow is strongly nonexpansive:

$$|x(t,\bar{x}) - x(t,\hat{x})| \leq e^{-\lambda t}|\bar{x} - \hat{x}|.$$

Entropy-entropy dissipation inequality

To prove convergence to equilibrium, we prove first the entropy-entropy dissipation inequality

$$F(x) - F(x_{\min}) \leq \frac{1}{2\lambda}|\nabla F(x)|^2.$$

We have indeed (by the 2-order mean value theorem and the uniform convexity condition with $\xi = x - x_{\min}$)

$$\begin{aligned}
F(x) - F(x_{\min}) &\leq \langle \nabla F(x), x - x_{\min} \rangle - \frac{\lambda}{2}|x - x_{\min}|^2 \\
&\leq |\nabla F|(x)|x - x_{\min}| - \frac{\lambda}{2}|x - x_{\min}|^2 \\
&\leq \frac{1}{2\lambda}|\nabla F(x)|^2 + \frac{\lambda}{2}|x - x_{\min}|^2 - \frac{\lambda}{2}|x - x_{\min}|^2.
\end{aligned}$$

Now the proof of the exponential rate of convergence is easy: we differentiate $F(x(t)) - F(x_{\min})$ and use the energy-energy dissipation inequality to get

$$\frac{d}{dt}\left[F\left(x(t)\right) - F(x_{\min})\right] = -\left|\nabla F\left(x(t)\right)\right|^2 \leq -2\lambda\left[F\left(x(t)\right) - F(x_{\min})\right].$$

By integration, $[F(x(t)) - F(x_{\min})] \leq [F(\bar{x}) - F(x_{\min})]e^{-2\lambda t}$. Finally, the uniform convexity condition easily yields the energy-distance bound:

$$F(x) - F(x_{\min}) \geq \frac{\lambda}{2}|x - x_{\min}|^2 \qquad \forall x \in \mathbb{R}^n.$$

Therefore, exponential convergence to $F(x(t))$ to its minimal value yields exponential convergence of $x(t)$ to $x_{\min}$.

Diffusion PDE's and functional inequalities

The previous analysis of convex gradient flows in Euclidean spaces can be extended with no problem to nonsmooth convex gradient flows, even in infinite-dimensional Hilbert spaces H (thanks to the work of Brezis, Komura, Benilan, Pazy in the '70).

However, it is rather surprising that this whole picture still holds for convex (along constant speed geodesics) functionals in the *infinite-dimensional and curved space* $\mathscr{P}(\mathbb{R}^n)$ (or $\mathscr{P}(H)$)

This fact has generated many new results and new proofs on convergence to equilibrium for diffusion PDE's, and functional inequalities.

The relative entropy functional

As a model case, we consider a uniformly convex function $V : \mathbb{R}^n \to [0, +\infty]$ with $\int e^{-V}\,dx = 1$, and the probability measure γ in $\mathbb{R}^n$ whose density w.r.t. $\mathscr{L}^n$ is e^{-V} (the Gaussian, for $V(x) = c(d) + |x|^2/2$). Then, we consider the relative entropy functional

$$\mathcal{E}_\gamma(f) := \int_{\mathbb{R}^n} (f \ln f + fV)\,dx = \int_{\mathbb{R}^n} u \ln u\,d\gamma,$$

where $u = e^V f$ represents the density of $f\mathscr{L}^n$ with respect to γ.
In this case $f_{\min} = e^{-V}$, so that $u_{\min} = 1$.
It turns out that the Fokker-Planck equation

$$\frac{df}{dt} = \Delta f + \mathrm{div}(f \nabla V) = \mathrm{div}\left(\nabla(\ln f + 1 + V) f\right)$$

corresponds to the gradient flow of $\mathcal{E}_\gamma$ with respect to W_2, according to the differential calculus on $\mathscr{P}(\mathbb{R}^n)$ described before (this is based on the fact that the blue term is the functional derivative of $\mathcal{E}_\gamma$ with respect to γ).

Talagrand and logarithmic Sobolev inequalities

In addition, because of the uniform convexity of V, $\mathcal{E}_\gamma$ is uniform convex as well:

$$\mathcal{E}_\gamma(\rho_t) \leq (1-t)\mathcal{E}_\gamma(\rho_0) + t\mathcal{E}_\gamma(\rho_1) - \frac{\lambda}{2} t(1-t) W_2^2(\mu_0, \mu_1).$$

Therefore the energy-distance bound and the energy-energy dissipation inequality apply.
The former corresponds, in the Gaussian case, to Talagrand's inequality

$$W_2^2(u\gamma, \gamma) \leq \frac{2}{\lambda} \int u \ln u \, d\gamma.$$

The latter corresponds to

$$\int f \ln f + fV \, dx \leq \frac{1}{2\lambda} \int |\nabla(\ln f + V)|^2 f \, dx.$$

With the change of variables $f = h^2 e^{-V}$ we get the logarithmic Sobolev inequality

$$\int h^2 \ln h^2 \, d\gamma \leq \frac{2}{\lambda} \int |\nabla h|^2 \, d\gamma + \left(\int h^2 \, d\gamma \right) \ln \left(\int h^2 \, d\gamma \right).$$

The same theory provides convergence and error estimates for implicit time discretization of gradient flows in $\mathscr{P}(X)$, even when the ambient space $X = \mathbb{R}^n$ is replaced by an infinite-dimensional Hilbert space. Hence, the infinite-dimensional versions of the Fokker-Planck equation, and even some non-linear variants, can be studied with these methods.