Stein's method, logarithmic Sobolev and transport inequalities

Ivan Nourdin (Luxembourg University)

This talk is mainly based on the material developed in the following two papers:

- I. Nourdin, G. Peccati and Y. Swan (2014): Entropy and the fourth moment phenomenon, *Journal of Functional Analysis* **266**, no. 5, 3170–3207.

- M. Ledoux, I. Nourdin and G. Peccati (2014): Stein's method, logarithmic Sobolev and transport inequalities. Submitted.

PART I: Stein's method

Let F be a given real random variable with, say, mean zero and variance one.

Let $N \sim N(0,1)$.

In many situations of interest, we may expect the law of F to be close of that of N, and one is interested in quantifying it. How to formalize this ?

Stein's lemma. Assume $h : \mathbb{R} \to [0,1]$ is continuous. Consider φ_h defined, for $x \in \mathbb{R}$, as

$$\begin{aligned} \varphi_h(x) &= e^{\frac{x^2}{2}} \int_{-\infty}^{x} (h(a) - E[h(N)]) e^{-\frac{a^2}{2}} da \\ &= -e^{\frac{x^2}{2}} \int_{x}^{\infty} (h(a) - E[h(N)]) e^{-\frac{a^2}{2}} da. \end{aligned}$$

Then φ_h is C^1, satisfies

$$\varphi_h'(x) - x\varphi_h(x) = h(x) - E[h(N)]$$

and is such that $\|\varphi_h'\|_\infty \le 2$.

A bound on the total variation distance. One has

$$
\begin{aligned}
d_{TV}(F,N) &:= \sup_{A\in\mathcal{B}(\mathbb{R})} \Big| P[F\in A] - P[N\in A] \Big| \\
&\leq \sup_{h:\mathbb{R}\to[0,1]} \Big| E[h(A)] - E[h(N)] \Big| \\
&= \sup_{h:\mathbb{R}\to[0,1]\in C^0} \Big| E[h(A)] - E[h(N)] \Big| \quad \text{(Lusin)} \\
&\leq \sup_{\varphi\in C^1:\,\|\varphi'\|_\infty\leq 2} \Big| E[\varphi'(F)] - E[F\varphi(F)] \Big| \quad \text{(Stein).}
\end{aligned}
$$

Now the question is: how to relate $E[\varphi'(F)]$ and $E[F\varphi(F)]$?

Definition. We say that $\tau_F : \mathbb{R} \to \mathbb{R}$ is a *Stein factor* for F if

$$E[F\varphi(F)] = E[\tau_F(F)\varphi'(F)]$$

for all test function φ. The *Stein discrepancy* is

$$S(F|N) = \sqrt{E[(1 - \tau_F(F))^2]} = \sqrt{\mathrm{Var}(\tau_F(F))}.$$

Theorem. One has $d_{TV}(F, N) \leq 2\, S(F|N)$.

Examples: 1. If $F \sim N(0,1)$ then $\tau_F(x) = 1$ is a Stein factor for F. We then have $S(F|N) = 0$.

2. If $F_n = \frac{1}{\sqrt{n}}(X_1 + \ldots + X_n)$ with X_i iid, centered and unit variance, then $\tau_{F_n} = \frac{1}{n}\sum_{i=1}^n E[\tau_X(X_i)|F_n]$ is a Stein factor for F_n, so that

$$S(F_n|N)^2 \leq \frac{1}{n}\mathrm{Var}(\tau_X(X)).$$

3. If $F = I_p(f)$ is a multiple integral of order p, then $\tau_F = E[\langle DF, -DL^{-1}F\rangle|F]$ is a Stein factor for F, implying in turn that

$$S(F|N)^2 \leq \frac{p-1}{3p}(E[F^4] - 3).$$

Multivariate extension. Let F be a centered random vector of $\mathbb{R}^d$. Let $N \sim N_d(0, \mathsf{Id})$.

Definition. 1) A measurable matrix-valued map on $\mathbb{R}^d$

$$x \mapsto \tau_F(x) = \left\{\tau_F^{ij}(x) : i, j = 1, \ldots, d\right\}$$

is said to be a *Stein matrix* for F if $\tau_F^{ij}(F) \in \mathsf{L}^1(\Omega)$ for every i, j and, for every smooth $\varphi : \mathbb{R}^d \to \mathbb{R}$,

$$E[F \cdot \nabla\varphi(F)] = E[\langle \tau_F(F), \mathsf{Hess}(\varphi)(F)\rangle_{\mathsf{HS}}],$$

with $\langle A, B\rangle_{\mathsf{HS}} = \sum_{i,j=1}^d a_{ij} b_{ij}$.

2) The *Stein discrepancy* is $\mathsf{S}\Big(F \,|\, N\Big)$, with

$$\mathsf{S}^2\Big(F \,|\, N\Big) = E\|\tau_F(F) - \mathsf{Id}\|_{\mathsf{HS}}^2.$$

PART II: Logarithmic Sobolev inequality

Let F be any random vector of $\mathbb{R}^d$ whose law, noted ν, is absolutely continuous (wrt Lebesgue).

Let $N \sim N_d(0, \mathrm{Id})$ and denote its law by γ, that is, $d\gamma(x) = (2\pi)^{-d/2} e^{-|x|^2/2} dx$ on $\mathbb{R}^d$.

Let $h = \frac{d\nu}{d\gamma}$.

Definitions. 1) The *relative entropy* of F with respect to N is

$$\mathsf{H}\big(F \,|\, N\big) = \int_{\mathbb{R}^d} h \log h \, d\gamma \, (= \mathrm{Ent}_\gamma(h)\,).$$

2) The *Fisher information* of F with respect to N is

$$\mathsf{I}\big(F \,|\, N\big) = \int_{\mathbb{R}^d} \frac{|\nabla h|^2}{h} \, d\gamma \, (= \mathsf{I}_\gamma(h)\,).$$

The classical **logarithmic Sobolev inequality** with respect to the standard Gaussian measure γ indicates that

$$\mathsf{H}\Big(F \,|\, N\Big) \, \leq \, \frac{1}{2}\,\mathsf{I}\Big(F \,|\, N\Big).$$

The HSI inequality (Ledoux, Nourdin, Peccati). One has

$$\mathsf{H}\Big(F \,|\, N\Big) \;\le\; \frac{1}{2}\,\mathsf{S}^2\Big(F \,|\, N\Big)\log\left(1+\frac{\mathsf{I}(F \,|\, N)}{\mathsf{S}^2(F \,|\, N)}\right).$$

Since $\log(1+x) \le x$, our inequality is a new improved form of the logarithmic Sobolev inequality.

PART III: Transport inequalities

Wasserstein quadratic distance. Given two probability measures ν and μ on the Borel sets of $\mathbb{R}^d$ whose marginals are square integrable, we define the *quadratic Wasserstein distance* between ν and μ as the quantity

$$\mathsf{W}_2(\nu, \mu) = \inf_{\pi} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi(x, y) \right)^{1/2}$$

where the infimum runs over all probability measures π on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals ν and μ.

Let F be any random vector of $\mathbb{R}^d$ whose law, noted ν, is absolutely continuous (wrt Lebesgue).

Let $N \sim N_d(0, \mathrm{Id})$ and denote its law by γ, that is, $d\gamma(x) = (2\pi)^{-d/2} e^{-|x|^2/2} dx$ on $\mathbb{R}^d$.

Talagrand inequality: One has the quadratic transportation cost inequality:

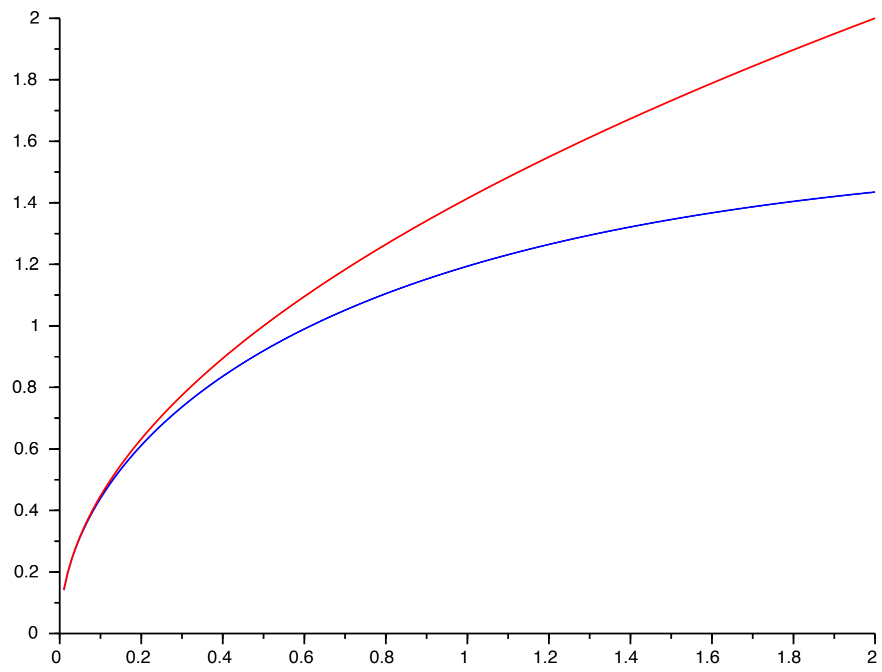
$$\mathsf{W}_2(F, N)^2 \leq 2\,\mathsf{H}\Big(F \,|\, N\Big).$$

WSH inequality (Ledoux, Nourdin, Peccati):

$$\mathsf{W}_2(F, N) \leq \mathsf{S}\Big(F \,|\, N\Big) \arccos\Big(e^{-\frac{\mathsf{H}(F|N)}{\mathsf{S}^2(F|N)}}\Big).$$

WSH implies Talagrand. This is because

$$\arccos(e^{-r}) \leq \sqrt{2r}, \quad r \geq 0.$$



Can we also recover the celebrated HWI inequality?

Otto and Villani:

$$\mathsf{H}(F|N) \leq \mathsf{W}_2(F,N)\sqrt{\mathsf{I}(F|N)} - \frac{1}{2}\mathsf{W}_2(F,N)^2.$$

(It implies the log-Sobolev inequality since $xy - \frac{1}{2}y^2 \leq \frac{1}{2}x^2$.)

Can we consider further distributions?

On the basis of the Gaussian example, we are able to address the issue of HSI inequalities for distributions on $\mathbb{R}^d$, $d \geq 1$, that are not necessarily Gaussian.

Our basic ingredient is a semigroup approach *à la* Bakry-Emery. As such, one can deal with the family of invariant measures of second order differential operators.

These include gamma and beta distributions, as well as families of log-concave measures as illustrations.

But we will not give the details here!

REST OF THE TALK:
some elements of proof

Denote by τ_F a Stein matrix associated with F, having distribution $d\nu = h d\gamma$.

For every $t > 0$, set $d\nu_t = P_t h \, d\gamma$, and write $v_t = \log P_t h$, with $(P_t)_{t \geq 0}$ the Ornstein-Uhlenbeck semigroup associated with $d\gamma(x) = (2\pi)^{-d/2} e^{-|x|^2/2} dx$ on $\mathbb{R}^d$.

We will make intensive use of the following key inequalities.

* **Integrated de Bruijn's formula**

$$\mathsf{H}\big(F \,|\, N\big) \;=\; \mathsf{Ent}_\gamma(h) \;=\; \int_0^\infty \mathsf{I}_\gamma(P_t h) dt.$$

* **Exponential decay of Fisher information**

$$\mathsf{I}_\gamma(P_t h) \;\leq\; e^{-2t}\, \mathsf{I}_\gamma(h) \;=\; e^{-2t}\, \mathsf{I}\big(F \,|\, N\big).$$

* **Linking I and** S (crucial!)

$$\mathsf{I}_\gamma(P_t h) = \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big[\big(\tau_F(x) - \mathsf{Id}\big) y \cdot \nabla v_t\big(e^{-t}x + \sqrt{1-e^{-2t}}\, y\big)\Big] d\nu(x) d\gamma(y).$$

As a consequence,

$$\mathsf{I}_\gamma(P_t h) \;=\; \int_{\mathbb{R}^d} |\nabla v_t|^2 d\nu \;\leq\; \frac{e^{-4t}}{1-e^{-2t}}\, \mathsf{S}^2\big(F \,|\, N\big).$$

* **Exponential decay of Stein discrepancy**

$$\mathsf{S}\big(\nu_t \,|\, \gamma\big) \;\leq\; e^{-2t}\, \mathsf{S}\big(F \,|\, N\big)\, \big(\leq\; \mathsf{S}\big(F \,|\, N\big)\big).$$

Proof of the HSI inequality. The idea is to bound the Fisher information $\mathsf{I}_\gamma(P_t h)$ differently for t around 0 and away from 0.

We can write, for every $u > 0$,

$$\begin{aligned}
\mathsf{H}\big(F\,|\,N\big) &= \int_0^u \mathsf{I}_\gamma(P_t h)dt + \int_u^\infty \mathsf{I}_\gamma(P_t h)dt \\
&\le \mathsf{I}\big(F\,|\,N\big)\int_0^u e^{-2t}dt + \mathsf{S}^2\big(F\,|\,N\big)\int_u^\infty \frac{e^{-4t}}{1-e^{-2t}}\,dt \\
&\le \frac{1}{2}\,\mathsf{I}\big(F\,|\,N\big)(1-e^{-2u}) + \frac{1}{2}\,\mathsf{S}^2\big(F\,|\,N\big)\Big(-e^{-2u}-\log(1-e^{-2u})\Big).
\end{aligned}$$

Optimizing in u (set $1-e^{-2u} = r \in (0,1)$) leads to the desired inequality

$$\mathsf{H}\big(F\,|\,N\big) \le \frac{1}{2}\,\mathsf{S}^2\big(F\,|\,N\big)\log\left(1+\frac{\mathsf{I}(F\,|\,N)}{\mathsf{S}^2(F\,|\,N)}\right).$$

Proof of the WSH inequality. For any $t \geq 0$, recall $d\nu_t = P_t h d\gamma$ (in particular, $\nu_0 = \nu$ and $\nu_t \to \gamma$ as $t \to \infty$). The HSI inequality applied to ν_t yields that

$$\mathsf{H}\big(\nu_t \,|\, \gamma\big) \;\leq\; \frac{1}{2}\,\mathsf{S}^2\big(\nu_t \,|\, \gamma\big)\, \log\left(1 + \frac{\mathsf{I}(\nu_t \,|\, \gamma)}{\mathsf{S}^2(\nu_t \,|\, \gamma)}\right).$$

Now, $\mathsf{S}^2(\nu_t \,|\, \gamma) \leq \mathsf{S}^2(\nu \,|\, \gamma)$ (due to the exponential decay of the Stein discrepancy) and $r \mapsto r \log\big(1 + \frac{s}{r}\big)$ is increasing for any fixed s. It follows that

$$\mathsf{H}\big(\nu_t \,|\, \gamma\big) \;\leq\; \frac{1}{2}\,\mathsf{S}^2\big(\nu \,|\, \gamma\big)\, \log\left(1 + \frac{\mathsf{I}(\nu_t \,|\, \gamma)}{\mathsf{S}^2(\nu \,|\, \gamma)}\right).$$

By exponentiating both sides, this inequality is equivalent to:

$$\sqrt{\mathsf{I}\Big(\nu_t\,|\,\gamma\Big)} \;\le\; \frac{\mathsf{I}(\nu_t\,|\,\gamma)}{\mathsf{S}(\nu\,|\,\gamma)\sqrt{e^{\frac{2\mathsf{H}(\nu_t|\gamma)}{\mathsf{S}^2(\nu|\gamma)}}-1}}\,.$$

But Otto and Villani showed in their pathbreaking paper that

$$\frac{d^+}{dt}\mathsf{W}_2(\nu,\nu_t) \;\le\; \sqrt{\mathsf{I}\Big(\nu_t\,|\,\gamma\Big)}.$$

Moreover, one has the de Bruijn identity:

$$\mathsf{I}\Big(\nu_t\,|\,\gamma\Big) = -\frac{d}{dt}\mathsf{H}(\nu_t\,|\,\gamma).$$

Plugging everything together leads to

$$\frac{d^+}{dt}\mathsf{W}_2(\nu,\nu_t) \le \sqrt{\mathsf{I}\big(\nu_t\,|\,\gamma\big)} \le -\frac{\frac{d}{dt}\mathsf{H}(\nu_t\,|\,\gamma)}{\mathsf{S}(\nu\,|\,\gamma)\sqrt{e^{\frac{2\mathsf{H}(\nu_t|\gamma)}{\mathsf{S}^2(\nu|\gamma)}}-1}}$$

$$= -\frac{d}{dt}\Bigg(\mathsf{S}\big(\nu\,|\,\gamma\big)\,\mathsf{arccos}\Big(e^{-\frac{H(\nu_t|\gamma)}{\mathsf{S}^2(\nu|\gamma)}}\Big)\Bigg).$$

In other words,

$$\frac{d^+}{dt}\Bigg(\mathsf{W}_2(\nu,\nu_t)+\mathsf{S}\big(\nu\,|\,\gamma\big)\,\mathsf{arccos}\Big(e^{-\frac{H(\nu_t|\gamma)}{S^2(\nu|\gamma)}}\Big)\Bigg) \le 0,$$

and we get the WSH inequality, namely

$$\mathsf{W}_2(F,N) \le \mathsf{S}\big(F\,|\,N\big)\,\mathsf{arccos}\Big(e^{-\frac{\mathsf{H}(F|N)}{\mathsf{S}^2(F|N)}}\Big),$$

after integrating between $t=0$ and $t=\infty$.