The dynamics of Schrödinger bridges

Giovanni Conforti

Stochastic processes and statistical machine learning

February, 15, 2018

Plan of the talk

- The Schrödinger problem and relations with Monge-Kantorovich problem
- Newton's law for entropic interpolation
- The entropy along the entropic interpolations

The talk is based on

- G. Conforti. A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost.
  Probability Theory and Related Fields(to appear)

Part I: The of Schrödinger problem and relations with the Monge-Kantorovich problem

Schrödinger's thought experiment

An old story from Schrödinger back in 1931...

" *Imaginez que vous observez un **système de particules en diffusion**, qui soient en **équilibre thermodynamique.** Admettons qu' à un instant donné* 0 *vous les ayez trouvées en **répartition à peu près uniforme** et quà* 1 *vous ayez trouvé un **écart spontané et considérable par rapport à cette uniformité**. On vous demande de quelle manière cet écart sest produit. Quelle en est la **manière la plus probable ?***"

A more recent story from C.Villani's textbook..

*Take a **perfect gas** in which particles do not interact, and ask him to move from a certain **prescribed density field** at time* $t = 0$, *to another prescribed density field at time* $t = 1$. *Since **the gas is red lazy**, he will find a way to do so that it needs a **minimal amount of work (least action path).***

Schrödinger's thought experiment

An old story from Schrödinger back in 1931...

" Imaginez que vous observez un ***système de particules en diffusion****, qui soient en* ***équilibre thermodynamique.*** *Admettons qu' à un instant donné* 0 *vous les ayez trouvées en* ***répartition à peu près uniforme*** *et quà* 1 *vous ayez trouvé un* ***écart spontané et considérable par rapport à cette uniformité****. On vous demande de quelle manière cet écart sest produit. Quelle en est la* ***manière la plus probable ?****"*

A more recent story from C.Villani's textbook..

Take a ***perfect gas*** *in which particles do not interact, and ask him to move from a certain* ***prescribed density field*** *at time* $\mathtt{t} = 0$*, to another prescribed density field at time* $\mathtt{t} = 1$*. Since* ***the gas is red lazy****, he will find a way to do so that it needs a* ***minimal amount of work (least action path).***

Schrödinger's thought experiment

An old story from Schrödinger back in 1931...

*" Imaginez que vous observez un **système de particules en diffusion**, qui soient en **équilibre thermodynamique.** Admettons qu' à un instant donné* 0 *vous les ayez trouvées en **répartition à peu près uniforme** et quà* 1 *vous ayez trouvé un **écart spontané et considérable par rapport à cette uniformité**. On vous demande de quelle manière cet écart sest produit. Quelle en est la **manière la plus probable ?**"*

A more recent story from C.Villani's textbook..

*Take a **perfect gas** in which particles do not interact, and ask him to move from a certain **prescribed density field** at time* $t = 0$*, to another prescribed density field at time* $t = 1$*. Since **the gas is lazy**, he will find a way to do so that it needs a **minimal amount of work (least action path).***

To model Schrödinger's experiment we need

- An **ambient space** $\hookrightarrow$ A Riemannian manifold M
- The **equilibrium dynamics** for the particles $\hookrightarrow$ stationary Brownian motion $\mathbf{P}$
- The **non-interacting** particles $\hookrightarrow X^1_\cdot, \ldots, X^N_\cdot$ independent stationary Brownian motions
- The **particle-configuration** $\hookrightarrow$ empirical measure μ^N

$$\mu^N(A) = \frac{1}{N}\mathrm{Card}(\{i : X^i_\cdot \in A\})$$

- The initial and final **prescribed configurations** $\hookrightarrow \nu_0, \nu_1$.

Schrödinger bridge; Problem formulation

To model Schrödinger's experiment we need

- An **ambient space** $\hookrightarrow$ A Riemannian manifold M
- The **equilibrium dynamics** for the particles $\hookrightarrow$ stationary Brownian motion $\mathbf{P}$
- The **non-interacting** particles $\hookrightarrow X^1_\cdot, \ldots, X^N_\cdot$ independent stationary Brownian motions
- The **particle-configuration** $\hookrightarrow$ empirical measure μ^N

$$\mu^N(A) = \frac{1}{N}\mathrm{Card}(\{i : X^i_\cdot \in A\})$$

- The initial and final **prescribed configurations** $\hookrightarrow \nu_0, \nu_1$.

Schrödinger bridge; Problem formulation

To model Schrödinger's experiment we need

- An **ambient space** $\hookrightarrow$ A Riemannian manifold M
- The **equilibrium dynamics** for the particles $\hookrightarrow$ stationary Brownian motion $\mathbf{P}$
- The **non-interacting** particles $\hookrightarrow X^1_\cdot, \ldots, X^N_\cdot$ independent stationary Brownian motions
- The **particle-configuration** $\hookrightarrow$ empirical measure μ^N

$$\mu^N(A) = \frac{1}{N}\mathrm{Card}(\{i : X^i_\cdot \in A\})$$

- The initial and final **prescribed configurations** $\hookrightarrow \nu_0, \nu_1$.

Schrödinger bridge; Problem formulation

To model Schrödinger's experiment we need

- An **ambient space** $\hookrightarrow$ A Riemannian manifold M
- The **equilibrium dynamics** for the particles $\hookrightarrow$ stationary Brownian motion $\mathbf{P}$
- The **non-interacting** particles $\hookrightarrow X^1_\cdot, \ldots, X^N_\cdot$ independent stationary Brownian motions
- The **particle-configuration** $\hookrightarrow$ empirical measure $\boldsymbol{\mu}^N$

$$\boldsymbol{\mu}^N(A) = \frac{1}{N}\mathrm{Card}(\{i : X^i_\cdot \in A\})$$

- The initial and final **prescribed configurations** $\hookrightarrow \nu_0, \nu_1$.

Schrödinger bridge; Problem formulation

To model Schrödinger's experiment we need

- An **ambient space** $\hookrightarrow$ A Riemannian manifold M
- The **equilibrium dynamics** for the particles $\hookrightarrow$ stationary Brownian motion $\mathbf{P}$
- The **non-interacting** particles $\hookrightarrow X^1_\cdot, \dots, X^N_\cdot$ independent stationary Brownian motions
- The **particle-configuration** $\hookrightarrow$ empirical measure $\boldsymbol{\mu}^N$

$$\boldsymbol{\mu}^N(A) = \frac{1}{N}\mathrm{Card}(\{i : X^i_\cdot \in A\})$$

- The initial and final **prescribed configurations** $\hookrightarrow \nu_0, \nu_1$.

Schrödinger bridge problem: dynamic formulation

We denote the law μ^N by $\mathscr{P}^N$

Sanov'Theorem

$$\frac{1}{N}\log \mathscr{P}^N\left(\mu^N = \mathbf{Q}\right) \asymp -\mathscr{H}^*(\mathbf{Q}|\mathbf{P})$$

Thus, the "most likely evolution" is found solving

Schrödinger Problem (SP)

$$\inf_{\mathbf{Q}} \mathscr{H}_{\text{path}}(\mathbf{Q}|\mathbf{P})$$

$$\mathbf{Q} \in \mathcal{P}(C([0,1],M)), \quad (X_0)_\#\mathbf{Q} = \nu_0, (X_1)_\#\mathbf{Q} = \nu_1$$

- $\mathscr{H}_{\text{path}}$ is the relative entropy for laws on the path space $C([0,1],M)$
- The **Schrödinger bridge** (SB) is the optimal solution of (SP)

Schrödinger bridge problem: dynamic formulation

We denote the law $\boldsymbol{\mu}^N$ by $\mathscr{P}^N$

Sanov'Theorem

$$\frac{1}{N} \log \mathscr{P}^N \left(\boldsymbol{\mu}^N = \mathbf{Q} \right) \asymp -\mathscr{H}^*(\mathbf{Q}|\mathbf{P})$$

Thus, the "most likely evolution" is found solving

Schrödinger Problem (SP)

$$\inf_{\mathbf{Q}} \mathscr{H}_{\text{path}}(\mathbf{Q}|\mathbf{P})$$

$$\mathbf{Q} \in \mathcal{P}(C([0,1],M)), \quad (X_0)_{\#}\mathbf{Q} = \nu_0, (X_1)_{\#}\mathbf{Q} = \nu_1$$

- $\mathscr{H}_{\text{path}}$ is the relative entropy for laws on the path space $C([0,1],M)$
- The **Schrödinger bridge** (SB) is the optimal solution of (SP)

Some notation and two classical results

- Marginal flow of SB **entropic interpolation** $\hookrightarrow (\mu_t)$
- The optimal value of SP is the **entropic cost**$\hookrightarrow \mathcal{T}_H(\nu_0, \nu_1)$

Theorem (fg-decompostion)

Under some mild regularity assumptions on M, ν_0, ν_1 *there exist non-negative functions* f_t, g_t *such*

$$\forall t \in [0,1], \quad \mu_t = f_t g_t$$

f_t,g_t *solve the equation*

$$\partial_t f_t = \frac{1}{2}\Delta f_t, \quad \partial_t g_t = -\frac{1}{2}\Delta g_t$$

Theorem (Vague statement)

In the ***small noise regime*** *(SP)* Γ*-converges to the Monge-Kantorovich problem.*

Some notation and two classical results

- Marginal flow of SB **entropic interpolation** $\hookrightarrow (\mu_t)$
- The optimal value of SP is the **entropic cost** $\hookrightarrow \mathcal{T}_H(\nu_0, \nu_1)$

Theorem (fg-decompostion)

Under some mild regularity assumptions on M, ν_0, ν_1 *there exist non-negative functions* f_t, g_t *such*

$$\forall t \in [0,1], \quad \mu_t = f_t g_t$$

f_t,g_t *solve the equation*

$$\partial_t f_t = \frac{1}{2}\Delta f_t, \quad \partial_t g_t = -\frac{1}{2}\Delta g_t$$

Theorem (Vague statement)

*In the **small noise regime** (SP)* Γ*-converges to the Monge-Kantorovich problem.*

What about Machine Learning?

The use of **Sinkhorn's algorithm** to compute (approximate) solutions of OT has led to a dramatic reduction in the computational cost, $O(d^2)$ vs. $O(d^3 \log d)$.

- Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport.
  In *Advances in neural information processing systems*, pages 2292–2300, 2013
- The regularized problem solved using Sinkhorn is a discrete Schrödinger problem!
- The reason why we can use Sinkhorn's algorithm is the fg decomposition Theorem

What about Machine Learning?

The use of **Sinkhorn's algorithm** to compute (approximate) solutions of OT has led to a dramatic reduction in the computational cost, $O(d^2)$ vs. $O(d^3 \log d)$.

- Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport.
  In *Advances in neural information processing systems*, pages 2292–2300, 2013

- The regularized problem solved using Sinkhorn is a discrete Schrödinger problem!
- The reason why we can use Sinkhorn's algorithm is the fg decomposition Theorem

Some references(incomplete list)

- E. Schrödinger. La théorie relativiste de l'électron et l' interprétation de la mécanique quantique.
  Ann. Inst Henri Poincaré, (2):269 – 310, 1932
- C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport.
  Discrete and Continuous Dynamical Systems, 34(4):1533–1574, 2014
- H. Föllmer. Random fields and diffusion processes.
  In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pages 101–203. Springer, 1988
- C. Léonard. From the Schrödinger problem to the Monge–Kantorovich problem.
  Journal of Functional Analysis, 262(4):1879–1920, 2012
- T. Mikami. Monges problem with a quadratic cost by the zero-noise limit of h-path processes.
  Probability Theory and Related Fields, 129(2):245–260, 2004

Motivating question

"What is the shape of the particle cloud at $t = \frac{1}{2}$*?"*

- **Entropy minimization** $\rightsquigarrow$ particles try to arrange according to the **equilibrium** configuration $\mathbf{m}$.
- **Prescribed initial and final configurations** $\rightsquigarrow$ particles are forced into a configuration far from equilibrium at $t = 0, 1$.

"Does $\mu_{1/2}$ *look like* $\mathbf{m}$*?"*

$$\Updownarrow$$

" How big is $\mathscr{H}_{\mathrm{marg}}(\mu_{1/2}|\mathbf{m})$*?"*

The key to answer the question is to view the entropic interpolation (μ_t) as a curve in a Riemannian manifold.

Motivating question

"What is the shape of the particle cloud at $t = \frac{1}{2}$*?"*

- **Entropy minimization** $\rightsquigarrow$ particles try to arrange according to the **equilibrium** configuration $\mathbf{m}$.
- **Prescribed initial and final configurations** $\rightsquigarrow$ particles are forced into a configuration far from equilibrium at $t = 0, 1$.

"Does $\mu_{1/2}$ *look like* $\mathbf{m}$*?"*

$$\Updownarrow$$

" How big is $\mathscr{H}_{\text{marg}}(\mu_{1/2}|\mathbf{m}) := S(\mu_{1/2})$*?"*

The key to answer the question is to view the entropic interpolation (μ_t) as a curve in a Riemannian manifold.

Part II: Newton's law for the entropic interpolation

The "Otto metric"

Formally, it is the Riemannian metric on $\mathcal{P}_2(M)$ whose associated geodesic distance is the Wasserstein distance W_2.

- The **tangent space** at $\mu \in \mathcal{P}_2(M)$ is identified with the gradient vector fields

$$\mathbf{T}_\mu = \overline{\{\nabla\varphi, \varphi \in \mathcal{C}_c^\infty\}}^{L^2(\mu)}$$

- We define the Riemannian metric on it

"Riemannian metric" on $\mathbf{T}_\mu$

$$\langle \nabla\varphi, \nabla\psi \rangle_{\mathbf{T}_\mu} := \int_M \langle \nabla\varphi, \nabla\psi \rangle d\mu.$$

- The **velocity** of an absolutely continuous curve (μ_t) is given by

Continuity equation

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0, \quad v_t \in \mathbf{T}_{\mu_t}$$

The "Otto metric"

Formally, it is the Riemannian metric on $\mathcal{P}_2(M)$ whose associated geodesic distance is the Wasserstein distance W_2.

- The **tangent space** at $\mu \in \mathcal{P}_2(M)$ is identified with the gradient vector fields

$$\mathbf{T}_\mu = \overline{\{\nabla\varphi, \varphi \in \mathcal{C}_c^\infty\}}^{L^2(\mu)}$$

- We define the Riemannian metric on it

"Riemannian metric" on $\mathbf{T}_\mu$

$$\langle \nabla\varphi, \nabla\psi \rangle_{\mathbf{T}_\mu} := \int_M \langle \nabla\varphi, \nabla\psi \rangle d\mu.$$

- The **velocity** of an absolutely continuous curve (μ_t) is given by

Continuity equation

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0, \quad v_t \in \mathbf{T}_{\mu_t}$$

Construction of the covariant derivative

The **Benamou-Brenier formula** tells indeed that the geodesic distance for this Riemannian metric is the Wasserstein distance.

Displacement interpolations are geodesics

$$W_2^2(\nu_0, \nu_1) = \inf_{\substack{(\mu,\nu) \\ \mu_0=\nu_0, \mu_1=\nu_1}} \int_0^1 |v_t|^2_{\mathbf{T}_{\mu_t}} \mathrm{d}t,$$

In a Riemannian manifold, the acceleration of a curve is the **covariant derivative** of its velocity

Acceleration of a curve

$$\nabla^{W_2}_{v_t} v_t = \partial_t v_t + \frac{1}{2} \nabla \left(|v_t|^2 \right)$$

Construction of the covariant derivative

The **Benamou-Brenier formula** tells indeed that the geodesic distance for this Riemannian metric is the Wasserstein distance.

Displacement interpolations are geodesics

$$W_2^2(\nu_0, \nu_1) = \inf_{\substack{(\mu,v) \\ \mu_0=\nu_0, \mu_1=\nu_1}} \int_0^1 |v_t|^2_{\mathbf{T}_{\mu_t}} \mathrm{d}t,$$

In a Riemannian manifold, the acceleration of a curve is the **covariant derivative** of its velocity

Acceleration of a curve

$$\nabla^{W_2}_{v_t} v_t = \partial_t v_t + \frac{1}{2}\nabla\left(|v_t|^2\right)$$

The acceleration of a SB

"Particles move from configuration ν_0 to configuration ν_1 minimizing relative entropy"

- The natural way of doing it would be to follow the gradient flow

Gradient flow

$$v_t = -\frac{1}{2}\nabla^{W_2} S(\mu_t), \quad \mu_0 = \nu_0.$$

- If particles go along the gradient flow $\mu_1 \neq \nu_1$
- IDEA: Modify the gradient flow equation **as little as possible** in order to be able to impose the terminal condition

The acceleration of a SB

"Particles move from configuration ν_0 to configuration ν_1 minimizing relative entropy"

- The natural way of doing it would be to follow the gradient flow

Gradient flow

$$v_t = -\frac{1}{2}\nabla^{W_2} S(\mu_t), \quad \mu_0 = \nu_0.$$

- If particles go along the gradient flow $\mu_1 \neq \nu_1$
- IDEA: Modify the gradient flow equation **as little as possible** in order to be able to impose the terminal condition

The acceleration of a SB

"Particles move from configuration ν_0 to configuration ν_1 minimizing relative entropy"

- The natural way of doing it would be to follow the gradient flow

Gradient flow

$$v_t = -\frac{1}{2}\nabla^{W_2} S(\mu_t), \quad \mu_0 = \nu_0.$$

- If particles go along the gradient flow $\mu_1 \neq \nu_1$
- IDEA: Modify the gradient flow equation **as little as possible** in order to be able to impose the terminal condition

The acceleration of a SB

"Particles move from configuration ν_0 to configuration ν_1 minimizing relative entropy"

- The natural way of doing it would be to follow the gradient flow

Gradient flow

$$v_t = -\frac{1}{2}\nabla^{W_2} S(\mu_t), \quad \mu_0 = \nu_0.$$

- If particles go along the gradient flow $\mu_1 \neq \nu_1$
- IDEA: Modify the gradient flow equation **as little as possible** in order to be able to impose the terminal condition

Tweaking a gradient flow

Gradient flow in $\mathbb{R}^d$

$$\dot{x}_t = -\nabla S(x_t)$$

Second order equation for the gradient flow

$$\begin{aligned}\ddot{x}_t &= -\nabla^2 S(x_t)\cdot \dot{x}_t \\ &= \nabla^2 S(x_t)\cdot \nabla S(x_t) \\ &= \frac{1}{2}\nabla\left(|\nabla S(x_t)|^2\right)\end{aligned}$$

Back to the OT setting (S = Relative entropy)

The **Fisher information** $\mathcal{I}$ is the norm squared of the gradient of the entropy

$$\mathcal{I}(\mu) = \left|\nabla^{W_2} S(\cdot)\right|^2_{\mathbf{T}_\mu}$$

Thus, we have a candidate equation...

Tweaking a gradient flow

Gradient flow in $\mathbb{R}^d$

$$\dot{x}_t = -\nabla S(x_t)$$

Second order equation for the gradient flow

$$\begin{aligned}\ddot{x}_t &= -\nabla^2 S(x_t) \cdot \dot{x}_t \\ &= \nabla^2 S(x_t) \cdot \nabla S(x_t) \\ &= \frac{1}{2}\nabla\left(|\nabla S(x_t)|^2\right)\end{aligned}$$

Back to the OT setting (S = Relative entropy)

The **Fisher information** $\mathcal{I}$ is the norm squared of the gradient of the entropy

$$\mathcal{I}(\mu) = \left|\nabla^{W_2} S(\cdot)\right|^2_{\mathbf{T}_\mu}$$

Thus, we have a candidate equation...

Tweaking a gradient flow

Gradient flow in $\mathbb{R}^d$

$$\dot{x}_t = -\nabla S(x_t)$$

Second order equation for the gradient flow

$$\begin{aligned} \ddot{x}_t &= -\nabla^2 S(x_t) \cdot \dot{x}_t \\ &= \nabla^2 S(x_t) \cdot \nabla S(x_t) \\ &= \frac{1}{2} \nabla \left(|\nabla S(x_t)|^2 \right) \end{aligned}$$

Back to the OT setting (S = Relative entropy)

The **Fisher information** $\mathcal{I}$ is the norm squared of the gradient of the entropy

$$\mathcal{I}(\mu) = \left| \nabla^{W_2} S(\cdot) \right|^2_{\mathbf{T}_\mu}$$

Thus, we have a candidate equation...

Tweaking a gradient flow

Gradient flow in $\mathbb{R}^d$

$$\dot{x}_t = -\nabla S(x_t)$$

Second order equation for the gradient flow

$$\begin{aligned}
\ddot{x}_t &= -\nabla^2 S(x_t) \cdot \dot{x}_t \\
&= \nabla^2 S(x_t) \cdot \nabla S(x_t) \\
&= \frac{1}{2} \nabla \left(|\nabla S(x_t)|^2 \right)
\end{aligned}$$

Back to the OT setting (S = Relative entropy)

The **Fisher information $\mathcal{J}$** is the norm squared of the gradient of the entropy

$$\mathcal{J}(\mu) = \left| \nabla^{W_2} S(\cdot) \right|^2_{\mathbf{T}_\mu}$$

Thus, we have a candidate equation...

Second order equation for entropic interpolation

Theorem (C.'17)

Let (μ_t) *be the entropic interpolation between* ν_0 *and* ν_1 *and* (v_t) *its velocity field. Under suitable regularity assumptions* (μ_t) *solves the equation*

$$\nabla^{W_2}_{v_t} v_t = \frac{1}{8} \nabla^{W_2} \mathcal{I}(\mu_t)$$

- The equation answers in a precise way

  "What kind of 2nd order equation the bridge of a diffusion satisfies?"

  and thus gives grounding to the intuition that the Brownian bridge is the stochastic version of a geodesic.
- To prove a rigorous statement, we took advantage of Gigli's rigorous version of the Otto calculus

Second order equation for entropic interpolation

Theorem (C.'17)

Let (μ_t) be the entropic interpolation between ν_0 and ν_1 and (v_t) its velocity field. Under suitable regularity assumptions (μ_t) solves the equation

$$\nabla^{W_2}_{v_t} v_t = \frac{1}{8} \nabla^{W_2} \mathcal{I}(\mu_t)$$

- The equation answers in a precise way

  "What kind of 2nd order equation the bridge of a diffusion satisfies?"

  and thus gives grounding to the intuition that the Brownian bridge is the stochastic version of a geodesic.
- To prove a rigorous statement, we took advantage of Gigli's rigorous version of the Otto calculus

Proof sketch

It relies on the representation $\mu_t = f_t g_t$.

Lemma (Representation of the velocity field)

The velocity field of (μ_t) *is* $\frac{1}{2}\nabla(\log g_t - \log f_t)$.

$\log f_t$(resp $\log g_t$) solve the forward(backward) HJB equation

HJB

$$\partial_t \log f_t = \frac{1}{2}\Delta \log f_t + \frac{1}{2}|\nabla \log f_t|^2,$$
$$\partial_t \log g_t = -\frac{1}{2}\Delta \log g_t - \frac{1}{2}|\nabla \log g_t|^2$$

Gradient of the Fisher information

We have

$$\nabla^{W_2}\mathcal{I}(\mu) = -2\nabla\Delta \log \mu - \nabla|\nabla \log \mu|^2$$

Proof sketch

Recall that $\nabla^{W_2}_{v_t} v_t = \partial_t v_t + \frac{1}{2}\nabla\left(|v_t|^2\right)$. We have, using HJB

$$
\begin{aligned}
\partial_t v_t &= -\frac{1}{2}\nabla\partial_t \log f_t + \frac{1}{2}\nabla\partial_t \log g_t \\
&\overset{\text{HJB}}{=} -\frac{1}{4}\nabla(\Delta \log f_t + \Delta \log g_t) - \frac{1}{4}[|\nabla \log f_t|^2 + |\nabla \log g_t|^2] \\
&\overset{\mu_t = f_t g_t}{=} -\frac{1}{4}\nabla\Delta \log \mu_t - \frac{1}{4}\nabla[|\nabla \log f_t|^2 + |\nabla \log g_t|^2] \\
&\overset{\text{polarization}}{=} -\frac{1}{4}\nabla\Delta \log \mu_t - \frac{1}{8}\nabla|\nabla \log f_t + \nabla \log g_t|^2 \\
&\quad -\frac{1}{8}\nabla|\nabla \log g_t - \nabla \log f_t|^2 \\
&= -\frac{1}{8}[2\nabla\Delta \log \mu_t - \nabla|\nabla \log \mu_t|^2] - \frac{1}{2}\nabla|v_t|^2 \\
&= \frac{1}{8}\nabla^{W_2}\mathcal{I}(\mu_t) - \frac{1}{2}|v_t|^2
\end{aligned}
$$

which (formally) concludes the proof.

Proof sketch

Recall that $\nabla^{W_2}_{v_t} v_t = \partial_t v_t + \frac{1}{2}\nabla\left(|v_t|^2\right)$. We have, using HJB

$$
\begin{aligned}
\partial_t v_t &= -\frac{1}{2}\nabla\partial_t \log f_t + \frac{1}{2}\nabla\partial_t \log g_t \\
&\overset{\text{HJB}}{=} -\frac{1}{4}\nabla(\Delta \log f_t + \Delta \log g_t) - \frac{1}{4}[|\nabla \log f_t|^2 + |\nabla \log g_t|^2] \\
&\overset{\mu_t = f_t g_t}{=} -\frac{1}{4}\nabla\Delta \log \mu_t - \frac{1}{4}\nabla[|\nabla \log f_t|^2 + |\nabla \log g_t|^2] \\
&\overset{\text{polarization}}{=} -\frac{1}{4}\nabla\Delta \log \mu_t - \frac{1}{8}\nabla|\nabla \log f_t + \nabla \log g_t|^2 \\
&\quad -\frac{1}{8}\nabla|\nabla \log g_t - \nabla \log f_t|^2 \\
&= -\frac{1}{8}[2\nabla\Delta \log \mu_t - \nabla|\nabla \log \mu_t|^2] - \frac{1}{2}\nabla|v_t|^2 \\
&= \frac{1}{8}\nabla^{W_2}\mathcal{I}(\mu_t) - \frac{1}{2}|v_t|^2
\end{aligned}
$$

which (formally) concludes the proof.

Proof sketch

Recall that $\nabla^{W_2}_{v_t} v_t = \partial_t v_t + \frac{1}{2}\nabla\left(|v_t|^2\right)$. We have, using HJB

$$
\begin{aligned}
\partial_t v_t &= -\frac{1}{2}\nabla\partial_t \log f_t + \frac{1}{2}\nabla\partial_t \log g_t \\
&\overset{\text{HJB}}{=} -\frac{1}{4}\nabla(\Delta\log f_t + \Delta\log g_t) - \frac{1}{4}[|\nabla\log f_t|^2 + |\nabla\log g_t|^2] \\
&\overset{\mu_t = f_t g_t}{=} -\frac{1}{4}\nabla\Delta\log\mu_t - \frac{1}{4}\nabla[|\nabla\log f_t|^2 + |\nabla\log g_t|^2] \\
&\overset{\text{polarization}}{=} -\frac{1}{4}\nabla\Delta\log\mu_t - \frac{1}{8}\nabla|\nabla\log f_t + \nabla\log g_t|^2 \\
&\quad -\frac{1}{8}\nabla|\nabla\log g_t - \nabla\log f_t|^2 \\
&= -\frac{1}{8}[2\nabla\Delta\log\mu_t - \nabla|\nabla\log\mu_t|^2] - \frac{1}{2}\nabla|v_t|^2 \\
&= \frac{1}{8}\nabla^{W_2}\mathcal{I}(\mu_t) - \frac{1}{2}|v_t|^2
\end{aligned}
$$

which (formally) concludes the proof.

Proof sketch

Recall that $\nabla^{W_2}_{v_t} v_t = \partial_t v_t + \frac{1}{2}\nabla\left(|v_t|^2\right)$. We have, using HJB

$$
\begin{aligned}
\partial_t v_t &= -\frac{1}{2}\nabla\partial_t \log f_t + \frac{1}{2}\nabla\partial_t \log g_t \\
&\overset{\text{HJB}}{=} -\frac{1}{4}\nabla(\Delta \log f_t + \Delta \log g_t) - \frac{1}{4}[|\nabla \log f_t|^2 + |\nabla \log g_t|^2] \\
&\overset{\mu_t = f_t g_t}{=} -\frac{1}{4}\nabla\Delta \log \mu_t - \frac{1}{4}\nabla[|\nabla \log f_t|^2 + |\nabla \log g_t|^2] \\
&\overset{\text{polarization}}{=} -\frac{1}{4}\nabla\Delta \log \mu_t - \frac{1}{8}\nabla|\nabla \log f_t + \nabla \log g_t|^2 \\
&\quad -\frac{1}{8}\nabla|\nabla \log g_t - \nabla \log f_t|^2 \\
&= -\frac{1}{8}[2\nabla\Delta \log \mu_t - \nabla|\nabla \log \mu_t|^2] - \frac{1}{2}\nabla|v_t|^2 \\
&= \frac{1}{8}\nabla^{W_2}\mathcal{I}(\mu_t) - \frac{1}{2}|v_t|^2
\end{aligned}
$$

which (formally) concludes the proof.

Proof sketch

Recall that $\nabla^{W_2}_{v_t} v_t = \partial_t v_t + \frac{1}{2}\nabla\left(|v_t|^2\right)$. We have, using HJB

$$
\begin{aligned}
\partial_t v_t \quad &= \quad -\frac{1}{2}\nabla\partial_t \log f_t + \frac{1}{2}\nabla\partial_t \log g_t \\
&\overset{\text{HJB}}{=} -\frac{1}{4}\nabla(\Delta \log f_t + \Delta \log g_t) - \frac{1}{4}[|\nabla \log f_t|^2 + |\nabla \log g_t|^2] \\
&\overset{\mu_t = f_t g_t}{=} -\frac{1}{4}\nabla\Delta \log \mu_t - \frac{1}{4}\nabla[|\nabla \log f_t|^2 + |\nabla \log g_t|^2] \\
&\overset{\text{polarization}}{=} -\frac{1}{4}\nabla\Delta \log \mu_t - \frac{1}{8}\nabla|\nabla \log f_t + \nabla \log g_t|^2 \\
&\qquad -\frac{1}{8}\nabla|\nabla \log g_t - \nabla \log f_t|^2 \\
&= \quad -\frac{1}{8}[2\nabla\Delta \log \mu_t - \nabla|\nabla \log \mu_t|^2] - \frac{1}{2}\nabla|v_t|^2 \\
&= \quad \frac{1}{8}\nabla^{W_2}\mathcal{I}(\mu_t) - \frac{1}{2}|v_t|^2
\end{aligned}
$$

which (formally) concludes the proof.

Proof sketch

Recall that $\nabla^{W_2}_{v_t} v_t = \partial_t v_t + \frac{1}{2}\nabla\left(|v_t|^2\right)$. We have, using HJB

$$
\begin{aligned}
\partial_t v_t &= -\frac{1}{2}\nabla\partial_t \log f_t + \frac{1}{2}\nabla\partial_t \log g_t \\
&\overset{\text{HJB}}{=} -\frac{1}{4}\nabla(\Delta \log f_t + \Delta \log g_t) - \frac{1}{4}[|\nabla \log f_t|^2 + |\nabla \log g_t|^2] \\
&\overset{\mu_t = f_t g_t}{=} -\frac{1}{4}\nabla\Delta \log \mu_t - \frac{1}{4}\nabla[|\nabla \log f_t|^2 + |\nabla \log g_t|^2] \\
&\overset{\text{polarization}}{=} -\frac{1}{4}\nabla\Delta \log \mu_t - \frac{1}{8}\nabla|\nabla \log f_t + \nabla \log g_t|^2 \\
&\quad -\frac{1}{8}\nabla|\nabla \log g_t - \nabla \log f_t|^2 \\
&= -\frac{1}{8}[2\nabla\Delta \log \mu_t - \nabla|\nabla \log \mu_t|^2] - \frac{1}{2}\nabla|v_t|^2 \\
&= \frac{1}{8}\nabla^{W_2}\mathcal{I}(\mu_t) - \frac{1}{2}|v_t|^2
\end{aligned}
$$

which (formally) concludes the proof.

Reference list (incomplete)

- Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, 2001
- N. Gigli. *Second Order Analysis on (P2(M),W2).* Memoirs of the American Mathematical Society
- Max-K von Renesse. An Optimal Transport view of Schrödinger's equation. *Canadian mathematical bulletin*, 55(4):858–869, 2012
- E. Nelson. *Dynamical theories of Brownian motion, volume 2.* Princeton university press Princeton, 1967

Part III: The entropy along entropic interpolations

First and second derivative of the entropy

We want to study how does the particle configuration evolves

"How much does $\mu_{1/2}$ *look like* $\mathbf{m}$*?"*

The relative entropy can be decomposed into

$$S(\mu_t) = \int_M \log f_t d\mu_t + \int_M \log g_t d\mu_t := \overrightarrow{h}(t) + \overleftarrow{h}(t)$$

$\overrightarrow{h}(t)$ is the **forward entropy**, $\overleftarrow{h}(t)$ the **backward entropy**.

First derivative -forward entropy

We have

$$\partial_t \overrightarrow{h}(t) = -\frac{1}{2}\left|v_t - \frac{1}{2}\nabla^{W_2} S\right|^2_{\mathbf{T}_{\mu_t}}$$

Second derivative-forward entropy

$$\partial_{tt} \overrightarrow{h}(t) = \frac{1}{2}\left\langle \nabla^{W_2}_{\xi_t} \nabla^{W_2} S, \xi_t \right\rangle_{\mathbf{T}_{\mu_t}}$$

with $\xi_t := \frac{1}{2}\nabla^{W_2} S - v_t$

First and second derivative of the entropy

We want to study how does the particle configuration evolves

"How much does $\mu_{1/2}$ look like $\mathbf{m}$?"

The relative entropy can be decomposed into

$$S(\mu_t) = \int_M \log f_t d\mu_t + \int_M \log g_t d\mu_t := \overrightarrow{h}(t) + \overleftarrow{h}(t)$$

$\overrightarrow{h}(t)$ is the **forward entropy**, $\overleftarrow{h}(t)$ the **backward entropy**.

First derivative -forward entropy

We have

$$\partial_t \overrightarrow{h}(t) = -\frac{1}{2}\Big|v_t - \frac{1}{2}\nabla^{W_2} S\Big|^2_{\mathbf{T}_{\mu_t}}$$

Second derivative-forward entropy

$$\partial_{tt} \overrightarrow{h}(t) = \frac{1}{2}\Big\langle \nabla^{W_2}_{\xi_t} \nabla^{W_2} S, \xi_t \Big\rangle_{\mathbf{T}_{\mu_t}}$$

with $\xi_t := \frac{1}{2}\nabla^{W_2} S - v_t$

Curvature enters the game

- Assume now that the **Ricci curvature** is bounded below, i.e. $\mathrm{Ric}_x(w,w) \geqslant \lambda |w|^2$ uniformly in $x, w \in T_x M$.
- A fundamental result of OT is that S is a λ-**convex** functional.

Differential inequality-forward entropy

$$\begin{aligned} \partial_{tt} \overrightarrow{h}(t) &= \frac{1}{2}\left\langle \nabla^{W_2}_{\xi_t} \nabla^{W_2} S, \xi_t \right\rangle_{\mathbf{T}_{\mu_t}} \\ &\geqslant \frac{\lambda}{2} |\xi_t|^2_{\mathbf{T}_{\mu_t}} \\ &= \frac{\lambda}{2} \left| \frac{1}{2} \nabla^{W_2} S - \nu_t \right|^2_{\mathbf{T}_{\mu_t}} \\ &= -\lambda \, \partial_t \overrightarrow{h}(t) \end{aligned}$$

The entropy along entropic interpolations

Theorem (C. '17)

Let M *be a compact manifold with Ricci curvature bounded below*

$$\forall x \in M, w \in T_x M, \quad \mathrm{Ric}_x(w,w) \geqslant \lambda |w|^2$$

Then, for all ν_0, ν_1 *and* $t \in [0,1]$ *the entropic interpolation* (μ_t) *satisfies:*

$$S(\mu_t) \leqslant \frac{1-\exp(-\lambda(1-t))}{1-\exp(-\lambda)} S(\nu_0) + \frac{1-\exp(-\lambda t)}{1-\exp(-\lambda)} S(\nu_1) \\ - \frac{\cosh(\frac{\lambda}{2}) - \cosh(-\lambda(t-\frac{1}{2}))}{\sinh(\frac{\lambda}{2})} \mathcal{T}_H(\nu_0,\nu_1).$$

About the entropy bound

- If M has a Ricci curvature bound, then the particle configuration at $t = \frac{1}{2}$ is very close to the equilibrium measure $\mathfrak{m}$, and we have a way to quantify this.
- In the small noise regime, the entropy estimate becomes the well known **convexity of the entropy along entropic interpolations.**
- There is a version of the Theorem when $\mathbf{P}$ is the Langevin dynamics

About the entropic transportation cost

- The entropic cost $\mathcal{T}_H(\mu, \mathbf{m})$ measures how difficult it is to **steer Brownian particles** which start "**out of equilibrium**" into the equilibrium configuration $\mathbf{m}$ in one unit of time.
- We expect that the more μ looks like $\mathbf{m}$, the smaller is $\mathcal{T}_H$

"How to bound $\mathcal{T}_H$? And with what?"

Theorem (C.'17)

Assume $\mathrm{Ric} \geqslant \lambda$. *Then for all* μ *we have*

$$\mathcal{T}_H(\mu, \mathbf{m}) \leqslant \frac{1}{1 - \exp(-\lambda)} \mathcal{S}(\mu)$$

*We call this an **entropy-entropy** inequality.*

About the entropic transportation cost

- The entropic cost $\mathcal{T}_H(\mu, \mathbf{m})$ measures how difficult it is to **steer Brownian particles** which start "**out of equilibrium**" into the equilibrium configuration $\mathbf{m}$ in one unit of time.
- We expect that the more μ looks like $\mathbf{m}$, the smaller is $\mathcal{T}_H$

"How to bound $\mathcal{T}_H$? And with what?"

Theorem (C.'17)

Assume $\mathrm{Ric} \geqslant \lambda$. *Then for all* μ *we have*

$$\mathcal{T}_H(\mu, \mathbf{m}) \leqslant \frac{1}{1 - \exp(-\lambda)} \mathcal{S}(\mu)$$

*We call this an **entropy-entropy** inequality.*

About the entropic transportation cost

- The entropic cost $\mathcal{T}_H(\mu, \mathbf{m})$ measures how difficult it is to **steer Brownian particles** which start "**out of equilibrium**" into the equilibrium configuration $\mathbf{m}$ in one unit of time.
- We expect that the more μ looks like $\mathbf{m}$, the smaller is $\mathcal{T}_H$

"How to bound $\mathcal{T}_H$? And with what?"

Theorem (C.'17)

Assume $\mathrm{Ric} \geqslant \lambda$. *Then for all* μ *we have*

$$\mathcal{T}_H(\mu, \mathbf{m}) \leqslant \frac{1}{1 - \exp(-\lambda)} \mathcal{S}(\mu)$$

*We call this an **entropy-entropy** inequality.*

About the entropy-entropy inequality

- The bound is useful since $\mathcal{T}_H(\mu, \mathbf{m})$ is hard to compute and $\mathcal{S}(\mu)$ is easy to compute (it is just an integral).
- In the **small noise regime** the entropy-entropy inequality becomes

Talagrand's transportation entropy inequality

$$W_2^2(\mu, \mathbf{m}) \leqslant 2\lambda \mathcal{S}(\mu)$$

- The inequality implies **concentration of measure** properties for $\mathbf{m}$ (work in progress)
- It allows to bound a **joint** entropy, $\mathcal{T}_H(\mu, \mathbf{m})$ with a **marginal** entropy ($\mathcal{S}(\mu)$)!

About the entropy-entropy inequality

- The bound is useful since $\mathcal{T}_H(\mu, \mathbf{m})$ is hard to compute and $\mathcal{S}(\mu)$ is easy to compute (it is just an integral).
- In the **small noise regime** the entropy-entropy inequality becomes

Talagrand's transportation entropy inequality

$$W_2^2(\mu, \mathbf{m}) \leqslant 2\lambda \mathcal{S}(\mu)$$

- The inequality implies **concentration of measure** properties for $\mathbf{m}$ (work in progress)
- It allows to bound a **joint** entropy, $\mathcal{T}_H(\mu, \mathbf{m})$ with a **marginal** entropy ($\mathcal{S}(\mu)$)!

Dual form of the entropy-entropy inequality

- Under the curvature condtion, it is known that the heat semigroup $(P_t)_{t \geqslant 0}$ is **hypercontractive**.
- For all $p, q \geqslant 1$ s.t. $\frac{q-1}{p-1} = \exp(2\lambda t)$ we have

$$\forall f \;\; s.t. \int f dm = 0, \quad \|P_t f\|_{L^q(m)} \leqslant \|f\|_{L^p(m)}$$

- It is known that **hypercontractivity** is equivalent to the **Logarithmic Sobolev inequality**.

Theorem (C.'18)

The following are equivalent

i) *The entropy entropy inequality with constant* $1/(1-\exp(-\lambda))$.

ii) *For all* $p \in (0,1), q < 1$ *s.t.* $\frac{q-1}{p-1} = \exp(2\lambda t)$, *and for all* f *s.t.* $\int f dm = 0$,

$$\|P_t f\|_{L^q(m)} \leqslant \|f\|_{L^p(m)}$$

References list(incomplete)

- Felix Otto and Cédric Villani. Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality.
  Journal of Functional Analysis, 173(2):361–400, 2000
- G. Conforti and M. Von Renesse. Couplings,gradient estimates and logarithmic Sobolev inequality for Langevin bridges.
  to appear in Probability Theory and Related Fields
- Christian Léonard. On the convexity of the entropy along entropic interpolations.
  arXiv preprint arXiv:1310.1274, 2013
- N. Gozlan, C. Roberto, P.M. Samson, and P. Tetali.

  Kantorovich duality for general transport costs and applications, to appear in j. funct. anal., preprint (2014)

Some thoughts for the future

- Is the entropy bound equivalent to curvature even if we do not look at the small noise regime?
- How close is the **entropic** interpolation to the **displacement** interpolation?
- How to construct a **Schrödinger bridge** for a system of **weakly interacting** particles system? Is there a **Netwon's law**?

Some thoughts for the future

- Is the entropy bound equivalent to curvature even if we do not look at the small noise regime?
- How close is the **entropic** interpolation to the **displacement** interpolation?
- How to construct a **Schrödinger bridge** for a system of **weakly interacting** particles system? Is there a **Netwon's law**?

Thank you very much!