Stein's inequality for multivariate Gaussian

Michel Ledoux

Abstract

The note briefly describes a couple of analogues of Stein's inequality for the multivariate Gaussian distribution in terms of the quadratic Kantorovich metric, which arose recently in the literature.

Stein's method [15] is a general device to achieve approximation of probability measures by a fixed target measure, typically the normal distribution.

Let $d\gamma(x) = e^{-x^2/2}\frac{dx}{\sqrt{2\pi}}$ be the standard normal distribution on the real line $\mathbb{R}$. The classical Stein lemma expresses that given a (bounded, measurable) function $\varphi : \mathbb{R} \to \mathbb{R}$, the equation

$$\psi' - x\,\psi \,=\, \varphi - \int_{\mathbb{R}} \varphi\, d\gamma \tag{1}$$

may be solved with a function ψ, bounded as well as its derivative. More precisely, ψ may be chosen so that $\|\psi\|_\infty \leq \sqrt{2\pi}\,\|\varphi\|_\infty$ and $\|\psi'\|_\infty \leq 4\,\|\varphi\|_\infty$ (see [15], and [5, 12, 4] for a detailed proof and related inequalities).

Recall the total variation distance between a probability measure μ on $\mathbb{R}$ and γ,

$$\|\mu - \gamma\|_{\mathrm{TV}} \,=\, \sup_{A \in \mathcal{B}(\mathbb{R})} \big[\mu(A) - \gamma(A)\big] \,=\, \frac{1}{2}\sup\bigg[\int_{\mathbb{R}} \varphi\, d\mu - \int_{\mathbb{R}} \varphi\, d\gamma\bigg]$$

where the supremum is taken over all bounded measurable $\varphi : \mathbb{R} \to \mathbb{R}$ with $\|\varphi\|_\infty \leq 1$. Stein's lemma may then be used to provide the basic approximation bound, called Stein's inequality,

$$\|\mu - \gamma\|_{\mathrm{TV}} \,\leq\, \sup\bigg|\int_{\mathbb{R}} \psi' d\mu - \int_{\mathbb{R}} x\,\psi\, d\mu\bigg| \tag{2}$$

where the supremum runs over all continuously differentiable functions $\psi : \mathbb{R} \to \mathbb{R}$ such that $\|\psi\|_\infty \leq \sqrt{\frac{\pi}{2}}$ and $\|\psi'\|_\infty \leq 2$. One of the main interests in (2) lies in the fact that only the measure μ is involved in the upper bound via explicit integrals. It has been used in a wide range of applications quantifying the convergence to a normal distribution (cf. [15, 5, 12, 4] and the references therein).

The question of analogues of the Stein inequality (2) in higher dimension has been of interest in the recent years, in particular due to the difficulty to handle equivalents of (1) in a multivariate setting. This short note emphasizes a couple of results in this direction in terms of the Kantorovich quadratic transportation distance. Given probability measures μ and ν on the Borel sets of $\mathbb{R}^d$ with a finite second moment, let

$$\mathrm{W}_2(\mu,\nu) \,=\, \inf_\pi \bigg(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi(x,y) \bigg)^{\frac{1}{2}},$$

where the infimum is taken over all couplings π on $\mathbb{R}^d\times\mathbb{R}^d$ with respective marginals μ and ν, be the quadratic Kantorovich (Wasserstein) distance between μ and ν.

Let γ_d denote the standard Gaussian measure on the Borel sets of $\mathbb{R}^d$ with density $(2\pi)^{-d/2}e^{-\frac{1}{2}|x|^2}$, $x\in\mathbb{R}^d$, with respect to the Lebesgue measure.

1 Stein kernel

The first highlighted bound is expressed in terms of a Stein kernel associated to the unknown distribution μ. Given a centered probability measure μ on $\mathbb{R}^d$, a Stein kernel of μ is a measurable matrix-valued map τ^μ on $\mathbb{R}^d$ such that for every smooth test function $\varphi:\mathbb{R}^d\to\mathbb{R}$,

$$\int_{\mathbb{R}^d} x\cdot\nabla\varphi\,d\mu \,=\, \int_{\mathbb{R}^d} \tau^\mu\cdot\nabla^2\varphi\,d\mu$$

where $\nabla\varphi$ stands for the gradient of φ, with the scalar product between vectors in $\mathbb{R}^d$, and $\nabla^2\varphi$ stands for the Hessian of φ, with the Hilbert-Schmidt scalar product between (symmetric) $d\times d$ matrices. The choice for φ of the coordinate maps $x\mapsto x_k$, $k=1,\ldots,d$, justifies the centering hypothesis. With respect to the differential equation (1), the picture here lies at a second differential order.

Stein kernels appear implicitly in the literature about Stein's method (see the original monograph [15, Lecture VI] of C. Stein, as well as [3, 4, 6, 7]...), while second order operators in a multivariate setting were considered in [2, 8]. They gained momentum in the past years, specially in connection with probabilistic approximations involving random variables living on a Gaussian (Wiener) space (see the monograph [12]).

According to the standard Gaussian integration by parts formula

$$\int_{\mathbb{R}^d} x\,\psi\,d\gamma_d \,=\, \int_{\mathbb{R}^d} \nabla\psi\,d\gamma_d$$

(as vector valued integrals) for smooth functions $\psi : \mathbb{R}^d \to \mathbb{R}$, the identity matrix Id in $\mathbb{R}^d$ is a Stein kernel for γ_d. The proximity of τ^μ with Id thus indicates that μ should be close to the Gaussian distribution γ_d. Therefore, whenever such a Stein kernel τ^μ exists, the quantity, called Stein discrepancy (of μ with respect to γ_d, and associated to the underlying kernel τ^μ),

$$\mathrm{S}_2\big(\mu \,|\, \gamma_d\big) \;=\; \bigg(\int_{\mathbb{R}^d} |\tau^\mu - \mathrm{Id}|^2 \, d\mu\bigg)^{\frac{1}{2}}$$

(with $|\cdot|$ the Hilbert-Schmidt norm) becomes relevant as a measure of the proximity of μ and γ_d.

In dimension one, Stein's inequality (2) (with $\psi = \varphi'$) precisely indicates that

$$\|\mu - \gamma_1\|_{\mathrm{TV}} \;\leq\; 2\int_{\mathbb{R}} |\tau^\mu - 1| d\mu,$$

and therefore, by Jensen's inequality,

$$\|\mu - \gamma_1\|_{\mathrm{TV}} \;\leq\; 2\,\mathrm{S}_2\big(\mu \,|\, \gamma_1\big),$$

justifying the interest in the Stein discrepancy.

It is the purpose of the following proposition from [10] to emphasize the corresponding inequality in $\mathbb{R}^d$ in terms of the Kantorovich metric W_2.

Proposition 1. *In the preceding notation,*

$$\mathrm{W}_2\big(\mu, \gamma_d\big) \;\leq\; \mathrm{S}_2\big(\mu \,|\, \gamma_d\big).$$

For such a result to be useful and of interest, it is necessary to determine and describe suitable kernels τ^μ of the probability μ to be approximated by the Gaussian distribution γ_d. In dimension $d = 1$, if μ has a density ρ with respect to the Lebesgue measure on $\mathbb{R}$, the Stein kernel τ^μ is uniquely determined (up to sets of zero Lebesgue measure), and under standard regularity assumptions on ρ, a version of τ^μ is given by

$$\tau^\mu(x) \;=\; \frac{1}{\rho(x)} \int_x^\infty y\rho(y)dy$$

for x inside the support of ρ. In higher dimension, Stein kernels are not always unique and even may not exist.

In several illustrations and applications however, the unknown probability measure μ is actually of more concrete nature, allowing for descriptions of a kernel. A typical instance is the example of the law μ of $F(X)$ where

$$F \;=\; (F_1, \ldots, F_d) : \mathbb{R}^n \;\to\; \mathbb{R}^d$$

is measurable and X is a (standard) Gaussian random vector on $\mathbb{R}^n$ (on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$). In other words, μ, as a probability measure on the Borel sets of $\mathbb{R}^d$, is the law of F under γ_n, and the following studies its proximity to the standard Gaussian distribution γ_d on $\mathbb{R}^d$. It is possible to consider more general distributions for X, such as the Wiener measure in infinite dimension [12], or the invariant measure of a Markov triple in the sense of [1], but for simplicity of this note, X is just normal in finite dimension (see [12, 10] for these extensions). In most applications, the function F is also assumed to be reasonably regular in order to perform a number of differentiation and integration by parts operations, smoothness that will always be implicit below (polynomials is a class of examples). Since μ should be centered, $\mathbb{E}(F(X)) = 0$.

Towards the analysis of such distributions, it will be critical, following [2, 8], to deal with the (Ornstein-Uhlenbeck) second order differential operator $\mathrm{L} = \Delta - x \cdot \nabla$, acting on smooth functions on $\mathbb{R}^n$, invariant and symmetric with respect to γ_n. By integration by parts, for smooth functions $f, g : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} f \, \mathrm{L}g \, d\gamma_n \; = \; - \int_{\mathbb{R}^n} \nabla f \cdot \nabla g \, d\gamma_n.$$

The operator L admits a spectral representation, with spectrum the negative integers and the Hermite polynomials as eigenfunctions. As such, it makes sense to consider the inverse operator $(-\mathrm{L})^{-1}$ acting on mean zero function of the underlying domain. The reader is referred to [1] for an account on the Ornstein-Uhlenbeck operator. The suitable smoothness and domain properties will be mostly understood below.

Given smooth functions

$$F = (F_1, \ldots, F_d), \; G = (G_1, \ldots, G_d) : \mathbb{R}^n \to \mathbb{R}^d,$$

denote by $\Gamma(F, G)$ the $d \times d$ matrix $(\nabla F_k \cdot \nabla G_\ell)_{1 \leq k, \ell \leq d}$. The operators L and $(-\mathrm{L})^{-1}$ are acting on a multivariate function $F = (F_1, \ldots, F_d)$ as $\mathrm{L}F = (\mathrm{L}F_1, \ldots, \mathrm{L}F_d)$ and similarly for $(-\mathrm{L})^{-1}$.

In this setting and with the latter notation, a Stein kernel τ^μ for the law μ of $F(X) = (F_1, \ldots, F_d)(X)$ on $\mathbb{R}^d$ may be represented by a regular version of the conditional (matrix-valued) expectation

$$\mathbb{E}\big(T(X) \,|\, F(X)\big)$$

of $T(X)$, where $T = \Gamma((-\mathrm{L})^{-1}F, F)$, with respect to the σ-field generated by $F(X)$. Indeed, for any smooth test function $\varphi : \mathbb{R}^d \to \mathbb{R}$ and every $k, \ell = 1, \ldots, \ell$, by definition of the conditional expectation,

$$\begin{aligned}\int_{\mathbb{R}^d} \tau^\mu_{k\ell}\, \partial^2_{k\ell}\varphi\, d\mu \;&=\; \mathbb{E}\Big(\mathbb{E}\big(T_{k\ell}(X)\,|\,F(X)\big)\partial^2_{k\ell}\varphi(F(X))\Big)\\ &=\; \mathbb{E}\big(T_{k\ell}(X)\,\partial^2_{k\ell}\varphi(F(X))\big)\\ &=\; \int_{\mathbb{R}^n} T_{k\ell}\,(\partial^2_{k\ell}\varphi)\circ F\, d\gamma_n\end{aligned}$$

since X has law γ_n. Hence, by definition of T,

$$\begin{aligned}\sum_{k,\ell=1}^d \int_{\mathbb{R}^d} \tau^\mu_{k\ell}\, \partial^2_{k\ell}\varphi\, d\mu \;&=\; \sum_{k,\ell=1}^d \int_{\mathbb{R}^n} \nabla((-\mathrm{L})^{-1}F_k)\cdot\nabla F_\ell\; (\partial^2_{k\ell}\varphi)\circ F\, d\gamma_n\\ &=\; \sum_{k=1}^d \int_{\mathbb{R}^n} \nabla((-\mathrm{L})^{-1}F_k)\cdot\nabla((\partial_k\varphi)\circ F))\, d\gamma_n\\ &=\; \sum_{k=1}^d \int_{\mathbb{R}^n} F_k\,(\partial_k\varphi)\circ F\, d\gamma_n\end{aligned}$$

where integration by parts for the operator L has been used in the last step. Since μ is the law of F under γ_n, $\sum_{k=1}^d \int_{\mathbb{R}^n} F_k\,(\partial_k\varphi)\circ F\, d\gamma_n = \sum_{k=1}^d \int_{\mathbb{R}^d} x_k\,\partial_k\varphi\, d\mu$, so that

$$\int_{\mathbb{R}^d} \tau^\mu\cdot\nabla^2\varphi\, d\mu \;=\; \sum_{k,\ell=1}^d \int_{\mathbb{R}^d} \tau^\mu_{k\ell}\,\partial^2_{k\ell}\varphi\, d\mu \;=\; \sum_{k=1}^d \int_{\mathbb{R}^d} x_k\,\partial_k\varphi\, d\mu \;=\; \int_{\mathbb{R}^d} x\cdot\nabla\varphi\, d\mu,$$

justifying the form of the kernel τ^μ.

Proposition 1 therefore yields

$$\begin{aligned}\mathrm{W}_2\big(\mu,\gamma_d\big)^2 \;&\le\; \int_{\mathbb{R}^d} |\tau^\mu - \mathrm{Id}|^2\, d\mu\\ &=\; \mathbb{E}\Big(\big|\mathbb{E}(T(X)\,|\,F(X)) - \mathrm{Id}\big|^2\Big)\\ &\le\; \mathbb{E}\big(|T(X)-\mathrm{Id}|^2\big) \;=\; \int_{\mathbb{R}^n} |T-\mathrm{Id}|^2 d\gamma_n\end{aligned}\tag{3}$$

after the use of Jensen's inequality in the conditional expectation.

Inequality (3) is relevant in a number of illustrations. At first sight, the upper bound might not appear so tractable due to the form of $T = \Gamma((-\mathrm{L})^{-1}F, F)$. However, several instances are easily handled. A typical illustration, that actually motivated this type of results [12], is the case of a function F consisting of eigenfunctions of L. That is, $-\mathrm{L}F_k = n_k F_k$ for some integers $n_k \ge 1$, $k = 1,\ldots,d$, so that T takes the simple form

$$T_{k\ell} \;=\; \frac{1}{n_k}\,\nabla F_k\cdot\nabla F_\ell, \quad k,\ell = 1,\ldots,d.$$

The inequality (3) then yields

$$\mathrm{W}_2\big(\mu,\gamma_d\big)^2 \,\leq\, \int_{\mathbb{R}^n} \sum_{k,\ell=1}^d \Big|\frac{1}{n_k}\,\nabla F_k\cdot\nabla F_\ell - \delta_{k\ell}\Big|^2 d\gamma_n \tag{4}$$

which may be suitably controlled given the (smooth) function $F=(F_1,\ldots,F_d)$.

In the present Gaussian setting, the eigenfunctions F_k, $k=1,\ldots,d$, are actually explicit Hermite polynomials, and computational bounds are therefore available. This procedure has been developed in [12] in the more general context of infinite dimensional Wiener chaos to provide in particular rates of convergence to the Gaussian distribution in the fourth moment theorem for stochastic integrals of [14].

In general, the inverse operator $(-\mathrm{L})^{-1}$ embedded into $T=\Gamma((-\mathrm{L})^{-1}F,F)$ may be described by the Ornstein-Ulenbeck semigroup $(P_t)_{t\geq 0}$ with infinitesimal generator L to provide handful expressions. The semigroup $(P_t)_{t\geq 0}$ namely admits the integral representation

$$P_tf(x) \,=\, \int_{\mathbb{R}^n} f\big(e^{-t}x+\sqrt{1-e^{-2t}}\,y\big)d\gamma_n(y), \quad x\in\mathbb{R}^n,\ t\geq 0,$$

and $(-\mathrm{L})^{-1}$ is given by $(-\mathrm{L})^{-1}=\int_0^\infty P_t dt$ (on centered functions in the suitable domain). For the further purposes, also observe that for smooth f,

$$\nabla P_tf \,=\, e^{-t}P_t(\nabla f)$$

(P_t acting coordinatewise on vector valued functions). Then, for every $k,\ell=1,\ldots,d$,

$$\begin{aligned} T_{k\ell} \,&=\, \nabla((-\mathrm{L})^{-1}F_k)\cdot\nabla F_\ell \\ &=\, \int_0^\infty \nabla(P_tF_k)\cdot\nabla F_\ell\,dt \\ &=\, \int_0^\infty e^{-t}\,P_t(\nabla F_k)\cdot\nabla F_\ell\,dt, \end{aligned} \tag{5}$$

and after a use of Jensen's inequality in $e^{-t}dt$, (3) yields

$$\mathrm{W}_2\big(\mu,\gamma_d\big)^2 \,\leq\, \int_0^\infty e^{-t}\sum_{k,\ell=1}^d \int_{\mathbb{R}^n} \big|P_t(\nabla F_k)\cdot\nabla F_\ell-\delta_{k\ell}\big|^2 d\gamma_n\,dt.$$

Again, this bound may be reasonably controlled for given functions F.

Another direction to make use of (3) is provided by the Poincaré inequality for γ_n,

$$\int_{\mathbb{R}^n} f^2 d\gamma_n \,\leq\, \int_{\mathbb{R}^n} |\nabla f|^2 d\gamma_n$$

for every smooth $f : \mathbb{R}^n \to \mathbb{R}$ with mean zero. Applied to the coordinates of $T = \Gamma((-\mathrm{L})^{-1}F, F)$,

$$\int_{\mathbb{R}^n} |T - \mathrm{Id}|^2 d\gamma_n \,=\, \sum_{k,\ell=1}^{d} \int_{\mathbb{R}^n} |T_{k\ell} - \delta_{k\ell}|^2 d\gamma_n \,\leq\, \sum_{k,\ell=1}^{d} \int_{\mathbb{R}^n} |\nabla T_{k\ell}|^2 d\gamma_n \tag{6}$$

provided that

$$\begin{aligned}\int_{\mathbb{R}^n} T_{k\ell}\, d\gamma_n \,&=\, \int_{\mathbb{R}^n} \nabla((-\mathrm{L})^{-1}F_k) \cdot \nabla F_\ell \, d\gamma_n \\ &=\, \int_{\mathbb{R}^n} F_k F_\ell \, d\gamma_n \\ &=\, \mathbb{E}\big(F_k(X)F_\ell(X)\big) \,=\, \delta_{k\ell}\end{aligned}$$

for every $k, \ell = 1, \ldots, d$. In other words, the covariance matrix of the random vector $F(X)$ should be the identity matrix.

Now it follows from the description (5) that

$$\nabla T_{k\ell} \,=\, \int_0^\infty e^{-2t}\, P_t(\nabla^2 F_k)\, \nabla F_\ell \, dt + \int_0^\infty e^{-t}\, \nabla^2 F_\ell \, P_t(\nabla F_k)\, dt$$

(the matrix $P_t(\nabla^2 F_k)$ acting on the vector ∇F_ℓ to produce a vector in $\mathbb{R}^d$, and similarly for the other term). Therefore, whenever the covariance of $F(X)$ is the identity matrix, (3) and (6) yield, after the use of the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathrm{W}_2\big(\mu, \gamma_d\big)^2 \,\leq\, &\int_0^\infty e^{-2t} \sum_{k,\ell=1}^{d} \int_{\mathbb{R}^n} \big|P_t(\nabla^2 F_k)\, \nabla F_\ell\big|^2 d\gamma_n \, dt \\ &+ 2\int_0^\infty e^{-t} \sum_{k,\ell=1}^{d} \int_{\mathbb{R}^n} \big|\nabla^2 F_\ell \, P_t(\nabla F_k)\big|^2 d\gamma_n \, dt.\end{aligned} \tag{7}$$

Using convexity on the integral representation of the operators P_t, $t \geq 0$, the preceding bound may be given a more tractable form. Namely, for every $k, \ell = 1, \ldots, d$,

$$\begin{aligned}\int_{\mathbb{R}^n} \big|\nabla^2 F_\ell \, P_t(\nabla F_k)\big|^2 d\gamma_n \,&=\, \int_{\mathbb{R}^n} \sum_{i=1}^{n} \bigg| \sum_{j=1}^{n} \partial_{ij} F_\ell \, P_t(\partial_j F_k) \bigg|^2 d\gamma_n \\ &\leq\, \sum_{i=1}^{n} \int_{\mathbb{R}^n} \sum_{j=1}^{n} (\partial_{ij} F_\ell)^2 \sum_{j=1}^{n} \big[P_t(\partial_j F_k)\big]^2 d\gamma_n \\ &=\, \int_{\mathbb{R}^n} |\nabla^2 F_\ell|^2 \sum_{j=1}^{n} \big[P_t(\partial_j F_k)\big]^2 d\gamma_n.\end{aligned}$$

Now, for every $j = 1, \ldots, n$, by Jensen's inequality on the integral representation of P_t,

$$\int_{\mathbb{R}^n} |\nabla^2 F_\ell|^2 \big[P_t(\partial_j F_k)\big]^2 d\gamma_n$$
$$\leq \int_{\mathbb{R}^n} |\nabla^2 F_\ell(x)|^2 \bigg(\int_{\mathbb{R}^n} (\partial_j F_k)^2 \big(e^{-t}x + \sqrt{1-e^{-2t}}\, y\big) d\gamma_n(y)\bigg) d\gamma_n(x),$$

so that

$$\int_{\mathbb{R}^n} |\nabla^2 F_\ell|^2 \sum_{j=1}^n \big[P_t(\partial_j F_k)\big]^2 d\gamma_n$$
$$\leq \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |\nabla^2 F_\ell(x)|^2 \, |\nabla F_k|^2 \big(e^{-t}x + \sqrt{1-e^{-2t}}\, y\big) d\gamma_n(x) d\gamma_n(y).$$

Finally, by another use of the Cauchy-Schwarz inequality in the tensor measure $\gamma_n \otimes \gamma_n$, for every $t \geq 0$,

$$\sum_{k,\ell=1}^d \int_{\mathbb{R}^n} \big|\nabla^2 F_\ell \, P_t(\nabla F_k)\big|^2 d\gamma_n$$
$$\leq \bigg(\int_{\mathbb{R}^n} \bigg[\sum_{k=1}^d |\nabla F_k|^2 \bigg]^2 d\gamma_n \bigg)^{1/2} \bigg(\int_{\mathbb{R}^n} \bigg[\sum_{\ell=1}^d |\nabla^2 F_\ell|^2 \bigg]^2 d\gamma_n \bigg)^{1/2}$$

(since $e^{-t}x + \sqrt{1-e^{-2t}}\, y$ under $d\gamma_n(x)d\gamma_n(y)$ has law γ_n). The term

$$\sum_{k,\ell=1}^d \int_{\mathbb{R}^n} \big|P_t(\nabla^2 F_k)\, \nabla F_\ell\big|^2 d\gamma_n$$

in (7) is handled similarly, leading finally to the following statement.

Proposition 2. *In the preceding notation, provided that $F(X)$, with law μ on the Borel sets of $\mathbb{R}^d$, has covariance matrix the identity, and that $F : \mathbb{R}^n \to \mathbb{R}^d$ is smooth enough,*

$$\mathrm{W}_2\big(\mu, \gamma_d\big)^2 \leq 3 \bigg(\int_{\mathbb{R}^n} \bigg[\sum_{k=1}^d |\nabla F_k|^2 \bigg]^2 d\gamma_n \bigg)^{1/2} \bigg(\int_{\mathbb{R}^n} \bigg[\sum_{\ell=1}^d |\nabla^2 F_\ell|^2 \bigg]^2 d\gamma_n \bigg)^{1/2}. \tag{8}$$

Arbitrary covariances are considered in [13]). The preceding statement applies in dimension $d = 1$ also for the total variation distance.

Inequalities such as (8) have been exploited in [13] to study central limit theorems on Wiener space and in [3] to control the distance of the law of traces of random matrices to the Gaussian distribution, under the terminology of second-order Poincaré inequalities.

2 An additive variant

The preceding analysis involving the inverse operator $(-\mathrm{L})^{-1}$ may be replaced by an additive variant in the following way. Consider again the law μ of $F(X)$ where $F : \mathbb{R}^n \to \mathbb{R}^d$ and X follows the standard Gaussian distribution on $\mathbb{R}^n$. Recall the Ornstein-Uhlenbeck generator L.

For positive numbers $\kappa_k > 0$, $k = 1, \ldots, d$, denote by K the diagonal matrix $K = \mathrm{diag}(\kappa_1, \ldots, \kappa_d)$. Given $F = (F_1, \ldots, F_d)$ smooth enough, recall $\Gamma(F, F) = \Gamma(F)$ the $d \times d$ matrix $(\nabla F_k \cdot \nabla F_\ell)_{1 \leq k,\ell \leq d}$. Introduce the quantities A and B defined by

$$A = \bigg(\int_{\mathbb{R}^n} \big| F + K^{-1}\,\mathrm{L}F \big|^2 d\gamma_n \bigg)^{\frac{1}{2}}$$

and

$$B = \bigg(\int_{\mathbb{R}^n} \big| \mathrm{Id} - K^{-1}\,\Gamma(F) \big|^2 \, d\gamma_n \bigg)^{\frac{1}{2}},$$

the norms $|\cdot|$ being understood as above in the Euclidean space $\mathbb{R}^d$ and in the space of $d \times d$ matrices (Hilbert-Schmidt norm). The expressions A and B thus depend on F and on $\kappa_k > 0$, $k = 1, \ldots, d$. While A involves the generator L, it may be given a form only involving derivatives of F (as for B). Namely,

$$\begin{aligned}
\int_{\mathbb{R}^n} \big| F + K^{-1}\,\mathrm{L}F \big|^2 d\gamma_n &= \int_{\mathbb{R}^n} |F|^2 d\gamma_n + 2 \sum_{k=1}^d \int_{\mathbb{R}^n} \kappa_k^{-1} F_k \, \mathrm{L}F_k \, d\gamma_n + \sum_{k=1}^d \int_{\mathbb{R}^n} \kappa_k^{-2} (\mathrm{L}F_k)^2 d\gamma_n \\
&= \int_{\mathbb{R}^n} |F|^2 d\gamma_n - 2 \sum_{k=1}^d \int_{\mathbb{R}^n} \kappa_k^{-1}\, \Gamma(F_k) \, d\gamma_n + \sum_{k=1}^d \int_{\mathbb{R}^n} \kappa_k^{-2}\, \Gamma_2(F_k) d\gamma_n
\end{aligned}$$

where $\Gamma(F_k) = |\nabla F_k|^2$ and $\Gamma_2(F_k) = |\nabla^2 F_k|^2 + |\nabla F_k|^2$ (cf. [1]).

The following approximation result is presented in [9].

Proposition 3. *In the preceding notation, for any choice of $\kappa_k > 0$, $k = 1, \ldots, d$,*

$$\mathrm{W}_2\big(\mu, \gamma_d\big) \leq A + B.$$

A form of this multivariate normal approximation result in the W_1 metric was developed in [11] relying on exchangeable pairs. The ingredients in the proof of Proposition 3 are similar in nature to the ones in Proposition 1. The conclusion may be used in different instances of interest. If $F_1, \ldots F_d$ are eigenfunctions of L with eigenvalues $-n_1, \ldots, -n_d$, for the choice of $\kappa_k = n_k$, $k = 1, \ldots, d$, the quantity A vanishes, and Proposition 3 amounts to (4). However, when $F_1, \ldots F_d$ are only approximate eigenfunctions in the sense that $\mathrm{L}F_k + n_k F_k$ is small in some L^2 sense, then Proposition 3 becomes of interest while (3) is not easily exploitable.

Proposition 3 has been illustrated in [9] via this device in the study of rates of convergence of linear statistics along polynomials of the spectral measure of random matrices from the Gaussian Unitary Ensemble.

References

[1] D. Bakry, I. Gentil, M. Ledoux. *Analysis and geometry of Markov diffusion operators.* Grundlehren der mathematischen Wissenschaften 348. Springer (2014).

[2] A. Barbour. Stein's method for diffusion approximations. *Probab. Theory Rel. Fields* 84, 297–322 (1990).

[3] S. Chatterjee. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* 143, 1–40 (2009).

[4] S. Chatterjee. A short survey of Stein's method (2014).

[5] L. H. Y. Chen, L. Goldstein, Q.-M. Shao. *Normal approximation by Stein's method.* Probability and its Applications. Springer (2011).

[6] L. Goldstein, G. Reinert. Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* 7, 935–952 (1997).

[7] L. Goldstein, G. Reinert. Zero biasing in one and higher dimensions, and applications. *Stein's method and applications,* Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. 5, 1–18. Singapore Univ. Press (2005).

[8] F. Götze. On the rate of convergence in the multivariate CLT. *Ann. Probab.* 19, 724–739 (1991).

[9] G. Lambert, M. Ledoux, C. Webb. Quantitative normal approximation of linear statistics of beta-ensembles. *Ann. Probab.* 47, 2619–2685 (2019).

[10] M. Ledoux, I. Nourdin, G. Peccati. Stein's method, logarithmic Sobolev and transport inequalities. *Geom. and Funct. Anal.* 25, 256–306 (2015).

[11] E. Meckes. On Stein's method for multivariate normal approximation. *High dimensional probability V: the Luminy volume,* 153–178, Inst. Math. Stat. Collection 5 (2009).

[12] I. Nourdin, G. Peccati. *Normal approximations with Malliavin calculus: from Stein's method to universality.* Cambridge Tracts in Mathematics. Cambridge University Press (2012).

[13] I. Nourdin, G. Peccati, G. Reinert. Second order Poincaré inequalities and CLTs on Wiener space. *J. Funct. Anal.* 257, 593–609 (2009).

[14] D. Nualart, G. Peccati. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* 33, 177–193 (2005).

[15] C. Stein. *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes – Monograph Series 7. Institute of Mathematical Statistics (1986).

Institut de Mathématiques de Toulouse

Université de Toulouse – Paul-Sabatier, F-31062 Toulouse, France

ledoux@math.univ-toulouse.fr