Stein's method, logarithmic Sobolev and transport inequalities

Michel Ledoux* Ivan Nourdin† Giovanni Peccati‡

October 22, 2014

Abstract

We develop connections between Stein's approximation method, logarithmic Sobolev and transport inequalities by introducing a new class of functional inequalities involving the relative entropy, the Stein kernel, the relative Fisher information and the Wasserstein distance with respect to a given reference distribution on $\mathbb{R}^d$. For the Gaussian model, the results improve upon the classical logarithmic Sobolev inequality and the Talagrand quadratic transportation cost inequality. Further examples of illustrations include multidimensional gamma distributions, beta distributions, as well as families of log-concave densities. As a by-product, the new inequalities are shown to be relevant towards convergence to equilibrium, concentration inequalities and entropic convergence expressed in terms of the Stein kernel. The tools rely on semigroup interpolation and bounds, in particular by means of the iterated gradients of the Markov generator with invariant measure the distribution under consideration. In a second part, motivated by the recent investigation by Nourdin, Peccati and Swan on Wiener chaoses, we address the issue of entropic bounds on multidimensional functionals F with the Stein kernel via a set of data on F and its gradients rather than on the Fisher information of the density. A natural framework for this investigation is given by the Markov Triple structure (E, μ, Γ) in which abstract Malliavin-type arguments may be developed and extend the Wiener chaos setting.

*Institut de Mathématiques de Toulouse, Université de Toulouse, F-31062 Toulouse, France, and Institut Universitaire de France ledoux@math.univ-toulouse.fr

†Faculté des Sciences, de la Technologie et de la Communication; UR en Mathématiques. Luxembourg University, 6, rue Richard Coudenhove-Kalergi, L-1359 Luxembourg ivan.nourdin@uni.lu; IN was partially supported by the ANR Grant ANR-10-BLAN-0121.

‡Faculté des Sciences, de la Technologie et de la Communication; UR en Mathématiques. Luxembourg University, 6, rue Richard Coudenhove-Kalergi, L-1359 Luxembourg giovanni.peccati@gmail.com; GP was partially supported by the Grant F1R-MTH-PUL-12PAMP (PAMPAS) from Luxembourg University.

Keywords: Entropy, Fisher Information, Stein Kernel and Discrepancy, Logarithmic Sobolev Inequality, Transport Inequality, Convergence to Equilibrium, Concentration Inequality, Normal Approximation, Gamma Calculus.

2000 Mathematics Subject Classification: 60E15, 26D10, 60B10

1 Introduction

The classical logarithmic Sobolev inequality with respect to the standard Gaussian measure $d\gamma(x) = (2\pi)^{-d/2} e^{-|x|^2/2} dx$ on $\mathbb{R}^d$ indicates that for every probability $d\nu = h d\gamma$ with (smooth) density $h : \mathbb{R}^d \to \mathbb{R}_+$ with respect to γ,

$$\mathrm{H}\big(\nu \,|\, \gamma\big) \,=\, \int_{\mathbb{R}^d} h \log h \, d\gamma \,\leq\, \frac{1}{2} \int_{\mathbb{R}^d} \frac{|\nabla h|^2}{h} \, d\gamma \,=\, \frac{1}{2} \, \mathrm{I}\big(\nu \,|\, \gamma\big) \tag{1.1}$$

where

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

is the relative entropy of $d\nu = h d\gamma$ with respect to γ and

$$\mathrm{I}\big(\nu \,|\, \gamma\big) \,=\, \int_{\mathbb{R}^d} \frac{|\nabla h|^2}{h} \, d\gamma \,=\, \mathrm{I}_\gamma(h)$$

is the Fisher information of ν (or h) with respect to γ, see e.g. [B-G-L, Chapter II.5] for a general discussion. (Throughout this work, $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$.)

Inspired by the recent investigation [N-P-S1], this work puts forward a new form of the logarithmic Sobolev inequality (1.1) by considering a further ingredient, namely the Stein discrepancy given by the Stein kernel of ν. A measurable matrix-valued map τ_ν on $\mathbb{R}^d$ is said to be a *Stein kernel* for the (centered) probability ν if for every smooth test function $\varphi : \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\mathbb{R}^d} x \cdot \nabla \varphi \, d\nu \,=\, \int_{\mathbb{R}^d} \big\langle \tau_\nu, \mathrm{Hess}(\varphi) \big\rangle_{\mathrm{HS}} \, d\nu$$

where $\mathrm{Hess}(\varphi)$ stands for the Hessian of φ, whereas $\langle \cdot, \cdot \rangle_{\mathrm{HS}}$ and $\| \cdot \|_{\mathrm{HS}}$ denote the usual Hilbert-Schmidt scalar product and norm, respectively. Note that while Stein kernels appear implicitly in the literature about Stein's method (see the original monograph [Ste, Lecture VI] of C. Stein, as well as [C2, C3, G-S1, G-S2]...), they gained momentum in recent years, specially in connection with probabilistic approximations involving random variables living on a Gaussian (Wiener) space (see the recent monograph [N-P2] for an overview of this emerging area). The terminology 'kernel' with respect to 'factor' seems the most appropriate to avoid confusion with related but different existing notions.

According to the standard Gaussian integration by parts formula from which $\tau_\gamma = \mathrm{Id}$, the identity matrix in $\mathbb{R}^d$, the proximity of τ_ν with Id indicates that ν should be close to the Gaussian distribution γ. Therefore, whenever such a Stein kernel τ_ν exists, the quantity, called *Stein discrepancy* (of ν with respect to γ),

$$\mathrm{S}\big(\nu\,|\,\gamma\big) \,=\, \bigg(\int_{\mathbb{R}^d} \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}}^2 \, d\nu\bigg)^{1/2}$$

becomes relevant as a measure of the proximity of ν and γ. This quantity is actually at the root of the Stein method [C-G-S, N-P2]. For example, in dimension one, the classical Stein bound expresses that the total variation distance $\mathrm{TV}(\nu,\gamma)$ between a probability measure ν and the standard Gaussian distribution γ is bounded from above as

$$\mathrm{TV}(\nu,\gamma) \,\leq\, \sup \bigg| \int_{\mathbb{R}} \varphi'(x) d\nu(x) - \int_{\mathbb{R}} x\varphi(x) d\nu(x) \bigg| \tag{1.2}$$

where the supremum runs over all continuously differentiable functions $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\|\varphi\|_\infty \leq \sqrt{\frac{\pi}{2}}$ and $\|\varphi'\|_\infty \leq 2$. In particular, by definition of τ_ν (and considering φ' instead of φ),

$$\mathrm{TV}(\nu,\gamma) \,\leq\, 2\int_{\mathbb{R}} |\tau_\nu - 1| d\nu \,\leq\, 2\,\mathrm{S}\big(\nu\,|\,\gamma\big)$$

justifying therefore the interest in the Stein discrepancy (see also [C-P-U]). It is actually a main challenge addressed in [N-P-S1] and this work to investigate the multidimensional setting in which inequalities such as (1.2) are no more available.

With the Stein discrepancy $\mathrm{S}(\nu\,|\,\gamma)$, we emphasize here the inequality, for every probability $d\nu = h d\gamma$,

$$\mathrm{H}\big(\nu\,|\,\gamma\big) \,\leq\, \frac{1}{2}\,\mathrm{S}^2\big(\nu\,|\,\gamma\big) \log\bigg(1 + \frac{\mathrm{I}(\nu\,|\,\gamma)}{\mathrm{S}^2(\nu\,|\,\gamma)}\bigg) \tag{1.3}$$

as a new improved form of the logarithmic Sobolev inequality (1.1). In addition, this inequality (1.3) transforms bounds on the Stein discrepancy into entropic bounds, hence allowing for entropic approximations (under finiteness of the Fisher information). Indeed as is classical, the relative entropy $\mathrm{H}(\nu\,|\,\gamma)$ is another measure of the proximity between two probabilities ν and γ (note that $\mathrm{H}(\nu\,|\,\gamma) \geq 0$ and $\mathrm{H}(\nu\,|\,\gamma) = 0$ if and only if $\nu = \gamma$), which is moreover stronger than the total variation distance by the Pinsker-Csizsár-Kullback inequality

$$\mathrm{TV}(\nu,\gamma) \,\leq\, \sqrt{\frac{1}{2}\,\mathrm{H}\big(\nu\,|\,\gamma\big)}$$

(see, e.g. [V, Remark 22.12]).

The proof of (1.3) is achieved by the classical interpolation scheme along the Ornstein-Uhlenbeck semigroup $(P_t)_{t\geq 0}$ towards the logarithmic Sobolev inequality, but modified for time t away from 0 by a further integration by parts involving the Stein kernel. Indeed, while the exponential decay $\mathrm{I}_\gamma(P_t h) \leq e^{-2t}\,\mathrm{I}_\gamma(h)$ of the Fisher information classically produces the logarithmic Sobolev inequality (1.1), the argument is supplemented by a different control of $\mathrm{I}_\gamma(P_t h)$ by the Stein discrepancy for $t > 0$.

We call the inequality (1.3) HSI, connecting entropy H, Stein discrepancy S and Fisher information I, by analogy with the celebrated Otto-Villani HWI inequality [O-V] relating entropy H, (quadratic) Wasserstein distance W (W_2) and Fisher information I. We actually provide in Section 3 a comparison between the HWI and HSI inequalities (suggesting even an HWSI inequality). Moreover, based on the approach developed in [O-V], we prove that

$$\mathrm{W}_2(\nu,\gamma) \,\leq\, \mathrm{S}\big(\nu\,|\,\gamma\big)\arccos\Big(e^{-\frac{\mathrm{H}(\nu\,|\,\gamma)}{\mathrm{S}^2(\nu\,|\,\gamma)}}\Big), \tag{1.4}$$

an inequality that improves upon the celebrated Talagrand quadratic transportation cost inequality [T]

$$\mathrm{W}_2^2(\nu,\gamma) \,\leq 2\,\mathrm{H}\big(\nu\,|\,\gamma\big)$$

(since $\arccos(e^{-r}) \leq \sqrt{2r}$ for every $r \geq 0$). We shall refer to (1.4) as the 'WSH inequality'. Note also that $\mathrm{W}_2(\nu,\gamma) \,\leq\, \mathrm{S}(\nu\,|\,\gamma)$ so that, as entropy, the Stein discrepancy is a stronger measurement than the Wasserstein metric W_2.

The new HSI inequality put forward in this work has a number of significant applications to exponential convergence to equilibrium and concentration inequalities. For example, the standard exponential decay of entropy $\mathrm{H}(\nu^t\,|\,\gamma) \,\leq\, e^{-2t}\,\mathrm{H}(\nu^0\,|\,\gamma)$ along the flow $d\nu^t = P_t h d\gamma$, $t \geq 0$ ($\nu^0 = \nu$, $\nu^\infty = \gamma$), which characterizes the logarithmic Sobolev inequality (1.1) may be strengthened under finiteness of the Stein discrepancy $\mathrm{S} = \mathrm{S}(\nu\,|\,\gamma) = \mathrm{S}(\nu^0\,|\,\gamma)$ into

$$\mathrm{H}\big(\nu^t\,|\,\gamma\big) \,\leq\, \frac{e^{-4t}}{e^{-2t} + \frac{1-e^{-2t}}{\mathrm{S}^2}\,\mathrm{H}(\nu^0\,|\,\gamma)}\,\mathrm{H}\big(\nu^0\,|\,\gamma\big) \,\leq\, \frac{e^{-4t}}{1-e^{-2t}}\,\mathrm{S}^2\big(\nu^0\,|\,\gamma\big) \tag{1.5}$$

(see Corollary 2.7 for a precise statement). On the other hand, logarithmic Sobolev inequalities are classically related to (Gaussian) concentration inequalities by means of the Herbst argument (cf. e.g. [L2, B-L-M]). Stein's method has also been used to this task in [C3], going back however to the root of the methodology of exchangeable pairs. The basic principle emphasized in this work actually allows us to directly quantify concentration properties of a probability ν on $\mathbb{R}^d$ in terms of its Stein discrepancy with respect to the standard Gaussian measure. As a result, for any 1-Lipschitz function $u : \mathbb{R}^d \to \mathbb{R}$ with mean zero, and any $p \geq 2$,

$$\bigg(\int_{\mathbb{R}^d} |u|^p d\nu\bigg)^{1/p} \,\leq\, C\Big(\mathrm{S}_p\big(\nu\,|\,\gamma\big) + \sqrt{p} + \sqrt{p}\,\sqrt{\mathrm{S}_p\big(\nu\,|\,\gamma\big)}\,\Big) \tag{1.6}$$

where $C > 0$ is numerical and

$$\mathrm{S}_p\big(\nu\,|\,\gamma\big) \,=\, \bigg(\int_{\mathbb{R}^d} \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}}^p\, d\nu\bigg)^{1/p}.$$

(When $\nu = \gamma$, the result fits the standard Gaussian concentration properties.) In other words, the growth of the Stein discrepancy $\mathrm{S}_p(\nu\,|\,\gamma)$ in p entails concentration properties of the measure ν in terms of the growth of its moments. This result is one very first instance showing how to directly transfer informations on the Stein kernel into concentration properties. It yields for example that if $T_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$ where $X_1, \ldots, X_n$ are independent with common distribution ν in $\mathbb{R}^d$ with mean zero and covariance matrix Id, for any 1-Lipschitz function $u : \mathbb{R}^d \to \mathbb{R}$ such that $\mathbb{E}(u(T_n)) = 0$,

$$\mathbb{P}\big(u(T_n) \geq r\big) \,\leq\, C\, e^{-r^2/C}$$

for all $0 \leq r \leq r_n$ where $r_n \to \infty$ according to the growth of $\mathrm{S}_p(\nu \,|\, \gamma)$ as $p \to \infty$.

While put forward for the Gaussian measure γ, the question of the validity of (a form of) the HSI and WSH inequalities for other reference measures should be addressed. Natural examples exhibiting HSI inequalities may be described as invariant measures of second order differential operators (on $\mathbb{R}^d$) in order to run the semigroup interpolation scheme. The prototypical example is of course the Ornstein-Uhlenbeck operator with the standard Gaussian measure as invariant measure. But gamma or beta distributions associated to Laguerre or Jacobi operators may be covered in the same way, as well as families of log-concave measures. It should be mentioned that the definition of Stein kernel has then to be adapted to the diffusion coefficient of the underlying differential operator. The use of second order differential operators in order to study multidimensional probabilistic approximations plays a fundamental role in the so-called *generator approach* to Stein's method, as introduced in the seminal references [Ba, G]; see also [R] for a survey on the subject. A convenient setting to work out this investigation is the one of Markov Triples (E, μ, Γ) and semigroups $(P_t)_{t \geq 0}$ as emphasized in [B-G-L] allowing for the Γ-calculus and the necessary heat kernel bounds in terms of the iterated gradients Γ_n. In particular, while the classical Bakry-Émery Γ_2 criterion [B-E, B-G-L] ensures the validity of the logarithmic Sobolev inequality in this context, it is worth mentioning that the analysis towards the HSI bound makes critical use of the associated Γ_3 operator, a rather new feature in the study of functional inequalities.

As alluded to above, the HSI inequality (1.3) is designed to yield entropic central limit theorems for sequences of probability measures of the form $d\nu_n = h_n d\gamma$, $n \geq 1$, such that $s_n = \mathrm{S}(\nu_n \,|\, \gamma) \to 0$ and

$$\log\bigg(1 + \frac{\mathrm{I}(\nu_n \,|\, \gamma)}{s_n^2}\bigg) \,=\, o(s_n^{-2}), \quad n \to \infty.$$

This is achieved, for instance, when the sequence $\mathrm{I}(\nu_n \,|\, \gamma)$, $n \geq 1$, is bounded. However, the principle behind the HSI inequality may actually be used to deduce entropic convergence (with explicit rates) in more delicate situations, including cases for which $\mathrm{I}(\nu_n \,|\, \gamma) \to \infty$. Indeed, it was one main achievement of the work [N-P-S1] in the context of Wiener chaoses to set up bounds involving entropy and the Stein discrepancy without conditions on the Fisher information. Specifically, it was proved in [N-P-S1] that the entropy with respect to the Gaussian measure γ of the distribution on $\mathbb{R}^d$ of a vector $F = (F_1, \ldots, F_d)$ of Wiener chaoses may be controlled by the Stein discrepancy, providing the first multidimensional entropic approximation results in this context. The key feature underlying the HSI inequality is the control as $t \to 0$ of the Fisher information $\mathrm{I}_\gamma(P_t h)$ along the semigroup (where h the density with respect to γ of the law of F) by the Stein discrepancy. The arguments in [N-P-S1] actually provide the suitable small time behavior of $\mathrm{I}_\gamma(P_t h)$ relying on specific properties of the functionals (Wiener chaoses) under investigation and tools from Malliavin calculus.

In the second part of the work, we therefore develop a general approach to cover the results of [N-P-S1] and to include a number of further potential instances of interest. As before, the setting of a Markov Triple (E, μ, Γ) provides a convenient abstract framework to achieve this goal in which the Γ-calculus appears as a kind of substitute to the Malliavin calculus in this context. Let Ψ be the function $1 + \log r$ on $\mathbb{R}_+$ but linearized by r on $[0, 1]$, that is, $\Psi(r) = 1 + \log r$ if $r \geq 1$ and $\Psi(r) = r$ if $0 \leq r \leq 1$ (note that $\Psi(r) \leq r$ for every $r \in \mathbb{R}_+$). A typical conclusion is a bound of the type

$$\mathrm{H}\big(\nu_F \,|\, \gamma\big) \;\leq\; C_F \, \mathrm{S}^2\big(\nu_F \,|\, \gamma\big) \, \Psi\bigg(\frac{\widetilde{C}_F}{\mathrm{S}^2(\nu_F \,|\, \gamma)}\bigg) \tag{1.7}$$

of the relative entropy of the distribution ν_F of a vector $F = (F_1, \ldots, F_d)$ on (E, μ, Γ) with respect to γ by the Stein discrepancy $\mathrm{S}(\nu_F \,|\, \gamma)$, where $C_F, \widetilde{C}_F > 0$ depend on integrability properties of F, the carré du champ operators $\Gamma(F_i, F_j)$, $i, j = 1, \ldots, d$, and the inverse of the determinant of the matrix $(\Gamma(F_i, F_j))_{1 \leq i,j \leq d}$. In particular, $\mathrm{H}(\nu_F \,|\, \gamma) \to 0$ as $\mathrm{S}(\nu_F \,|\, \gamma) \to 0$ providing therefore entropic convergence under the Stein discrepancy. The general results obtained here cover not only normal approximation but also gamma approximation.

The inequality (1.7) thus transfers bounds on the Stein discrepancy to entropic bounds. The issue of controlling the Stein discrepancy $\mathrm{S}(\nu_F \,|\, \gamma)$ itself (in terms of moment conditions for example) is not addressed here, and has been the subject of numerous recent studies around the so-called Nualart-Peccati fourth moment theorem (cf. [N-P2]). This investigation is in particular well adapted to functionals $F = (F_1, \ldots, F_d)$ whose coordinates are eigenfunctions of the underlying Markov generator. See [A-C-P, A-M-P, L3] for several results in this direction and [N-P2, Chapters 5-6] for a detailed discussion of estimates on $\mathrm{S}(\nu_F \,|\, \gamma)$ that are available for random vectors F living on the Wiener space.

The structure of the paper thus consists of two main parts, the first one devoted to the new HSI and WSH inequalities, the second one to an investigation of entropic bounds via the Stein discrepancy. Section 2 is devoted to the proof and discussions of the HSI inequality in the Gaussian case, with a first sample of illustrations and applications to convergence to equilibrium and measure concentration. In Section 3, we investigate connections between the Stein discrepancy, Wasserstein distances and transportation cost inequalities, in particular the HWI inequality, and establish the WSH inequality. Extensions of the HSI inequality to more general distributions arising as invariant probability measures of second order differential operators are addressed in

Section 4. The second part consists of Section 5 which develops a general methodology (in the context of Markov Triples) to reach entropic bounds on densities of families of functionals under conditions which do not necessarily involve the Fisher information.

2 Logarithmic Sobolev inequality and Stein discrepancy

Throughout this section, we fix an integer $d \geq 1$ and let $\gamma = \gamma^d$ indicate the standard Gaussian measure on the Borel sets of $\mathbb{R}^d$.

2.1 Stein kernel and discrepancy

Let ν be a probability measure on the Borel sets of $\mathbb{R}^d$. In view of the forthcoming definitions, we shall always assume (without loss of generality) that ν is centered, that is, $\int_{\mathbb{R}^d} x_j \, d\nu(x) = 0$, $j = 1, \dots, d$.

As alluded to in the introduction, a measurable matrix-valued map on $\mathbb{R}^d$

$$x \mapsto \tau_\nu(x) = \big\{ \tau_\nu^{ij}(x) : i, j = 1, \dots, d \big\}$$

is said to be a Stein kernel for ν if $\tau_\nu^{ij} \in \mathrm{L}^1(\nu)$ for every i, j and, for every smooth $\varphi : \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\mathbb{R}^d} x \cdot \nabla \varphi \, d\nu = \int_{\mathbb{R}^d} \big\langle \tau_\nu, \mathrm{Hess}(\varphi) \big\rangle_{\mathrm{HS}} \, d\nu. \tag{2.1}$$

Observe from (2.1) that, without loss of generality, one may and will assume in the sequel that $\tau_\nu^{ij}(x) = \tau_\nu^{ji}(x)$ ν-a.e., $i, j = 1, \dots, d$. Also, by choosing $\varphi = x_i$, $i = 1, \dots, d$, in (2.1) one sees that, if ν admits a Stein kernel, then ν is necessarily centered. Moreover, by selecting $\varphi = x_i x_j$, $i, j = 1, \dots, d$, and since $\tau_\nu^{ij} = \tau_\nu^{ji}$,

$$\int_{\mathbb{R}^d} x_i x_j \, d\nu = \int_{\mathbb{R}^d} \tau_\nu^{ij} d\nu, \quad i, j = 1, \dots, d$$

(and in particular ν has finite second moments).

Remark 2.1. (a) Let $d = 1$ and assume that ν has a density ρ with respect to the Lebesgue measure on $\mathbb{R}$. In this case, it is easily seen that, whenever it exists, the Stein kernel τ_ν is uniquely determined (up to sets of zero Lebesgue measure). Moreover, under standard regularity assumptions on ρ, one deduces from integration by parts that a version of τ_ν is given by

$$\tau_\nu(x) = \frac{1}{\rho(x)} \int_x^\infty y \rho(y) dy \tag{2.2}$$

for x inside the support of ρ.

(b) In dimension $d \geq 2$, a Stein kernel τ_ν may not be unique – see [N-P-S2, Appendix A].

(c) It is important to notice that, in dimension $d \geq 2$, the definition (2.1) of Stein kernel is actually weaker than the one used in [N-P-S1, N-P-S2]. Indeed, in those references a Stein kernel τ_ν is required to satisfy the stronger 'vector' (as opposed to the trace identity (2.1)) relation

$$\int_{\mathbb{R}^d} x \, \varphi \, d\nu \, = \int_{\mathbb{R}^d} \tau_\nu \nabla \varphi \, d\nu \tag{2.3}$$

for every smooth test function $\varphi : \mathbb{R}^d \to \mathbb{R}$. The definition (2.1) of a Stein kernel adopted in the present paper allows one to establish more transparent connections between normal and non-normal approximations, such as the ones explored in Section 4. Observe that it will be nevertheless necessary to use Stein kernels in the strong sense (2.3) when dealing with Wasserstein distances of order $\neq 2$ in Section 3.2.

Definition (2.1) is directly inspired by the Gaussian integration by parts formula according to which

$$\int_{\mathbb{R}^d} x \cdot \nabla \varphi \, d\gamma \, = \int_{\mathbb{R}^d} \Delta \varphi \, d\gamma \, = \int_{\mathbb{R}^d} \big\langle \mathrm{Id}, \mathrm{Hess}(\varphi) \big\rangle_{\mathrm{HS}} \, d\nu \tag{2.4}$$

so that the proximity of τ_ν with the identity matrix Id indicates that ν should be close to γ. In particular, it should be clear that the notion of Stein kernel in the sense of (2.1) is motivated by normal approximation. Section 4 will introduce analogous definitions adapted to the target measure in the context of the generator approach to Stein's method. Whenever a Stein kernel exists, we consider to this task the quantity, called Stein discrepancy of ν with respect to γ in the introduction,

$$\mathrm{S}\big(\nu \,|\, \gamma\big) \, = \, \|\tau_\nu - \mathrm{Id}\|_{2,\nu} \, = \, \bigg(\int_{\mathbb{R}^d} \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}}^2 \, d\nu \bigg)^{1/2}.$$

(Note that $\mathrm{S}(\nu \,|\, \gamma)$ may be infinite if one of the τ_ν^{ij}'s is not in $\mathrm{L}^2(\nu)$.) Whenever $\mathrm{S}(\nu \,|\, \gamma) = 0$, then $\nu = \gamma$ since τ_ν is the identity matrix (see e.g. [N-P2, Lemma 4.1.3]). Observe also that if C denotes the covariance matrix of ν, then

$$\mathrm{S}^2\big(\nu \,|\, \gamma\big) \, = \sum_{i,j=1}^{d} \mathrm{Var}_\nu \, (\tau_\nu^{ij}) + \|C - \mathrm{Id}\|_{\mathrm{HS}}^2, \tag{2.5}$$

where Var_ν indicates the variance under the probability measure ν.

2.2 The Gaussian HSI inequality

As before, write $d\nu = hd\gamma$ to indicate a centered probability measure on $\mathbb{R}^d$ which is absolutely continuous with density h with respect to the standard Gaussian distribution γ. We assume that there exists a Stein kernel τ_ν for ν as defined in (2.1) of the preceding section.

The following result emphasizes the Gaussian HSI inequality connecting entropy H, Stein discrepancy S and Fisher information I. In the statement, we use the conventions $0 \log(1 + \frac{s}{0}) = 0$ and $\infty \log(1 + \frac{s}{\infty}) = s$ for every $s \in [0, \infty]$, and $r \log(1 + \frac{\infty}{r}) = \infty$ for every $r \in (0, \infty)$.

Theorem 2.2 (Gaussian HSI inequality)**.** *For any centered probability measure $d\nu = hd\gamma$ on $\mathbb{R}^d$ with smooth density h with respect to γ,*

$$\mathrm{H}\big(\nu \,|\, \gamma\big) \,\leq\, \frac{1}{2}\, \mathrm{S}^2\big(\nu \,|\, \gamma\big) \log\left(1 + \frac{\mathrm{I}(\nu \,|\, \gamma)}{\mathrm{S}^2(\nu \,|\, \gamma)}\right). \tag{2.6}$$

Since $r \log\big(1 + \frac{s}{r}\big) \leq s$ for every $r > 0$, $s \geq 0$, the HSI inequality (2.6) improves upon the standard logarithmic Sobolev inequality (1.1). It may be observed also that the HSI inequality immediately produces the (classical) equality case in this logarithmic Sobolev inequality. Indeed, due to the centering hypothesis, equality is achieved only for the Gaussian measure γ itself (if not, recenter first ν so that the only extremals of (1.1) have densities $e^{m \cdot x - |m|^2/2}$, $m \in \mathbb{R}^d$, with respect to γ). To this task, assume by contradiction that $\mathrm{S}(\nu \,|\, \gamma) > 0$. Then, if $\mathrm{H}(\nu \,|\, \gamma) = \frac{1}{2}\, \mathrm{I}(\nu \,|\, \gamma)$, the HSI inequality (2.6) yields

$$\frac{\mathrm{I}(\nu \,|\, \gamma)}{\mathrm{S}^2(\nu \,|\, \gamma)} \,\leq\, \log\left(1 + \frac{\mathrm{I}(\nu \,|\, \gamma)}{\mathrm{S}^2(\nu \,|\, \gamma)}\right)$$

from which $\mathrm{I}(\nu \,|\, \gamma) = 0$, and therefore $\nu = \gamma$, which is in contrast with the assumption $\mathrm{S}(\nu \,|\, \gamma) > 0$. As a consequence, we infer that $\mathrm{S}(\nu \,|\, \gamma) = 0$, from which it follows that $\nu = \gamma$.

The HSI inequality (2.6) may be extended to the case of a centered Gaussian distribution on $\mathbb{R}^d$ with a general non-degenerate covariance matrix C. We denote such a measure by γ_C, so that $\gamma = \gamma_{\mathrm{Id}}$. We also denote by $\|C\|_{\mathrm{op}}$ the operator norm of C, that is, $\|C\|_{\mathrm{op}}$ is the largest eigenvalue of C.

Corollary 2.3 (Gaussian HSI inequality, general covariance)**.** *Let γ_C be as above (with C non-singular), and let $d\nu = hd\gamma_C$ be centered with smooth probability density h with respect to γ_C. Assume that ν admits a Stein kernel τ_ν in the sense of* (2.1)*. Then,*

$$\mathrm{H}\big(\nu \,|\, \gamma_C\big) \,\leq\, \frac{1}{2} \left\|C^{-\frac{1}{2}} \tau_\nu \, C^{-\frac{1}{2}} - \mathrm{Id}\right\|_{2,\nu}^2 \log\left(1 + \frac{\|C\|_{\mathrm{op}}\, \mathrm{I}(\nu \,|\, \gamma_C)}{\|C^{-\frac{1}{2}} \tau_\nu \, C^{-\frac{1}{2}} - \mathrm{Id}\|_{2,\nu}^2}\right),$$

where $C^{-\frac{1}{2}}$ denotes the unique symmetric non-singular matrix such that $\big(C^{-\frac{1}{2}}\big)^2 = C^{-1}$.

Corollary 2.3 is easily deduced from Theorem 2.2 and details are left to the reader. The argument simply uses that if M is the unique non-singular symmetric matrix such that $C = M^2$, then $\mathrm{H}(\nu \,|\, \gamma_C) = \mathrm{H}(\nu^0 \,|\, \gamma)$ where $d\nu^0(x) = h(Mx)d\gamma(x)$.

2.3 Proof of the Gaussian HSI inequality

According to our conventions, if either $\mathrm{S}(\nu \,|\, \gamma)$ or $\mathrm{I}(\nu \,|\, \gamma)$ is infinite, then (2.6) coincides with the logarithmic Sobolev inequality (1.1). On the other hand, if $\mathrm{S}(\nu \,|\, \gamma)$ or $\mathrm{I}(\nu \,|\, \gamma)$ equals zero, then $\nu = \gamma$, and therefore $\mathrm{H}(\nu \,|\, \gamma) = 0$. It follows that, in order to prove (2.6), we can assume without loss of generality that $\mathrm{S}(\nu \,|\, \gamma)$ and $\mathrm{I}(\nu \,|\, \gamma)$ are both non-zero and finite.

The proof of Theorem 2.2 is based on the heat flow interpolation along the Ornstein-Uhlenbeck semigroup. We recall a few basic facts in this regard, and refer the reader to e.g. [B-G-L, Section 2.7.1] for any unexplained definition or result. Let thus $(P_t)_{t\geq 0}$ be the Ornstein-Uhlenbeck semigroup on $\mathbb{R}^d$ with infinitesimal generator

$$\mathcal{L}f \,=\, \Delta f - x\cdot\nabla f \,=\, \sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2} - \sum_{i=1}^d x_i \frac{\partial f}{\partial x_i} \tag{2.7}$$

(acting on smooth functions f), invariant and symmetric with respect to γ. We shall often use the fact that the action of P_t on smooth functions $f : \mathbb{R}^d \to \mathbb{R}$ admits the integral representation (sometimes called *Mehler's formula*)

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

The semigroup is trivially extended to vector-valued functions $f : \mathbb{R}^d \to \mathbb{R}^d$. In particular, if $f : \mathbb{R}^d \to \mathbb{R}$ is smooth enough,

$$\nabla P_t f \,=\, e^{-t} P_t(\nabla f). \tag{2.8}$$

One technical important property (part of the much more general Bismut formulas in a geometric context [Bi, B-G-L]) is the identity, between vectors in $\mathbb{R}^d$,

$$P_t(\nabla f)(x) \,=\, \frac{1}{\sqrt{1-e^{-2t}}} \int_{\mathbb{R}^d} y\, f\big(e^{-t}x + \sqrt{1-e^{-2t}}\, y\big) d\gamma(y), \tag{2.9}$$

owing to a standard integration by parts of the Gaussian density.

The generator $\mathcal{L}$ is a diffusion and satisfies the integration by parts formula

$$\int_{\mathbb{R}^d} f\, \mathcal{L}g\, d\gamma \,=\, -\int_{\mathbb{R}^d} \nabla f\cdot\nabla g\, d\gamma \tag{2.10}$$

on smooth functions $f, g : \mathbb{R}^d \to \mathbb{R}$. In particular, given the smooth probability density h with respect to γ,

$$\mathrm{I}_\gamma(h) \,=\, \int_{\mathbb{R}^d} \frac{|\nabla h|^2}{h}\, d\gamma \,=\, \int_{\mathbb{R}^d} |\nabla(\log h)|^2 h d\gamma \,=\, -\int_{\mathbb{R}^d} \mathcal{L}(\log h) h d\gamma.$$

As $d\nu = hd\gamma$, setting $v = \log h$,

$$\mathrm{I}\big(\nu\,|\,\gamma\big) \,=\, \mathrm{I}_\gamma(h) \,=\, \int_{\mathbb{R}^d} |\nabla v|^2 d\nu \,=\, -\int_{\mathbb{R}^d} \mathcal{L}v\, d\nu. \tag{2.11}$$

(These expressions should actually be considered for $h + \varepsilon$ as $\varepsilon \to 0$.) Using $P_t h$ instead of h in the previous relations and writing $v_t = \log P_t h$, one deduces from the symmetry of P_t that

$$\mathrm{I}\big(\nu^t\,|\,\gamma\big) \,=\, \mathrm{I}_\gamma(P_t h) \,=\, \int_{\mathbb{R}^d} \frac{|\nabla P_t h|^2}{P_t h}\, d\gamma \,=\, -\int_{\mathbb{R}^d} \mathcal{L}v_t\, P_t h d\gamma \,=\, -\int_{\mathbb{R}^d} \mathcal{L}P_t v_t\, d\nu. \tag{2.12}$$

Recall finally that if $d\nu^t = P_t h d\gamma$, $t \geq 0$ (with $\nu^0 = \nu$ and $\nu^t \to \gamma$), the classical *de Bruijn's formula* (see e.g. [B-G-L, Proposition 5.2.2]) indicates that

$$\frac{d}{dt}\, \mathrm{H}\big(\nu^t\,|\,\gamma\big) = -\,\mathrm{I}\big(\nu^t\,|\,\gamma\big). \tag{2.13}$$

Theorem 2.2 will follow from the next Proposition 2.4. In this proposition, *(i)* corresponds to the integral version of (2.13) whereas *(ii)* describes the well-known exponential decay of the Fisher information along the Ornstein-Uhlenbeck semigroup. This decay actually yields the logarithmic Sobolev inequality (1.1), see [B-G-L, Section 5.7]. The new third point *(iii)* is a reformulation of [N-P-S1, Theorem 2.1] for which we provide a self-contained proof. It describes an alternate bound on the Fisher information along the semigroup in terms of the Stein discrepancy for values of $t > 0$ away from 0. It is the combination of *(ii)* and *(iii)* which will produce the HSI inequality. Point (*iv*) will be needed in the forthcoming proof of the WSH inequality (1.4), as well as in the proof of Proposition 3.1 providing a direct bound of the Wasserstein distance W_2 by the Stein discrepancy.

Proposition 2.4. *Under the above notation and assumptions, denote by* τ_ν *a Stein kernel of* $d\nu = hd\gamma$. *For every* $t > 0$, *recall* $d\nu^t = P_t h\, d\gamma$, *and write* $v_t = \log P_t h$. *Then,*

(i) (Integrated de Bruijn's formula)

$$\mathrm{H}\big(\nu\,|\,\gamma\big) \,=\, \mathrm{Ent}_\gamma(h) \,=\, \int_0^\infty \mathrm{I}_\gamma(P_t h) dt. \tag{2.14}$$

(ii) (Exponential decay of Fisher information) *For every* $t \geq 0$,

$$\mathrm{I}\big(\nu^t\,|\,\gamma\big) \,=\, \mathrm{I}_\gamma(P_t h) \,\leq\, e^{-2t}\, \mathrm{I}_\gamma(h) \,=\, e^{-2t}\, \mathrm{I}\big(\nu^0\,|\,\gamma\big). \tag{2.15}$$

(iii) *For every* $t > 0$,

$$\mathrm{I}_\gamma(P_t h) \;=\; \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Big[\big(\tau_\nu(x)-\mathrm{Id}\big)y\cdot\nabla v_t\big(e^{-t}x+\sqrt{1-e^{-2t}}\,y\big)\Big]d\nu(x)d\gamma(y). \tag{2.16}$$

As a consequence, for every $t > 0$,

$$\mathrm{I}\big(\nu^t\,|\,\gamma\big) \;=\; \mathrm{I}_\gamma(P_t h) \;\leq\; \frac{e^{-4t}}{1-e^{-2t}}\,\|\tau_\nu-\mathrm{Id}\|_{2,\nu} \;=\; \frac{e^{-4t}}{1-e^{-2t}}\,\mathrm{S}^2\big(\nu^0\,|\,\gamma\big). \tag{2.17}$$

(iv) (Exponential decay of Stein discrepancy) *For every* $t \geq 0$,

$$\mathrm{S}\big(\nu^t\,|\,\gamma\big) \;\leq\; e^{-2t}\,\mathrm{S}\big(\nu^0\,|\,\gamma\big). \tag{2.18}$$

Proof. In view of the preceding discussion, only the proofs of *(iii)* and *(iv)* need to be detailed. Throughout the various analytical arguments below, it may be assumed that the density h is regular enough, the final conclusions being then reached by approximation arguments as e.g. in [O-V, B-G-L]. Starting with *(iii)*, use (2.12) and the definition (2.1) of τ_ν to write, for any $t > 0$,

$$\begin{aligned}\mathrm{I}_\gamma(P_t h) \;=\; -\int_{\mathbb{R}^d}\mathcal{L}P_t v_t\,d\nu \;&=\; -\int_{\mathbb{R}^d}\big[\Delta P_t v_t - x\cdot\nabla P_t v_t\big]d\nu \\ &=\; \int_{\mathbb{R}^d}\big\langle\tau_\nu-\mathrm{Id},\mathrm{Hess}(P_t v_t)\big\rangle_{\mathrm{HS}}\,d\nu.\end{aligned} \tag{2.19}$$

Now, for all $i, j = 1, \ldots, d$, by (2.8) and (2.9),

$$\partial_{ij}P_t v_t(x) = e^{-2t}P_t(\partial_{ij}v_t)(x) \;=\; \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\int_{\mathbb{R}^d} y_i\,\frac{\partial v_t}{\partial x_j}\big(e^{-t}x+\sqrt{1-e^{-2t}}\,y\big)d\gamma(y).$$

Hence

$$\begin{aligned}&\int_{\mathbb{R}^d}\big\langle\tau_\nu-\mathrm{Id},\mathrm{Hess}(P_t v_t)\big\rangle_{\mathrm{HS}}\,d\nu \\ &\qquad= \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Big[\big(\tau_\nu(x)-\mathrm{Id}\big)y\cdot\nabla v_t\big(e^{-t}x+\sqrt{1-e^{-2t}}\,y\big)\Big]d\nu(x)d\gamma(y)\end{aligned}$$

which is (2.16). To deduce the estimate (2.17), it suffices to apply (twice) the Cauchy-Schwarz inequality to the right-hand side of (2.16) in such a way that, by integrating out the y variable,

$$\begin{aligned}\mathrm{I}_\gamma(P_t h) \;&\leq\; \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\big|\big(\tau_\nu(x)-\mathrm{Id}\big)y\big|\big|\nabla v_t\big(e^{-t}x+\sqrt{1-e^{-2t}}\,y\big)\big|d\nu(x)d\gamma(y) \\ &\leq\; \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\bigg(\int_{\mathbb{R}^d}\|\tau_\nu-\mathrm{Id}\|_{\mathrm{HS}}^2\,d\nu\bigg)^{1/2}\bigg(\int_{\mathbb{R}^d}P_t\big(|\nabla v_t|^2\big)d\nu\bigg)^{1/2}.\end{aligned}$$

Since

$$\int_{\mathbb{R}^d} P_t\big(|\nabla v_t|^2\big)d\nu \,=\, \int_{\mathbb{R}^d} P_t\big(|\nabla v_t|^2\big)hd\gamma \,=\, \int_{\mathbb{R}^d} |\nabla v_t|^2 P_t h d\gamma \,=\, \mathrm{I}_\gamma(P_t h)$$

by symmetry of P_t, the proof of (2.17) is complete.

Let us now turn to the proof of (2.18). For any smooth test function φ on $\mathbb{R}^d$, by symmetry of $(P_t)_{t\geq 0}$, for any $t \geq 0$,

$$\int_{\mathbb{R}^d} x\cdot\nabla\varphi\, d\nu^t \,=\, \int_{\mathbb{R}^d} x\cdot\nabla\varphi\; P_t h\, d\gamma \,=\, \int_{\mathbb{R}^d} P_t(x\cdot\nabla\varphi)hd\gamma \,=\, \int_{\mathbb{R}^d} P_t(x\cdot\nabla\varphi)d\nu.$$

By the integral representation of P_t,

$$\begin{aligned}\int_{\mathbb{R}^d} P_t(x\cdot\nabla\varphi)d\nu \,=\, & e^{-t}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} x\cdot\nabla\varphi\big(e^{-t}x+\sqrt{1-e^{-2t}}\,y\big)d\nu(x)d\gamma(y)\\ &+\sqrt{1-e^{-2t}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} y\cdot\nabla\varphi\big(e^{-t}x+\sqrt{1-e^{-2t}}\,y\big)d\nu(x)d\gamma(y).\end{aligned}$$

Use now the definition of τ_ν in the x variable and integration by parts in the y variable to get that

$$\begin{aligned}\int_{\mathbb{R}^d} P_t(x\cdot\nabla\varphi)d\nu \,=\, & e^{-2t}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\big\langle\tau_\nu(x),\big(\mathrm{Hess}(\varphi)\big)\big(e^{-t}x+\sqrt{1-e^{-2t}}\,y\big)\big\rangle_{\mathrm{HS}}\, d\nu(x)d\gamma(y)\\ &+(1-e^{-2t})\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Delta\varphi\big(e^{-t}x+\sqrt{1-e^{-2t}}\,y\big)d\nu(x)d\gamma(y)\\ =\, & e^{-2t}\int_{\mathbb{R}^d}\big\langle\tau_\nu, P_t\big(\mathrm{Hess}(\varphi)\big)\big\rangle_{\mathrm{HS}}\, d\nu+(1-e^{-2t})\int_{\mathbb{R}^d} P_t(\Delta\varphi)d\nu\\ =\, & e^{-2t}\int_{\mathbb{R}^d}\big\langle P_t(h\tau_\nu),\mathrm{Hess}(\varphi)\big\rangle_{\mathrm{HS}}\, d\gamma+(1-e^{-2t})\int_{\mathbb{R}^d}\Delta\varphi\, P_t h\, d\gamma.\end{aligned}$$

As a consequence, a Stein kernel for ν^t is

$$\tau_{\nu^t} \,=\, e^{-2t}\,\frac{P_t(h\tau_\nu)}{P_t h}+(1-e^{-2t})\,\mathrm{Id}. \tag{2.20}$$

Therefore,

$$\int_{\mathbb{R}^d}\|\tau_{\nu^t}-\mathrm{Id}\|^2_{\mathrm{HS}}\, d\nu^t \,=\, e^{-4t}\int_{\mathbb{R}^d}\frac{\|P_t(h(\tau_\nu-\mathrm{Id}))\|^2_{\mathrm{HS}}}{P_t h}\, d\gamma.$$

By the Cauchy-Schwarz inequality along P_t,

$$\big\|P_t\big(h(\tau_\nu-\mathrm{Id})\big)\big\|^2_{\mathrm{HS}} \,\leq\, P_t\big(h\|\tau_\nu-\mathrm{Id}\|^2_{\mathrm{HS}}\big)P_t h.$$

Hence,

$$\begin{aligned}\int_{\mathbb{R}^d}\|\tau_{\nu^t}-\mathrm{Id}\|^2_{\mathrm{HS}}\, d\nu^t \,\leq\, & e^{-4t}\int P_t\big(h\|\tau_\nu-\mathrm{Id}\|^2_{\mathrm{HS}}\big)d\gamma\\ =\, & e^{-4t}\int_{\mathbb{R}^d}\|\tau_\nu-\mathrm{Id}\|^2_{\mathrm{HS}}\, hd\gamma \,=\, e^{-4t}\int_{\mathbb{R}^d}\|\tau_\nu-\mathrm{Id}\|^2_{\mathrm{HS}}\, d\nu,\end{aligned}$$

that is the announced result *(iv)*. Proposition 2.4 is established.

□

Remark 2.5. For every $t > 0$, it is easily checked that the mapping $x \mapsto \tau_{\nu^t}(x)$ appearing in (2.20) admits the probabilistic representation

$$\tau_{\nu^t}(x) \,=\, \mathbb{E}\big[e^{-2t}\tau_\nu(F) + (1-e^{-2t})\,\mathrm{Id}\,|\,F_t = x\big] \quad d\nu^t(x)\text{–a.e.}, \tag{2.21}$$

where, on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, F has distribution ν and $F_t = e^{-t}F + \sqrt{1-e^{-2t}}Z$, with Z a d-dimensional vector with distribution γ, independent of F.

We are now in a position to prove Theorem 2.2.

Proof of Theorem 2.2. As announced, on the basis of the interpolation (2.14), we apply (2.15) and (2.17) respectively to bound the Fisher information $\mathrm{I}_\gamma(P_t h)$ for t around 0 and away from 0. We thus get, for every $u > 0$,

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

Optimizing in u (set $1 - e^{-2u} = r \in (0,1)$) concludes the proof. □

Remark 2.6. It is worth mentioning that a slight modification of the proof of *(iii)* in Proposition 2.4 leads to the improved form of the exponential decay (2.15) of the Fisher information

$$\mathrm{I}\big(\nu^t\,|\,\gamma\big) \,\leq\, \frac{e^{-2t}\,\mathrm{S}^2(\nu\,|\,\gamma)\,\mathrm{I}(\nu\,|\,\gamma)}{\mathrm{S}^2(\nu\,|\,\gamma) + (e^{2t}-1)\,\mathrm{I}(\nu\,|\,\gamma)}\,. \tag{2.22}$$

As for the classical logarithmic Sobolev inequality, the inequality (2.22) may be integrated along de Bruijin's formula (2.13) towards the better, although less tractable, HSI inequality

$$\mathrm{H} \,\leq\, \frac{\mathrm{S}^2\,\mathrm{I}}{2(\mathrm{S}^2 - \mathrm{I})}\left(1 + \frac{\mathrm{I}}{\mathrm{S}^2 - \mathrm{I}}\,\log\Big(\frac{\mathrm{I}}{\mathrm{S}^2}\Big)\right)$$

(understood in the limit as $\mathrm{S}^2 = \mathrm{I}$), where $\mathrm{H} = \mathrm{H}(\nu\,|\,\gamma)$, $\mathrm{S} = \mathrm{S}(\nu\,|\,\gamma)$ and $\mathrm{I} = \mathrm{I}(\nu\,|\,\gamma)$.

Together with the de Bruijn identity (2.13), the classical logarithmic Sobolev inequality (1.1) ensures the exponential decay in $t \geq 0$ of the relative entropy

$$\mathrm{H}\big(\nu^t\,|\,\gamma\big) \,\leq\, e^{-2t}\,\mathrm{H}\big(\nu^0\,|\,\gamma\big) \tag{2.23}$$

along the Ornstein-Uhlenbeck semigroup (cf. e.g. [B-G-L, Theorem 5.2.1]). The new HSI produces a reinforcement of this exponential convergence to equilibrium under finiteness of the Stein discrepancy.

Corollary 2.7 (Exponential decay of entropy from HSI). *Let ν with Stein discrepancy* $\mathrm{S}(\nu \,|\, \gamma) = \mathrm{S}$. *For any $t \geq 0$,*

$$\mathrm{H}\big(\nu^t \,|\, \gamma\big) \,\leq\, \frac{e^{-4t}}{e^{-2t} + \frac{1-e^{-2t}}{\mathrm{S}^2}\,\mathrm{H}(\nu^0 \,|\, \gamma)}\,\mathrm{H}\big(\nu^0 \,|\, \gamma\big) \,\leq\, \frac{e^{-4t}}{1-e^{-2t}}\,\mathrm{S}^2\big(\nu^0 \,|\, \gamma\big). \tag{2.24}$$

Proof. Together with (2.18) and since $r \mapsto r \log\big(1 + \frac{s}{r}\big)$ is increasing for any fixed s, the HSI inequality applied to ν^t implies that

$$\mathrm{H}\big(\nu^t \,|\, \gamma\big) \,\leq\, \frac{e^{-4t}\,\mathrm{S}^2}{2}\,\log\bigg(1 + \frac{e^{4t}\,\mathrm{I}(\nu^t \,|\, \gamma)}{\mathrm{S}^2}\bigg).$$

Set $U(t) = \frac{e^{4t}}{\mathrm{S}^2}\,\mathrm{H}(\nu^t \,|\, \gamma)$, $t \geq 0$, so that by (2.13), $U' = 4U - \frac{e^{4t}}{\mathrm{S}^2}\,\mathrm{I}(\nu^t \,|\, \gamma)$. The latter inequality therefore rewrites as

$$e^{2U} - 1 - 4U \,\leq\, -U'. \tag{2.25}$$

Since $e^r - 1 - r \geq \frac{r^2}{2}$ for $r \geq 0$, this inequality may be relaxed into $-2U + 2U^2 \,\leq\, -U'$. Setting $V(t) = e^{-2t}U(t)$, $t \geq 0$, it follows that $2e^{2t}V^2(t) \leq -V'(t)$ so that, after integration,

$$e^{2t} - 1 \,\leq\, \frac{1}{V(t)} - \frac{1}{V(0)}\,.$$

By definition of V, this inequality amounts to the conclusion of Corollary 2.7 and the proof is complete. □

2.4 Stein discrepancy and concentration inequalities

This paragraph investigates another feature of Stein's discrepancy applied to concentration inequalities. It is of course by now classical that logarithmic Sobolev inequalities may be used as a robust tool towards (Gaussian) concentration inequalities (cf. e.g. [L2, B-L-M]). For example, for the standard Gaussian measure γ itself, the Herbst argument yields that for any 1-Lipschitz function $u : \mathbb{R}^d \to \mathbb{R}$ with mean zero,

$$\gamma(u \geq r) \,\leq\, e^{-r^2/2}, \quad r \geq 0. \tag{2.26}$$

Equivalently (up to numerical constants) in terms of moment growth,

$$\bigg(\int_{\mathbb{R}^d} |u|^p d\gamma\bigg)^{1/p} \,\leq\, C\sqrt{p}\,, \quad p \geq 1. \tag{2.27}$$

Here, we describe how to directly implement Stein's discrepancy into such concentration inequalities on the basis of the principle leading to the HSI inequality. If ν is a probability measure on the Borel sets of $\mathbb{R}^d$ with Stein kernel τ_ν, set for $p \geq 1$,

$$\mathrm{S}_p\big(\nu \,|\, \gamma\big) \,=\, \bigg(\int_{\mathbb{R}^d} \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}}^p \, d\nu\bigg)^{1/p}.$$

Hence $\mathrm{S}_2(\nu\,|\,\gamma) = \mathrm{S}(\nu\,|\,\gamma)$ is the Stein discrepancy as defined earlier. Recall $\|\cdot\|_{\mathrm{op}}$ the operator norm on the $d \times d$ matrices.

Theorem 2.8 (Moment bounds and Stein discrepancy)**.** *Let ν have Stein kernel τ_ν. There exists a numerical constant $C > 0$ such that for every 1-Lipschitz function $u : \mathbb{R}^d \to \mathbb{R}$ with $\int_{\mathbb{R}^d} u d\nu = 0$, and every $p \geq 2$,*

$$\bigg(\int_{\mathbb{R}^d} |u|^p d\nu \bigg)^{1/p} \leq C \bigg[\mathrm{S}_p\big(\nu\,|\,\gamma\big) + \sqrt{p} \, \bigg(\int_{\mathbb{R}^d} \|\tau_\nu\|_{\mathrm{op}}^{p/2} \, d\nu \bigg)^{1/p} \bigg]. \tag{2.28}$$

Before turning to the proof of this result, let us comment on its measure concentration content. One first important aspect is that the constant C is dimension free. When $\nu = \gamma$, (2.28) exactly fits the Gaussian case (2.27). In general, the moment growth in p describes various concentration regimes of ν (cf. [L2, Section 1.3], [B-L-M, Chapter 14]) according to the growth of the p-Stein discrepancy $\mathrm{S}_p(\nu\,|\,\gamma)$.

In view of the elementary estimate $\|\tau_\nu\|_{\mathrm{op}} \leq 1 + \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}}$, the conclusion (2.28) immediately yields the moment growth (1.6) emphasized in the introduction

$$\bigg(\int_{\mathbb{R}^d} |u|^p d\nu \bigg)^{1/p} \leq C \Big(\mathrm{S}_p\big(\nu\,|\,\gamma\big) + \sqrt{p} + \sqrt{p} \, \sqrt{\mathrm{S}_p\big(\nu\,|\,\gamma\big)} \, \Big).$$

Note that there is already an interest to write this bound for $p = 2$,

$$\mathrm{Var}_\nu(u) \leq C \Big(1 + \mathrm{S}\big(\nu\,|\,\gamma\big) + \mathrm{S}^2\big(\nu\,|\,\gamma\big) \Big).$$

Together with E. Milman's Lipschitz characterization of Poincaré inequalities for log-concave measures [M], it shows that the Stein discrepancy $\mathrm{S}(\nu\,|\,\gamma)$ with respect to the standard Gaussian measure is another control of the spectral properties in this class of measures.

Similar inequalities hold for arbitrary covariances by suitably adapting the Stein kernel as in Corollary 2.3.

A main example of illustration of Theorem 2.8 concerns sums of independent random vectors. Consider X a mean zero random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^d$, and $X_1, \ldots, X_n$ independent copies of X. Assume that the law ν of X admits a Stein kernel τ_ν. Setting $T_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$, it is easily seen by independence that, as matrices, a Stein kernel τ_{ν_n} of the law ν_n of T_n satisfies

$$\tau_{\nu_n}(T_n) = \mathbb{E}\bigg[\frac{1}{n} \sum_{k=1}^n \tau_\nu(X_k) \,\Big|\, T_n \bigg].$$

Hence,

$$\mathrm{S}_p\big(\nu_n\,|\,\gamma\big) \leq \mathbb{E}\bigg[\bigg\| \frac{1}{n} \sum_{k=1}^n \big[\tau_\nu(X_k) - \mathrm{Id}\big] \bigg\|_{\mathrm{HS}}^p \bigg]^{1/p}.$$

By the triangle inequality, the latter is bounded from above by

$$\mathbb{E}\big[\|\tau_\nu(X) - \mathrm{Id}\|_{\mathrm{HS}}^p\big]^{1/p} \,=\, \mathrm{S}_p\big(\nu\,|\,\gamma\big) \,=\, \mathrm{S}_p$$

which produces a first bound of interest. If it is assumed in addition that the covariance matrix of X is the identity, we may use classical inequalities for sums of independent centered random vectors (in Euclidean space) to the family $\tau_\nu(X_k) - \mathrm{Id}$, $k = 1, \ldots, n$. Hence, by for example Rosenthal's inequality (see e.g. [B-L-M, M-J-C-F-T]), for $p \geq 2$,

$$\mathbb{E}\bigg[\bigg\|\frac{1}{n}\sum_{k=1}^{n}\big[\tau_\nu(X_k) - \mathrm{Id}\big]\bigg\|_{\mathrm{HS}}^p\bigg]^{1/p} \,\leq\, K_p\, n^{-1/2}\, \mathrm{S}_p.$$

Together with (1.6), it yields a growth control of the moments of $u(T_n)$ for any Lipschitz function u, and therefore concentration of the law of T_n. More precisely, and since it is known that $K_p = O(p)$, for any 1-Lipschitz function $u : \mathbb{R}^d \to \mathbb{R}$ such that $\mathbb{E}(u(T_n)) = 0$,

$$\mathbb{E}\big[\big|u(T_n)\big|^p\big]^{1/p} \,\leq\, C\sqrt{p}\,\Big(1 + n^{-1/2}\sqrt{p}\,\mathrm{S}_p + n^{-1/4}\sqrt{p\,\mathrm{S}_p}\,\Big)$$

for some numerical $C > 0$. Note that the bound is optimal both for $\nu = \gamma$ and as $n \to \infty$ describing the standard Gaussian concentration (2.27). By Markov's inequality, optimizing in $p \geq 2$, one deduces that for some numerical $C' > 0$,

$$\mathbb{P}\big(u(T_n) \geq r\big) \,\leq\, C'\, e^{-r^2/C'} \tag{2.29}$$

for all $0 \leq r \leq r_n$ where $r_n \to \infty$ according to the growth of S_p as $p \to \infty$. For example, if $\mathrm{S}_p = O(p^\alpha)$ for some $\alpha > 0$ (see below for such illustrations), then

$$\mathbb{E}\big[\big|u(T_n)\big|^p\big]^{1/p} \,\leq\, C\sqrt{p}$$

(for some possibly different numerical $C > 0$) for every $p \leq n^{\frac{1}{2\alpha+2}}$. By Markov's inequality in this range of p,

$$\mathbb{P}\big(\big|u(T_n)\big| \geq r\big) \,\leq\, \Big(\frac{C\sqrt{p}}{r}\Big)^p,$$

and with $p \sim \frac{r^2}{4C^2}$, the claims follows with r_n of the order of $n^{\frac{1}{4\alpha+4}}$.

For the applications of the concentration inequality (2.29), it is therefore useful to provide a handy set of conditions ensuring a suitable control of (the growth in p of) $\mathrm{S}_p = S_p(\nu\,|\,\gamma)$, that is of the moments of the Stein kernel $\tau_\nu(X)$ of a given random variable X with law ν. The following remark collects families of examples in dimension one. Together with this remark, (2.29) therefore produces with the Stein methodology concentration properties for measures not necessarily satisfying a logarithmic Sobolev inequality. For example, the conclusion may be applied to a vector X with independent coordinates in $\mathbb{R}^d$ each of them of the Pearson class as described in (b) of the following Remark 2.9.

Remark 2.9. For concreteness, we describe two classes of one-dimensional distributions such that the associated Stein kernel has finite moments of all orders. Denote by X a centered real-valued random variable with law ν and Stein kernel τ_ν. Recall from (2.2) of Remark 2.1, that if ν has density ρ with respect to the Lebesgue measure, a version of τ_ν is given by $\tau_\nu(x) = \rho(x)^{-1}\int_x^\infty y\rho(y)dy$ for x inside the support of ρ.

(a) Assume that $\rho(x) = q(x)\frac{e^{-x^2/2}}{\sqrt{2\pi}}$, $x \in \mathbb{R}$, where q is smooth and satisfies the uniform bounds $q(x) \geq c > 0$ et $|q'(x)| \leq C < \infty$ for constants $c, C > 0$. Therefore,

$$\tau_\nu(x) \,=\, 1 + \frac{e^{-x^2/2}}{q(x)}\int_x^\infty q'(y)e^{-y^2/2}dy \,=\, 1 - \frac{e^{-x^2/2}}{q(x)}\int_{-\infty}^x q'(y)e^{-y^2/2}dy.$$

Studying separately the two cases $x > 0$ and $x < 0$, it easily follows that $|\tau_\nu(x) - 1| \leq \frac{\sqrt{2\pi}C}{c}$, and consequently $\mathbb{E}[|\tau_\nu(X)|^r] < \infty$ for every $r > 0$.

(b) Assume that the support of ρ coincides with an open interval of the type (a, b), with $-\infty \leq a < b \leq +\infty$. Say then that the law ν of X is a (centered) member of the *Pearson family* of continuous distributions if the density ρ satisfies the differential equation

$$\frac{\rho'(x)}{\rho(x)} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2}\,, \quad x \in (a, b), \tag{2.30}$$

for some real numbers a_0, a_1, b_0, b_1, b_2. We refer the reader e.g. to [D-Z, Sec. 5.1] for an introduction to the Pearson family. It is a well-known fact that there are basically five families of distributions satisfying (2.30): the centered normal distributions, centered gamma and beta distributions, and distributions that are obtained by centering densities of the type $\rho(x) = Cx^{-\alpha}e^{-\beta/x}$ or $\rho(x) = C(1+x)^{-\alpha}\exp(\beta\arctan(x))$ (C being a suitable normalizing constant). According to [Ste, Theorem 1, p. 65], if τ_ν satisfies

$$\int_0^b \frac{y}{\tau_\nu(y)}\,dy \,=\, +\infty \quad \text{and} \quad \int_a^0 \frac{y}{\tau_\nu(y)}\,dy \,=\, -\infty, \tag{2.31}$$

then $\tau_\nu(x) = \alpha x^2 + \beta x + \gamma$, $x \in (a, b)$ (with α, β, γ real constants) if and only if ν is a member of the Pearson family in the sense that ρ satisfies (2.30) for every $x \in (a, b)$ with $a_0 = \beta$, $a_1 = 2\alpha + 1$, $b_0 = \gamma$, $b_1 = \beta$ and $b_2 = \alpha$. It follows that if ν is centered member of the Pearson family such that (2.31) is satisfied and X has finite moments of all orders, so has $\tau_\nu(X)$. This includes the case of Gaussian, gamma and beta distribution for example.

Further illustrations of Theorem 2.8 may be developed in the general context of eigenfunctions on abstract Markov Triples (E, μ, Γ) as addressed in the forthcoming Section 5. Indeed, let $F : E \to \mathbb{R}$ be an eigenfunction of the underlying diffusion operator

L with eigenvalues $\lambda > 0$ with distribution ν and normalized such that $\int_E F^2 d\mu = 1$. Then, according to Proposition 5.1 below, a version of the Stein kernel is given by

$$\tau_\nu \, = \, \frac{1}{\lambda} \, \mathbb{E}_\mu \big[\Gamma(F) \, | \, F \big]$$

so that

$$\mathrm{S}_p^p \big(\nu \, | \, \gamma \big) \, \leq \, \int_E \bigg| \frac{\Gamma(F)}{\lambda} - 1 \bigg|^p d\mu.$$

In concrete instances, such as Wiener chaos for example, the latter expression may be easily controled so to yield concentration properties of the underlying distribution of F. For example, in the setting of the recent [A-C-P], it may be shown by hypercontractive means that for the Hermite, Laguerre or Jacobi (or mixed ones) chaos structures, for any $p \geq 2$,

$$\mathrm{S}_p^p \big(\nu \, | \, \gamma \big) \, \leq \, C_{p,\lambda} \bigg(\int_E F^4 d\mu - 3 \bigg)^{p/2}.$$

According to the respective growth in p of $C_{p,\lambda}$, concentration properties on F may be achieved.

We turn to the proof of Theorem 2.8.

Proof of Theorem 2.8. We only prove the result for p an even integer, the general case following similarly with some further technicalities. We may also replace the assumption $\int_{\mathbb{R}^d} u d\nu = 0$ by $\int_{\mathbb{R}^d} u d\gamma = 0$ by a simple use of the triangle inequality. Indeed, by Jensen's inequality,

$$\bigg| \int_{\mathbb{R}^d} u \, d\nu - \int_{\mathbb{R}^d} u \, d\gamma \bigg|^p \, \leq \, \int_{\mathbb{R}^d} \bigg| u - \int_{\mathbb{R}^d} u \, d\gamma \bigg|^p d\nu$$

so that if the conclusion (2.28) holds for u satisfying $\int_{\mathbb{R}^d} u d\gamma = 0$, it holds similarly for u satisfying $\int_{\mathbb{R}^d} u d\nu = 0$ with maybe $2C$ instead of C.

We run as in the preceding section the Ornstein-Uhlenbeck semigroup $(P_t)_{t \geq 0}$ with infinitesimal generator $\mathcal{L} = \Delta - x \cdot \nabla$. Let $u : \mathbb{R}^d \to \mathbb{R}$ be 1-Lipschitz, assumed furthermore to be smooth and bounded after a cut-off argument (cf. [L2, Section 1.3] for standard technology in this regard). Let thus $q \geq 1$ be an integer, and set

$$\phi(t) \, = \, \int_{\mathbb{R}^d} (P_t u)^{2q} d\nu, \quad t \geq 0.$$

Under the centering hypothesis $\int_{\mathbb{R}^d} u d\gamma = 0$, $\phi(\infty) = 0$. Differentiating along $(P_t)_{t \geq 0}$

together with the definition of Stein kernel τ_ν yields

$$\begin{aligned}
\phi'(t) &= 2q \int_{\mathbb{R}^d} (P_t u)^{2q-1} \, \mathcal{L} P_t u \, d\nu \\
&= 2q \int_{\mathbb{R}^d} (P_t u)^{2q-1} \, \Delta P_t u \, d\nu - \int_{\mathbb{R}^d} x \cdot \nabla (P_t u)^{2q} d\nu \\
&= 2q \int_{\mathbb{R}^d} (P_t u)^{2q-1} \, \Delta P_t u \, d\nu - \int_{\mathbb{R}^d} \big\langle \tau_\nu, \mathrm{Hess}\big((P_t u)^{2q}\big)\big\rangle_{\mathrm{HS}} \, d\nu \\
&= 2q \int_{\mathbb{R}^d} (P_t u)^{2q-1} \big\langle \mathrm{Id} - \tau_\nu, \mathrm{Hess}(P_t u) \rangle_{\mathrm{HS}} \, d\nu \\
&\quad - 2q(2q-1) \int_{\mathbb{R}^d} (P_t u)^{2q-2} \langle \tau_\nu, \nabla P_t u \otimes \nabla P_t u \rangle_{\mathrm{HS}} \, d\nu.
\end{aligned}$$

As in the proof of Proposition 2.4,

$$\begin{aligned}
&\int_{\mathbb{R}^d} \big(P_t u)^{2q-1} \langle \tau_\nu - \mathrm{Id}, \mathrm{Hess}(P_t u)\big\rangle_{\mathrm{HS}} \, d\nu \\
&= \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (P_t u)^{2q-1}(x) \, \big(\tau_\nu(x) - \mathrm{Id}\big) y \cdot \nabla u\big(e^{-t} x + \sqrt{1-e^{-2t}} \, y\big) d\nu(x) d\gamma(y)
\end{aligned}$$

Using that $|\nabla u| \leq 1$ (since u is 1-Lipschitz) and furthermore

$$|\nabla P_t u| \, \leq \, e^{-t} P_t\big(|\nabla u|\big) \, \leq \, e^{-t},$$

it easily follows as in the previous section that for every t,

$$\begin{aligned}
-\phi'(t) \, &\leq \, \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} \int_{\mathbb{R}^d} 2q |P_t u|^{2q-1} \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}} \, d\nu \\
&\quad + e^{-2t} \int_{\mathbb{R}^d} 2q(2q-1) (P_t u)^{2q-2} \|\tau_\nu\|_{\mathrm{op}} \, d\nu.
\end{aligned} \tag{2.32}$$

By the Young-Hölder inequality,

$$2q |P_t u|^{2q-1} \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}} \, \leq \, \frac{1}{\alpha} \, \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}}^{\alpha} + \frac{1}{\beta} \, \big[2q |P_t u|^{(2q-1)}\big]^{\beta}$$

where $\alpha = 2q$ and $(2q-1)\beta = 2q$, and

$$2q(2q-1)(P_t u)^{2q-2} \|\tau_\nu\|_{\mathrm{op}} \, \leq \, \frac{1}{\alpha'} \, \big[(2q-1) \|\tau_\nu\|_{\mathrm{op}}\big]^{\alpha'} + \frac{1}{\beta'} \, \big[2q (P_t u)^{(2q-2)}\big]^{\beta'}$$

where $\alpha' = q$ and $(2q-2)\beta' = 2q$. Therefore (2.32) implies that, for every t,

$$-\phi'(t) \, \leq \, C(t) \, \phi\,(t) + D(t)$$

where

$$C(t) \,=\, \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\, (2q)^{\beta} + e^{-2t}\, (2q)^{\beta'}$$

and

$$D(t) \,=\, \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} \int_{\mathbb{R}^d} \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}}^{2q}\, d\nu + e^{-2t} \int_{\mathbb{R}^d} \big[(2q-1)\|\tau_\nu\|_{\mathrm{op}}\big]^q d\nu.$$

Integrating this differential inequality yields that

$$\phi(t) \,\leq\, e^{\widetilde{C}(t)} \int_t^\infty e^{-\widetilde{C}(s)} D(s) ds$$

where $\widetilde{C}(t) = \int_t^\infty C(s)ds$, $t \geq 0$. It follows that $\phi(0) \leq e^{\widetilde{C}(0)} \int_0^\infty D(s)ds$ and therefore

$$\int_{\mathbb{R}^d} |u|^{2q} d\nu \,=\, \phi(0) \,\leq\, e^{\widetilde{C}(0)} \bigg(\int_{\mathbb{R}^d} \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}}^{2q}\, d\nu + \int_{\mathbb{R}^d} \big[(2q-1)\|\tau_\nu\|_{\mathrm{op}}\big]^q d\nu \bigg).$$

Since $\widetilde{C}(0)$ is bounded above by Cq for some numerical $C > 0$, the announced claim follows. The proof of Theorem 2.8 is therefore complete. □

2.5 On the rate of convergence in the entropic central limit theorem

In this last paragraph, we provide a brief and simple application of the HSI inequality to (yet non optimal) rates in the entropic central limit theorem. Let X be a real-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with mean zero and variance one. Let also $X_1, \ldots, X_n$ be independent copies of X and set

$$T \,=\, \sum_{k=1}^n a_k X_k$$

where $\sum_{k=1}^n a_k^2 = 1$.

Assume that the law ν of X has a density h with respect to the standard Gaussian measure γ on $\mathbb{R}$ with (finite) Fisher information $\mathrm{I}(\nu \,|\, \gamma)$ and a Stein kernel τ_ν with discrepancy $\mathrm{S}(\nu \,|\, \gamma)$. Let ν_T be the law of T. The classical Blachman-Stam inequality (cf. [Sta, Bl, V]) indicates that

$$\mathrm{I}\big(\nu_T \,|\, \gamma\big) \,\leq\, \mathrm{I}\big(\nu \,|\, \gamma\big).$$

On the other hand, as in the previous paragraph,

$$\tau_{\nu_T}(T) \,=\, \mathbb{E}\bigg[\sum_{k=1}^n a_k^2 \tau_\nu(X_k) \,\bigg|\, T \bigg]$$

so that

$$\mathrm{S}^2\big(\nu_T\,|\,\gamma\big) \,\leq\, \alpha(a)\,\mathrm{S}^2\big(\nu\,|\,\gamma\big)$$

where $\alpha(a) = \sum_{i=1}^n a_i^4$.

As a consequence therefore of the HSI inequality of Theorem 2.2,

$$\mathrm{H}\big(\nu_T\,|\,\gamma\big) \,\leq\, \frac{1}{2}\,\alpha(a)\,\mathrm{S}^2\big(\nu\,|\,\gamma\big)\log\bigg(1 + \frac{\mathrm{I}(\nu\,|\,\gamma)}{\alpha(a)\,\mathrm{S}^2(\nu\,|\,\gamma)}\bigg). \tag{2.33}$$

This result has to be compared with the works [A-B-B-N] and [B-J] (cf. [J]) which produce the bound

$$\mathrm{H}\big(\nu_T\,|\,\gamma\big) \,\leq\, \frac{\alpha(a)}{c/2 + (1-c/2)\alpha(a)}\,\mathrm{H}\big(\nu\,|\,\gamma\big) \tag{2.34}$$

under the hypothesis that ν satisfies a Poincaré inequality with constant $c > 0$.

For the classical average $T_n = \frac{1}{\sqrt{n}}\sum_{k=1}^n X_k$, (2.34) yields a rate $O(\frac{1}{n})$ in the entropic central limit theorem while (2.33) only produces $O(\frac{\log n}{n})$, however at a cheap expense and under potentially different conditions as described in Remark 2.9. For this classical average, the recent works [B-C-G1, B-C-G2] actually provide a complete picture with rate $O(\frac{1}{n})$ under a fourth-moment condition on X based on local central limit theorems and Edgeworth expansions. General sums $T = \sum_{k=1}^n a_k X_k$ are studied in [B-C-G3] as a particular case of sums of independent non-identically distributed random variables. Vector-valued random variables may be considered similarly.

3 Transport distances and Stein discrepancy

In this section, we develop further inequalities involving the Stein discrepancy, this time in relation with Wasserstein distances. A new improved form of the Talagrand quadratic transportation cost inequality, called WSH, is emphasized, and comparison between the HSI inequality and the Talagrand and Otto-Villani HWI inequalities is provided. Let again $\gamma = \gamma^d$ denote the standard Gaussian measure on $\mathbb{R}^d$.

Fix $p \geq 1$. Given two probability measures ν and μ on the Borel sets of $\mathbb{R}^d$ whose marginals have finite absolute moments of order p, define the *Wasserstein distance* (of order p) between ν and μ as the quantity

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

where the infimum runs over all probability measures π on $\mathbb{R}^d\times\mathbb{R}^d$ with marginals ν and μ. Relevant information about Wasserstein (or Kantorovich) distances can be found, e.g. in [V, Section I.6].

We shall subdivide the analysis into two parts. In Section 3.1, we deal with the special case of the quadratic Wasserstein distance W_2, for which we use the definition

(2.1) of a Stein kernel. In Section 3.2, we deal with general Wasserstein distances W_p possibly of order $p \neq 2$, for which it seems necessary to use the stronger definition (2.3) adopted in [N-P-S1, N-P-S2].

3.1 The case of the Wasserstein distance W_2

We provide here a dimension-free estimate on the Wasserstein W_2 distance expressed in terms of the Stein discrepancy. In the forthcoming statement, denote by ν a centered probability measure on $\mathbb{R}^d$ admitting a Stein kernel τ_ν (that is, τ_ν verifies (2.1) for every smooth test function φ). It is not assumed that ν admits a density with respect to the Lebesgue measure on $\mathbb{R}^d$ (in particular, ν can have atoms). As already observed, the existence of a Stein kernel for ν implies that ν has finite moments of order 2.

Proposition 3.1 (Wasserstein distance and Stein discrepancy)**.** *For every centered probability measure ν on $\mathbb{R}^d$,*

$$\mathrm{W}_2(\nu,\gamma) \,\leq\, \mathrm{S}\big(\nu\,|\,\gamma\big). \tag{3.1}$$

Proof. Assume first that $d\nu = hd\gamma$ where h is a smooth density with respect to the standard Gaussian measure γ on $\mathbb{R}^d$. As in Section 2, write $v_t = \log P_t h$ and $d\nu^t = P_t h d\gamma$. We shall rely on the estimate, borrowed from [O-V, Lemma 2] (cf. also [V, Theorem 24.2(iv)]),

$$\frac{d^+}{dt}\,\mathrm{W}_2(\nu,\nu^t) \,\leq\, \bigg(\int_{\mathbb{R}^d} |\nabla v_t|^2\, d\nu^t\bigg)^{1/2}. \tag{3.2}$$

Note that (3.2) is actually the central argument in the Otto-Villani theorem [O-V] asserting that a logarithmic Sobolev inequality implies a Talagrand transport inequality. Here, by making use of (3.2) and then (2.17) we get that

$$\mathrm{W}_2(\nu,\gamma) \,\leq\, \int_0^\infty \bigg(\int_{\mathbb{R}^d} |\nabla v_t|^2 d\nu^t\bigg)^{1/2} dt \,\leq\, \mathrm{S}\big(\nu\,|\,\gamma\big)\int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\,dt$$

which is the result in this case.

The general case is obtained by a simple regularization procedure which is best presented in probabilistic terms. Fix $\varepsilon > 0$ and introduce the auxiliary random variable $F_\varepsilon = e^{-\varepsilon}F + \sqrt{1-e^{-2\varepsilon}}Z$ where F and Z are independent with respective laws ν and γ. It is immediately checked that: (a) the distribution of F_ε, denoted by ν^ε, admits a smooth density h_ε with respect to γ (of course, this density coincides with $P_\varepsilon h$ whenever the distribution of F admits a density h with respect to γ as in the first part of the proof); (b) a Stein kernel for ν^ε is given by

$$\tau_{\nu^\varepsilon}(x) \,=\, \mathbb{E}\big[e^{-2\epsilon}\tau_\nu(F) + (1-e^{-2\varepsilon})\,\mathrm{Id}\,|\,F_\varepsilon = x\big] \quad d\nu^\varepsilon(x)\text{–a.e.}$$

(consistent with (2.21)); (c) $\mathrm{S}(\nu^\varepsilon\,|\,\gamma) \leq e^{-2\varepsilon}\,\mathrm{S}(\nu\,|\,\gamma)$; (d) as $\varepsilon \to 0$, F_ε converges to F in L^2, so that, in particular, $\mathrm{W}_2(\nu^\varepsilon,\gamma) \to \mathrm{W}_2(\nu,\gamma)$. One therefore infers that

$$\mathrm{W}_2(\nu,\gamma) \,=\, \lim_{\varepsilon\to 0}\mathrm{W}_2(\nu^\varepsilon,\gamma) \,\leq\, \limsup_{\varepsilon\to 0}\,\mathrm{S}\big(\nu^\varepsilon\,|\,\gamma\big) \,\leq\, \mathrm{S}\big(\nu\,|\,\gamma\big),$$

and the proof is concluded. □

The inequality (3.1) may of course be compared to the Talagrand quadratic transportation cost inequality [T, V, B-G-L]

$$\mathrm{W}_2^2(\nu,\gamma) \,\leq\, 2\,\mathrm{H}\big(\nu\,|\,\gamma\big). \tag{3.3}$$

As announced in the introduction, one can actually further refine (3.1) in order to deduce an improvement of (3.3) in the form of a WSH inequality. The refinement relies on the HSI inequality itself.

Theorem 3.2 (Gaussian WSH inequality)**.** *Let $d\nu = hd\gamma$ be a centered probability measure on $\mathbb{R}^d$ with smooth density h with respect to γ. Assume further that $\mathrm{S}(\nu\,|\,\gamma)$ and $\mathrm{H}(\nu\,|\,\gamma)$ are both positive and finite. Then*

$$\mathrm{W}_2(\nu,\gamma) \,\leq\, \mathrm{S}\big(\nu\,|\,\gamma\big)\arccos\Big(e^{-\frac{\mathrm{H}(\nu|\gamma)}{\mathrm{S}^2(\nu|\gamma)}}\Big).$$

Proof. 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 (2.6) applied to ν^t yields that

$$\mathrm{H}\big(\nu^t\,|\,\gamma\big) \,\leq\, \frac{1}{2}\,\mathrm{S}^2\big(\nu^t\,|\,\gamma\big)\,\log\bigg(1+\frac{\mathrm{I}(\nu^t\,|\,\gamma)}{\mathrm{S}^2(\nu^t\,|\,\gamma)}\bigg).$$

Now, $\mathrm{S}^2(\nu^t\,|\,\gamma) \leq \mathrm{S}^2(\nu\,|\,\gamma)$ by (2.18) and $r \mapsto r\log\big(1+\frac{s}{r}\big)$ is increasing for any fixed s from which it follows that

$$\mathrm{H}\big(\nu^t\,|\,\gamma\big) \,\leq\, \frac{1}{2}\,\mathrm{S}^2\big(\nu\,|\,\gamma\big)\,\log\bigg(1+\frac{\mathrm{I}(\nu^t\,|\,\gamma)}{\mathrm{S}^2(\nu\,|\,\gamma)}\bigg).$$

By exponentiating both sides, this inequality is equivalent to

$$\sqrt{\mathrm{I}\big(\nu^t\,|\,\gamma\big)} \,\leq\, \frac{\mathrm{I}(\nu^t\,|\,\gamma)}{\mathrm{S}(\nu\,|\,\gamma)\sqrt{e^{\frac{2\mathrm{H}(\nu^t|\gamma)}{\mathrm{S}^2(\nu|\gamma)}}-1}}\,.$$

Combining with (3.2) and recalling (2.13) leads to

$$\begin{aligned}\frac{d^+}{dt}\,\mathrm{W}_2(\nu,\nu^t) \,\leq\, \sqrt{\mathrm{I}\big(\nu^t\,|\,\gamma\big)} &\leq -\frac{\frac{d}{dt}\mathrm{H}(\nu^t\,|\,\gamma)}{\mathrm{S}(\nu\,|\,\gamma)\sqrt{e^{\frac{2\mathrm{H}(\nu^t|\gamma)}{\mathrm{S}^2(\nu|\gamma)}}-1}}\\ &= -\frac{d}{dt}\bigg(\mathrm{S}\big(\nu\,|\,\gamma\big)\arccos\Big(e^{-\frac{H(\nu^t|\gamma)}{\mathrm{S}^2(\nu|\gamma)}}\Big)\bigg).\end{aligned}$$

In other words,

$$\frac{d}{dt}\bigg(\mathrm{W}_2(\nu,\nu^t)+\mathrm{S}\big(\nu\,|\,\gamma\big)\arccos\Big(e^{-\frac{H(\nu^t|\gamma)}{S^2(\nu|\gamma)}}\Big)\bigg) \,\leq\, 0.$$

The desired conclusion is achieved by integrating between $t = 0$ and $t = \infty$. The proof of Theorem 3.2 is complete. □

Proposition 3.1 and Theorem 3.2 raise a number of observations.

Remark 3.3. (a) Since $\arccos(e^{-r}) \leq \sqrt{2r}$ for every $r \geq 0$, the WSH inequality thus represents an improvement upon the Talagrand inequality (3.3). Moreover, as for the HSI inequality, the WSH inequality produces the case of equality in (3.3) since $\arccos(e^{-r}) \leq \sqrt{2r}$ is an equality only at $r = 0$.

(b) The Talagrand inequality may combined with the HSI inequality of Theorem 2.2 to yield the bound

$$\mathrm{W}_2^2(\nu, \gamma) \leq \mathrm{S}^2\big(\nu \,|\, \gamma\big) \log\bigg(1 + \frac{\mathrm{I}(\nu \,|\, \gamma)}{\mathrm{S}^2(\nu \,|\, \gamma)}\bigg). \tag{3.4}$$

(c) (HWI inequality). As described in the introduction, a fundamental estimate connecting entropy H, Wassertein distance W_2 and Fisher information I is the so-called HWI inequality of Otto and Villani [O-V] stating that, for all $d\nu = hd\gamma$ with density h with respect to γ,

$$\mathrm{H}\big(\nu \,|\, \gamma\big) \leq \mathrm{W}_2(\nu, \gamma) \sqrt{\mathrm{I}\big(\nu \,|\, \gamma\big)} - \frac{1}{2} \mathrm{W}_2^2(\nu, \gamma) \tag{3.5}$$

(see, e.g. [V, pp. 529-542] or [B-G-L, Section 9.3.1] for a general discussion). Recall that the HWI inequality (3.5) improves upon both the logarithmic Sobolev inequality (1.1) and the Talagrand inequality (3.3). It is natural to look for a more general inequality, involving all four quantities H, W_2, I and the Stein discrepancy S, and improving both the HSI and HWI inequalities. One strategy towards this task would be to follow again the heat flow approach of the proof of Theorem 2.2 and write, for $0 < u \leq t$,

$$\begin{aligned}
\mathrm{Ent}_\gamma(h) &= \int_0^t \mathrm{I}_\gamma(P_s h) ds + \mathrm{Ent}_\gamma(P_t h) \\
&\leq \mathrm{I}_\gamma(h) \int_0^u e^{-2s} ds + \mathrm{S}^2\big(\nu \,|\, \gamma\big) \int_u^t \frac{e^{-4s}}{1 - e^{-2s}}\, ds + \frac{e^{-2t}}{2(1 - e^{-2t})} \mathrm{W}_2^2(\nu, \gamma).
\end{aligned}$$

Here, we used (2.15) and (2.17), as well as the known reverse Talagrand inequality along the semigroup given by

$$\mathrm{Ent}_\gamma(P_t h) \leq \frac{e^{-2t}}{2(1 - e^{-2t})} \mathrm{W}_2^2(\nu, \gamma)$$

(cf. e.g. [B-G-L, p. 446]). Setting $\alpha = 1 - e^{-2u} \leq 1 - e^{-2t} = \beta$, the preceding estimate yields

$$\mathrm{H}\big(\nu \,|\, \gamma\big) \leq \inf_{0 < \alpha \leq \beta \leq 1} \Phi(\alpha, \beta)$$

where

$$\Phi(\alpha, \beta) = \alpha\, \mathrm{I}\big(\nu \,|\, \gamma\big) + (\alpha - \log \alpha)\, \mathrm{S}^2\big(\nu \,|\, \gamma\big) + \frac{1 - \beta}{\beta} \mathrm{W}_2^2(\nu, \gamma) + (\log \beta - \beta)\, \mathrm{S}^2\big(\nu \,|\, \gamma\big).$$

However, elementary computations show that, unless the rather unnatural inequality $2\mathrm{W}_2(\nu,\gamma) \leq \mathrm{S}(\nu\,|\,\gamma)$ is verified, the minimum in the above expression is attained at a point (α,β) such that either $\alpha=\beta$ (and in this case one recovers HWI) or $\beta=1$ (yielding HSI). Hence, at this stage, it seems difficult to outperform both HWI and HSI estimates with a single 'HWSI' inequality. In the subsequent point (d), we provide an elementary explicit example in which the HSI estimate perform better than the HWI inequality.

(d) In this item, we thus compare the HWI and HSI inequalities on a specific example in dimension $d=1$. For every $n\geq 1$, consider the probability measure $d\nu_n(x) = \rho_n(x)dx$ with density

$$\rho_n(x) \;=\; \frac{1}{\sqrt{2\pi}}\big[(1-a_n)e^{-x^2/2} + na_n e^{-n^2x^2/2}\big], \quad x\in\mathbb{R},$$

where $(a_n)_{n\geq1}$ is such that $a_n\in[0,1]$ for every $n\geq1$, $a_n = o\big(\frac{1}{\log n}\big)$ and $n^{2/3}a_n\to\infty$. A direct computation easily shows that $\mathrm{H}(\nu_n\,|\,\gamma)\to 0$. Also, since

$$\rho_n'(x) \;=\; -\frac{x}{\sqrt{2\pi}}\big[(1-a_n)e^{-x^2/2} + n^3a_n e^{-x^2n^2/2}\big],$$

one may show after simple (but a bit lengthy) computations that

$$\mathrm{I}\big(\nu_n\,|\,\gamma\big) \;=\; \int_{\mathbb{R}} \frac{\rho_n'(x)^2}{\rho_n(x)}dx - 1 \;\sim\; n^2a_n \quad \text{as} \quad n\to\infty.$$

We next examine the Stein discrepancy $\mathrm{S}(\nu_n\,|\,\gamma)$ and Wassertein distance $\mathrm{W}_2(\nu_n,\gamma)$. Since a Stein kernel τ_n of ν_n is given by

$$\tau_n(x) \;=\; \frac{1}{\sqrt{2\pi}\,\rho_n}\Big[(1-a_n)e^{-x^2/2} + \frac{a_n}{n}\,e^{-n^2x^2/2}\Big],$$

it is easily seen that

$$\mathrm{S}^2\big(\nu_n\,|\,\gamma\big) \;=\; \int_{\mathbb{R}} \big(\tau_n(x)-1\big)^2\rho_n(x)dx \;\leq\; a_n \;\to\; 0.$$

Concerning the Wasserstein distance, from the inequality (3.1), we deduce that $\mathrm{W}_2(\nu_n,\gamma) \leq \sqrt{a_n}$. On the other hand, by the Lipschitz characterization of W_1 (specializing to the Lipschitz function $x\mapsto|\cos(x)|$), cf. e.g. [V, Remark 6.5]),

$$\mathrm{W}_2(\nu_n,\gamma) \;\geq\; \mathrm{W}_1(\nu_n,\gamma) \;\geq\; \bigg|\int_{\mathbb{R}} \big|\cos(x)\big|d\nu_n(x) - \int_{\mathbb{R}}\big|\cos(x)\big|d\gamma(x)\bigg|.$$

Now, the right-hand side of this inequality multiplied by $\frac{1}{a_n}$ is equal to

$$\begin{aligned}\bigg|n\int_{\mathbb{R}}\big|\cos(x)\big|e^{-n^2x^2/2}\,\frac{dx}{\sqrt{2\pi}} - \int_{\mathbb{R}}\big|\cos(x)\big|d\gamma(x)\bigg| \\ = \bigg|\int_{\mathbb{R}}\Big[\big|\cos(\tfrac{x}{n})\big| - \big|\cos(x)\big|\Big]d\gamma(x)\bigg|\end{aligned}$$

which, by dominated convergence, converges to a non-zero limit. As a consequence, there exists $c > 0$ such that, for n large enough, $\mathrm{W}_2(\nu_n, \gamma) \geq c\, a_n$.

Summarizing the conclusions, the quantity

$$\mathrm{W}_2\big(\nu_n, \gamma\big)\sqrt{\mathrm{I}\big(\nu_n \,|\, \gamma\big)} - \frac{1}{2}\,\mathrm{W}_2^2(\nu_n, \gamma)$$

is bigger than a sequence of the order of $n a_n^{3/2} = (n^{2/3} a_n)^{3/2}$, which (by construction) diverges to infinity as $n \to \infty$. This fact implies that, in this specific case, the bound in the HWI inequality diverges to infinity, whereas $\mathrm{H}(\nu_n \,|\, \gamma) \to 0$. On the other hand, the HSI bound converges to zero, since

$$\mathrm{S}^2\big(\nu_n \,|\, \gamma\big) \log\left(1 + \frac{\mathrm{I}(\nu_n \,|\, \gamma)}{\mathrm{S}^2(\nu_n \,|\, \gamma)}\right) \,\leq\, a_n \log(1 + n^2) \,\sim\, 2a_n \log n \,\to\, 0.$$

3.2 General Wasserstein distances under a stronger notion of Stein kernel

In this part, we obtain bounds in terms of Stein discrepancies on the Wasserstein distance W_p of any order p between a centered probability measure ν on $\mathbb{R}^d$ and the standard Gaussian distribution γ. As in Proposition 3.1, we shall consider probabilities ν not necessarily admitting a density with respect to γ. However, it will be assumed that ν has a Stein kernel τ_ν verifying the stronger 'vector' relation (2.3). The reason for this is that, in order to deal with Wasserstein distances of the type W_p, $p \neq 2$, one needs to have access to the explicit expression of the score function $\nabla(\log P_t h)$ along the Ornstein-Uhlenbeck semigroup, as proved in [N-P-S1, Lemma 2.9] in the framework of Stein kernels verifying (2.3). Recall that the existence of τ_ν implies that ν has finite moments of order 2.

Proposition 3.4 (W_p distance and Stein discrepancy)**.** *Let ν be a centered probability measure on $\mathbb{R}^d$ with Stein kernel τ_ν in the sense of* (2.3)*. For every $p \geq 1$, set*

$$\|\tau_\nu - \mathrm{Id}\|_{p,\nu} \,=\, \bigg(\sum_{i,j=1}^{d} \int_{\mathbb{R}^d} \big|\tau_\nu^{ij} - \delta_{ij}\big|^p d\nu\bigg)^{1/p}$$

(where $\delta_{ij} = 1$ if $i = j$ and 0 if not), possibly infinite if $\tau_\nu^{ij} \notin \mathrm{L}^p(\nu)$. In particular, $\|\tau_\nu - \mathrm{Id}\|_{2,\nu} = \mathrm{S}(\nu \,|\, \gamma)$.

(i) Let $p \in [1, 2)$. Then,

$$\mathrm{W}_p(\nu, \gamma) \,\leq\, C_p\, d^{1-1/p} \|\tau_\nu - \mathrm{Id}\|_{p,\nu} \tag{3.6}$$

where $C_p^p = \int_{\mathbb{R}} |x|^p d\gamma^1(x)$.

(ii) Let $p \in [2,\infty)$. If ν has finite moments of order p, then (with the same C_p as in (i))

$$\mathrm{W}_p(\nu,\gamma) \,\leq\, C_p\, d^{1-2/p}\, \|\tau_\nu - \mathrm{Id}\|_{p,\nu}. \tag{3.7}$$

In particular, for $p=2$ we recover (3.1).

Proof. Owing to an approximation argument analogous to the one rehearsed at end of the proof of Proposition 3.1, it is sufficient to consider the case $d\nu = h\, d\gamma$ where h is a smooth density. Write as before $v_t = \log P_t h$ and $d\nu^t = P_t h d\gamma$. By virtue of [N-P-S1, Lemma 2.9], under thus the strengthened assumption (2.3), a version of ∇v_t, $t>0$, is given by

$$x \,\mapsto\, \nabla v_t(x) \,=\, \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\, \mathbb{E}\big[\big(\tau_\nu(F) - \mathrm{Id}\big) Z \,|\, F_t = x\big], \quad x \in \mathbb{R}^d,$$

where, as in Remark 2.5, F and Z are independent with respective law ν and γ, and $F_t = e^{-t}F + \sqrt{1-e^{-2t}}\,Z$. Moreover, one can straightforwardly modify the proof of [O-V, Lemma 2] (cf. also [V, Theorem 24.2(iv)]) in order to obtain the general estimate

$$\frac{d^+}{dt}\, \mathrm{W}_p(\nu,\nu^t) \,\leq\, \bigg(\int_{\mathbb{R}^d} |\nabla v_t|^p\, d\nu^t\bigg)^{1/p}. \tag{3.8}$$

It follows that

$$\begin{aligned}
\mathrm{W}_p(\nu,\gamma) \,&\leq\, \int_0^\infty \bigg(\int_{\mathbb{R}^d} |\nabla v_t|^p d\nu^t\bigg)^{1/p} dt \\
&= \int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\, \mathbb{E}\Bigg[\bigg(\sum_{i=1}^d \mathbb{E}\bigg[\sum_{j=1}^d \big(\tau_\nu^{ij}(F) - \delta_{ij}\big) Z_j \bigg| F_t\bigg]^2\bigg)^{p/2}\Bigg]^{1/p} dt.
\end{aligned}$$

Now, if $1 \leq p < 2$,

$$\begin{aligned}
\mathrm{W}_p(\nu,\gamma) \,&\leq\, \int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\, dt \,\bigg(\sum_{i=1}^d \mathbb{E}\bigg[\bigg|\sum_{j=1}^d \big(\tau_\nu^{ij}(F) - \delta_{ij}\big) Z_j\bigg|^p\bigg]\bigg)^{1/p} \\
&\leq\, C_p\, d^{1-1/p} \bigg(\sum_{i,j=1}^d \mathbb{E}\big[\big|\tau_\nu^{ij}(F) - \delta_{ij}\big|^p\big]\bigg)^{1/p}
\end{aligned}$$

yielding *(i)*. On the other hand, if $p \geq 2$, then

$$\begin{aligned} \mathrm{W}_p(\nu,\gamma) &\leq \int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\, \mathbb{E}\bigg[\bigg(\sum_{i=1}^d \mathbb{E}\bigg[\bigg(\sum_{j=1}^d \big(\tau_\nu^{ij}(F)-\delta_{ij}\big)Z_j\bigg)^2\bigg|F_t\bigg]\bigg)^{p/2}\bigg]^{1/p} dt \\ &\leq d^{1/2-1/p}\bigg(\sum_{i=1}^d \mathbb{E}\bigg[\bigg|\sum_{j=1}^d \big(\tau_\nu^{ij}(F)-\delta_{ij}\big)Z_j\bigg|^p\bigg]\bigg)^{1/p} \\ &= C_p\, d^{1/2-1/p}\bigg(\sum_{i=1}^d \bigg(\sum_{j=1}^d \mathbb{E}\Big[\big(\tau_\nu^{ij}(F)-\delta_{ij}\big)^2\Big]\bigg)^{p/2}\bigg)^{1/p} \\ &\leq C_p\, d^{1-2/p}\bigg(\sum_{i,j=1}^d \mathbb{E}\Big[\big|\tau_\nu^{ij}(F)-\delta_{ij}\big|^p\Big]\bigg)^{1/p} \end{aligned}$$

which immediately yields *(ii)*. The proof of Proposition 3.4 is complete. □

Remark 3.5. Specializing (3.6) to the case $p=1$ yields the estimate

$$\mathrm{W}_1(\nu,\gamma) \leq \sqrt{\frac{2}{\pi}}\, \|\tau_\nu - \mathrm{Id}\|_{1,\nu} \tag{3.9}$$

which improves previous dimensional bounds obtained by an application of the multidimensional Stein method (cf. the proof of [N-P2, Theorem 6.1.1]). It is important to note that, apart from the results obtained in the present paper, there is no other version of Stein's method allowing one to deal with Wasserstein distances of order $p > 1$. Observe that coupling results from [C3] (that are based on completely different methods) may be used to deduce analogous estimates in the case when $d = 1$ and the Stein kernel τ_ν is bounded.

4 HSI inequalities for further distributions

On the basis of the Gaussian example of Section 2, we next address the issue of HSI inequalities for distributions on $\mathbb{R}^d$, $d \geq 1$, that are not necessarily Gaussian. In order to reach the basic semigroup ingredients towards such HSI inequalities put forward in Proposition 2.4, a convenient family of measures to deal with is 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. As such, the investigation is part of the generator approach to Stein's method as developed in [Ba, G, R]. We present it here in the framework of Markov Triples as developed in [B-G-L] and, for simplicity, only consider operators and measures on $\mathbb{R}^d$.

4.1 A general statement

Let E be a domain of $\mathbb{R}^d$ and consider a family of real-valued C^∞-functions $a^{ij}(x)$ and $b^i(x)$, $i,j=1,\dots,d$, defined on E. We assume that the matrix $a(x)=(a^{ij}(x))_{1\leq i,j\leq d}$ is symmetric and positive definite for any $x\in E$. For every $x\in E$, we let $a^{\frac{1}{2}}(x)$ be the unique symmetric non-singular matrix such that $(a^{\frac{1}{2}}(x))^2=a(x)$. Let $\mathcal{A}$ denote the algebra of C^∞-functions on E and $\mathcal{L}$ be the second order differential operator given on functions $f\in\mathcal{A}$ by

$$\mathcal{L}f \,=\, \big\langle a, \mathrm{Hess}(f)\big\rangle_{\mathrm{HS}} + b\cdot\nabla f \,=\, \sum_{i,j=1}^d a^{ij}\,\frac{\partial^2 f}{\partial x_i\partial x_j} + \sum_{i=1}^d b^i\,\frac{\partial f}{\partial x_i}\,. \tag{4.1}$$

The operator $\mathcal{L}$ satisfies the chain rule formula and defines a diffusion operator. We assume that $\mathcal{L}$ is the generator of a symmetric Markov semigroup $(P_t)_{t\geq 0}$, where the symmetry is with respect to an invariant probability measure μ.

A central object of interest in this context is the carré du champ operator Γ defined from the generator $\mathcal{L}$ by

$$\Gamma(f,g) \,=\, \frac{1}{2}\big[\mathcal{L}(fg)-f\mathcal{L}g-g\mathcal{L}f\big] = \sum_{i,j=1}^d a^{ij}\,\frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_j}$$

for all $(f,g)\in\mathcal{A}\times\mathcal{A}$. Note that Γ is bilinear and symmetric and $\Gamma(f,f)\geq 0$. Moreover, the integration by parts property for $\mathcal{L}$ with respect to the invariant measure μ is expressed by the fact that, for functions $f,g\in\mathcal{A}$,

$$\int_E f\,\mathcal{L}g\,d\mu \,=\, -\int_E \Gamma(f,g)d\mu.$$

The structure (E,μ,Γ) then defines a Markov Triple in the sense of [B-G-L] to which we refer for the necessary background.

The requested semigroup analysis toward HSI inequalities will actually involve in addition the iterated gradient operators Γ_n, $n\geq 1$, defined inductively for $(f,g)\in\mathcal{A}\times\mathcal{A}$ via the relations $\Gamma_0(f,g)=fg$ and

$$\Gamma_n(f,g) \,=\, \frac{1}{2}\big[\mathcal{L}\,\Gamma_{n-1}(f,g)-\Gamma_{n-1}(f,\mathcal{L}g)-\Gamma_{n-1}(g,\mathcal{L}f)\big], \quad n\geq 1.$$

In particular $\Gamma_1=\Gamma$ and the operators Γ_n, $n\geq 1$, are similarly symmetric and bilinear. In what follows, we shall often adopt the shorthand notation $\Gamma_n(f)$ instead of $\Gamma_n(f,f)$. The Γ_2 operator is part of the famous Bakry-Émery criterion for logarithmic Sobolev inequalities [B-E], [B-G-L, Section 5.7]. As a new feature of the analysis here, the iterated gradient Γ_3 will turn essential towards a suitable analogue of *(iii)* in Proposition 2.4.

A prototypical example of this setting is of course the Ornstein-Uhlenbeck operator $\mathcal{L}=\Delta-x\cdot\nabla$ on $\mathbb{R}^d$ considered earlier, with the standard Gaussian measure γ as symmetric and invariant measure. In this case, the carré du champ operator is simply given by $\Gamma(f) = |\nabla f|^2$ on smooth functions f. It is easily seen that, for example (cf. [L1]),

$$\Gamma_2(f) = \sum_{i,j=1}^{d} \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2 + \Gamma(f)$$

and

$$\Gamma_3(f) = \sum_{i,j,k=1}^{d} \left(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k} \right)^2 + 3\,\Gamma_2(f) - 2\,\Gamma(f).$$

Given thus the preceding Markov Triple (E, μ, Γ) associated to the second order differential operator $\mathcal{L}$ of (4.1), let $d\nu = h d\mu$ where h is a smooth probability density with respect to μ. As in the Gaussian case, the relative entropy of ν with respect to μ is the quantity

$$\mathrm{H}\big(\nu \,|\, \mu\big) = \mathrm{Ent}_\mu(h) = \int_E h \log h \, d\mu.$$

Similarly, the Fisher information of ν (or h) with respect to μ is defined as

$$\mathrm{I}\big(\nu \,|\, \mu\big) = \mathrm{I}_\mu(h) = \int_E \frac{\Gamma(h)}{h} \, d\mu = \int_E \Gamma(\log h) h d\mu = -\int_E \mathcal{L}(\log h) d\nu. \tag{4.2}$$

The (integrated) *de Bruijn's identity* (cf. Proposition 5.2.2 in [B-G-L]) reads as in *(i)* of Proposition 2.4,

$$\mathrm{H}\big(\nu \,|\, \mu\big) = \int_0^\infty \mathrm{I}_\mu(P_t h) d\mu.$$

Let $\mathcal{M}_{d\times d}$ denote the class of $d \times d$ matrices with real entries. Analogously to the definition of Stein kernel of Section 2.1, we shall say that a matrix-valued mapping $\tau_\nu : \mathbb{R}^d \to \mathcal{M}_{d\times d}$ satisfying $\tau_\nu^{ij} \in \mathrm{L}^1(\nu)$ for every $i, j = 1, \ldots, d$, and

$$-\int_E b \cdot \nabla f \, d\nu = \int_E \big\langle \tau_\nu, \mathrm{Hess}(f) \big\rangle_{\mathrm{HS}} \, d\nu, \quad f \in \mathcal{A}, \tag{4.3}$$

is a Stein kernel for the probability ν on E with respect to the generator of $\mathcal{L}$ of (4.1), where $b = (b_i(x))_{1\leq i\leq d}$ is part of the definition of $\mathcal{L}$. For the Ornstein-Uhlenbeck operator $\mathcal{L} = \Delta - x \cdot \nabla$, the definition corresponds to (2.1). Since $\int_E \mathcal{L} f \, d\mu = 0$, observe that a is a Stein kernel for μ. The main result in this section is an HSI inequality that relates $\mathrm{H}(\nu \,|\, \mu)$, $\mathrm{I}(\nu \,|\, \mu)$ and the Stein discrepancy of ν with respect to μ

$$\mathrm{S}\big(\nu \,|\, \mu\big) = \bigg(\int_E \big\| a^{-\frac{1}{2}} \tau_\nu a^{-\frac{1}{2}} - \mathrm{Id} \big\|_{\mathrm{HS}}^2 \, d\nu \bigg)^{1/2} \tag{4.4}$$

that we regard, as in the Gaussian case of Section 2, as a measure of the distance between ν and μ (since $\tau_\mu = a$). Note that choosing $a = C$ in (4.4), with C non-singular, yields the quantity arising in Corollary 2.3. It should also be mentioned that the Stein discrepancy (4.4) is somewhat in contrast with the bounds one customarily obtains when applying Stein's method (see e.g. [N-P1] for the specific example of the one-dimensional Gamma distribution, or [R] for a general reference), which typically involve quantities of the type $\int_E \|\tau_\nu - a\|_{\mathrm{HS}}^2 \, d\nu$. The appearance of the inverse matrices $a^{-\frac{1}{2}}$ seems to be inextricably connected with the fact that we deal with information-theoretical functionals.

The following general statement collects the necessary assumptions on the iterated gradients Γ, Γ_2 and Γ_3 to achieve the expected HSI inequality by the semigroup interpolation scheme. The next paragraphs will provide illustrations in various concrete instances of interest. In Theorem 4.1 below, *(i)* amounts to the Bakry-Émery Γ_2 criterion to ensure the logarithmic Sobolev inequality (cf. [B-G-L, Section 5.7]) while condition *(ii)* linking the Γ_2 and Γ_3 operators will provide (together with *(iii)*) the suitable semigroup bound for the time control of $\mathrm{I}(P_t h)$ away from 0. Recall $\Psi(r) = 1 + \log r$ if $r \geq 1$ and $\Psi(r) = r$ if $0 \leq r \leq 1$.

Theorem 4.1 (General HSI inequality)**.** *In the preceding context, let $d\nu = h d\mu$ where h is a smooth density with Stein kernel τ_ν with respect to μ. Assume that there exists $\rho, \kappa, \sigma > 0$ such that, for any $f \in \mathcal{A}$,*

(i) $\Gamma_2(f) \geq \rho \, \Gamma(f)$;

(ii) $\Gamma_3(f) \geq \kappa \, \Gamma_2(f)$;

(iii) $\Gamma_2(f) \geq \sigma \, \|a^{\frac{1}{2}} \, \mathrm{Hess}(f) \, a^{\frac{1}{2}}\|_{\mathrm{HS}}^2$ *(with a as in (4.1)).*

Then,

$$\mathrm{H}(\nu \,|\, \mu) \, \leq \, \frac{1}{2\sigma} \, \mathrm{S}^2(\nu \,|\, \mu) \, \Psi\bigg(\frac{\sigma \max(\rho, \kappa) \, \mathrm{I}(\nu \,|\, \mu)}{\rho\kappa \, \mathrm{S}^2(\nu \,|\, \mu)} \bigg).$$

Note that in the Ornstein-Uhlenbeck example, $\rho = \kappa = \sigma = 1$ from which we recover the HSI inequality (2.6), however in a slightly weaker formulation.

Proof. It is therefore a classical fact (see e.g. [B-G-L, (5.7.4)]) that *(i)* ensures the exponential decay of the Fisher information along the semigroup

$$\mathrm{I}_\mu(P_t h) \, \leq \, e^{-2\rho t} \, \mathrm{I}_\mu(h) \, = \, e^{-2\rho t} \, \mathrm{I}\big(\mu \,|\, \nu\big) \tag{4.5}$$

for every $t \geq 0$ (and then yields a logarithmic Sobolev inequality for μ.) Now, fix $t > 0$ and let $f \in \mathcal{A}$. The Γ-calculus as developed in [B-G-L], but at the level of the Γ_2 and Γ_3 operators, yields on $[0, t]$ (by the very definition of Γ_3 from Γ_2),

$$\begin{aligned} \frac{d}{ds}\Big(P_s\big(\Gamma_2(P_{t-s}f)\big) e^{-2\kappa s} \Big) \, &= \, 2 e^{-2\kappa s} \Big(P_s\big(\Gamma_3(P_{t-s}f)\big) - \kappa P_s\big(\Gamma_2(P_{t-s}f)\big) \Big) \\ &= \, 2 e^{-2\kappa s} P_s\big((\Gamma_3 - \kappa \, \Gamma_2)(P_{t-s}f) \big). \end{aligned}$$

By *(ii)*, the latter is non-negative so that the map $s \mapsto P_s(\Gamma_2(P_{t-s}f))e^{-2\kappa s}$ is increasing on $[0,t]$, and thus

$$P_t\big(\Gamma(f)\big) - \Gamma\big(P_t(f)\big) \,=\, 2\int_0^t P_s\big(\Gamma_2(P_{t-s}f)\big)ds$$
$$\geq\, 2\,\Gamma_2(P_tf)\int_0^t e^{2\kappa s}ds \,=\, \frac{1}{\kappa}\,(e^{2\kappa t}-1)\,\Gamma_2(P_tf).$$

Together with *(iii)*, it then follows that

$$P_t\big(\Gamma(f)\big) \,\geq\, P_t\big(\Gamma(f)\big) - \Gamma\big(P_t(f)\big) \,\geq\, \frac{\sigma}{\kappa}\,(e^{2\kappa t}-1)\,\big\|a^{\frac{1}{2}}\,\mathrm{Hess}(P_tf)a^{\frac{1}{2}}\big\|_{\mathrm{HS}}^2. \tag{4.6}$$

We shall apply (4.6) to $v_t = \log P_th$ (with h regular enough). First, by symmetry of μ with respect to $(P_t)_{t\geq 0}$,

$$\mathrm{I}_\mu(P_th) \,=\, -\int_E \mathcal{L}v_t\,P_th\,d\mu \,=\, -\int_E \mathcal{L}P_tv_t\,hd\mu \,=\, -\int_E \mathcal{L}P_tv_t\,d\nu. \tag{4.7}$$

Hence, by (4.1) and (4.3),

$$\mathrm{I}_\mu(P_th) \,=\, -\int_E \big\langle a, \mathrm{Hess}(P_tv_t)\big\rangle_{\mathrm{HS}}\,d\nu - \int_E b\cdot\nabla P_tv_t\,d\nu$$
$$=\, \int_E \big\langle \tau_\nu - a, \mathrm{Hess}(P_tv_t)\big\rangle_{\mathrm{HS}}\,d\nu.$$

Now, by the Cauchy-Schwarz inequality,

$$\mathrm{I}_\mu(P_th) \,=\, \int_E \big\langle a^{-\frac{1}{2}}\tau_\nu a^{-\frac{1}{2}} - \mathrm{Id}, a^{\frac{1}{2}}\,\mathrm{Hess}(P_tv_t)a^{\frac{1}{2}}\big\rangle_{\mathrm{HS}}\,d\nu$$
$$\leq\, \bigg(\int_E \big\|a^{-\frac{1}{2}}\tau_\nu a^{-\frac{1}{2}} - \mathrm{Id}\big\|_{\mathrm{HS}}^2\,d\nu\bigg)^{1/2}\bigg(\int_E \big\|a^{\frac{1}{2}}\,\mathrm{Hess}(P_tv_t)a^{\frac{1}{2}}\big\|_{\mathrm{HS}}^2\,d\nu\bigg)^{1/2}$$
$$\leq\, \mathrm{S}\big(\nu\,|\,\mu\big)\bigg(\frac{\kappa}{\sigma(e^{2\kappa t}-1)}\int_E P_t\big(\Gamma(v_t)\big)\big)d\nu\bigg)^{1/2}$$

where the last step follows from (4.6). Since

$$\int_E P_t\big(\Gamma(v_t)\big)d\nu \,=\, \int_E P_t\big(\Gamma(v_t)\big)hd\mu \,=\, \int_E \Gamma(v_t)P_thd\mu \,=\, \mathrm{I}_\mu(P_th),$$

it follows that

$$\mathrm{I}_\mu(P_th) \,\leq\, \frac{\kappa}{\sigma(e^{2\kappa t}-1)}\,\mathrm{S}^2\big(\nu\,|\,\mu\big). \tag{4.8}$$

Finally, using (4.5) for small t and (4.8) for large t, one deduces that, for every $u > 0$,

$$\mathrm{H}\big(\nu\,|\,\mu\big) \,\leq\, \mathrm{I}\big(\nu\,|\,\mu\big)\int_0^u e^{-2\rho t}dt + \mathrm{S}^2\big(\nu\,|\,\mu\big)\int_u^\infty \frac{\kappa}{\sigma(e^{2\kappa t}-1)}\,dt$$
$$=\, \frac{\mathrm{I}(\nu\,|\,\mu)}{2\rho}\,(1-e^{-2\rho u}) - \frac{\mathrm{S}^2(\nu\,|\,\mu)}{2\sigma}\,\log(1-e^{-2\kappa u}).$$

Setting $r = e^{-2u}$,

$$\mathrm{H}\big(\nu\,|\,\mu\big) \,\leq\, \inf_{0<r<1} \bigg\{ \frac{\mathrm{I}(\nu\,|\,\mu)}{2\rho}\,(1-r^\rho) - \frac{\mathrm{S}^2(\nu\,|\,\mu)}{2\sigma}\,\log(1-r^\kappa)\bigg\}.$$

Now, using that $1 - r^\rho \leq \max(1, \frac{\rho}{\kappa})(1-r^\kappa)$ for $r \in (0,1)$, a simple (non-optimal) optimization yields the desired conclusion. The proof of Theorem 4.1 is complete. □

Remark 4.2. It should be pointed out that, on the basis of (4.8), transport inequalities as studied in Section 3 may be investigated similarly in the preceding general context, and with similar illustrations as developed below. For example, as an analogue of (3.1),

$$\mathrm{W}_2(\nu,\mu) \,\leq\, \frac{2}{\sqrt{\kappa\sigma}}\,\mathrm{S}\big(\nu\,|\,\mu\big).$$

In order not to expand too much the exposition, we leave the details to the reader.

The next paragraphs present various illustrations of Theorem 4.1.

4.2 Multivariate gamma distribution

As a first example of illustration of the preceding general result, we consider the case of the multidimensional Laguerre operator, which is the product on $\mathbb{R}_+^d$ of one-dimensional Laguerre operators of parameters $p_i > 0$, $i = 1,\ldots,d$, that is,

$$\mathcal{L}f \,=\, \sum_{i=1}^d x_i\,\frac{\partial^2 f}{\partial x_i^2} + \sum_{i=1}^d (p_i - x_i)\,\frac{\partial f}{\partial x_i}\,.$$

In particular, $a(x) = (x_i\delta_{ij})_{1\leq i,j\leq d}$ in (4.1). It is a standard fact that the invariant measure μ associated with $\mathcal{L}$ has a density with respect to the Lebesgue measure given by the tensor product of d gamma densities of the type $\Gamma(p_i)^{-1}x_i^{p_i-1}e^{-x_i}$, $x_i \in \mathbb{R}_+$, $i = 1,\ldots,d$. For reasons that will become clear later on, we assume that $p_i \geq \frac{3}{2}$, $i = 1,\ldots,d$.

After some easy but cumbersome calculations, it may be checked that, along suitable smooth functions f,

$$\begin{aligned}
\Gamma(f) &= \sum_{i=1}^{d} x_i \left(\frac{\partial f}{\partial x_i}\right)^2 \\
\Gamma_2(f) &= \sum_{i,j=1}^{d} x_i x_j \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)^2 + \sum_{i=1}^{d} x_i \,\frac{\partial f}{\partial x_i}\,\frac{\partial^2 f}{\partial x_i^2} + \frac{1}{2}\sum_{i=1}^{d}(p_i + x_i)\left(\frac{\partial f}{\partial x_i}\right)^2 \\
\Gamma_3(f) &= \sum_{i,j,k=1}^{d} x_i x_j x_k \left(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}\right)^2 + 3\sum_{i,j=1}^{d} x_i x_j \,\frac{\partial^2 f}{\partial x_i \partial x_j}\,\frac{\partial^3 f}{\partial x_i^2 \partial x_j} \\
&\quad + \frac{3}{2}\sum_{i,j=1}^{d}(p_i + x_i) x_j \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)^2 + \frac{3}{2}\sum_{i=1}^{d} x_i \left(\frac{\partial^2 f}{\partial x_i^2}\right)^2 \\
&\quad + \frac{3}{2}\sum_{i=1}^{d} x_i \,\frac{\partial f}{\partial x_i}\,\frac{\partial^2 f}{\partial x_i^2} + \frac{1}{4}\sum_{i=1}^{d}(3p_i + x_i)\left(\frac{\partial f}{\partial x_i}\right)^2 .
\end{aligned}$$

Note that (recall $x_i, x_j, x_k \geq 0$)

$$\begin{aligned}
&\sum_{i,j,k=1}^{d} x_i x_j x_k \left(\frac{\partial^3 f}{\partial x_i \partial x_j\, \partial x_k}\right)^2 + 3\sum_{i,j=1}^{d} x_i x_j \frac{\partial^2 f}{\partial x_i \partial x_j}\,\frac{\partial^3 f}{\partial x_i^2 \partial x_j} \\
&\qquad \geq \sum_{i,j=1}^{d} x_i^2 x_j \left(\frac{\partial^3 f}{\partial x_i^2 \partial x_j}\right)^2 + 3\sum_{i,j=1}^{d} x_i x_j \,\frac{\partial^2 f}{\partial x_i \partial x_j}\,\frac{\partial^3 f}{\partial x_i^2 \partial x_j} \\
&\qquad \geq -\frac{9}{4}\sum_{i,j=1}^{d} x_j \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)^2 .
\end{aligned}$$

Therefore

$$\begin{aligned}
\Gamma_3(f) \;\geq\; & \frac{3}{2}\sum_{i,j=1}^{d}\Big(p_i - \frac{3}{2} + x_i\Big) x_j \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)^2 + \frac{3}{2}\sum_{i=1}^{d} x_i \left(\frac{\partial^2 f}{\partial x_i^2}\right)^2 \\
& + \frac{3}{2}\sum_{i=1}^{d} x_i \,\frac{\partial f}{\partial x_i}\,\frac{\partial^2 f}{\partial x_i^2} + \frac{1}{4}\sum_{i=1}^{d}(3p_i + x_i)\left(\frac{\partial f}{\partial x_i}\right)^2 .
\end{aligned}$$

Since $p_i \geq \frac{3}{2}$, it follows at once that $\Gamma_3(f) \geq \frac{1}{2}\,\Gamma_2(f)$. Analogous computations lead to

$$\sum_{i=1}^{d} x_i \,\frac{\partial f}{\partial x_i}\,\frac{\partial^2 f}{\partial x_i^2} \;\geq\; -\frac{1}{2}\sum_{i=1}^{d} x_i^2 \left(\frac{\partial^2 f}{\partial x_i^2}\right)^2 - \frac{1}{2}\sum_{i=1}^{d}\left(\frac{\partial f}{\partial x_i}\right)^2 ,$$

implying that

$$\Gamma_2(f) \;\geq\; \frac{1}{2}\sum_{i=1}^{d} x_i^2 \left(\frac{\partial^2 f}{\partial x_i^2}\right)^2 + \frac{1}{2}\sum_{i=1}^{d}(p_i - 1 + x_i)\left(\frac{\partial f}{\partial x_i}\right)^2 \;\geq\; \frac{1}{2}\,\Gamma(f).$$

Finally, one has

$$
\begin{aligned}
\frac{1}{2}\left\|\sqrt{a}\,\mathrm{Hess}(f)\,\sqrt{a}\right\|_{\mathrm{HS}}^2 &= \frac{1}{2}\sum_{i,j=1}^d x_i x_j \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)^2 \\
&\leq \frac{1}{2}\sum_{i,j=1}^d x_i x_j \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)^2 + \frac{1}{2}\sum_{i=1}^d \left(x_i \frac{\partial^2 f}{\partial x_i^2} + \frac{\partial f}{\partial x_i}\right)^2 \\
&\leq \Gamma_2(f).
\end{aligned}
$$

As a consequence, Theorem 4.1 applies with $\rho = \kappa = \sigma = \frac{1}{2}$ to yield the following result (the numerical constants there are not sharp). The restrictions $p_i \geq \frac{3}{2}$, $i = 1, \ldots, d$, are probably not optimal. For example, it is not difficult to see from the preceding computations that in the one-dimensional case $d = 1$, it is actually enough to assume that $p \geq \frac{1}{2}$.

Proposition 4.3 (HSI inequality for gamma distribution)**.** *Let μ be the product measure of gamma distributions $\Gamma(p_i)^{-1} x_i^{p_i - 1} e^{-x_i} dx_i$ on $\mathbb{R}_+^d$ with $p_i \geq \frac{3}{2}$, $i = 1, \ldots, d$. Then, for any $d\nu = h d\mu$ where h is a smooth probability density,*

$$
\mathrm{H}\big(\nu \,|\, \mu\big) \,\leq\, \mathrm{S}^2\big(\nu \,|\, \mu\big)\,\Psi\bigg(\frac{\mathrm{I}(\nu \,|\, \mu)}{\mathrm{S}^2(\nu \,|\, \mu)}\bigg).
$$

4.3 One-dimensional uniform distribution on $[-1, +1]$

In this section, we examine the case of the one-dimensional Jacobi operator of parameters $\alpha = \beta = 1$, that is,

$$
\mathcal{L}f \,=\, (1 - x^2) f'' - 2x f',
$$

whose associated invariant measure μ is uniform distribution on $[-1, +1]$. The general family of parameters with the beta distributions as invariant measures (cf. [B-G-L, Section 2.7.4]) may be considered similarly, at the expense however of tedious computations, as well as multivariate (product) versions. For simplicity, we only detail this case to better illustrate the conclusion.

Easy calculations lead to, for a smooth function f on $[-1, +1]$,

$$
\begin{aligned}
\Gamma(f) &= (1 - x^2) f'^2 \\
\Gamma_2(f) &= (1 + x^2) f'^2 + (1 - x^2)^2 f''^2 - 2x(1 - x^2) f' f'' \\
\Gamma_3(f) &= (1 - x^2)^3 f'''^2 - 6x(1 - x^2)^2 f'' f''' - 2(1 - x^2)^2 f' f''' \\
&\quad + 3(1 - x^2)(1 + 3x^2) f''^2 + 6x(1 - x^2) f' f'' + (3 - x^2) f'^2.
\end{aligned}
$$

Observe that

$$
\Gamma_2(f) \,=\, f'^2 + \big(x f' - (1 - x^2) f''\big)^2 \,\geq\, \Gamma(f).
$$

Furthermore,

$$\begin{aligned}\Gamma_3(f) - \Gamma_2(f) \,&=\, (1-x^2)\Big[(1-x^2)^2 f'''^2 - 6x(1-x^2)f''f''' \\ &\quad - 2(1-x^2)f'f''' + 2(1+5x^2)f''^2 + 8xf'f'' + 2f'^2\Big] \\ &=\, (1-x^2)\Big[\big((1-x^2)f''' - 3xf'' - f'\big)^2 + \big(f' + xf''\big)^2 + 2f''^2\Big] \,\geq\, 0\end{aligned}$$

so that $\Gamma_3(f) \geq \Gamma_2(f)$. Also,

$$\begin{aligned}\Gamma_2(f) \,&\geq\, (1+x^2)f'^2 + (1-x^2)^2 f''^2 - 2x^2 f'^2 - \frac{1}{2}(1-x^2)^2 f''^2 \\ &= (1-x^2)f'^2 + \frac{1}{2}(1-x^2)^2 f''^2 \\ &\geq \frac{1}{2}(1-x^2)^2 f''^2.\end{aligned}$$

Hence, Theorem 4.1 applies with $\rho = \kappa = 1$ and $\sigma = \frac{1}{2}$ (note that $a(x) = 1 - x^2$) to yield the following conclusion. Again, the numerical constants are not sharp.

Proposition 4.4 (HSI inequality for the uniform distribution)**.** *Let μ be uniform probability measure on $[-1,+1]$. Then, for any $d\nu = hd\mu$ where h is a smooth probability density,*

$$\mathrm{H}\big(\nu\,|\,\mu\big) \,\leq\, \mathrm{S}^2\big(\nu\,|\,\mu\big)\,\Psi\bigg(\frac{\mathrm{I}(\nu\,|\,\mu)}{2\,\mathrm{S}^2(\nu\,|\,\mu)}\bigg)$$

4.4 Families of log-concave distributions

We consider here a diffusion operator on the line of the type

$$\mathcal{L}f \,=\, f'' - u'f'$$

associated with a symmetric invariant probability measure $d\mu = e^{-u}dx$, where u is a smooth potential on $\mathbb{R}$. The Gaussian model corresponds to the quadratic potential $u(x) = \frac{x^2}{2}$.

We have, for smooth functions f,

$$\begin{aligned}\Gamma(f) \,&=\, f'^2 \\ \Gamma_2(f) \,&=\, f''^2 + u''f'^2 \\ \Gamma_3(f) \,&=\, f'''^2 + 3u'''f'f'' + 3u''f''^2 + \frac{1}{2}\,\big(u^{(4)} - u'u''' + 2u''^2\big)f'^2.\end{aligned}$$

Assume that there exists $c > 0$ such that, uniformly, $u'' \geq c$,

$$u^{(4)} - u'u''' + 2u''^2 - 6cu'' \,\geq\, 0 \tag{4.9}$$

and

$$3u'''^2 \leq 2(u''-c)\big(u^{(4)} - u'u''' + 2u''^2 - 6cu''\big). \tag{4.10}$$

Then $\Gamma_2(f) \geq c\,\Gamma(f)$, $\Gamma_2(f) \geq f''^2$ and $\Gamma_3(f) \geq 3c\,\Gamma_2(f)$ for every f. Hence, Theorem 4.1 applies with $\rho = c$, $\kappa = 3c$ and $\sigma = 1$.

Proposition 4.5 (HSI inequality for log-concave distribution)**.** *Let $d\mu = e^{-u}dx$ on $\mathbb{R}$ where u is a smooth potential on $\mathbb{R}$ such that for some $c > 0$, $u'' \geq c$ and* (4.9) *and* (4.10) *hold. Then, for any $d\nu = hd\mu$ where h is a smooth probability density,*

$$\mathrm{H}\big(\nu\,|\,\mu\big) \leq \frac{1}{2}\,\mathrm{S}^2\big(\nu\,|\,\mu\big)\,\Psi\bigg(\frac{\mathrm{I}(\nu\,|\,\mu)}{c\,\mathrm{S}^2(\nu\,|\,\mu)}\bigg).$$

Recall that in this context, the only condition $u'' \geq c > 0$ ensures the logarithmic Sobolev inequality for μ [B-G-L, Corollary 5.7.2]. It is not difficult to find (simple) examples outside the Gaussian model (corresponding to $c = \frac{1}{3}$) such that conditions (4.9) and (4.10) are fulfilled. For example, if $u(x) = \frac{x^2}{2} + \varepsilon x^4$, it is easily seen that these hold for $c = \frac{1}{4}$ and $\varepsilon = \frac{1}{12}$ (for instance). In the Gaussian case, the estimate obtained in this proposition is somewhat worse than the HSI inequality of Theorem 2.2. At the expenses of more involved conditions (4.9) and (4.10), multidimensional versions may be considered similarly.

5 Entropy bounds on laws of functionals

As emphasized in the introduction, the new HSI inequalities described in the preceding sections provide entropic bounds on probability measures ν which may be used towards convergence in entropy via the Stein discrepancy $S(\nu\,|\,\mu)$. Now, these bounds assume that the Fisher information $\mathrm{I}_\mu(h)$ of the density h of ν with respect to μ is finite (in order to control $\mathrm{I}_\mu(P_t h)$ in small time), which may or may not hold in specific illustrations. The goal pursued in the second part of this work is actually to overcome this difficulty and to describe conditions (integrability and tail behavior) on the initial data itself of a multidimensional functional $F = (F_1, \ldots, F_d)$ with distribution $\nu = \nu_F$ (on $\mathbb{R}^d$) in order to control the Fisher information $\mathrm{I}_\mu(P_t h)$ in small time. This investigation was initiated in [N-P-S1] in Wiener space towards the first normal approximation results in entropy for Wiener chaos distributions. Here, we consider distributions of functionals on a Markov Triple structure (E, μ, Γ) already put forward in the preceding section, and describe how the associated Γ-calculus may be developed towards normal (as well as gamma) approximations in the entropic sense.

Referring as before to [B-G-L] for a complete account, we thus deal with a Markov Triple (E, μ, Γ) on a probability space $(E, \mathcal{E}, \mu)$, with Markov semigroup $(P_t)_{t\geq 0}$ with symmetric and invariant probability measure μ, infinitesimal generator L, associated carré du champ operator Γ and underlying algebra of (smooth) functions $\mathcal{A}$. Integration by parts expresses that

$$\int_E f\,\mathrm{L}g\,d\mu \,=\, -\int_E \Gamma(f,g)d\mu \tag{5.1}$$

for every $f, g \in \mathcal{A}$.

The second order differential operators of Section 4 provide instances of this general framework. Gaussian and Wiener spaces with associated Ornstein-Uhlenbeck semigroup and generator are a prototypical example for the illustrations. Note in particular that Wiener chaoses as investigated in [N-P-S1] are eigenfunctions of the Ornstein-Uhlenbeck generator. Eigenfunctions of the underlying operator L are actually of special interest in the context of the Stein method as illustrated in Section 5.1.

For $d \geq 1$, let $F = (F_1, \ldots, F_d)$ be a vector defined on $(E, \mathcal{E}, \mu)$, where each F_i is centered and square-integrable, and denote by ν_F the law of F. Common to the three Sections 5.1–5.3 below, assume that the distribution ν_F of F admits a density h with respect to the standard Gaussian distribution γ on $\mathbb{R}^d$ (in particular, ν_F is absolutely continuous with respect to the Lebesgue measure). In the first part, we describe the Stein kernel and discrepancy for vectors of eigenfunctions of L. Next, we address some direct bounds on the Fisher information $\mathrm{I}_\gamma(h)$ in terms of the data of the functional F and its gradients. Then, we develop the results on entropic normal approximations, extending the conclusions in [N-P-S1], by an analysis of the small time behavior of $\mathrm{I}_\gamma(P_t h)$. Finally, we address similar issues in the context of one-dimensional gamma approximation.

5.1 Stein kernel and discrepancy for eigenfunctions

The first statement shows that, whenever the vector F is composed of eigenfunctions of L, a Stein kernel τ_{ν_F} of ν_F with respect to γ as defined in (2.1) can be expressed in terms of the carré du champ operator Γ.

Proposition 5.1 (Stein kernel for eigenfunctions)**.** *Let $F = (F_1, \ldots, F_d)$ on $(E, \mathcal{E}, \mu)$ such that, for every $i = 1, \ldots, d$, the random variable F_i is an eigenfunction of $-\mathrm{L}$, with eigenvalue $\lambda_i > 0$. Assume moreover that $\Gamma(F_i, F_j) \in \mathrm{L}^1(\mu)$ for every $i, j = 1, \ldots, d$. Then, the matrix-valued map τ_{ν_F} defined as*

$$\tau_{\nu_F}^{ij}(x_1, \ldots, x_d) \,=\, \frac{1}{\lambda_i}\,\mathbb{E}_\mu\Big[\Gamma(F_i, F_j)\,\big|\, F = (x_1, \ldots, x_d)\Big], \quad i, j = 1, \ldots, d, \tag{5.2}$$

is a Stein kernel for ν_F, that is, it satisfies (2.1)*. (The right-hand side of* (5.2) *indicates a version of the conditional expectation of $\Gamma(F_i, F_j)$ with respect to F under the probability measure μ.)*

Proof. Use integration by parts with respect to L to get that, for every smooth test function φ on $\mathbb{R}^d$ and every $i = 1, \ldots, d$,

$$\lambda_i \int_E F_i\,\varphi(F)d\mu \,=\, -\int_E \mathrm{L}F_i\,\varphi(F)d\mu \,=\, \sum_{j=1}^d \int_E \Gamma(F_i, F_j)\,\frac{\partial\varphi}{\partial x_j}(F)d\mu.$$

The proof is concluded by taking conditional expectations. □

As a consequence, together with (2.5) and Jensen's inequality,

$$\mathrm{S}^2\big(\nu_F \,|\, \gamma\big) \,\leq\, \sum_{i,j=1}^{d} \frac{1}{\lambda_i^2}\,\mathrm{Var}_\mu\big(\Gamma(F_i, F_j)\big) + \big\|C - \mathrm{Id}\big\|_{\mathrm{HS}}^2 \,=\, \mathrm{V}^2 \tag{5.3}$$

where C denotes the covariance matrix of ν_F, providing therefore a tractable way to control the Stein discrepancy in this case. In addition, combining with the HSI inequality of Theorem 2.2 immediately yields the following statement.

Corollary 5.2. *Under the assumptions and notation of Proposition 5.1,*

$$\mathrm{H}\big(\nu_F \,|\, \gamma\big) \,\leq\, \mathrm{V}^2 \log\left(1 + \frac{\mathrm{I}(\nu_F \,|\, \gamma)}{\mathrm{V}^2}\right). \tag{5.4}$$

Example 5.3. A typical example of a Markov Triple for which the quantity V^2 appearing in the above bound can be estimated explicitly corresponds to the case where $(E, \mathcal{E}, \mu)$ is a probability space supporting an isonormal Gaussian process $X = \{X(h) : h \in \mathfrak{H}\}$ over some real separable Hilbert space $\mathfrak{H}$, and L is the generator of the associated Ornstein-Uhlenbeck semigroup. In this case, $\Gamma(F, G) = \langle DF, DG\rangle_{\mathfrak{H}}$ for smooth functionals F and G, where D stands for the Malliavin derivative operator, and the eigenspaces of $-\mathrm{L}$ are the so-called *Wiener chaoses* $\{C_k : k \geq 0\}$ of X. For $k = 0, 1, 2, \ldots$, the eigenvalue of C_k is given by k. A detailed discussion about how to bound a quantity such as V^2 in the case of random vectors with components inside a Wiener chaos can be found in [N-P2, Chapter 6]. In particular, if $d = 1$ and F belongs to C_k, then V^2 can be controlled by the second and fourth moments of F as

$$\begin{aligned}\mathrm{V}^2 \,&=\, \big(\mathbb{E}[F^2] - 1\big)^2 + \frac{1}{k^2}\,\mathrm{Var}\big(\|DF\|_{\mathfrak{H}}^2\big) \\ &\leq\, \big(\mathbb{E}[F^2] - 1\big)^2 + \frac{k-1}{3k}\,\big(\mathbb{E}[F^4] - 3\,\mathbb{E}[F^2]^2\big).\end{aligned}$$

In particular, such an estimate provides a proof of the famous 'fourth moment theorem' for chaotic random variables, cf. [N-P2, Theorem 5.2.7].

Remark 5.4. While eigenfunctions appear as functionals of particular interest for the control of the Stein discrepancy itself, the Γ-calculus actually provides a formal description of Stein kernels of a given functional F on (E, μ, Γ) (in dimension one for simplicity) as the conditional expectation with respect to F of $\Gamma(F, \mathrm{L}^{-1}F)$ (where $\mathrm{L}^{-1}F = \int_0^\infty P_t F dt$). This observation further expands on the preceding example, allowing for a rather general analysis.

5.2 Bounds on the Fisher information

When dealing with the upper-bound (5.4), the Fisher information $\mathrm{I}(\nu_F \,|\, \gamma) = \mathrm{I}_\gamma(h)$ of the density h of the law ν_F of F cannot always be explicitly deduced from the data concerning the random vector F. The task of this paragraph is therefore to deduce some useful bounds on $\mathrm{I}(\nu_F \,|\, \gamma)$ in terms of F and its gradients.

Let $F = (F_1, \ldots, F_d)$ be general vector of centered and square-integrable random variables (that need not necessarily be eigenfunctions of $-\mathrm{L}$). Recall that the distribution ν_F of F is assumed to admit a (smooth) density h with respect to the standard Gaussian distribution γ on $\mathbb{R}^d$. It is furthermore implicitly assumed that all the F_i's are in $\mathcal{A}$ (or some extended algebra in the sense of [B-G-L]) allowing for the formal computations developed next. These assumptions should then be verified on the concrete examples of interest (such as Wiener chaoses).

Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be smooth enough. By integration by parts (5.1) with respect to L, for every $w \in \mathcal{A}$, and every $i, j = 1, \ldots, d$,

$$\sum_{k=1}^d \int_E w\, \Gamma(F_i, F_k) \frac{\partial^2 \phi}{\partial x_k \partial x_j}(F) d\mu \;=\; -\int_E \mathrm{L} F_i \, w \, \frac{\partial \phi}{\partial x_j}(F) d\mu - \int_E \Gamma(F_i, w) \frac{\partial \phi}{\partial x_j}(F) d\mu.$$

Let $\widetilde{\Gamma}$ be the symmetric matrix with entries $\Gamma(F_i, F_j)$, $i, j = 1, \ldots, d$. Applying the latter to $w = w_{ij}$, symmetric in i, j, yields

$$\begin{aligned}
&\int_E \mathrm{Tr}\big(W \widetilde{\Gamma}\, \mathrm{Hess}(\phi)(F)\big) d\mu \\
&\qquad = -\sum_{i,j=1}^d \int_E \mathrm{L} F_i \, w_{ij} \, \frac{\partial \phi}{\partial x_j}(F) d\mu - \sum_{i,j=1}^d \int_E \Gamma(F_i, w_{ij}) \, \frac{\partial \phi}{\partial x_j}(F) d\mu
\end{aligned} \tag{5.5}$$

where $W = (w_{ij})_{1 \leq i,j \leq d}$. Provided it exists, set $W = \widetilde{\Gamma}^{-1}$, so that the left-hand side in the previous identity is just $\int_E \Delta\phi(F) d\mu$. Recalling from (2.7) the Ornstein-Uhlenbeck generator $\mathcal{L} = \Delta - x \cdot \nabla$ associated with the standard Gaussian distribution γ on $\mathbb{R}^d$, it follows that

$$\begin{aligned}
-\int_E \mathcal{L}\phi(F) d\mu \;=\; &\sum_{i,j=1}^d \int_E \mathrm{L} F_i \, (\widetilde{\Gamma}^{-1})_{ij} \, \frac{\partial \phi}{\partial x_j}(F) d\mu \\
&+ \sum_{i,j=1}^d \int_E \Gamma\big(F_i, (\widetilde{\Gamma}^{-1})_{ij}\big) \frac{\partial \phi}{\partial x_j}(F) d\mu + \sum_{j=1}^d \int_E F_j \, \frac{\partial \phi}{\partial x_j}(F) d\mu.
\end{aligned}$$

In more compact notation, if

$$V \;=\; \bigg(\sum_{i=1}^d \Gamma\big(F_i, (\widetilde{\Gamma}^{-1})_{ij}\big) \bigg)_{1 \leq j \leq d} \qquad \text{and} \qquad U \;=\; \widetilde{\Gamma}^{-1} \mathrm{L} F + V + F,$$

then

$$-\int_E \mathcal{L}\phi(F)d\mu \,=\, \int_E U\cdot\nabla\phi(F)d\mu.$$

Applied to $\phi = v = \log h$, by the Cauchy-Schwarz inequality and (4.2),

$$\mathrm{I}_\gamma(h) \,\leq\, \int_E |U|^2 d\mu.$$

The consequences of the previous computations are gathered together in the next statement, where we point out a set of sufficient conditions on F and its gradients $\Gamma(F_i, F_j)$ ensuring that the random variable U is indeed square-integrable.

Proposition 5.5 (Bound on the Fisher information)**.** *Let $F = (F_1, \ldots, F_d)$ be a vector of elements of $\mathcal{A}$ on (E, μ, Γ). Assume that all the F_i, $\mathrm{L}F_i$, $\Gamma(F_i, F_j)$, $i, j = 1, \ldots, d$, and $\frac{1}{\det(\widetilde{\Gamma})}$ are in $\mathrm{L}^p(\mu)$ for every $p \geq 1$. Then, $\int_E |U|^2 d\mu < \infty$ and*

$$\mathrm{I}\big(\nu_F\,|\,\gamma\big) \,\leq\, \int_E |U|^2 d\mu. \tag{5.6}$$

The condition on $\frac{1}{\det(\widetilde{\Gamma})}$ in Proposition 5.5 has some similarity with basic assumptions in Malliavin calculus (cf. [N, N-P2]).

Example 5.6. One may of course wonder whether the bound (5.6) is of any interest. Here is a simple example showing that there are instances where $\mathrm{I}_\gamma(h)$ might be quite intricate to handle directly on the density h of the distribution of F while U has a clear description. On $E = \mathbb{R}^{2n}$ with the standard Gaussian measure $\mu = \gamma$ and $\Gamma(f) = |\nabla f|^2$ the standard carré du champ operator, let

$$F(x) \,=\, x_1x_2 + x_3x_4 + \cdots + x_{2n-1}x_{2n}, \quad x = (x_1, \ldots, x_{2n}) \in \mathbb{R}^{2n}.$$

It is classical that the distribution of the product of two independent standard normal has a density (with respect to the Lebesgue measure on $\mathbb{R}$) given by a Bessel function. The density h of the distribution of F is thus rather involved. On the other hand, it is easily seen that

$$\mathrm{L}F \,=\, -2F \quad \text{and} \quad \Gamma(F) \,=\, x_1^2 + \cdots + x_{2n}^2 \,=\, R^2$$

so that

$$U \,=\, F\Big(-\frac{2}{R^2} - \frac{4}{R^4} + 1\Big).$$

By using polar coordinates, it is immediately seen that $\int_{\mathbb{R}^{2n}} U^2 d\mu < \infty$ as soon as $n \geq 5$.

5.3 Fisher information growth and normal approximation

One evident drawback of Proposition 5.5 of the previous paragraph is that, since the quantity $|U|$ is singular as the determinant of $\widetilde{\Gamma}$ is close to 0, one is forced to assume that $\frac{1}{\det(\widetilde{\Gamma})}$ is in all $\mathrm{L}^p(\mu)$ spaces (or at least for some p large enough depending on d). This assumption is in general too strong, and very difficult to check in concrete situations. The idea developed in this section (which generalizes the approach initiated in [N-P-S1]) is that, under weaker moment assumptions, while the Fisher information $\mathrm{I}_\gamma(h)$ might be infinite, it is nevertheless possible to control the growth as $t \to 0$ of $\mathrm{I}_\gamma(P_t h)$. Together with the control in terms of the Stein discrepancy for large time achieved in Section 2, one may then reach entropic bounds which can be handled in concrete examples (such as those of random vectors whose components belong to some Wiener chaos).

As before, let $F = (F_1, \ldots, F_d)$ be a general vector of centered and square-integrable random variables (in the algebra $\mathcal{A}$ or some natural extension), with distribution $d\nu_F = h d\gamma$. As a crucial assumption, ν_F has a Stein kernel τ_{ν_F} with respect to γ as defined in (2.1) (see also Proposition 5.1 and Remark (5.4)). Recall the matrix $\widetilde{\Gamma}$ with entries $\Gamma(F_i, F_j)$, $i, j = 1, \ldots, d$. Also, in what follows we use the convention that, if $\widetilde{\Gamma}$ is singular, then the matrix $\det(\widetilde{\Gamma})\,\widetilde{\Gamma}^{-1}$ must be understood as the transpose of usual adjugate matrix operator of $\widetilde{\Gamma}$ (both quantities being of course equal for non-singular matrices).

With the notation of the preceding section, given $\varepsilon > 0$, write first, again for a smooth function ϕ on $\mathbb{R}^d$ and $\mathcal{L}$ the Ornstein-Uhlenbeck operator in $\mathbb{R}^d$,

$$\begin{aligned}
\int_E \mathcal{L}\phi(F) d\mu &= \int_E \Delta\phi(F) d\mu - \int_E F \cdot \nabla\phi(F) d\mu \\
&= \int_E \frac{\det(\widetilde{\Gamma})}{\det(\widetilde{\Gamma}) + \varepsilon}\, \Delta\phi(F) d\mu + \int_E \frac{\varepsilon}{\det(\widetilde{\Gamma}) + \varepsilon}\, \Delta\phi(F) d\mu \\
&\quad - \int_E F \cdot \nabla\phi(F) d\mu.
\end{aligned}$$

Choose $W = \frac{\det(\widetilde{\Gamma})\,\widetilde{\Gamma}^{-1}}{\det(\widetilde{\Gamma}) + \varepsilon}$ in (5.5), so that

$$\int_E \frac{\det(\widetilde{\Gamma})}{\det(\widetilde{\Gamma}) + \varepsilon}\, \Delta\phi(F) d\mu = -\int_E \left(\frac{\det(\widetilde{\Gamma})\,\widetilde{\Gamma}^{-1}\mathrm{L}F + V_1}{\det(\widetilde{\Gamma}) + \varepsilon} - \frac{V_2}{(\det(\widetilde{\Gamma}) + \varepsilon)^2} \right) \cdot \nabla\phi(F) d\mu$$

where

$$V_1 = \left(\sum_{i=1}^d \Gamma\big(F_i, \det(\widetilde{\Gamma})(\widetilde{\Gamma}^{-1})_{ij}\big) \right)_{1 \le j \le d}$$

and

$$V_2 = \left(\sum_{i=1}^d \det(\widetilde{\Gamma})(\widetilde{\Gamma}^{-1})_{ij}\, \Gamma\big(F_i, \det(\widetilde{\Gamma})\big) \right)_{1 \le j \le d}$$

Apply now the preceding to $\phi = P_t v_t$, $v_t = \log P_t h$, $t > 0$. Since $\nabla P_t v_t(F) = e^{-t} P_t(\nabla v_t)$ and

$$\mathrm{I}_\gamma(P_t h) \,=\, \int_E P_t\big(|\nabla v_t|^2\big)(F) d\mu,$$

by the Cauchy-Schwarz inequality, assuming for simplicity that $0 < \varepsilon \leq 1$,

$$\begin{aligned}&\bigg| \int_E \frac{\det(\widetilde{\Gamma})}{\det(\widetilde{\Gamma}) + \varepsilon} \, \Delta P_t v_t(F) d\mu - \int_E F \cdot \nabla P_t v_t(F) d\mu \bigg| \\ &\qquad\qquad \leq \, \frac{e^{-t}}{\varepsilon^2} \bigg(\int_E \Big[\big| \det(\widetilde{\Gamma}) \, \widetilde{\Gamma}^{-1} \mathrm{L} F \big| + |V_1| + |V_2| + |F| \Big]^2 d\mu \bigg)^{1/2} \mathrm{I}_\gamma(P_t h)^{1/2}\end{aligned}$$

On the other hand, using the same semigroup computations as in Section 2,

$$\Delta P_t v_t(F) \,=\, \frac{e^{-2t}}{\sqrt{1 - e^{-2t}}} \int_{\mathbb{R}^d} y \cdot \nabla v_t\big(e^{-t} F + \sqrt{1 - e^{-2t}}\, y \big) d\gamma(y)$$

so that

$$\bigg| \int_E \frac{\varepsilon}{\det(\widetilde{\Gamma}) + \varepsilon} \, \Delta P_t v_t(F) d\mu \bigg| \,\leq\, \frac{\sqrt{d}\, \varepsilon \, e^{-2t}}{\sqrt{1 - e^{-2t}}} \bigg(\int_E \frac{1}{(\det(\widetilde{\Gamma}) + \varepsilon)^2} \, d\mu \bigg)^{1/2} \mathrm{I}_\gamma(P_t h)^{1/2}.$$

Assume now that

$$\int_E \Big[\big| \det(\widetilde{\Gamma}) \, \widetilde{\Gamma}^{-1} \mathrm{L} F \big| + |V_1| + |V_2| + |F| \Big]^2 d\mu \,=\, A_F < \infty \tag{5.7}$$

and that

$$\int_E \frac{1}{(\det(\widetilde{\Gamma}) + \varepsilon)^2} \, d\mu \,\leq\, \delta(\varepsilon). \tag{5.8}$$

Collecting the preceding bounds and recalling from (4.7) that

$$\int_E \mathcal{L} P_t v_t(F) d\mu \,=\, \int_E \mathcal{L} P_t v_t \, h d\mu \,=\, - \mathrm{I}_\gamma(P_t h)$$

yields that, for $t > 0$ and $0 < \varepsilon \leq 1$,

$$\mathrm{I}_\gamma(P_t h) \,\leq\, 2 e^{-2t} \bigg(\frac{A_F}{\varepsilon^4} + \frac{d \, \varepsilon^2 \delta(\varepsilon)}{1 - e^{-2t}} \bigg). \tag{5.9}$$

In the following statement, we determine a handy set of sufficient conditions on F and its gradients ensuring that, for some choice of $\varepsilon = \varepsilon(t) > 0$, the function on the right-hand side of (5.9) is integrable for the small values of $t > 0$. Combined with (2.19) for the large values of $t > 0$, a control of the entropy of ν_F in terms of the Stein discrepancy $\mathrm{S}(\nu_F \,|\, \gamma)$ may then be produced. Recall the function Ψ on $\mathbb{R}_+$ given by $\Psi(r) = 1 + \log r$ if $r \geq 1$ and $\Psi(r) = r$ if $0 \leq r \leq 1$.

Theorem 5.7 (normal entropic approximation via Stein discrepancy)**.** *Let $F = (F_1, \dots, F_d)$ be a vector of centered elements of $\mathcal{A}$ on (E, μ, Γ). Assume that all the F_i, $\mathrm{L}F_i$, $\Gamma(F_i, F_j)$, $\Gamma(\Gamma(F_i, F_j), F_k)$, $i, j, k = 1, \dots, d$, are in $\mathrm{L}^p(\mu)$ for every $p \geq 1$, and that*

$$B_F \,=\, \int_E \frac{1}{\det(\widetilde{\Gamma})^\alpha} \, d\mu < \infty \tag{5.10}$$

for some $\alpha > 0$. Then, $A_F < \infty$ (as defined in (5.7)*) and*

$$\mathrm{H}\big(\nu_F \,|\, \gamma\big) \,\leq\, \frac{\mathrm{S}^2(\nu_F \,|\, \gamma)}{2(1 - 4\kappa)} \, \Psi\bigg(\frac{2(A_F + d(B_F + 1))}{\mathrm{S}^2(\nu_F \,|\, \gamma)}\bigg) \tag{5.11}$$

where $\kappa = \frac{2+\alpha}{2(4+3\alpha)}$ $(< \frac{1}{4})$.

Proof. First of all, we have that the parameter A_F is finite, since the expressions $\det(\widetilde{\Gamma})\widetilde{\Gamma}^{-1}\mathrm{L}F$, V_1 and V_2 only involve products of F_i, $\mathrm{L}F_i$ and $\Gamma(F_i, F_j)$, $\Gamma(\Gamma(F_i, F_j), F_k)$, $i, j, k = 1, \dots, d$.

Now, for every $\varepsilon > 0$ and $r > 0$,

$$\int_E \frac{1}{(\det(\widetilde{\Gamma}) + \varepsilon)^2} \, d\mu \,\leq\, \frac{1}{\varepsilon^2} \, \mu\big(\det(\widetilde{\Gamma}) \leq r\big) + \frac{1}{r^2} \,\leq\, \frac{B_F r^\alpha}{\varepsilon^2} + \frac{1}{r^2} \, . \tag{5.12}$$

The choice of $r = \varepsilon^{\frac{2}{\alpha+2}}$ yields (5.8) with $\delta(\varepsilon) = (B_F + 1)\varepsilon^{-\frac{4}{2+\alpha}}$. Let then $\varepsilon = \varepsilon(t) = (1 - e^{-2t})^\kappa$, $t \geq 0$, for $\kappa = \frac{2+\alpha}{2(4+3\alpha)}$ $(< \frac{1}{4})$. Then

$$\frac{A_F}{\varepsilon^4} + \frac{d\varepsilon^2 \delta(\varepsilon)}{1 - e^{-2t}} \,\leq\, \frac{A_F + d(B_F + 1)}{(1 - e^{-2t})^{4\kappa}}$$

from which, as a consequence of (5.9), for every $t > 0$,

$$\mathrm{I}_\gamma(P_t h) \,\leq\, 2\big[A_F + d(B_F + 1)\big] \, \frac{e^{-2t}}{(1 - e^{-2t})^{4\kappa}} \, . \tag{5.13}$$

To conclude, recall, as in the proof of Theorem 2.2, the decomposition for every $u > 0$,

$$\mathrm{H}\big(\nu_F \,|\, \gamma\big) \,\leq\, \int_0^u \mathrm{I}_\gamma(P_t h) dt + \mathrm{S}^2\big(\nu_F \,|\, \gamma\big) \int_u^\infty \frac{e^{-4t}}{1 - e^{-2t}} \, dt.$$

Therefore, by (5.13),

$$\begin{aligned} \mathrm{H}\big(\nu_F \,|\, \gamma\big) \,&\leq\, \frac{A_F + d(B_F + 1)}{1 - 4\kappa} \, (1 - e^{-2u})^{1 - 4\kappa} \\ &\quad\; + \frac{1}{2} \, \mathrm{S}^2\big(\nu_F \,|\, \gamma\big)\big(- e^{-2u} - \log(1 - e^{-2u})\big) \\ &\leq\, \frac{A_F + d(B_F + 1)}{1 - 4\kappa} \, (1 - e^{-2u})^{1 - 4\kappa} - \frac{1}{2} \, \mathrm{S}^2\big(\nu_F \,|\, \gamma\big) \log(1 - e^{-2u}), \end{aligned}$$

and the bound (5.11) in the statement follows by optimizing in $u > 0$ (set $(1 - e^{-2u})^{1-4\kappa} = r \in (0, 1)$.) Theorem 5.7 is established. □

Since $\Psi(r) \leq r$ for every $r \in \mathbb{R}_+$, observe from (5.11) that

$$\mathrm{H}\big(\nu_F \,|\, \gamma\big) \;\leq\; \frac{A_F + d(B_F + 1)}{(1 - 4\kappa)}$$

so that, under the assumptions of Theorem 5.7, one also has that $\mathrm{H}(\nu_F \,|\, \gamma) < \infty$, a conclusion of independent interest.

The quantity A_F of (5.7) involves integrability conditions on F and its gradients (they may actually be weakened according to the precise expression of A_F). On the other hand, B_F of (5.10) is rather concerned with a small ball behavior. For a vector $F = (F_1, \ldots, F_d)$ of eigenvectors of the underlying Markov generator L, Theorem 5.7 may be combined with (5.3) to fully control the relative entropy in terms of F and its gradients as now illustrated in some instances.

Typically, consider $F_n = (F_{1,n}, \ldots, F_{d,n})$, $n \in \mathbb{N}$, a sequence of vectors of centered elements of $\mathcal{A}$ on (E, μ, Γ) such that

(C_1) $F_{i,n}$, $\mathrm{L}F_{i,n}$, $\Gamma(F_{i,n}, F_{j,n})$, $\Gamma(\Gamma(F_{i,n}, F_{j,n}), F_{k,n})$, $i, j, k = 1, \ldots, d$, $n \in \mathbb{N}$, are uniformly bounded in $\mathrm{L}^p(\mu)$ for every $p \geq 1$;

(C_2) $\sup_{n\in\mathbb{N}} \int_E \det(\widetilde{\Gamma}_n)^{-\alpha}\, d\mu < \infty$ for some $\alpha > 0$.

Since (C_1) implies that $\sup_{n\in\mathbb{N}} A_{F_n} < \infty$ and (C_2) that $\sup_{n\in\mathbb{N}} B_{F_n} < \infty$, it follows from Theorem 5.7 that $\mathrm{H}(\nu_{F_n}|\gamma) \to 0$ provided that $\mathrm{S}(\nu_{F_n}|\gamma) \to 0$.

Example 5.8. We describe, in part following [N-P-S1], how this setting may be applied to concrete examples of interest.

(a) In the Wiener space framework of Example 5.3, fix some integers $k_1, \ldots, k_d \geq 1$ and, for any $i = 1, \ldots, d$, assume that the sequence $(F_{i,n})_{n\in\mathbb{N}}$ belongs to the Wiener chaos C_{k_i}. In particular, $\widetilde{\Gamma}_n = (\Gamma(F_{i,n}, F_{j,n}))_{1\leq i,j\leq d} = (\langle DF_{i,n}, DF_{j,n}\rangle_{\mathfrak{H}})_{1\leq i,j\leq d}$. Assume furthermore that the distribution of F_n converges to γ as $n \to \infty$. From [N-R, Lemma 2.1], it follows that $\sup_{n\in\mathbb{N}} \mathbb{E}[F_{i,n}^2] < \infty$ for every i. Since inside a Wiener chaos all the L^p-norms are equivalent (see e.g. [N-P2, Corollary 2.8.14]), the condition (C_1) is satisfied. Turning to (C_2), as a consequence of the main result in [N-OL], $\widetilde{\Gamma}_n \to \mathrm{Id}$ in L^2 as $n \to \infty$. Again using that inside a Wiener chaos all the L^p-norms are equivalent, it follows that $\mathbb{E}[\det(\widetilde{\Gamma}_n)] \to 1$ as $n \to \infty$. In particular, for n large enough ($n \geq n_0$ say), $\mathbb{E}[\det(\widetilde{\Gamma}_n)] \geq \frac{1}{2}$. Now, as a consequence of the Carbery-Wright inequality [C-W] (see [N-P-S1]), for some universal constant $c > 0$ and every $r > 0$ and $n \geq n_0$,

$$\mathbb{P}\big(\det(\widetilde{\Gamma}_n) \leq r\big) \;\leq\; cNr^{1/N}\mathbb{E}\big[\det(\widetilde{\Gamma}_n)\big]^{-1/N} \;\leq\; cN(2r)^{1/N}, \tag{5.14}$$

where $N \geq 1$ is an integer related to the degrees of the F_i's (cf. [N-P-S1, Lemma 4.3] for further details). With the preceding, condition (C_2) then clearly holds for any $\alpha < \frac{1}{N}$.

(b) It may be observed that the same bounds (5.14) hold true when the F_i's are polynomials under a log-concave measure $d\mu = e^{-u}dx$ on $\mathbb{R}^n$, at least when u is a polynomial or such that $|\nabla u| \in \mathrm{L}^p(\mu)$ for every $p \geq 1$. Indeed, the determinant $\det(\widetilde{\Gamma})$ is then also of this form, and the seminal result from [C-W] applies similarly. This observation allows for an extension of the conclusions of Theorem 5.7 far away the Gaussian framework.

5.4 Fisher information growth and gamma approximation

This final section develops the analogous investigation towards gamma approximation, for simplicity one-dimensional. Denote by γ_p the gamma distribution (on the positive real line) with parameter $p > 0$, invariant measure of the Laguerre operator

$$\mathcal{L}_p f \,=\, xf'' + (p-x)f'. \tag{5.15}$$

Consider a random variable $F \geq 0$ with law $d\nu_F = h d\gamma_p$ absolutely continuous with respect to γ_p. Assume that ν_F admits a Stein kernel τ_{ν_F} with respect to γ_p, that is, according to (4.3) (taking into account the diffusion coefficient $a(x) = x$ in (5.15)), τ_{ν_F} is a mapping on $\mathbb{R}_+$ verifying

$$\int_{\mathbb{R}_+} (x-p)\,\varphi\, d\nu_F \,=\, \int_{\mathbb{R}_+} \tau_{\nu_F}\, \varphi'\, d\nu_F$$

for every smooth test function φ. In particular, $\int_E F d\mu = p$. Note that, in this case,

$$\mathrm{S}^2\big(\nu_F \,|\, \gamma_p\big) \,=\, \int_E \bigg(\frac{\tau_{\nu_F}(F)}{F} - 1\bigg)^2 d\mu.$$

From the study of Gaussian chaoses for example, and as already mentioned earlier, it appears that the latter $\mathrm{S}(\nu_F \,|\, \gamma_p)$ might not always be the relevant quantity of interest (cf. [N-P1, R]). Indeed, for an eigenfunction F with eigenvalue $-\lambda$, $\lambda > 0$, the Stein kernel $\tau_\nu(F)$ may be identified with the conditional expectation of $\lambda^{-1}\Gamma(F)$ knowing F. Now, for such a functional, moment conditions on F may be used to rather control the variance of $\lambda^{-1}\Gamma(F) - F$, and similarly higher moments (cf. [A-C-P, A-M-P, L3]). Of course, by Hölder's inequality,

$$\bigg(\int_E \bigg(\frac{\Gamma(F)}{\lambda F} - 1\bigg)^2 d\mu\bigg)^{1/2} \,\leq\, \bigg(\int_E F^{-2r} d\mu\bigg)^{1/r} \bigg(\int_F \bigg|\frac{\Gamma(F)}{\lambda} - F\bigg|^{2s} d\mu\bigg)^{1/s}$$

for $r > 1$, $\frac{1}{r} + \frac{1}{s} = 1$. Provided it may be ensured that $\int_E F^{-2r} d\mu < \infty$ for some $r > 1$, the results here are nevertheless still of interest.

We assume below that $p \geq \frac{1}{2}$ so that the estimates (4.6) and (4.8) are verified, with the choice of parameters $d = 1$ and $\rho = \kappa = \sigma = \frac{1}{2}$ (see the comment preceding Proposition 4.3). The proof of the following statement will follow the one developed for Theorem 5.7.

Theorem 5.9 (Gamma entropic approximation via Stein discrepancy)**.** *On* (E, μ, Γ)*, let* $F \geq 0$ *in* $\mathcal{A}$*. Assume that* F*,* $\mathrm{L}F$*,* $\Gamma(F)$ *and* $\Gamma(F, \Gamma(F))$ *are in* $\mathrm{L}^q(\mu)$ *for every* $q \geq 1$ *and that*

$$B_F \,=\, \int_E \frac{1}{\Gamma(F)^\alpha}\, d\mu < \infty$$

for some $\alpha > 0$*. Then*

$$A_F \,=\, \int_E \frac{1}{F} \Big[F|\mathrm{L}F| + \Gamma(F) + F \big| \Gamma\big(F, \Gamma(F)\big) \big| + p + F \Big]^2 d\mu < \infty$$

and

$$\mathrm{H}\big(\nu_F \,|\, \gamma_p\big) \,\leq\, \frac{\mathrm{S}^2(\nu_F \,|\, \gamma_p)}{2(1 - 4\kappa)}\, \Psi\bigg(\frac{2(A_F + B_F + 1)}{\mathrm{S}^2(\nu_F \,|\, \gamma_p)} \bigg)$$

where $\kappa = \frac{2+\alpha}{2(4+3\alpha)}$ $(< \frac{1}{4})$.

Proof. Denoting by $(P_t)_{t \geq 0}$ the semigroup with infinitesimal generator $\mathcal{L}_p$, we have as in (4.7),

$$\mathrm{I}_{\gamma_p}(P_t h) \,=\, -\int_{\mathbb{R}_+} \mathcal{L}_p P_t v_t \, h d\gamma_p \,=\, -\int_E \mathcal{L}_p P_t v_t(F) d\mu$$

where $v_t = \log P_t h$. Now, for every $\varepsilon > 0$,

$$\begin{aligned}
\int_E \mathcal{L}_p P_t v_t(F) d\mu \,&=\, \int_E F (P_t v_t)''(F) d\mu + \int_E (p - F)(P_t v_t)'(F) d\mu \\
&=\, \int_E F (P_t v_t)''(F)\, \frac{\Gamma(F)}{\Gamma(F) + \varepsilon}\, d\mu + \int_E F (P_t v_t)''(F)\, \frac{\varepsilon}{\Gamma(F) + \varepsilon}\, d\mu \\
&\quad + \int_E (p - F)(P_t v_t)'(F) d\gamma_p.
\end{aligned}$$

By integration by parts,

$$\int_E F (P_t v_t)''(F)\, \frac{\Gamma(F)}{\Gamma(F) + \varepsilon}\, d\mu \,=\, \int_E (P_t v_t)'(F) \bigg[\frac{F(-\mathrm{L}F)}{\Gamma(F) + \varepsilon} - \Gamma\bigg(F, \frac{F}{\Gamma(F) + \varepsilon} \bigg) \bigg] d\mu.$$

Using that

$$\Gamma\bigg(F, \frac{F}{\varepsilon + \Gamma(F)} \bigg) \,=\, \frac{\Gamma(F)}{\Gamma(F) + \varepsilon} - \frac{F\, \Gamma(F, \Gamma(F))}{(\Gamma(F) + \varepsilon)^2}$$

it follows that

$$\int_E \mathcal{L}_p P_t v_t(F) d\mu \,=\, \int_E \sqrt{F} (P_t v_t)'(F) W_\varepsilon(F) d\mu + \int_E F (P_t v_t)''(F)\, \frac{\varepsilon}{\Gamma(F) + \varepsilon}\, d\mu$$

with

$$W_\varepsilon(F) \,=\, \frac{\sqrt{F}(-\mathrm{L}F)}{\Gamma(F) + \varepsilon} - \frac{\Gamma(F)}{\sqrt{F}(\Gamma(F) + \varepsilon)} + \sqrt{F}\, \frac{\Gamma(F, \Gamma(F))}{(\Gamma(F) + \varepsilon)^2} + \frac{p}{\sqrt{F}} - \sqrt{F}.$$

Now, for every $0 < \varepsilon \leq 1$,

$$\big|W_\varepsilon(F)\big| \,\leq\, \frac{1}{\varepsilon^2\sqrt{F}}\Big[F|\mathrm{L}F| + \Gamma(F) + F\big|\Gamma\big(F,\Gamma(F)\big)\big| + p + F\Big].$$

As a consequence, with the notation introduced in the statement,

$$\int_E W_\varepsilon^2(F)d\mu \,\leq\, \frac{A_F}{\varepsilon^4}\,.$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}\bigg|\int_E \sqrt{F}(P_tv_t)'(F)W_\varepsilon(F)d\gamma_p\bigg| &\leq \bigg(\int_E W_\varepsilon^2(F)d\mu\bigg)^{1/2}\bigg(\int_E F(P_tv_t)'^2(F)d\mu\bigg)^{1/2}\\ &= \bigg(\int_E W_\varepsilon^2(F)d\mu\bigg)^{1/2}\bigg(\int_{\mathbb{R}_+}\Gamma(P_tv_t)hd\gamma_p\bigg)^{1/2}.\end{aligned}$$

Since $\Gamma(P_tv_t) \leq e^{-t}P_t(\Gamma(v_t))$ (Theorem 3.2.4 in [B-G-L]),

$$\begin{aligned}\bigg|\int_E \sqrt{F}(P_tv_t)'(F)W_\varepsilon(F)d\gamma_p\bigg| &\leq e^{-t/2}\bigg(\int_E W_\varepsilon^2(F)d\mu\bigg)^{1/2}\bigg(\int_{\mathbb{R}_+} P_t\big(\Gamma(v_t)\big)hd\gamma_p\bigg)^{1/2}\\ &\leq \frac{e^{-t/2}}{\varepsilon^2}\,A_F^{1/2}\,\mathrm{I}_{\gamma_p}(P_th)^{1/2}.\end{aligned}$$

On the other hand, the estimate (4.6) yields the bound

$$\begin{aligned}\int_E F^2(P_tv_t)''(F)^2d\mu &= \int_{\mathbb{R}_+} x^2(P_tv_t)''^2hd\gamma_p\\ &\leq \frac{1}{e^t-1}\int_{\mathbb{R}_+} P_t\big(\Gamma(v_t)\big)hd\gamma_p\\ &= \frac{1}{e^t-1}\int_{\mathbb{R}_+}\Gamma(v_t)P_thd\gamma_p \,=\, \frac{1}{e^t-1}\,\mathrm{I}_{\gamma_p}(P_th).\end{aligned}$$

This in turn implies that

$$\bigg|\int_E F(P_tv_t)''(F)\,\frac{\varepsilon}{\Gamma(F)+\varepsilon}\,d\mu\bigg| \,\leq\, \frac{1}{\sqrt{e^t-1}}\,\mathrm{I}_{\gamma_p}(P_th)^{1/2}\bigg(\int_E\bigg(\frac{\varepsilon}{\Gamma(F)+\varepsilon}\bigg)^2d\mu\bigg)^{1/2}.$$

Gathering together all the previous estimates, we deduce that, for every $0 < \varepsilon \leq 1$ and $t > 0$,

$$\mathrm{I}_{\gamma_p}(P_th) \,\leq\, \frac{2e^{-t}A_F}{\varepsilon^4} + \frac{2}{e^t-1}\int_E\bigg(\frac{\varepsilon}{\Gamma(F)+\varepsilon}\bigg)^2d\mu.$$

On the basis of this estimate, we then conclude exactly as in the proof of Theorem 5.7. □

Acknowledgment: We are grateful to an anonymous referee for a thorough reading of the manuscript and a number of helpful suggestions.

References

[A-B-B-N] S. Artstein, K. Ball, F. Barthe, A. Naor (2004). On the rate of convergence in the entropic central limit theorem *Probab. Theor. Rel. Fields* **129**, 381–390.

[A-C-P] E. Azmoodeh, S. Campese, G. Poly (2013). Fourth moment theorems for Markov diffusion generators. *J. Func. Anal.*, to appear.

[A-M-P] E. Azmoodeh, D. Malicet, G. Poly (2013). Generalization of the Nualart-Peccati criterion. *Preprint.*

[B-E] D. Bakry, M. Émery (1985). Diffusions hypercontractives. Séminaire de Probabilités XIX, Lecture Notes in Math. **1123**, 177–206. Springer.

[B-G-L] D. Bakry, I. Gentil, M. Ledoux (2014). *Analysis and geometry of Markov diffusion operators.* Grundlehren der Mathematischen Wissenschaften **348**. Springer.

[Ba] A. D. Barbour (1990). Stein's method for diffusion approximations. *Probab. Theory Rel. Fields* **84**(3), 297–322.

[B-J] A. Barron, O. Johnson (2004). Fisher information inequalities and the central limit theorem. *Probab. Theor. Rel. Fields* **129**, 391–409.

[Bi] J.-M. Bismut (1984). *Large deviations and the Malliavin calculus.* Progress in Mathematics **45**. Birkhäuser.

[Bl] N. Blachman (1965). The convolution inequality for entropy powers. *IEEE Trans. Information Theory* **11**, 267–271.

[B-C-G1] S. Bobkov, G. Chistyakov, F. Götze (2013). Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.* **4**, 2479–2512.

[B-C-G2] S. Bobkov, G. Chistyakov, F. Götze (2014). Fisher information and the central limit theorem. *Probab. Theor. Rel. Fields* **159**, 1–59.

[B-C-G3] S. Bobkov, G. Chistyakov, F. Götze (2014) Berry-Esseen bounds in the entropic central limit theorem. *Probab. Theor. Rel. Fields* **159**, 435–478.

[B-L-M] S. Boucheron, G. Lugosi, P. Massart (2013). *Concentration inequalities: A nonasymptotic theory of independence.* Oxford University Press.

[C-P-U] T. Cacoullos, V. Papathanasiou, S. Utev (1994). Variational inequalities with examples and an application to the central limit theorem. *Ann. Prob.* **22**, 1607–1618.

[C-W] A. Carbery, J. Wright (2001). Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Math. Res. Letters* **8**, 233–248.

[C1] S. Chatterjee (2007). Stein's method for concentration inequalities. *Probab. Theory Rel. Fields* **138**, 305–321.

[C2] S. Chatterjee (2009). Fluctuation of eigenvalues and second order Poincaré inequalities. *Probab. Theory Rel. Fields* **143**, 1–40.

[C3] S. Chatterjee (2012). A new approach to strong embeddings. *Probab. Theory Rel. Fields* **152**, 231–264.

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

[D-Z] P. Diaconis, S. Zabell (1991). Closed form summation for classical distributions: variations on a theme of de Moivre. *Stat. Science* **6**, 284–302

[G-S1] L. Goldstein, G. Reinert (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7**, 935–952.

[G-S2] L. Goldstein, G. Reinert (2005). 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.

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

[J] O. Johnson (2004). *Information theory and the central limit theorem.* Imperial College Press.

[L1] M. Ledoux (1995). L'algèbre de Lie des gradients itérés d'un générateur markovien—développements de moyennes et entropies. *Ann. Sci. École Norm. Sup.* **28**, 435–460.

[L2] M. Ledoux (2001). *The concentration of measure phenomenon.* Mathematical Surveys and Monographs **89**, American Mathematical Society.

[L3] M. Ledoux (2012). Chaos of a Markov operator and the fourth moment condition. *Ann. Probab.* **40**, 2439–2459.

[M-J-C-F-T] L. Mackey, M. I. Jordan, R. Y. Chen, B. Farrell, J. A. Tropp (2014). Matrix concentration inequalities via the method of exchangeable pairs. *Ann. Probab.* **42**, 906–945.

[M] E. Milman (2009). On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177**, 1–43.

[N-P1] I. Nourdin, G. Peccati (2009). Non-central convergence of multiple integrals. *Ann. Probab.* **37**, 1412–1426.

[N-P2] I. Nourdin, G. Peccati (2012). *Normal approximations with Malliavin calculus: from Stein's method to universality.* Cambridge Tracts in Mathematics. Cambridge University Press.

[N-P-S1] I. Nourdin, G. Peccati, Y. Swan (2013). Entropy and the fourth moment phenomenon. *J. Funct. Anal.* **266**, 3170–3207.

[N-P-S2] I. Nourdin, G. Peccati, Y. Swan (2013). Integration by parts and representation of information functionals. Preprint.

[N-R] I. Nourdin, J. Rosiński (2014). Asymptotic independence of multiple Wiener-Itô integrals and the resulting limit laws. *Ann. Probab.* **42**, 497–526.

[N] D. Nualart (2006). *The Malliavin calculus and related topics.* Second Edition. Probability and its Applications. Springer.

[N-OL] D. Nualart, S. Ortiz-Latorre (2008). Central limit theorems for multiple stochastic integrals and Malliavin calculus. *Stoch. Proc. Appl.* **118**, 614–628.

[O-V] F. Otto, C. Villani (2000). Generalization of an inequality by Talagrand, and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173**, 361–400.

[R] G. Reinert (2005). Three general approaches to Stein's method. An introduction to Stein's method, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. 4, 183–221. Singapore Univ. Press.

[Sta] A. Stam (1959). Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control* **2**, 101–112.

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

[T] M. Talagrand (1996). Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **6**, 587–600.

[V] C. Villani (2009). *Optimal transport. Old and new.* Grundlehren der Mathematischen Wissenschaften **338**. Springer.