HAL open science

Higher-order Stein kernels for Gaussian approximation

Max Fathi

To cite this version:

Max Fathi. Higher-order Stein kernels for Gaussian approximation. Studia Mathematica, In press. hal-02350138

HAL Id: hal-02350138

https://hal.science/hal-02350138v1

Submitted on 5 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Higher-order Stein kernels for Gaussian approximation

Max Fathi
CNRS & Institut de Mathématiques de Toulouse
Université de Toulouse
118 route de Narbonne
31068 Toulouse, France
E-mail: max.fathi@math.univ-toulouse.fr

Abstract

We introduce higher-order Stein kernels relative to the standard Gaussian measure, which generalize the usual Stein kernels by involving higher-order derivatives of test functions. We relate the associated discrepancies to various metrics on the space of probability measures and prove new functional inequalities involving them. As an application, we obtain new explicit improved rates of convergence in the classical multidimensional CLT under higher moment and regularity assumptions.

1 Introduction

Stein's method is a set of techniques, originating in works of Stein [37, 38], to bound distances between probability measures. We refer to [14, 35] for a recent overview of the field. The purpose of this work is a generalization of one particular way of implementing Stein's method when the target measure is Gaussian, which is known as the Stein kernel approach.

Let μ be a probability measure on $\mathbb{R}^d$. A matrix-valued function $\tau_\mu : \mathbb{R}^d \longrightarrow \mathcal{M}_d(\mathbb{R})$ is said to be a *Stein kernel* for μ (with respect to the

2010 *Mathematics Subject Classification*: Primary 60F05; Secondary 60E15, 26D10, 94A17.

Key words and phrases: Stein's method, Central limit theorem, Functional inequalities

standard Gaussian measure γ on $\mathbb{R}^d$) if for any smooth test function φ taking values in $\mathbb{R}^d$, we have

$$\int x \cdot \varphi d\mu = \int \langle \tau_\mu, \nabla\varphi \rangle_{\mathrm{HS}} d\mu. \tag{1.1}$$

For applications, it is generally enough to consider the restricted class of test functions φ satisfying $\int (|\varphi|^2 + \|\nabla\varphi\|_{\mathrm{HS}}^2) d\mu < \infty$, in which case both integrals in (1.1) are well-defined as soon as $\tau_\mu \in L^2(\mu)$, provided μ has finite second moments.

The motivation behind the definition is that, since the standard centered Gaussian measure γ is the only probability distribution on $\mathbb{R}^d$ satisfying the integration by parts formula

$$\int x \cdot \varphi d\gamma = \int \mathrm{div}(\varphi) d\gamma, \tag{1.2}$$

the Stein kernel τ_μ coincides with the identity matrix, denoted by Id, if and only if the measure μ is equal to γ. Hence, a Stein kernel can be used to control how far μ is from being a standard Gaussian measure in terms of how much it violates the integration by parts formula (1.2). This notion appears implicitly in many works on Stein's method [39], and has recently been the topic of more direct investigations [3, 13, 33, 27, 17].

However, (1.2) is not the only integration by parts formula that characterizes the Gaussian measure. For example, in dimension one, the standard Gaussian measure is characterized by the relation

$$\int H_k(x) f(x) d\gamma(x) = \int H_{k-1}(x) f'(x) d\gamma(x)$$

for all smooth test functions f, where the H_k are the Hermite polynomials $H_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}$. The case $k = 1$ corresponds to the standard formula (1.2). While these are not the only integration by parts formulas one could state, they are in some sense the most natural ones, due to the role Hermite polynomials play as eigenfunctions of the Ornstein-Uhlenbeck generator. Before defining higher-order Stein kernels, we must define a few notations. $\mathcal{T}_{k,d}$ shall denote the space of k-tensors on $\mathbb{R}^d$, that is k-dimensional arrays of size d, and $\mathcal{T}_{k,d}^{sym}$ the subspace of symmetric tensors, that is arrays A such that for any permutation $\sigma \in \mathcal{S}_k$ and $i_1, ..., i_k \in \{1, ..., d\}$ we have $A_{i_1 \dots i_k} = A_{i_{\sigma(1)} \dots \sigma(k)}$. In particular, differentials of order k of smooth functions belong to $\mathcal{T}_{k,d}^{sym}$, $\mathcal{T}_{1,d} = \mathbb{R}^d$ and $\mathcal{T}_{2,d} = \mathcal{M}_d(\mathbb{R})$. We equip these spaces with their natural Euclidean structure and L^2 norm, which we shall respectively denote by $\langle \cdot, \cdot \rangle$ and $|| \cdot ||_2$.

In dimension $d \geq 1$, Hermite polynomials are defined as follows:

Definition 1.1 (Multi-dimensional Hermite polynomials)**.** For $k \geq 1$ and indices $i_1, .., i_d \in \mathbb{N}$ such that $i_1 + .. + i_d = k$, we define the Hermite polynomial $H_k^{i_1,..,i_d} \in \mathbb{R}[X_1, .., X_d]$ as

$$H_k^{i_1,..,i_d}(x_1, .., x_d) := (-1)^k e^{|x|^2/2} \frac{d^k}{dx_1^{i_1}..dx_d^{i_d}} e^{-|x|^2/2}.$$

We also define $\bar{H}_k := (-1)^k e^{|x|^2/2} d^k(e^{-|x|^2/2})$, so that the coefficients of the k-tensor $\bar{H}_k$ are the $H_k^{i_1,..,i_d}$.

The natural generalization of the notion of Stein kernels with respect to the integration by parts formulas defined via multidimensional Hermite polynomials would be to say that a $(k+1)$-tensor τ_k is a k-th order Stein kernel for μ is for any smooth $f : \mathbb{R}^d \longrightarrow \mathcal{T}_{k,d}$ we have

$$\int \langle \bar{H}_k, f \rangle d\mu = \int \langle \tau_k, Df \rangle d\mu.$$

In this formula, Df stands for the differential of f, viewed as a $(k+1)$-tensor. However, it turns out that for the applications we shall describe below, it is more convenient to define higher-order Stein kernels in a different way:

Definition 1.2 (Higher-order Stein kernels)**.** We define Stein kernels of order k $\bar{\tau}_k$ (as long as they exist) as any symmetric $(k+1)$-tensor satisfying

$$\int \langle \bar{\tau}_k, D^k f \rangle d\mu = \int x \cdot f - \mathrm{Tr}(\nabla f) d\mu \tag{1.3}$$

for all smooth vector-valued f such that $x \cdot f - \mathrm{Tr}(\nabla f)$ and $||D^k f||^2$ are integrable with respect to μ.

Note that $\bar{\tau}_1 = \tau - \mathrm{Id}$ where τ is a classical Stein kernel. The choice of restricting the definition to symmetric tensors is non-standard when $k = 1$. It is motivated by the fact that since we only test out the relation on tensors of the form $D^{k+2} f$, which are symmetric, and will allow us to easily relate the expectation of such kernels to moments of the underlying measure.

In the sequel, we shall consider the weighted Sobolev space with respect to μ

$$H^k(\mu) := \left\{ f : \mathbb{R}^d \longrightarrow \mathbb{R}; \int |D^k f|^2 d\mu < \infty \right\}.$$

In the sequel, we shall always make the following structural assumptions on μ:

Assumption. We assume the following:

1. The sequence of spaces $(H^j(\mu))_{j\leq k}$ is decreasing.
2. The Dirichlet form $\mathcal{E} : f \longrightarrow \int |\nabla f|^2 d\mu$ is closable on the set of smooth, compactly supported test functions, that is for any sequence (f_n) of such functions, if $||f_n||_{L^2(\mu} \longrightarrow 0$ and $\mathcal{E}(f_n - f_\ell) \longrightarrow 0$ as n, ℓ go to infinity, then $\mathcal{E}(f_n) \longrightarrow 0$.
3. The space of smooth, compactly supported functions is dense in $H^k(\mu)$, with respect to the norm $\left(\sum_{j\leq k} ||D^j f||^2_{L^2(\mu)}\right)^{1/2}$.

When these assumptions are satisfied, $H^k(\mu)$ endowed with the norm defined above is a Hilbert space. We shall mainly consider measures that satisfy a Poincaré inequality, which implies the first assumption. If μ has a smooth positive density with respect to the Lebesgue measure on $\mathbb{R}^d$, the last two assumptions are satisfied. However, for more general measures it may be an issue to handle the influence of the boundary of the support, or of some singular part [25]. The second assumption is a classical assumption in the study of Dirichlet forms [18]. The third one could be avoided, but then the integration by parts formula for Stein kernels would only hold for the completion of the set of smooth test functions with respect to the Sobolev norm. We refer to [4, Section 3] for background on the role of such assumptions in the theory of Markov semigroups and functional inequalities.

In some sense, the point of view we develop here is very close to the one developed in [22], where approximate Stein identities with higher-order derivatives are used, in the framework of the zero-bias transform. See also [23] for connections between distributional transforms and orthogonal polynomials applied with Stein's method in a more general context. The main advantage of the functional-analytic framework presented here is to allow more explicit estimates in the multivariate setting, albeit under strong regularity conditions. A particular upside of our estimates is that the dependence on the dimension will be very explicit.

A first remark is that we have the iterative relation

$$\int \langle \bar{\tau}_k, D^k f\rangle d\mu = \int \langle \bar{\tau}_{k-1}, D^{k-1} f\rangle d\mu \tag{1.4}$$

As we shall later see in Lemma 2.1, for $\bar{\tau}_k$ to exist, we must have $\int P(x_1, .., x_d)d\mu = \int P(x_1, .., x_d)d\gamma$ for any $P \in \mathbb{R}_{k+1}[X_1, .., X_d]$. Of course, this is not a sufficient condition. These kernels are in some sense centered, so that μ is Gaussian iff $\bar{\tau}_k = 0$. For $k = 1$, this does not exactly match with the usual definition, which is not centered, but this shift will make notations much lighter.

These Stein kernels can be related to kernels associated with Hermite polynomials via linear combinations. For example, if τ_2 (resp. τ_1) is a kernel associated with Hermite polynomials of degree 2 (resp. 1), then $\bar{\tau}_2 = \tau_2 - x \otimes \tau_1$ is a second-order Stein kernel in the sense of (1.3).

As for classical Stein kernels, we can then define the associated discrepancy, which measures how far a given probability measure is from satisfying the associated Gaussian integration by parts formula.

Definition 1.3. The k-th order Stein discrepancy is defined by

$$\bar{S}_k(\mu) := \inf \int ||\bar{\tau}_k||^2 d\mu,$$

where the infimum is over all possible Stein kernels of order k for μ, since they may not be unique.

Remark 1.4. The abstract setting we use here is not restricted to Hermite polynomials or higher-order derivatives. For example, it would be possible to define a kernel by considering *any* tensor-valued function $u : \mathbb{R}^d \longrightarrow \mathcal{T}_{k,d}$ and looking for a function $K_\mu(u) : \mathbb{R}^d \longrightarrow \mathcal{T}_{k+1,d}$ such that for any smooth function $f : \mathbb{R}^d \longrightarrow \mathcal{T}_{k,d}$ we would have

$$\int \langle u, f \rangle d\mu = \int \langle K_\mu(u), \nabla f \rangle d\mu.$$

This more general point of view is related to the one developed in [30]. Existence would be treated in the same way as we shall implement in this work, but we do not have any other example leading to meaningful applications at this point.

The main application of these higher-order Stein kernels to the rate of convergence in the classical CLT is the following decay estimate, made precise in Corollary 4.4: if the random variables (X_i) are iid, isotropic, centered and have mixed moments of order three equal to zero, then if μ_n is the law of the renormalized sum in the CLT we have an estimate of the form

$$W_2(\mu_n, \gamma) \leq \frac{Cd^{1/2}(1 + \log n)}{n}$$

where C is a constant we shall make precise, that depends on a regularity condition on the law of the X_i. This seems to be the first improved rate of convergence in the multidimensional CLT in W_2 distance.

The plan of the sequel is as follows: in Section 2, we shall establish basic properties of higher-order Stein kernels, including existence and some

first results on what distances the associated discrepancies control. In Section 3, we shall establish some functional inequalities relating Wasserstein distances, entropy and Fisher information. Finally, in Section 4, we shall derive various improved bounds on the rate of convergence in the central limit theorem under moment constraints.

2 Properties

2.1 Existence

Before studying these higher-order Stein kernels and their applications, the first question to ask is when do they actually exist? As for classical Stein kernels, there must be some condition beyond normalizing the moments, since they may not exist for measures with purely atomic support.

The first condition we can point out is that existence of Stein kernels constrains the values of certain moments:

Lemma 2.1. *Assume that μ admits Stein kernels $\bar{\tau}_j$ up to order $k \geq 1$, and has finite moments up to order $k+1$. Then for any polynomial P in d variables of degree $\ell \leq k$ we have $\int P(x)d\mu = \int P(x)d\gamma$, and moreover if this is also true for polynomials of degree $k+1$ then*

$$\int (\bar{\tau}_k)_{i_1,\dots i_{k+1}} d\mu = 0$$

for any indices $i_1, ..., i_k$.

Proof. We prove this statement by induction on k. The case $k = 1$ can be readily checked by testing the Stein identity on coordinates $x_1, .., x_d$. Assume the statement holds for $k \geq 1$. To prove the statement for $k+1$, it is enough to check it for monomials of degree k, by the induction assumption. Up to relabeling, we can restrict to the case where the degree in x_1 is positive. Let $\alpha_1, ..., \alpha_d$ be such that $\sum \alpha_i = k$ and $P(x) = x_1^{\alpha_1} x_2^{\alpha_2} ... x_d^{\alpha_d}$. Define $E = \{(i_1, ..., i_k); \forall \ell \, |\{j; i_j = \ell\} = \alpha_\ell\}$. We have

$$\begin{aligned}
\int x_1^{\alpha_1+1} ... x_n^{\alpha_n} d\mu &= \int \langle \bar{\tau}_{k,1}, D^k P \rangle d\mu + \sum_i \int (\alpha_i x_i^{\alpha_i - 1}) \prod_{j \neq i} x_j^{\alpha_j} d\mu \\
&= \prod_{j=1}^{d} (\alpha_j!) \sum_{(i_1,\dots,i_k) \in E} \int (\bar{\tau}_k)_{1,i_1,\dots,i_k} d\mu + \sum_i \int (\alpha_i x_i^{\alpha_i - 1}) \prod_{j \neq i} x_j^{\alpha_j} d\gamma \\
&= \left(\prod_{j=1}^{d} (\alpha_j!) \right) k! \int \bar{\tau}_{1,1\dots1,2\dots d} d\mu + \sum_i \int (\alpha_i x_i^{\alpha_i - 1}) \prod_{j \neq i} x_j^{\alpha_j} d\gamma
\end{aligned}$$

where we have used the symmetry of $\bar{\tau}_k$, and the moment assumption to match the second term. The indices in the last line correspond to having α_i times the indice i, and the order does not matter by symmetry of $\bar{\tau}_k$. Since for a Gaussian measure the two integrals of moments match, the integral of the kernel must be zero as soon as the moment assumption is satisfied. □

In dimension one, when μ has a nice density p with respect to the Lebesgue measure, we can give explicit formulas in terms of p:

Proposition 2.2. *Let $\mu(dx) = p(x)dx$ be a probability measure on $\mathbb{R}$ with connected support, such that $\int x^j d\mu = \int x^j d\gamma$ for all $j \leq k$. Then the iterative formula*

$$\bar{\tau}_k(x) = -\frac{1}{p(x)} \int_x^\infty \bar{\tau}_{k-1}(y)p(y)dy$$

defines Stein kernels, with $\bar{\tau}_1 = -\frac{1}{p(x)} \int_x^\infty yp(y)dy - 1$ the usual explicit formula for classical Stein kernels in dimension one.

We refer to [36] for a detailed study of 1st order kernels in dimension one. We shall not develop this point of view further, and focus on the situation in higher dimension, where this formula is no longer available. It turns out that, up to extra moment conditions, the arguments used in [16] for standard Stein kernels also apply. Before stating the conditions, we must first define Poincaré inequalities:

Definition 2.3. A probability measure μ on $\mathbb{R}^d$ satisfies a Poincaré inequality with constant C_P if for all locally lipschitz function f with $\int f d\mu = 0$ we have

$$\int f^2 d\mu \leq C_P \int |\nabla f|^2 d\mu.$$

Poincaré inequalities are a standard family of inequalities in stochastic analysis, with many applications, such as concentration inequalities and rates of convergence to equilibrium for stochastic processes. They can also be reinterpreted as a lower bound on the eigenvalues of a certain Markov generator with invariant measure μ. They are satisfied by a wide class of measures, including log-concave, and are stable by tensorisation (which makes them a well-adapted tool for high-dimensional problems) and bounded perturbations. In dimension one, measures that satisfy a Poincaré inequality are characterized by the Muckenhoupt criterion [31]. In higher dimension, we do not have such a characterization, but many sufficient conditions are known. For example, [5] proves that if the measure has a smooth density of the form e^{-V} and that $\epsilon|\nabla V|^2 - \Delta V$ is bounded from below by a positive constant

at infinity, for some $\epsilon > 0$, then a Poincaré inequality holds. See [4], the lecture notes [26] and references therein for more background information and other conditions ensuring such an inequality holds. Additionally, if a distribution is compactly supported, then its convolution with a Gaussian distribution with arbitrarily small variance satisfies a Poincaré inequality (with a constant that depends on the variance and the size of the support) [41, 8].

Connections between such functional inequalities and Stein's method go back to [39], and were more recently explored in [27, 16]. Our basic existence result for higher-order kernels is the following:

Theorem 2.4. *Assume that μ satisfies a Poincaré inequality with constant C_P, and that its moments of order less than or equal to $\max(k, 2)$ match with those of the standard Gaussian. Then a Stein kernel of order k exists, and moreover $\bar{S}_k(\mu)^2 \leq C_P^{k-1}(C_P - 1)d$.*

This theorem yields a sufficient condition for existence, but it is not necessary. Even in the case $k = 1$, we do not know of a useful full characterization of the situations where Stein kernels exist. Actually, [16] uses a more general type of functional inequality to ensure existence of a 1st order Stein kernel, but its extension to higher order kernels is a bit cumbersome, since the condition would iteratively require previous kernels to have a finite 2nd moment after multiplication with an extra weight.

Proof. We proceed by induction. The case $k = 1$ was proven in [16]. Assume that the statement is true for some k, and that μ has moments of order less than $k+1$ matching with those of the Gaussian. Let $\bar{\tau}_k$ be a Stein kernel of order k for μ, which exists by the induction assumption. We wish to prove existence of $\bar{\tau}_{k+1}$. Consider the functional

$$J(f) = \frac{1}{2}\int |D^{k+1}f|^2 d\mu - \int \langle \bar{\tau}_k, D^k f\rangle d\mu$$

defined for $f : \mathbb{R}^d \longrightarrow \mathbb{R}^d$. It is easy to check that, from the Euler-Lagrange equation for J, if g is a minimizer of J, then $\bar{\tau}_{k+1} := D^{k+1}g$ satisfies (1.3). Indeed, if we consider the first variation of J, we have

$$J(g + \epsilon h) - J(g) = \epsilon \left(\int \langle D^{k+1}g, F^{k+1}h\rangle d\mu - \int \langle \bar{\tau}_k, D^k h\rangle d\mu \right) + O(\epsilon^2)$$

so that g is a critical point of J iff $D^{k+1}g$ can be taken as a Stein kernel in (1.3).

Note that if two functions f and g are such that $D^{k+1}f = D^{k+1}g$ (or equivalently, that $D^k(f-g)$ is a constant), then since $\bar{\tau}_k$ is centered, $J(f) = J(g)$. Therefore we can view J as being defined on the quotient of $H^k(\mu)$ by $\{(f,g); D^{k+1}f = D^{k+1}g\}$. Because of the Poincaré inequality and our assumptions on μ, this quotient space is a Hilbert space when endowed with the norm $\left(\int |D^{k+1}f|^2 d\mu\right)^{1/2}$.

From the Poincaré inequality and the fact that $\bar{\tau}_k$ is centered due to the moment assumption, we have

$$\begin{aligned}\left|\int \langle \bar{\tau}_k, D^k f\rangle d\mu\right|^2 &= \left|\int \langle \bar{\tau}_k, \left(D^k f - \int D^k f d\mu\right)\rangle d\mu\right|^2 \\ &\leq \left(\int |\bar{\tau}_k|^2 d\mu\right)\left(\int \left|D^k f - \int D^k f d\mu\right|^2 d\mu\right) \\ &\leq C_P \left(\int |\bar{\tau}_k|^2 d\mu\right)\left(\int |D^{k+1}f|^2 d\mu\right).\end{aligned}$$

Therefore $f \longrightarrow \int \langle \bar{\tau}_k, D^{k+1}f\rangle d\mu$ is a continuous linear form w.r.t. the norm $\int |D^{k+1}f|^2 d\mu$ on the quotient space we defined above. Hence from the Lax-Milgram theorem (or Riesz representation theorem) we deduce existence (and uniqueness) of a centered global minimizer g, and $\bar{\tau}_{k+1} = D^{k+1}g$ is a suitable Stein kernel, and satisfies the symmetry assumption. Moreover,

$$\begin{aligned}-\frac{1}{2}\int |\bar{\tau}_{k+1}|^2 d\mu &= \frac{1}{2}\int |D^{k+1}g|^2 d\mu - \int \langle \bar{\tau}_k, g\rangle d\mu \\ &\geq \frac{1}{2}\int |\nabla g|^2 d\mu - \frac{C_P}{2}\bar{S}_k(\mu)^2 - \frac{1}{2C_P}\int \left|g - \int g d\mu\right|^2 d\mu \\ &\geq -\frac{C_P}{2}\bar{S}_k(\mu).\end{aligned}$$

The induction assumption then yields $\bar{S}_{k+1}(\mu)^2 \leq C_P^k (C_P - 1)d$. □

2.2 Topology

In this section, we are interested in studying what distances between a probability measure and a Gaussian are controlled by our discrepancies. As is classical in Stein's method, we seek to control a distance of the form

$$d(\mu,\nu) = \sup_{f\in\mathcal{F}} \int f d\mu - \int f d\nu$$

where the class of test functions $\mathcal{F}$ should be symmetric, and large enough to indeed separate probability measures. The total variation distance corresponds to the set of functions bounded by one, while the L^1 Kantorovitch-Wasserstein distance is obtained when considering the set of 1-lipschitz functions, thanks to the Kantorovitch-Rubinstein duality formula [40].

To relate such distances to Stein's method, we introduce the Poisson equation

$$\Delta h - x \cdot \nabla h = f - \int f d\gamma. \tag{2.1}$$

The classical implementation is that if the solution satisfies a suitable regularity bound, then we can control

$$\int f d\mu - \int f d\gamma = \int (\Delta h - x \cdot \nabla h) d\mu$$

by a type of Stein discrepancy. Due to the elliptic nature of the Ornstein-Uhlenbeck generator, the solution h gains some regularity compared to f. For example, if f is 1-lipschitz, h is $C^{3-\epsilon}$ [19]. Here, to control it by a Stein discrepancy of order k, we shall have to differentiate several times the solution, and require it to satisfy a bound of the form $||D^{k+1}h||_\infty \leq C$. In particular, solutions to the Poisson equation should be smooth enough, which typically requires f to be C^k (this will be explained in more details in the proof of Theorem 2.5 below).

We can now state a first result on the topology controlled by higher-order Stein discrepancies.

Theorem 2.5. *Let μ be a probability measure on $\mathbb{R}^d$ whose k first mixed moments match with those of a d-dimensional standard centered Gaussian. Then*

$$d_{Zol,k}(\mu, \gamma) := \sup_{||D^k f|| \leq 1} \int f d\mu - \int f d\gamma \leq \bar{S}_k(\mu).$$

The controlled distance $d_{Zol,k}$ can be thought of as a generalization of the L^1 Kantorovitch-Wasserstein distance, which corresponds to $k = 1$. It is known as the Zolotarev distance of order k, and it controls the same topology as the L^k Kantorovitch-Wasserstein distance [9], that is weak convergence and convergence of moments up to order k.

Proof. We first derive a regularity estimate for solutions of the Poisson equation. The scheme of proof below is a straightforward extension of the regularity bound of [15] in the case $k = 1$. Similar regularity bounds, in operator norm, for arbitrary k where derived in [20]. In the case where f is lipschitz, better regularity bounds (namely, $C^{2,1-}$ bounds) were obtained in [19], and it should be possible to get better regularity bounds for general k. However, for our purpose it is not clear that improved bounds would further help us here.

As pointed out by Barbour [7], a solution of the Poisson equation (2.1) is given by

$$h_f(x) = \int_0^1 \frac{1}{2t} \int (f(\sqrt{t}x + \sqrt{1-t}y) - f(y)) d\gamma(y) dt. \tag{2.2}$$

and after integrating by parts with respect to the Gaussian measure, its gradient can be represented as

$$\nabla h_f(x) = \int_0^1 \frac{1}{2\sqrt{t(1-t)}} \int y f(\sqrt{t}x + \sqrt{1-t}y) d\gamma(y) dt$$

and hence higher-order derivatives are given by

$$D^{k+1} h_f(x) = \int_0^1 \frac{t^{k/2}}{2\sqrt{t(1-t)}} \int y \otimes D^k f(\sqrt{t}x + \sqrt{1-t}y) d\gamma(y) dt.$$

We then have for any $A \in \mathcal{T}_{k+1,d}$

$$\begin{aligned}
\langle D^{k+1} h_f(x), A \rangle &= \int_0^1 \frac{t^{k/2}}{2\sqrt{t(1-t)}} \int \langle A, y \otimes D^k f(\sqrt{t}x + \sqrt{1-t}y) \rangle d\gamma(y) dt \\
&= \int_0^1 \frac{t^{k/2}}{2\sqrt{t(1-t)}} \int \langle Ay, D^k f(\sqrt{t}x + \sqrt{1-t}y) \rangle d\gamma(y) dt \\
&\leq \sup_z ||D^k f(z)||_2 \int_0^1 \frac{t^{k/2}}{2\sqrt{t(1-t)}} dt \int ||Ay||_2 d\gamma(y) \\
&\leq \sup_z ||D^k f(z)||_2 \left(\int \sum_{i_1,..,i_k} \left(\sum_j A_{i_1,..,i_k,j} y_j \right)^2 d\gamma(y) \right)^{1/2} \\
&= \sup_z ||D^k f(z)||_2 ||A||_2.
\end{aligned}$$

Therefore

$$\sup_x ||D^{k+1} h_f(x)||_2 \leq \sup_x ||D^k f(x)||_2.$$

We then have, for any function f satisfying $\sup_x ||D^k f(x)||_2 \leq 1$,

$$\begin{aligned}
\int f d\mu - \int f d\gamma &= \int \Delta h_f - x \cdot \nabla h_f d\mu \\
&= \int \langle \bar{\tau}_1, D^2 h_f \rangle d\mu \\
&= \int \langle \bar{\tau}_k, D^{k+1} h_f \rangle d\mu \\
&\leq \bar{S}_k(\mu).
\end{aligned}$$

This concludes the proof. □

Remark 2.6. In dimension one, the Ornstein-Uhlenbeck enjoys strictly better regularization properties, which would allow to control stronger distances.

3 Functional inequalities

Our first functional inequality is a generalization of the HSI inequality of [27]. Given a probability measure $\mu = f d\gamma$ on $\mathbb{R}^d$, we denote by $H(\mu) = \int f \log f d\gamma$ its relative entropy with respect to the Gaussian measure, and $I(\mu) = \int |\nabla \log f|^2 f d\gamma$ its Fisher information.

Theorem 3.1 (HSI inequalities)**.** *Let $k \geq 2$. We have*

$$H(\mu) \leq \frac{1}{2} \min \left(I(\mu), k I(\mu)^{(k-1)/k} \bar{S}_k(\mu)^{2/k} \right).$$

This inequality improves on the classical Gaussian logarithmic Sobolev inequality of Gross [24]. Such inequalities are used to establish concentration of measure inequalities, as well as to measure rates of convergence to equilibrium for stochastic processes [4].

We introduce the Ornstein-Uhlenbeck semigroup

$$P_t f(x) = \mathbb{E}[f(e^{-t}x + \sqrt{1-e^{-2t}}G)]$$

where G is a standard Gaussian random variable. We also denote by μ_t the law of $e^{-t}X + \sqrt{1-e^{-2t}}G$, where X is distributed according to μ, independently of G. The properties of this semigroup have been well-studied. In particular, as time goes to infinity, $P_t f$ converges to $\int f d\gamma$, and the entropy and Fisher information are related by De Brujin's formula:

$$H(f) = \int_0^\infty I(P_t f) dt. \tag{3.1}$$

The key lemma at the core of our results is the following estimate on Fisher information along the flow:

Lemma 3.2. *For any $t > 0$, we have*

$$I(\mu_t) \leq \frac{e^{-2(k+1)t}}{(1-e^{-2t})^k} k! \bar{S}_k(\mu)^2.$$

When $k = 1$, this estimate corresponds to the main result of [33], and played a core role in the proofs of the functional inequalities of [27]. This extension to higher orders will allow us to get more precise estimates when higher-order Stein kernels exist, i.e. under moment constraints.

Proof. We have the commutation relation

$$\partial_i (P_t f) = e^{-t} P_t (\partial_i f). \tag{3.2}$$

Let $v_t := \log(d\mu_t/d\gamma)$. Following [27], we have a representation formula for the Fisher information along the Ornstein-Uhlenbeck flow:

$$\begin{aligned} I(\mu_t) &= \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} \int\int \langle (\tau(x) - \mathrm{Id})y, \nabla v_t(e^{-t}x + \sqrt{1-e^{-2t}}y)\rangle d\mu(x)d\gamma(y) \\ &= \frac{e^{-(k+1)t}}{\sqrt{1-e^{-2t}}} \int\int \langle \bar{\tau}_k(x)y, D^k v_t(e^{-t}x + \sqrt{1-e^{-2t}}y)\rangle d\mu(x)d\gamma(y) \\ &= \frac{e^{-(k+1)t}}{(1-e^{-2t})^{k/2}} \int\int \langle \bar{\tau}_k(x)\bar{H}_k(y), \nabla v_t(e^{-t}x + \sqrt{1-e^{-2t}}y)\rangle d\mu(x)d\gamma(y) \end{aligned}$$

Applying the Cauchy-Schwarz inequality and integrating out in y, we get the result. □

Proof of Theorem 3.1. From (3.1) and the decay property of the Fisher information $I(\mu_t) \le e^{-2t} I(\mu)$, we deduce that for any $t \ge 0$ we have

$$H(\mu) \le \frac{1-e^{-2t}}{2} I(\mu) + \int_t^\infty I(\mu_s) ds.$$

Using Lemma 3.2 on the second term, we get

$$\begin{aligned} H(\mu) &\le \frac{1-e^{-2t}}{2} I(\mu) + k!\bar{S}_k(\mu)^2 \int_t^\infty \frac{e^{-2(k+1)s}}{(1-e^{-2s})^k} ds \\ &\le \frac{1-e^{-2t}}{2} I(\mu) + \frac{k!}{2(k-1)} \bar{S}_k(\mu)^2 \frac{e^{-2kt}}{(1-e^{-2t})^{k-1}} \\ &\le \frac{1-e^{-2t}}{2} I(\mu) + \frac{k!\bar{S}_k(\mu)^2}{2(1-e^{-2t})^{k-1}}. \end{aligned}$$

We optimize by taking t such that $1 - e^{-2t} = \left(\frac{k!\bar{S}_k^2}{I}\right)^{1/k}$ if possible, and otherwise $t = \infty$ (which boils down to the usual logarithmic Sobolev inequality), and we get the result. We used the easy bound $(k!)^{1/k} \le k$ to simplify the expression. □

We can also obtain functional inequalities controlling the W_2 distance, which is defined as

$$W_2(\mu,\nu)^2 := \inf_\pi \int |x-y|^2 d\pi(x,y)$$

with the infimum taken over the set of all couplings π of the two measures μ and ν. Recall that in the case $k = 1$, [27] established the inequality

$$W_2(\mu,\gamma)^2 \le \bar{S}_1(\mu)^2,$$

which itself reinforced classical bounds on the W_1 distance via Stein's method, and allows to get simple proofs of CLTs in W_2 distance, since Stein discrepancies turn out to be easier to estimate in some situations. Our result is the following variant involving higher-order discrepancies:

Theorem 3.3 (L^2 transport inequalities)**.** *For $k = 2$, we have*

$$W_2(\mu, \gamma) \leq \max(\bar{S}_2(\mu)(1 - \log(\bar{S}_2/\bar{S}_1)), \bar{S}_2(\mu)).$$

For $k \geq 3$, we have $W_2(\mu, \gamma) \leq 2k\bar{S}_1(\mu)^{1-1/(k-1)}\bar{S}_k(\mu)^{1/(k-1)}$.

The first inequality will allow to improve the rate of convergence in the CLT in W_2 distance for measures having its moments of order 3 equal to zero. As we will later see, when $k \geq 3$, these inequalities are not satisfactory for applications to CLTs.

Proof. As pointed out in [34], we have

$$W_2(\mu, \gamma) \leq \int_0^\infty I(\mu_s)^{1/2} ds.$$

For $k = 2$, we have for any t

$$\begin{aligned}
W_2(\mu, \gamma) &\leq \int_0^t I(\mu_s)^{1/2} ds + \int_t^\infty I(\mu_s)^{1/2} ds \\
&\leq \bar{S}_1(\mu) \int_0^t \frac{e^{-2s}}{\sqrt{1 - e^{-2s}}} ds + \bar{S}_2(\mu) \int_t^\infty \frac{e^{-3s}}{1 - e^{-2s}} ds \\
&\leq \sqrt{1 - e^{-2t}}\bar{S}_1(\mu) - \frac{1}{2} e^{-t} \log(1 - e^{-2t})\bar{S}_2(\mu) \\
&\leq \sqrt{1 - e^{-2t}}\bar{S}_1(\mu) - \frac{1}{2} \log(1 - e^{-2t})\bar{S}_2(\mu)
\end{aligned}$$

Optimizing in t then leads to choosing t such that $\sqrt{1 - e^{-2t}} = \min(\bar{S}_2/\bar{S}_1, 1)$. If $\bar{S}_2 \leq \bar{S}_1$, we end up with the bound $W_2 \leq \bar{S}_1(\mu)(1 - \log(\bar{S}_2/\bar{S}_1))$, and this upper bound, is larger than $\bar{S}_2$. Otherwise, we bound it by $\bar{S}_2 \geq \bar{S}_1$, and the desired bound holds either way.

For $k \geq 3$, we similarly have

$$\begin{aligned}
W_2(\mu, \gamma) &\leq \int_0^t I(\mu_s)^{1/2} ds + \int_t^\infty I(\mu_s)^{1/2} ds \\
&\leq \bar{S}_1(\mu) \int_0^t \frac{e^{-2s}}{\sqrt{1 - e^{-2s}}} ds + \sqrt{k!}\bar{S}_k(\mu) \int_t^\infty \frac{e^{-(k+1)s}}{(1 - e^{-2s})^{k/2}} ds \\
&\leq \sqrt{1 - e^{-2t}}\bar{S}_1(\mu) + \sqrt{k!}\bar{S}_k(\mu) \frac{e^{-(k-1)t}}{k - 2} \left(\frac{1}{1 - e^{-2t}}\right)^{(k-2)/2} \\
&\leq \sqrt{1 - e^{-2t}}\bar{S}_1(\mu) + \sqrt{k!}\bar{S}_k(\mu) \left(\frac{1}{1 - e^{-2t}}\right)^{(k-2)/2}
\end{aligned}$$

and taking $\sqrt{1-e^{-2t}} = (\sqrt{k!}\bar{S}_k(\mu)/\bar{S}_1(\mu))^{1/(k-1)}$ yields the result. The inequality could be improved, at the cost of clarity, but as far as we can see the sharper inequality obtained by this method does not significantly improve the outcomes in the applications. □

4 Improved rates of convergence in the classical CLT

We are interested in the rate of convergence of the law of (normalized) sums of iid random variables $n^{-1/2}\sum_{i=1}^n X_i$ to their Gaussian limit. It is known that the rate of convergence in Wasserstein distance W_2 is of order $\sqrt{n}$ in general, as soon as the fourth moment is finite [12]. However, it is possible to do a Taylor expansion of the distance as n goes to infinity, and see that under moment constraints, the asymptotic rate of decay may improve. More precisely, [10, 11] shows that in dimension one, if the first $k+1$ moments of the random variables match with those of the standard Gaussian, then the Wasserstein distance (and the stronger relative entropy and Fisher information) asymptotically decays like $n^{-k/2}$. Non-asymptotic rates in dimension one were obtained in [22] using a variant of Stein's method, and strong entropic rates under a Poincaré inequality and after regularization by convolution with a Gaussian measure were obtained in [29], still in dimension one. [2] gives a sharp non-improved rate of convergence in the entropic CLT in dimension one in the classical case (i.e. without the extra moment constraints satisfied), without any regularization. See also [6] for a multi-dimensional extension when the measure is additionally assumed to be log-concave.

It is possible to use Stein's method to give simple proofs of this decay rate [35, 27]. In particular, [16] proves a monotone decay of the Stein discrepancy, which immediately implies the quantitative CLT as soon as the Stein discrepancy of a single variable is finite.

We consider the usual setting for the classical CLT: a sequence (X_i) of iid random variables with distribution μ, and the normalized sum

$$U_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$$

whose law we shall denote by μ_n.

The aim of this section is to show similar results for higher-order discrepancies. The starting point is the following construction of Stein kernels

of the second type for sums of independent random variables, which is an immediate generalization of the same result for $k = 1$.

Lemma 4.1. *Let $\bar{\tau}_{k,1}$ be a k-th order Stein kernel for μ. Then*

$$\bar{\tau}_{k,n}(m) := n^{-(k+1)/2}\mathbb{E}\left[\sum \bar{\tau}_{k,1}(X_i)|n^{-1/2}\sum X_i = m\right]$$

is a k-th order Stein kernel for μ_n.

Proof. This can easily be checked by induction on k via (1.4). The case $k = 1$ is well-known [27]. □

As a consequence, we obtain bounds on the rate of convergence of the Stein discrepancies:

Corollary 4.2. *Assume that all the mixed moments of order up to $k+1$ of X_1 are the same as those of the standard Gaussian. Then*

$$\bar{S}_k(\mu_n)^2 \leq n^{-k}\bar{S}_k(\mu)^2.$$

We then obtain a rate of convergence in the multivariate CLT for the Zolotarev distances $d_{Zol,k}$ as an immediate consequence of the comparison from Theorem 2.5:

Corollary 4.3. *Assume that μ satisfies a Poincaré inequality with constant C_P, and that all of its mixed moments of order up to $k+1$ match with those of the standard Gaussian measure. Let μ_n be the law of U_n. Then*

$$d_{Zol,k}(\mu_n, \gamma) \leq \frac{\sqrt{C_P^{k-1}(C_P - 1)d}}{n^{-k/2}}.$$

Such results have been obtained in dimension one (and for random vectors with independent coordinates) in [20, 21]. See also [22] for related results.

Combined with the logarithmic Sobolev inequality and Lemma 3.2, this also yields a multi-dimensional extension of a result of [29] on improved entropic CLTs for regularized measures, with more explicit quantitative prefactors.

In the case $k = 3$, due to Theorem 3.3, we can upgrade the distance to W_2, losing however a logarithmic factor:

Corollary 4.4. *Assume that all of the mixed moments of order less than three of X_1 are the same as those of the standard Gaussian, and that its law satisfies a Poincaré inequality with constant C_P. Then*

$$W_2(\mu_n, \gamma) \leq \frac{\sqrt{dC_P(C_P-1)}}{n}\left(1 + \frac{1}{2}\log n + \frac{1}{2}\log C_P\right)$$

as soon as $n \geq \sqrt{dC_P(C_P-1)}$. If additionally the mixed fourth moments match with those of the Gaussian, we get

$$W_2(\mu_n, \gamma) \leq \frac{2\sqrt{dC_P(C_P-1)}}{(k-1)n}.$$

Proof. The first inequality is obtained by plugging the upper bounds on discrepancies in the bounds of Theorem 3.3, while using the fact that $x \to x(1-\log x)$ is increasing on $[0,1]$. The second inequality is obtained by using the 2nd order kernels, and with our estimates using even higher order kernels does not improve the bounds. □

When $k = 2$, we only miss the sharp asymptotic rate of [10] by a logarithmic factor. However, under higher moment constraints we know that the asymptotic rate is much better than n^{-1} (at least in dimension one), so this result is not satisfactory.

For the entropy without regularization, we obtain the following rates under the assumption that mixed third moments are equal to zero:

Proposition 4.5. *Assume that the law of the X_i satisfies a Poincaré inequality and that the moments of order less than three agree with those of the standard Gaussian measure. Then*

$$\mathrm{Ent}_\gamma(\mu_n) \leq \frac{2C_P\sqrt{d}}{n} I(\mu)^{1/2}.$$

This eliminates a logarithmic factor from previous results of [27] in this particular case, but once again does not give the expected sharp decay rate under the moment assumptions.

Proof. This estimate is obtained by applying the HSI inequality with $k = 2$ and the fact that Fisher information is monotone along the CLT [1]. □

Acknowledgments: This work was supported by the Projects MESA (ANR-18-CE40-006) and EFI (ANR-17-CE40-0030) of the French National Research Agency (ANR), ANR-11-LABX-0040-CIMI within the program

ANR-11-IDEX-0002-02 and the France-Berkeley Fund. I would also like to thank Guillaume Cébron, Thomas Courtade, Michel Ledoux and Gesine Reinert for discussions on this topic, as well as an anonymous referee for telling me about [23]. I also thank Benjamin Arras and an anonymous referee for valuable comments, and in particular for pointing out some flaws in how regularity issues were handled in a previous version of this work.

References

[1] S. Artstein, K. Ball, F. Barthe and A. Naor, Solution of Shannon's Problem on the Monotonicity of Entropy, *J. Amer. Math. Soc.* 17, 975-982 (2004).

[2] S. Artstein, K. Ball, F. Barthe and A. Naor, On the Rate of Convergence in the Entropic Central Limit Theorem. *Probab. Theory Relat. Fields* 129, 381-390 (2004).

[3] H. Airault, P. Malliavin, and F. Viens. Stokes formula on the Wiener space and n-dimensional Nourdin-Peccati analysis. *J. Funct. Anal.*, 258(5):1763–1783, 2010.

[4] D. Bakry, I. Gentil and M. Ledoux, Analysis and geometry of Markov diffusion operators. Springer, Grundlehren der mathematischen Wissenschaften, Vol. 348, xx+552 (2014).

[5] D. Bakry, F. Barthe, P. Cattiaux and A. Guillin, A simple proof of the Poincaré inequality in a large class of probability measures including log-concave cases. *Elec. Comm. Prob.* Vol. 13 60–66, 2008.

[6] K. Ball and V. H. Nguyen, Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Math.* 213, 1, 2012.

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

[8] J.-B. Bardet, N. Gozlan, F. Malrieu and P.-A. Zitt, Functional inequalities for Gaussian convolutions of compactly supported measures: explicit bounds and dimension dependence. *Bernoulli* (2018) Vol. 24, No. 1, 333–353

[9] N. Belili and H. Heinich, Distances de Wasserstein et de Zolotarev. *C. R. Acad. Sci. Paris*, t. 330, SÃľrie I, p. 811–814, 2000.

[10] S. G. Bobkov, G. P. Chistyakov and F. Götze, Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.* 41 (2013), no. 4, pp. 2479–2512.

[11] S. G. Bobkov, G. P. Chistyakov and F. Götze, Fisher information and the central limit theorem. *Probab. Theory Relat. Fields* 159 (2014), issue 1-2, pp. 1–59.

[12] T. Bonis, Rates in the Central Limit Theorem and diffusion approximation via Stein's method. *Arxiv preprint* 2018.

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

[14] S. Chatterjee, A short survey of Stein's method. *Proceedings of ICM 2014,* Vol IV, 1–24, 2014.

[15] S. Chatterjee and E. Meckes. Multivariate normal approximation using exchangeable pairs. *Alea*, 4, 257–283, 2008.

[16] T. Courtade, M. Fathi and A. Pananjady, Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Ann. I.H.P. Probab. Stat.*, 55, 2, (2019), 777–790.

[17] M. Fathi, Stein kernels and moment maps. *Ann. Probab.*, 47, 4, (2019), 2172–2185.

[18] M. Fukushima, Dirichlet Formas and Markov processes, volume 23 of North-Holland Mathematical Library, North-Holland Publishing Co., Amsterdam-New York; Kodansha, Ltd., Tokyo (1980).

[19] T. Gallouët, G. Mijoule and Y. Swan, Regularity of solutions of the Stein equation and rates in the multivariate central limit theorem. *Arxiv preprint* (2018).

[20] R. E. Gaunt. Rates of convergence in normal approximation under moment conditions via new bounds on solutions of the Stein equation. *Journal of Theoretical Probability*, 29(1):231–247, 2016.

[21] R. E. Gaunt, Stein's method for functions of multivariate normal random variables. *Arxiv preprint*, 2015.

[22] L. Goldstein and G. Reinert, Stein's Method and the Zero Bias Transformation with Application to Simple Random Sampling. *Ann. Appl. Probab.*, vol. 7, (1997), pp. 935-952.

[23] L. Goldstein and G. Reinert. Distributional transformations, orthogonal polynomials, and Stein characterizations. *Journal of Theoretical Probability*, 18 (2005), 185-208.

[24] L. Gross, Logarithmic Sobolev inequalities. *Amer. J. Math.* 97 (1975), no. 4, 1061–1083.

[25] A. Kufner and B. Opic, How to define reasonably weighted Sobolev spaces. *Comment. Math. Univ. Carolin.* 25 (1984), no. 3, 537–554.

[26] M. Ledoux, Concentration of measure and logarithmic Sobolev inequalities SÃľminaire de Probabilités XXXIII. Lecture Notes in Math. 1709, 120-216. Springer (1999)

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

[28] C. Ley, G. Reinert and Y. Swan, Stein's method for comparison of univariate distributions. *Probability Surveys* 14, 1-52 (2017).

[29] L. Miclo. Notes on the speed of entropic convergence in the central limit theorem. Stochastic inequalities and applications, Birkhäuser, Basel. Progress in Probability, 56, 129–156, 2003.

[30] G. Mijoule, G. Reinert and Y. Swan, Stein operators, kernels and discrepancies for multivariate continuous distributions. *Arxiv preprint* (2018).

[31] B. Muckenhoupt, HardyâĂŹs inequality with weights, *Studia Math.*, 44, p. 31–38 (1972).

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

[33] I. Nourdin, G. Peccati and Y. Swan, Entropy and the fourth moment phenomenon. *J. Funct. Anal.* 266, no. 5, 3170–3207 (2014).

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

[35] N. Ross, Fundamentals of Stein's method. *Probability Surveys* Vol. 8 (2011) 210–293.

[36] A. Saumard, Weighted Poincaré inequalities, concentration inequalities and tail bounds related to the behavior of the Stein kernel in dimension one. *Arxiv preprint*, 2018.

[37] C. Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, pages 583–602, Berkeley, Calif., 1972. Univ. California Press.

[38] C. Stein. Approximate computation of expectations. Institute of Mathematical Statistics Lecture Notes - Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986.

[39] S.A. Utev, Probability problems connected with a certain integrodifferential inequality. *Sib. Math. J.* 30(3):490-493, 1989.

[40] C. Villani, Topics in optimal transportation. Vol. 58 of Graduate Studies in Mathematics, Amer. Math. Soc., Providence, RI, 2003.

[41] D. Zimmermann. Logarithmic Sobolev inequalities for mollified compactly supported measures. *J. Funct. Anal.* 265.6 (2013), pp. 1064–1083.