Limit distributions in EOT when $\varepsilon \to 0$

Gilles Mordant

Institut für Mathematische Stochastik
Universität Göttingen

February 21, 2025

Regularised optimal transport (yes, that slide again!)

We consider two probability measures μ, ν on a compact subset of $\mathbb{R}^d$, a.c. with respect to the Lebesgue measure.

$$\mathrm{EOT}(\mu, \nu) := \inf_{\pi \in \Gamma(\mu,\nu)} \int \|x - y\|^2 \mathrm{d}\pi(x, y) + \varepsilon \, \mathrm{KL}(\pi | \mu \otimes \nu),$$

where $\Gamma(\mu, \nu)$ is the set of couplings of μ, ν.

The Monge-Ampère equation

In the unregularised case, the Brenier–McCann theorem states that $T = \nabla \phi$.

The Monge-Ampère equation

In the unregularised case, the Brenier–McCann theorem states that $T = \nabla\phi$.
The Monge–Ampère equation is

$$\det \nabla^2 \phi = \frac{f}{g \circ \nabla \phi}.$$

(Recall the change of variable formula)

Duality

$$\frac{\mathrm{d}\pi_\varepsilon}{\mathrm{d}(\mu \otimes \nu)}(x, y) = \exp\left(\frac{1}{\varepsilon}\left(f_\varepsilon(x) + g_\varepsilon(y) - \frac{1}{2}\|x - y\|^2\right)\right) \quad a.e.$$

where

$$f_\varepsilon(x) = -\varepsilon \log \int \exp\left(\frac{1}{\varepsilon}\left(g_\varepsilon(y) - \frac{1}{2}\|x - y\|^2\right)\right) \nu(y)\mathrm{d}y$$

$$g_\varepsilon(y) = -\varepsilon \log \int \exp\left(\frac{1}{\varepsilon}\left(f_\varepsilon(x) - \frac{1}{2}\|x - y\|^2\right)\right) \mu(x)\mathrm{d}x.$$

Maps from EOT

The *entropic map* between μ and ν is simply the barycentric projection of π_ε :

$$T_\varepsilon(x) := \int y \mathrm{d}\pi_\varepsilon^x(y) = \mathbb{E}_{\pi_\varepsilon}[Y \mid X = x] .$$

or

$$T_\varepsilon(x) := \frac{\int y e^{\frac{1}{\varepsilon}(g_\varepsilon(y) - \frac{1}{2}\|x-y\|^2)} \mathrm{d}\nu(y)}{\int e^{\frac{1}{\varepsilon}(g_\varepsilon(y) - \frac{1}{2}\|x-y\|^2)} \mathrm{d}\nu(y)} .$$

It can also be verified from these optimality conditions that that $T_\varepsilon = \mathrm{id} - \nabla f_\varepsilon$.

The one measure case

The dual conditions simplify to

$$\exp\left(-\frac{f_\varepsilon(x)}{\varepsilon}\right) = \int \exp\left(\frac{1}{\varepsilon}\left(f_\varepsilon(y) - \frac{1}{2}\|x-y\|^2\right)\right)\rho(y)\mathrm{d}y.$$

The one measure case

The dual conditions simplify to

$$\exp\left(-\frac{f_\varepsilon(x)}{\varepsilon}\right) = \int \exp\left(\frac{1}{\varepsilon}\left(f_\varepsilon(y) - \frac{1}{2}\|x-y\|^2\right)\right)\rho(y)\mathrm{d}y.$$

A fixed point equation!

The fixed point equation

Shuffling a bit, rewrite

$$u(x) = \int \frac{1}{(2\pi\varepsilon)^{d/2}} \exp\left(-\frac{\|x-y\|^2}{2\varepsilon}\right) \frac{\rho(y)}{u(y)} \mathrm{d}y.$$

The fixed point equation

Shuffling a bit, rewrite

$$u(x) = \int \frac{1}{(2\pi\varepsilon)^{d/2}} \exp\left(-\frac{\|x-y\|^2}{2\varepsilon}\right) \frac{\rho(y)}{u(y)} \mathrm{d}y.$$

So that

$$\begin{aligned} \mathcal{F}[u](\xi) &= \mathcal{F}\left[\frac{\rho}{u}\right](\xi) \times \exp\left(-\frac{\varepsilon\|\xi\|^2}{2}\right) \\ &= \mathcal{F}\left[\frac{\rho}{u}\right](\xi) \times \left(1 - \frac{\varepsilon\|\xi\|^2}{2} + \cdots\right). \end{aligned}$$

The fixed point equation

Shuffling a bit, rewrite

$$u(x) = \int \frac{1}{(2\pi\varepsilon)^{d/2}} \exp\left(-\frac{\|x-y\|^2}{2\varepsilon}\right) \frac{\rho(y)}{u(y)} \mathrm{d}y.$$

So that

$$\begin{aligned}\mathcal{F}[u](\xi) &= \mathcal{F}\left[\frac{\rho}{u}\right](\xi) \times \exp\left(-\frac{\varepsilon\|\xi\|^2}{2}\right) \\ &= \mathcal{F}\left[\frac{\rho}{u}\right](\xi) \times \left(1 - \frac{\varepsilon\|\xi\|^2}{2} + \cdots\right).\end{aligned}$$

$$\mathcal{F}[\Delta f](\xi) = -\|\xi\|^2\, \mathcal{F}[f](\xi) \qquad \textit{(Reminder)}$$

Expansion for the potentials as $\varepsilon \to 0$

Under suitable regularity assumptions, as $\varepsilon \to 0$,

$$\exp\left(-\frac{2f_\varepsilon(x)}{\varepsilon}\right) = \rho(x)(2\pi\varepsilon)^{d/2}\left(1 + \varepsilon\,\mathrm{tr}\left(\frac{\nabla^2\rho(x)}{4\rho(x)} + \frac{-1}{8}\left(\nabla \log \rho(x)\right)\left(\nabla \log \rho(x)\right)^\top\right) + \mathrm{o}(\varepsilon)\right).$$

Expansion for the potentials as $\varepsilon \to 0$

Under suitable regularity assumptions, as $\varepsilon \to 0$,

$$\exp\left(-\frac{2f_\varepsilon(x)}{\varepsilon}\right) = \rho(x)(2\pi\varepsilon)^{d/2}\left(1 + \varepsilon\,\mathrm{tr}\left(\frac{\nabla^2\rho(x)}{4\rho(x)} + \frac{-1}{8}\left(\nabla \log \rho(x)\right)\left(\nabla \log \rho(x)\right)^\top\right) + \mathrm{o}(\varepsilon)\right).$$

First takeaway:

$$-\frac{f_\varepsilon(x)}{\varepsilon} = \tfrac{1}{2}\log \rho(x) + C_\varepsilon + \mathrm{O}(\varepsilon).$$

Do we care?

Langevin equation

$$\mathrm{d}X_t = \nabla \log \rho(X_t)\ \mathrm{d}t + \sqrt{2}\ \mathrm{d}W_t.$$

Do we care?

Langevin equation

$$\mathrm{d}X_t = \nabla \log \rho(X_t)\ \mathrm{d}t + \sqrt{2}\ \mathrm{d}W_t.$$

Workhorse of denoising diffusion models. Need to estimate the score function in practice.

The main question

Given a sample $X_1, \ldots X_n \overset{i.i.d.}{\sim} \rho$, can I estimate $\nabla \log \rho$ with

$$-\frac{2\nabla \hat{f}_{\varepsilon_n}(x)}{\varepsilon_n},$$

choosing suitably $\varepsilon_n \to 0$ as $n \to \infty$?

Prerequisites

Introduce

$$h \mapsto K_\varepsilon[h](x) := \int h(y)\pi_\varepsilon(x,y)\mathrm{d}\rho(y).$$

Prerequisites

Introduce

$$h \mapsto K_\varepsilon[h](x) := \int h(y)\pi_\varepsilon(x,y)\mathrm{d}\rho(y).$$

(A) The density ρ is supported on a compact set K and C^2 on its domain. Furthermore, $\forall x \in K$, we have that $\ell \leq \rho(x) \leq L$, for $0 < \ell \leq L$.

(Reg) $$\left| f_\varepsilon(x) + \frac{\varepsilon}{2}\log\rho(x) + \frac{\varepsilon d}{4}\log(2\pi\varepsilon)\right| \leq \varepsilon^2 C_f \|x\|^2,$$

(Conv) $$\frac{1}{\varepsilon_n}\|\hat{f}_{\varepsilon_n} - f_{\varepsilon_n}\|_\infty \to 0,$$ (No worries, we'll chat about this one!)

with $\sqrt{n}\varepsilon_n^{d/4}/\sqrt{\log n} \to \infty$, $\varepsilon_n \to 0$ as $n \to \infty$.

First big result

Proposition

Let f_{ε_n} be the entropic self-potential for ρ satisfying **(A)** *and* **(Reg)***, and let $\hat{f}_{\varepsilon_n}$ be its empirical counterpart based on n i.i.d. samples from the distribution ρ. Assume* **(Conv)** *and fix $x \in \mathrm{int}(\mathrm{supp}(\rho))$. Choosing $\sqrt{n}\varepsilon_n^{d/4}/\sqrt{\log n} \to \infty$, as $n \to \infty$, there exists a sequence a_n with $a_n\varepsilon_n \to \infty$, as $n \to \infty$, such that*

$$a_n\big(\hat{f}_{\varepsilon_n}(x) - f_{\varepsilon_n}(x)\big) + a_n K_{\varepsilon_n}\Big[\big(\hat{f}_{\varepsilon_n} - f_{\varepsilon_n}\big)\big(1 + \mathrm{o}(1)\big)\Big](x)$$
$$= -\frac{a_n\varepsilon_n}{\sqrt{n}} \int \frac{\exp\left(-\frac{1}{2\varepsilon_n}\|y - x\|^2\right)}{(2\pi\varepsilon_n)^{d/2}\rho^{1/2}(y)\rho^{1/2}(x)(1 + \mathrm{o}(1))} \mathrm{d}\mathbb{G}_n(y) + \mathrm{o}_p(1),$$

A CLT

Theorem (Limit distribution for the empirical potentials)

Consider $X_1, \ldots, X_n \sim \rho$, as above, and suppose **(A)**, **(Reg)** *and* **(Conv)** *hold. Then, for any $x_1, \ldots, x_m \in \operatorname{int}(\operatorname{supp}(\rho))$ with $m \in \mathbb{N}$ fixed,*

$$\sqrt{n}\,\varepsilon_n^{-1+d/4} \begin{pmatrix} (\hat{f}_{\varepsilon_n}(x_1) - f_{\varepsilon_n}(x_1)) \\ \vdots \\ \hat{f}_{\varepsilon_n}(x_m) - f_{\varepsilon_n}(x_m) \end{pmatrix} \xrightarrow{D} \mathcal{N}\left(0_m, C_3 \operatorname{diag}\left(\rho(x_1)^{-1}, \ldots, \rho(x_m)^{-1}\right)\right),$$

as $n \to \infty$ provided that $\sqrt{n}\varepsilon_n^{d/4}/\sqrt{\log n} \to \infty$, where

$$C_3 := \sum_{0 \le \kappa, \kappa' \le \infty} 2^{-\kappa-\kappa'} \sum_{\eta=0}^{\kappa} \sum_{\eta'=0}^{\kappa'} \binom{\kappa}{\eta} (-1)^{\eta} (-1)^{\eta'} \binom{\kappa'}{\eta'} (\eta + \eta' + 2)^{-d/2}.$$

A word on the proof

$$a_n\left(\hat{f}_{\varepsilon_n}(x) - f_{\varepsilon_n}(x)\right) \asymp \frac{a_n \varepsilon_n}{\sqrt{n}} \int (\mathrm{id} + K_{\varepsilon_n})^{-1} \left[\frac{\exp\left(-\frac{1}{2\varepsilon_n}\|y - \cdot\|^2\right)}{(2\pi\varepsilon_n)^{d/2}\rho^{1/2}(y)\rho^{1/2}(\cdot)} \right](x)\mathrm{d}\mathbb{G}_n(y)$$

$$(\mathrm{id} + K_{\varepsilon_n})^{-1} = \left(2\,\mathrm{id} + \left(K_{\varepsilon_n} - \mathrm{id}\right)\right)^{-1} = \frac{1}{2}\sum_{\kappa=0}^{\infty}(\mathrm{id} - K_{\varepsilon_n})^{\kappa} 2^{-\kappa}$$

A word on the proof II

$(\mathrm{id} - K_{\varepsilon_n})^\kappa$ treated with the binomial formula.
Recall $K_\varepsilon[h](x) := \int h(y)\pi_\varepsilon(x,y)\mathrm{d}\rho(y)$.
Thus,

$$K_\varepsilon[h](x) = \int h(y)\frac{1}{(2\pi\varepsilon)^{d/2}\sqrt{\rho(x)\rho(y)}(1+\mathrm{o}(1))}\exp\left(-\frac{\|x-y\|^2}{2\varepsilon}\right)\rho(y)\mathrm{d}y$$
$$= \frac{1}{\sqrt{\rho(x)}}\int h(y)\frac{1}{(2\pi\varepsilon)^{d/2}(1+\mathrm{o}(1))}\exp\left(-\frac{\|x-y\|^2}{2\varepsilon}\right)\sqrt{\rho(y)}\mathrm{d}y.$$

The (Conv) assumption

Start from

$$\frac{\hat{f}(x)}{\varepsilon_n} = -\log \frac{1}{n} \sum_{i=1}^{n} \frac{\exp(-\frac{\|x-X_i\|^2}{2\varepsilon_n})}{(2\pi\varepsilon_n)^{d/2}} \exp\left(\frac{\hat{f}(X_i)}{\varepsilon_n}\right).$$

Now, let us replace $\hat{f}/\varepsilon_n$ by $-\log(\rho)/2$ in the identity above. This raises the question

$$\rho^{1/2}(x) \stackrel{?}{=} \frac{1}{n} \sum_{i=1}^{n} \frac{\exp(-\frac{\|x-X_i\|^2}{2\varepsilon_n})}{(2\pi\varepsilon_n)^{d/2}\rho^{1/2}(X_i)}.$$

The (Conv) assumption II

By Laplace's method

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\exp(-\frac{\|x-X_i\|^2}{2\varepsilon_n})}{(2\pi\varepsilon_n)^{d/2}\rho^{1/2}(X_i)}\right]=\rho^{1/2}(x)(1+\mathrm{o}(1)),\ \text{as } \varepsilon\to 0.$$

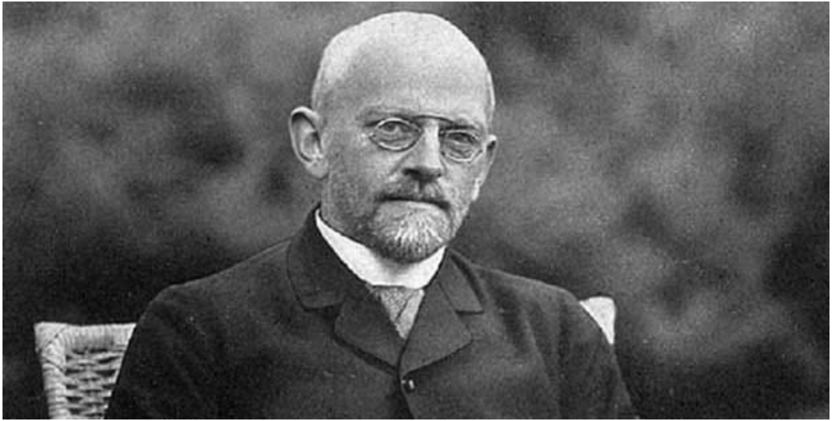
It further holds that

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{\exp(-\frac{\|x-X_i\|^2}{2\varepsilon_n})}{(2\pi\varepsilon_n)^{d/2}\rho^{1/2}(X_i)}-\mathbb{E}\left[\frac{\exp(-\frac{\|x-X\|^2}{\varepsilon_n})}{(2\pi\varepsilon_n)^{d/2}\rho^{1/2}(X)}\right]\right\|_\infty\to 0 \quad a.s.,$$

as $n\to\infty, \varepsilon_n\to 0$ and $\sqrt{n}\varepsilon_n^{d/4}/\sqrt{\log n}\to\infty$.

Hilbert’s mindset

“Wir müssen wissen, wir werden wissen”



Towards the two-measure case

If

$$a_n\big(\hat{f}_{\varepsilon_n}(x) - f_{\varepsilon_n}(x)\big) + K^{\nu}_{\varepsilon_n}\left[a_n\big(\hat{g}_{\varepsilon_n}(\cdot) - g_{\varepsilon_n}(\cdot)\big)\right](x) = -\frac{a_n\varepsilon_n}{\sqrt{n}}\int \pi_{\varepsilon_n}(x,y)\mathrm{d}\mathbb{G}^{\nu}_n(y) + \mathrm{o}_p(1),$$

then

$$\begin{aligned} & a_n\begin{pmatrix} \hat{f}_{\varepsilon_n} - f_{\varepsilon_n} \\ \hat{g}_{\varepsilon_n} - g_{\varepsilon_n} \end{pmatrix} \\ & = -\frac{a_n\varepsilon_n}{\sqrt{n}}\begin{pmatrix} (\mathrm{id} - K^{\nu}_{\varepsilon_n}K^{\mu}_{\varepsilon_n})^{-1} & -K^{\nu}_{\varepsilon_n}(\mathrm{id} - K^{\mu}_{\varepsilon_n}K^{\nu}_{\varepsilon_n})^{-1} \\ -(\mathrm{id} - K^{\mu}_{\varepsilon_n}K^{\nu}_{\varepsilon_n})^{-1}K^{\mu}_{\varepsilon_n} & (\mathrm{id} - K^{\mu}_{\varepsilon_n}K^{\nu}_{\varepsilon_n})^{-1} \end{pmatrix}\begin{pmatrix} \int \pi_{\varepsilon_n}(\cdot,y)\mathrm{d}\mathbb{G}^{\nu}_n(y) \\ \int \pi_{\varepsilon_n}(x,\cdot)\mathrm{d}\mathbb{G}^{\mu}_n(x) \end{pmatrix} + \mathrm{o}_p(1). \end{aligned}$$

This is similar to expansion in A. González-Sanz and S. Hundrieser' works.

Bregman divergence and Kim-McCann geometry

$$f_0(x) + g_0(y) - \frac{1}{2}\|x - y\|^2 =: \psi_0(x) + \phi_0(y) - \langle x, y \rangle =: D(x, y)$$

Assume

$$\left\| \phi_0(y) + \psi_0(x) - \langle x, y \rangle - \tfrac{1}{2}(y - x^*)^\top \nabla^2 \phi_0(x)(y - x^*) \right\| \leq C\|y - x^*\|^3.$$

Flavien Léger's expansions

We assume

$$\left| f_\varepsilon(x) - f_0(x) + \frac{\varepsilon}{2}\log\mu(x) + \frac{\varepsilon d}{4}\log(2\pi\varepsilon) \right| \le \varepsilon^2 C_f \|x\|^2$$

$$\left| g_\varepsilon(y) - g_0(y) + \frac{\varepsilon}{2}\log\nu(y) + \frac{\varepsilon d}{4}\log(2\pi\varepsilon) \right| \le \varepsilon^2 C_g \|y\|^2,$$

for all $\varepsilon \le \varepsilon_0$, where f_0, g_0 are the unregularized dual potentials.

About the why...

Start from the dual condition

$$1 = \int \pi_\varepsilon(x, y)\nu(y)\mathrm{d}y.$$

Then plugging-in the assumed expansion, we get

$$\begin{aligned}
1 &= \int \exp\left(\frac{1}{\varepsilon_n}\left[f_\varepsilon(x) + g_\varepsilon(y) - \tfrac{1}{2}\|x-y\|^2\right]\right)\nu(y)\mathrm{d}y \\
&= \int \frac{1}{(2\pi\varepsilon)^{d/2}\sqrt{\mu(x)\nu(y)}} \exp\left(\frac{1}{\varepsilon}\left[f_0(x) + g_0(y) - \tfrac{1}{2}\|x-y\|^2\right] + \mathrm{o}(1)\right)\nu(y)\mathrm{d}y \\
&= \int \frac{1}{(2\pi\varepsilon)^{d/2}\sqrt{\mu(x)\nu(y)}} \exp\left(-\frac{1}{2\varepsilon}\left[(y-x^*)^\top\nabla^2\phi_0(x)(y-x^*)\right] + \mathrm{o}(1)\right)\nu(y)\mathrm{d}y \\
&= \int \frac{\sqrt{\det[\nabla^2\varphi_0(x)]}}{(2\pi\varepsilon)^{d/2}} \exp\left(-\frac{1}{2\varepsilon}\left[(y-x^*)^\top\nabla^2\phi_0(x)(y-x^*)\right] + \mathrm{o}(1)\right)\frac{\nu^{1/2}(y)}{\sqrt{\nu(x^*)}}\mathrm{d}y.
\end{aligned}$$

Composition of Sinkhorn operators

Theorem

Under regularity assumptions, it holds that

$$
\begin{aligned}
h \mapsto\; & K^{\mu}_{\varepsilon_n}\left[K^{\nu}_{\varepsilon_n}[h]\right](y_2) \\
&= \frac{1+\mathrm{o}(1)}{\nu^{1/2}(y_2)} \int \frac{1}{(2\pi\varepsilon_n)^{d/2}}\, h(y) \sqrt{\frac{\det[\nabla^2\phi_0^*(y_2)]\,\det[\nabla^2\phi_0^*(y)]}{\det[\nabla^2\phi_0^*(y_2)+\nabla^2\phi_0^*(y)]}} \\
&\qquad \times \exp\left(-\frac{1}{4\varepsilon_n}(y-y_2)^\top\left[\nabla^2\phi_0^*(y_2)\right](y-y_2)+\mathrm{o}\left(\frac{\|y-y_2\|^2}{\varepsilon_n}\right)\right)\nu^{1/2}(y)\mathrm{d}y,
\end{aligned}
\tag{1}
$$

as $\varepsilon_n \to 0$.

Two-measure only one sample

$$a_n(\hat{f}_{\varepsilon_n}(x) - f_{\varepsilon_n}(x)) = -\frac{a_n \varepsilon_n}{\sqrt{n}} \int (\mathrm{id} - K^{\nu}_{\varepsilon_n} K^{\mu}_{\varepsilon_n})^{-1}[\pi_{\varepsilon_n}(x, y)] \mathrm{d}\mathbb{G}^{\mu}_n(y) + \mathrm{o}_p(1),$$

Two-measure only one sample

$$a_n(\hat{f}_{\varepsilon_n}(x) - f_{\varepsilon_n}(x)) = -\frac{a_n \varepsilon_n}{\sqrt{n}} \int (\mathrm{id} - K^{\nu}_{\varepsilon_n} K^{\mu}_{\varepsilon_n})^{-1} [\pi_{\varepsilon_n}(x,y)] \mathrm{d}\mathbb{G}^{\mu}_n(y) + \mathrm{o}_p(1),$$

We aim at finding $h_x(y)$ such that

$$h_x(y) - K^{\nu}_{\varepsilon_n} K^{\mu}_{\varepsilon_n}\big[h_x(\cdot)\big](y) = \left[\frac{1}{(2\pi\varepsilon_n)^{d/2}\nu^{1/2}(y)\mu^{1/2}(x)} \exp\left(-\frac{1}{\varepsilon_n} D(x,y) + \mathrm{o}(1)\right)\right].$$

Constant curvature case

In the case where $\nabla^2 \phi_0^*(y) = A, \forall y$, the equation of previous slide becomes, setting $h_x(y) = \frac{\tilde{h}_x(y)}{\nu^{1/2}(y)\mu^{1/2}(x)}$

$$\tilde{h}_x(y) - (\gamma_2(y) * \tilde{h}_x)(y) \asymp \left[\frac{1}{(2\pi\varepsilon_n)^{d/2}} \exp\left(-\frac{1}{\varepsilon_n} D(x,y) + \mathrm{o}(1) \right) \right],$$

where

$$\gamma_2(y) := \frac{\sqrt{\det A}}{(4\pi\varepsilon_n)^{d/2}} \exp\left(-\frac{1}{4\varepsilon_n} y^\top A y \right).$$

Constant curvature case II

Denoting by $\mathcal{F}$ the Fourier transform, we thus get that

$$h_x(y) \asymp \frac{1}{\nu^{1/2}(y)\mu^{1/2}(x)} \mathcal{F}^{-1}\left[\frac{(2\pi\varepsilon_n)^{-d/2}\mathcal{F}\left[e^{-\frac{1}{\varepsilon_n}D(x,\cdot)+\mathrm{o}(1)}\right]}{1-\mathcal{F}[\gamma_2(\cdot)]}\right](y).$$

Noticing that

$$\mathcal{F}[\gamma_2](\xi) = \exp\left(-\varepsilon_n \xi^\top A^{-1}\xi\right),$$

we can even rewrite

$$h_x(y) \asymp \frac{1}{\nu^{1/2}(y)\mu^{1/2}(x)} \mathcal{F}^{-1}\left[\frac{(2\pi\varepsilon_n)^{-d/2}\mathcal{F}\left[e^{-\frac{1}{\varepsilon_n}D(x,\cdot)+\mathrm{o}(1)}\right]}{\varepsilon_n\ \xi^\top A^{-1}\xi + \mathrm{o}(\varepsilon_n)}\right](y)$$

Conclusion so far

Clear description of what is happening!
In the constant curvature case,

$$a_n(\hat{f}_{\varepsilon_n}(x) - f_{\varepsilon_n}(x))$$
$$= -\frac{a_n\varepsilon_n}{\sqrt{n}} \int \frac{1}{\nu^{1/2}(y)\mu^{1/2}(x)} \mathcal{F}^{-1}\left[\frac{(2\pi\varepsilon_n)^{-d/2}\mathcal{F}\left[e^{-\frac{1}{\varepsilon_n}D(x,\cdot)+\mathrm{o}(1)}\right]}{\varepsilon_n\ \xi^\top A^{-1}\xi + \mathrm{o}(\varepsilon_n)}\right](y)\mathrm{d}\mathbb{G}_n^\mu(y) + \mathrm{o}_p(1).$$

Is it reasonable?

T. Manole's idea

The idea of Tudor is the following:
Solve the Monge–Ampère equation with kernel density estimators $\hat{p}_h, \hat{q}_h$ based on a kernel K, i.e.,

$$\det\left(\nabla^2\phi(x)\right) = \frac{\hat{p}_h(x)}{\hat{q}_h \circ \nabla\phi(x)}.$$

Set

$$\widehat{T}_n(x) := \nabla\phi(x).$$

T. Manole's idea

From

$$\det\left(\nabla^2\hat{\phi}(x)\right) = \frac{p(x)}{\hat{q}_h \circ \nabla\hat{\phi}(x)},$$

a linearisation yields

$$L(\hat{\phi} - \phi) \approx \hat{q}_h - q,$$

where $Lu := -\operatorname{div}(q\nabla u(\nabla\phi_0^*))$
Solving a stochastic PDE (Boundary conditions!).

T. Manole's result

For fixed $x \in \mathbb{T}^d$,

$$\sqrt{nh_n^{d-2}}\Big(\widehat{T}_n(x) - T_0(x)\Big) \xrightarrow{D} \mathcal{N}\Big(0, \Sigma(x)\Big)$$

where

$$\Sigma(x) = \frac{1}{p(x)} \int_{\mathbb{R}^d} \xi\xi^\top \left(\frac{\mathcal{F}[K](M(x)\xi)}{2\pi\langle M(x)\xi, \xi\rangle}\right)^2 \mathrm{d}\xi, \qquad M(x) := \nabla^2\phi_0^*\big(\nabla\phi_0(x)\big).$$

Conclusion

1. EOT with small regularisation parameter is like optimal transport between slightly smoothed densities, up to picking the right kernels (somehow conjectured in Feydy's thesis)!
2. The proof does not required any boundary conditions, EOT is taking care of them on its own.
3. Beautiful mathematical picture.

Proof

$$
\begin{aligned}
a_n(\hat{f}(x) - f(x)) &= a_n\varepsilon_n \left(\log \int \exp\left(\frac{1}{\varepsilon_n} f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho - \log \int \exp\left(\frac{1}{\varepsilon_n}\hat{f}\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho_n \right) \\
&= a_n\varepsilon_n \log \left(\int \exp\left(\frac{1}{\varepsilon_n} f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho \Big/ \int \exp\left(\frac{1}{\varepsilon_n} f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho_n \right) \\
&+ a_n\varepsilon_n \log \left(\int \exp\left(\frac{1}{\varepsilon_n} f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho_n \Big/ \int \exp\left(\frac{1}{\varepsilon_n}\hat{f}\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho_n \right) \\
&=: C(x) + D(x)\,.
\end{aligned}
$$

Proof II

$$
\begin{aligned}
C(x) &= -a_n\varepsilon_n \log\left(\frac{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho_n}{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho}\right) \\
&= -a_n\varepsilon_n \log\left(\frac{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho + \frac{1}{\sqrt{n}}\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\big(\sqrt{n}(\rho_n-\rho)\big)}{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho}\right) \\
&= -\frac{a_n\varepsilon_n}{\sqrt{n}}\frac{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\mathbb{G}_n}{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho} + \mathrm{o}\left(-\frac{a_n\varepsilon_n}{\sqrt{n}}\frac{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\mathbb{G}_n}{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho}\right) \\
&= -\frac{a_n\varepsilon_n}{\sqrt{n}}\int \pi_{\varepsilon_n}\mathrm{d}\mathbb{G}_n + \mathrm{o}\left(-\frac{a_n\varepsilon_n}{\sqrt{n}}\frac{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\mathbb{G}_n}{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\rho}\right).
\end{aligned}
$$

Proof III

$$
\begin{aligned}
D(x) &= -a_n\varepsilon_n \log\left(\frac{\int \exp\left(\frac{1}{\varepsilon_n}\hat{f}\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\hat{\rho}_n}{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\hat{\rho}_n}\right) \\
&= -a_n\varepsilon_n \log\left(\frac{\int \exp\left(\frac{1}{\varepsilon_n}f + \frac{a_n}{\varepsilon_n a_n}(\hat{f}-f)\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\hat{\rho}_n}{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\hat{\rho}_n}\right) \\
&= -a_n\varepsilon_n \log\left(\frac{\int \exp\left(\frac{1}{\varepsilon_n}f\right)\left(1 + \frac{a_n}{\varepsilon_n a_n}(\hat{f}-f) + \mathrm{O}(\varepsilon_n^{-2}(\hat{f}-f)^2)\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\hat{\rho}_n}{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\hat{\rho}_n}\right) \\
&= -a_n\varepsilon_n \log\left(1 + \frac{1}{\varepsilon_n a_n}\frac{\int \exp\left(\frac{1}{\varepsilon_n}f\right)\left(a_n(\hat{f}-f) + \mathrm{O}(a_n\varepsilon_n^{-1}(\hat{f}-f)^2)\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\hat{\rho}_n}{\int \exp\left(\frac{1}{\varepsilon_n}f\right) k_{\varepsilon_n}(x,\cdot)\mathrm{d}\hat{\rho}_n}\right)
\end{aligned}
$$

Proof IV

$$D(x) \asymp -\frac{\int \exp\left(\frac{1}{\varepsilon_n} f\right)\left(a_n(\hat{f} - f) + \mathrm{O}(a_n \varepsilon_n^{-1}(\hat{f} - f)^2)\right) k_{\varepsilon_n}(x, \cdot)\mathrm{d}\hat{\rho}_n}{\int \exp\left(\frac{1}{\varepsilon_n} f\right) k_{\varepsilon_n}(x, \cdot)\mathrm{d}\hat{\rho}_n}$$

$$\asymp -\frac{\int \exp\left(\frac{1}{\varepsilon_n} f\right)\left(a_n(\hat{f} - f) + \mathrm{O}(a_n \varepsilon_n^{-1}(\hat{f} - f)^2)\right) k_{\varepsilon_n}(x, \cdot)\mathrm{d}\hat{\rho}_n}{\int \exp\left(\frac{1}{\varepsilon_n} f\right) k_{\varepsilon_n}(x, \cdot)\mathrm{d}\rho + \frac{1}{\sqrt{n}} \int \exp\left(\frac{1}{\varepsilon_n} f\right) k_{\varepsilon_n}(x, \cdot)\mathrm{d}\mathbb{G}_n}$$

The denominator

$$\frac{1}{\sqrt{n\rho(x)}} \int \frac{k_{\varepsilon_n}(x,y)}{(2\pi\varepsilon_n)^{d/2}\sqrt{\rho(y)}} \mathrm{d}\mathbb{G}_n(y) + \mathrm{o}(1)$$

converges to zero if $\sqrt{n}\varepsilon_n^{d/4} \to \infty$.

The numerator

$$\int \exp\left(\frac{1}{\varepsilon_n} f\right) \left(a_n(\hat{f} - f) + \mathrm{O}(a_n \varepsilon_n^{-1} (\hat{f} - f)^2)\right) k_{\varepsilon_n}(x, \cdot) \mathrm{d}\hat{\rho}_n$$

$$= \int \exp\left(\frac{1}{\varepsilon_n} f\right) \left(a_n(\hat{f} - f) + \mathrm{O}(a_n \varepsilon_n^{-1} (\hat{f} - f)^2)\right) k_{\varepsilon_n}(x, \cdot) \mathrm{d}\rho$$

$$+ \frac{1}{\sqrt{n}} \int \exp\left(\frac{1}{\varepsilon_n} f\right) \left(a_n(\hat{f}(y) - f(y)) + \mathrm{o}_p(1)\right) k_{\varepsilon_n}(x, y) \mathrm{d}\mathbb{G}_n(y)$$

$$= \int \exp\left(\frac{1}{\varepsilon_n} f\right) \left(a_n(\hat{f} - f) + \mathrm{O}(a_n \varepsilon_n^{-1} (\hat{f} - f)^2)\right) k_{\varepsilon_n}(x, \cdot) \mathrm{d}\rho$$

$$+ \frac{a_n \varepsilon_n}{\sqrt{n}} \int \frac{1}{(2\pi\varepsilon_n)^{d/4} \sqrt{\rho(y)}} \left(\frac{1}{\varepsilon_n}(\hat{f}(y) - f(y)) + \mathrm{o}_p(1)\right) k_{\varepsilon_n}(x, y) \mathrm{d}\mathbb{G}_n(y)$$