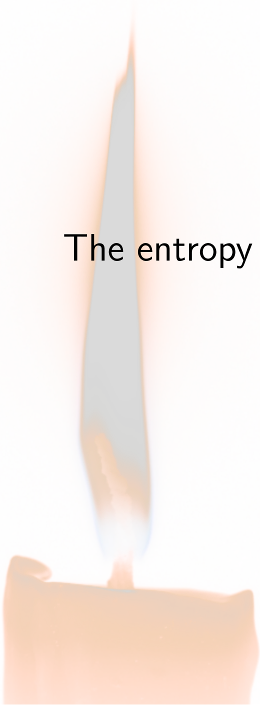
The Fokker-Planck equation as a gradient flow with respect to the Wasserstein metric

Theo Koppenhöfer

December 16, 2024



The entropy of a closed system never decreases

2nd law of Thermodynamics (physics convention)

The heatflow seeks to maximise entropy
(with respect to the Wasserstein metric)

Paraphrased summary of (Jordan, Kinderlehrer, and Otto, 1998)

Overview

The Fokker-Planck equation

Let $X = \mathbb{R}^d$ and $T = \mathbb{R}_{\geq 0}$. The *Fokker-Planck equation* is given by

$$\partial_t \rho = \operatorname{div}(\nabla\Psi\,\rho) + \frac{1}{\beta}\Delta\rho \qquad \text{on } X \times T$$
$$\rho(\cdot,0) = \rho^0 \qquad \text{on } X$$

with

- a.e. $\rho\colon X \times T \to \mathbb{R}_{\geq 0}$, s.t. ρ is a probability density at a.e. time
- smooth potential $\Psi\colon X \to \mathbb{R}_{\geq 0}$, s.t. $|\nabla\psi| \lesssim \Psi + 1$ on X
- parameter $\beta > 0$. As in (Jordan, Kinderlehrer, and Otto, 1998) set $\beta = 1$.
- initial probability density $\rho^0\colon X \to \mathbb{R}_{\geq 0}$

Wasserstein metric

Definition (Wasserstein metric)

Let μ_i be probability measures on X such that

$$M(\mu_i) = \int_X |x|^2 \,\mathrm{d}\mu_i < \infty \qquad \text{(finite second moments)}$$

Define the *set of transport plans* $\Pi(\mu_1, \mu_2)$ to be the probability measures p on $X \times X$ such that $p(\cdot \times X) = \mu_1$ and $p(X \times \cdot) = \mu_2$ on Borel sets. The *Wasserstein metric* is given by

$$d^2(\mu_1, \mu_2) := \inf_{p \in \Pi(\mu_1, \mu_2)} \int_{X^2} |x_1 - x_2|^2 \,\mathrm{d}p(x)$$

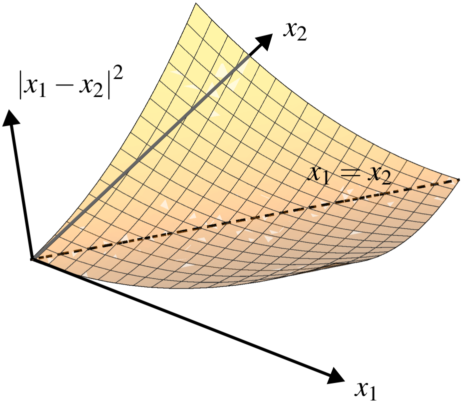


Figure: Motivation of the Wasserstein metric

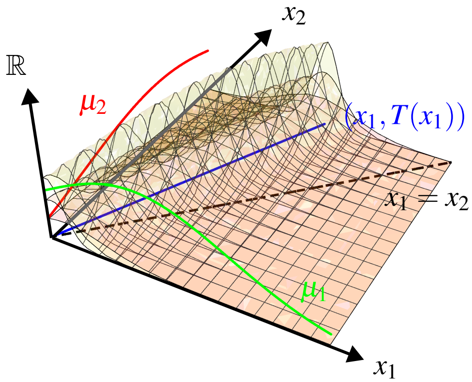


Figure: Motivation of the Wasserstein metric with μ_i gaussian densities

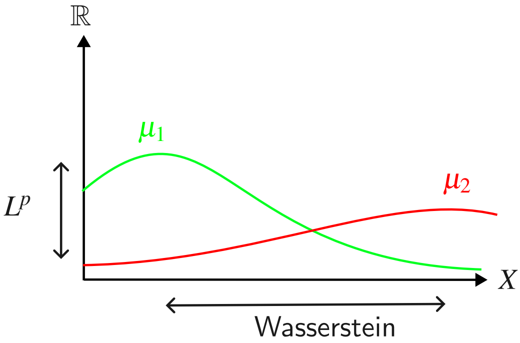


Figure: Difference between Wasserstein and L^p distance

Discretisation in time

Define the *(Helmholtz) free energy* as

internal/potential energy

Gibbs-Boltzmann entropy

$$F(\rho) := \underbrace{\int_X \Psi\rho}_{=:E(\rho)} + \underbrace{\int_X \rho\log(\rho)}_{=:S(\rho)}.$$

Let K denote the set of probability densities ρ on X, s.t.

$$M(\rho) = \int_X |x|^2 \,\mathrm{d}\rho < \infty.$$

Given ρ^{k-1} and $\tau > 0$ define the next iterate ρ^k as the minimiser over K of

$$\rho \mapsto \frac{1}{2}d^2\left(\rho^{k-1},\rho\right) + \tau F(\rho).$$

Well-posedness

Proposition (Well-posedness, (Jordan, Kinderlehrer, and Otto, 1998, Proposition 4.1))

Let $\rho_{k-1} \in K$ *then there exists a unique minimiser over* K *of*

$$\rho \mapsto \frac{1}{2} d^2\left(\rho^{k-1}, \rho\right) + \tau F(\rho)\,.$$

Interpolation in time

Define the step-wise interpolation

$$\rho_\tau(t) := \sum_k \rho^k \mathbb{1}_{\tau[k,k+1)}(t)\,.$$

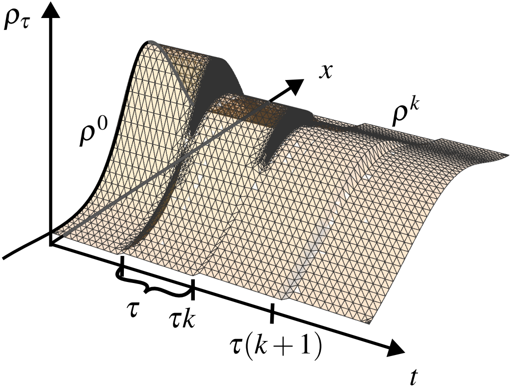


Figure: An example for ρ_τ.

Proposition (A priori estimates, (Jordan, Kinderlehrer, and Otto, 1998, Theorem 5.1))

Let $\rho^0 \in K$ *with* $F(\rho^0) < \infty$. *Fix an end time* $t_1 \in T$. *For all* $k_1 \tau \leq t_1$ *and* $t \leq t_1$ *one has*

$$M(\rho_\tau)\big|_t = \int_X |x|^2 \rho_\tau\big|_t \lesssim 1 \qquad \int_X (\rho_\tau \log(\rho_\tau))^+\big|_t \lesssim 1$$

$$E(\rho_\tau)\big|_t = \int_X \Psi \rho_\tau\big|_t \lesssim 1 \qquad \sum_{k \leq t_1/\tau} d\big(\rho_\tau^{k-1}, \rho_\tau^k\big)^2 \lesssim \tau\,.$$

where $f^+ := \max(f, 0)$.

Corollary (Weak convergence, (Jordan, Kinderlehrer, and Otto, 1998, Theorem 5.1))

Let $\rho^0 \in K$ *with* $F(\rho^0) < \infty$. *Then for a subsequence* $\rho_\tau \rightharpoonup \rho$ *weakly in* $L^1((0,t_1) \times X)$ *for all finite end times* t_1. *Additionally* $\rho \in K$ *for a.e. time and* $M(\rho), E(\rho) \in L^\infty((0,t_1))$.

Taking the limit $\tau \to 0$

Proposition (ρ solves weak Fokker-Planck, (Jordan, Kinderlehrer, and Otto, 1998, Theorem 5.1))

Let $\rho_\tau \rightharpoonup \rho$ *weakly in* $L^1((0,t_1)\times X)$ *for all finite times* t_1*. Then* ρ *solves the weak Fokker-Planck equation.*

Classical arguments (e.g. convolution with a heat kernel) then yield that this solution is strong, smooth and unique. See Jordan, Kinderlehrer, and Otto, 1998, Theorem 5.1.

Proof.

The proof comes from (Jordan, Kinderlehrer, and Otto, 1998, Theorem 5.1). The main idea is to perturb around ρ^k. For this let $\Phi_s : X \to X$ be such that

$$\begin{aligned} \partial_s \Phi_s &= \nabla\zeta \\ \Phi_0 &= \mathrm{Id} \end{aligned} \qquad \text{(flux)}$$

for $\nabla\zeta \in C_0^\infty(X;X)$.
Define the push forward

$$\rho_s := J(\Phi_s)\,\rho^k \circ \Phi_s^{-1}$$

where $J(f) := \det \nabla f$ is chosen such that the pertubation remains in K.

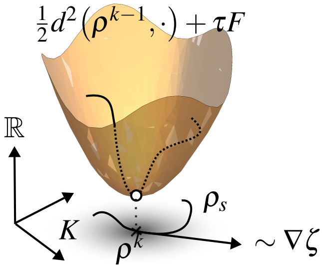


Figure: The situation at hand

Proof.

We perturb

$$0 \overset{\rho^k \text{ minimises } \frac{1}{2\tau}d^2(\rho^{k-1},\cdot)+F}{\leq} \frac{1}{s}\left(\left(\frac{1}{2\tau}d^2\left(\rho^{k-1},\rho_s\right)+F(\rho_s)\right)-\left(\frac{1}{2\tau}d^2\left(\rho^{k-1},\rho^k\right)+F\left(\rho^k\right)\right)\right)$$

$$\overset{F=E+S}{=} \underbrace{\frac{1}{2\tau s}\left(d^2\left(\rho^{k-1},\rho_s\right)-d^2\left(\rho^{k-1},\rho^k\right)\right)}_{=:(I)_s^k}+$$

$$+\underbrace{\frac{1}{s}\left(E(\rho_s)-E\left(\rho^k\right)\right)}_{=:(II)_s^k}+\underbrace{\frac{1}{s}\left(S(\rho_s)-S\left(\rho^k\right)\right)}_{=:(III)_s^k}$$

Proof.

We inspect the pertubation of the internal energy

$$
\begin{aligned}
(II)_s^k &= \frac{1}{s}\big(E(\rho_s) - E\big(\rho^k\big)\big) \\
&\qquad E(\rho) = \textstyle\int_X \Psi\rho \text{ and } \rho_s = J(\Phi_s)\rho^k \circ \Phi_s^{-1} \\
&= \frac{1}{s}\int_X \Psi\left(J(\Phi_s)\,\rho^k \circ \Phi_s^{-1} - \rho^k\right) \\
&\qquad \text{change of variables} \\
&= \int_X \frac{1}{s}(\Psi \circ \Phi_s - \Psi \circ \Phi_0)\rho^k \\
&\qquad \partial_s \Phi_s = \nabla \zeta \\
&\xrightarrow{s \to 0} \int_X \nabla\Psi \cdot \nabla\zeta\,\rho^k
\end{aligned}
$$

Proof.

The pertubation of the Gibbs-Boltzman entropy becomes

$$
\begin{aligned}
(III)_s^k &= \frac{1}{s}\big(S(\rho_s) - S\big(\rho^k\big)\big) \\
&\overset{S(\rho)=\int_X \rho\log(\rho) \text{ and } \rho_s = J(\Phi_s)\rho^k\circ\Phi_s^{-1}}{=} \frac{1}{s}\int_X J(\Phi_s)\rho^k\circ\Phi_s^{-1}\log\big(J(\Phi_s)\rho^k\circ\Phi_s^{-1}\big) - \rho^k\log\big(\rho^k\big) \\
&\overset{\text{change of variables}}{=} \frac{1}{s}\int_X \rho^k\big(\log\big(J(\Phi_s)\rho^k\big) - \log\big(\rho^k\big)\big) \\
&= \int_X \rho^k \frac{1}{s}(\log(\det\nabla\Phi_s) - \log(\det\nabla\Phi_0)) \\
&\overset{\text{Jacobi's formula and } \partial_s\Phi_s = \nabla\zeta}{\xrightarrow{s\to 0}} \int_X \rho^k \operatorname{div}\nabla\zeta = \int_X \rho^k \Delta\zeta
\end{aligned}
$$

Proof.

For the pertubation of the Wasserstein term let p be such that

$$d^2\left(\rho^{k-1}, \rho^k\right) = \int_{X^2} |x_1 - x_2|^2 \,\mathrm{d}p(x)\,.$$

Define the push forward measure $p_s \in \Pi\left(\rho^{k-1}, \rho_s\right)$ through

$$\int_{X^2} f(x_1, x_2)\,\mathrm{d}p_s = \int_{X^2} f(x_1, \Phi_s(x_2))\,\mathrm{d}p$$

for all measurable $f\colon X^2 \to \mathbb{R}$.

Proof.

so we obtain

$$
\begin{aligned}
\tau(I)_s^k &= \frac{1}{2s}\left(d^2(\rho^{k-1},\rho_s) - d^2(\rho^{k-1},\rho^k)\right) \\
&\leq \frac{1}{2s}\left(\int_{X^2} |x_2 - x_1|^2 \,\mathrm{d}p_s - \int_{X^2} |x_2 - x_1|^2 \,\mathrm{d}p\right) \\
&= \frac{1}{2s}\int_{X^2}\left(|\Phi_s(x_2) - x_1|^2 - |\Phi_0(x_2) - x_1|^2\right)\mathrm{d}p \\
&\xrightarrow{s\to 0} \int_{X^2} (x_2 - x_1)\cdot\nabla\zeta \,\mathrm{d}p \\
&\overset{\text{Taylor}}{=} \int_{X^2} \zeta(x_1) - \zeta(x_2)\,\mathrm{d}p + \mathscr{O}\left(\frac{|\nabla^2\zeta|_{L^\infty}}{2}\int_{X^2}|x_2 - x_1|^2\,\mathrm{d}p\right) \\
&= \int_X \zeta(\rho^{k-1} - \rho^k) + \mathscr{O}\left(d^2(\rho^{k-1},\rho^k)\right)
\end{aligned}
$$

Proof.

Putting things together for finite time $T=[0,t_1]$

$$
\begin{aligned}
0 &\le \int_T \limsup_{s\to 0} \sum_k \Big((I)_s^{k+1} + (II)_s^{k+1} + (III)_s^{k+1} \Big) \mathbb{1}_{\tau[k,k+1)}(t) \\
&\le \int_{T\times X} \zeta \frac{\rho_\tau - \rho_\tau(\cdot + \tau)}{\tau} + \nabla\Psi\cdot\nabla\zeta\,\rho_\tau + \rho_\tau \Delta\zeta \\
&\qquad + |T|\,\mathscr{O}\left(\sum_{k\le t_1/\tau} d^2\left(\rho^{k-1},\rho^k\right) \right) \\
&= \int_{T\times X} \rho_\tau \left(\frac{\zeta - \zeta(\cdot - \tau)}{\tau} + \nabla\Psi\cdot\nabla\zeta + \Delta\zeta \right) - \int_X \rho^0 \zeta(0,\cdot) + \dots
\end{aligned}
$$

A priori estimate

$$
\xrightarrow{\tau\to 0} \int_{T\times X} \rho\left(\partial_t\zeta - \nabla\Psi\cdot\nabla\zeta + \Delta\zeta\right) - \int_X \rho^0\zeta(0,\cdot)
$$

for all $\zeta \in C_0^\infty(\mathbb{R}\times X)$. This means that ρ fulfills the weak Fokker-Planck equation. $\square$

Summary

We have (in some sense) $\rho_\tau \to \rho$ where

- ρ solves the Fokker-Planck equation $\partial_t \rho = \mathrm{div}(\nabla \Psi \rho) + \Delta \rho$
- $\rho_\tau(t) = \sum_k \rho^k \mathbb{1}_{\tau[k,k+1)}(t)$
- ρ^k minimises $\frac{1}{2} d^2(\rho^{k-1}, \cdot) + \tau F$ over K
- F is the (Helmholtz) free energy $\int_X \Psi \rho + \int_X \rho \log(\rho)$
- d is the Wasserstein metric given by
  $d^2(\mu_1, \mu_2) = \inf_{p \in \Pi(\mu_1, \mu_2)} \int_{X^2} |x_1 - x_2|^2$

Thank you for your attention.

Sources I

Ambrosio, Luigi, Nicola Gigli, and Giuseppe Savaré (2005). *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, pp. viii+333. ISBN: 978-3-7643-2428-5; 3-7643-2428-7.

Ball, John (2024). *Lecture notes: Gradient Flows*.

Jordan, Richard, David Kinderlehrer, and Felix Otto (1998). "The variational formulation of the Fokker-Planck equation". In: *SIAM J. Math. Anal.* 29.1, pp. 1–17. ISSN: 0036-1410,1095-7154. DOI: 10.1137/S0036141096303359. URL: https://doi.org/10.1137/S0036141096303359.

Santambrogio, Filippo (2017). "{Euclidean, metric, and Wasserstein} gradient flows: an overview". In: *Bull. Math. Sci.* 7.1, pp. 87–154. ISSN: 1664-3607,1664-3615. DOI: 10.1007/s13373-017-0101-1. URL: https://doi.org/10.1007/s13373-017-0101-1.

Image Credits I

Wikimedia Commons contributors (2006). *a candle (eine Kerze)*. [Accessed on 2024-12-15]. URL: https://upload.wikimedia.org/wikipedia/commons/4/4b/Candle.jpg.