Existence of Stein Kernels under a Spectral Gap, and Discrepancy Bounds

Thomas A. Courtade[*], Max Fathi[†] and Ashwin Pananjady[*]

[*]UC Berkeley, Department of Electrical Engineering and Computer Sciences
[†]CNRS & Université Paul Sabatier, Institut de Mathématiques de Toulouse

April 11, 2017

Abstract

We establish existence of Stein kernels for probability measures on $\mathbb{R}^d$ satisfying a Poincaré inequality, and obtain bounds on the Stein discrepancy of such measures. Applications to quantitative central limit theorems are discussed, including a new CLT in Wasserstein distance W_2 with optimal rate and dependence on the dimension. As a byproduct, we obtain a stability version of an estimate of the Poincaré constant of probability measures under a second moment constraint. The results extend more generally to the setting of converse weighted Poincaré inequalities. The proof is based on simple arguments of calculus of variations.

Further, we establish two general properties enjoyed by the Stein discrepancy, holding whenever a Stein kernel exists: Stein discrepancy is strictly decreasing along the CLT, and it controls the skewness of a random vector.

1 Introduction

What is known as Stein's method is a vast array of concepts and techniques for proving quantitative convergence of sequences of random variables to some limit. These ideas originated in the work of Stein [47, 48], and have found many applications in the study of quantitative central limit theorems, Poisson and geometric approximation, concentration of measure, random matrix theory and free probability. We refer to the survey [45] for an overview of the topic.

In this work, we shall be interested in one particular concept used in this setting: Stein kernels (also known as Stein factors) and their use in proving quantitative central limit theorems. To this end, let ν be a probability measure on $\mathbb{R}^d$. A matrix-valued function $\tau_\nu : \mathbb{R}^d \longrightarrow \mathcal{M}_d(\mathbb{R})$ is said to be a *Stein kernel* for ν (with respect to the standard Gaussian measure γ on $\mathbb{R}^d$) if for any smooth test function φ taking values in $\mathbb{R}^d$, we have

$$\int x \cdot \varphi d\nu = \int \langle \tau_\nu, \nabla \varphi \rangle_{\mathrm{HS}} d\nu. \tag{1}$$

1

For applications, it generally suffices to consider the restricted class of test functions φ satisfying $\int(|\varphi|^2 + \|\nabla\varphi\|_{\mathrm{HS}}^2)d\nu < \infty$, in which case both integrals in (1) are well-defined as soon as $\tau_\nu \in L^2(\nu)$, provided ν has finite second moments. We shall adopt this convention throughout.

In parts of the literature, the notion of Stein kernel is replaced by the relation

$$\int x \cdot \nabla\varphi d\nu = \int \langle \tau_\nu, \mathrm{Hess}\,\varphi \rangle_{\mathrm{HS}} d\nu \tag{2}$$

for all smooth real-valued function φ. This notion is slightly weaker compared to (1) since it only requires test functions that are gradients, but for some applications it still suffices. Our results will hold for either definition, but we shall adopt the stronger notion (1) throughout since the improvement comes for free.

The motivation behind the definition is that, since the Gaussian measure is the only probability distribution satisfying the integration by parts formula

$$\int x \cdot \varphi d\gamma = \int \mathrm{div}(\varphi)d\gamma, \tag{3}$$

the Stein kernel τ_ν coincides with the identity matrix, denoted by Id, if and only if the measure ν is equal to γ. In this way, the Stein kernel can be seen as a measure of how far ν is from being a standard Gaussian measure in terms of how much it violates the integration by parts formula (3). Those kernels appear implicitly in many works on Stein's method, and have recently been the topic of more direct investigations [4, 40, 41, 42, 33].

The question of when a Stein kernel exists for a particular measure ν is a nontrivial one, and only a few results are known along this direction. In dimension one, it suffices to have a density with connected support to ensure existence. Indeed, if ν has a density p that does not vanish on the (possibly infinite) interval (a, b), then the Stein kernel τ_ν is unique up to sets of measure zero, and is given by

$$\tau_\nu(x) := \frac{1}{p(x)} \int_x^\infty yp(y)dy. \tag{4}$$

In general, however, Stein kernels are not necessarily unique when they exist. A more detailed study of the one-dimensional case and its generalizations to non-Gaussian reference measures can be found in [34].

In higher dimension, existence of Stein kernels has been previously studied using the tools of Malliavin calculus [39]. In particular, if a random variable can be realized as the image of a Gaussian random variable by a C^∞ function with derivatives of at most polynomial growth, then a Stein kernel exists. Another explicit formula for one-dimensional random variables that arise as smooth functions of some gaussian vector was also obtained in [22]. However, given a probability distribution, it may be difficult to find such a smooth function. For example, Brenier's theorem in optimal transport [20] tells us that under fairly general assumptions there exists a map sending a Gaussian random variable onto the distribution considered, but in general it is not smooth enough to apply the arguments of [39].

Our main results are roughly divided into two categories: sufficient conditions for existence of Stein kernels in arbitrary dimension, and general bounds on the so-called *Stein discrepancy* which hold whenever a Stein kernel exists. Specifically, we first show that if ν satisfies a Poincaré inequality, or more generally a converse weighted Poincaré inequality, then a Stein kernel exists. This affirmatively answers a question raised in [42]. In doing so, we obtain bounds on the associated Stein discrepancy for measures satisfying a Poincaré inequality. These estimates are dimension-free and depend only on the second moment and the Poincaré constant. We further establish two properties enjoyed by Stein discrepancy that hold in general, whenever a Stein kernel exists. First, like entropy and Fisher information, Stein discrepancy is monotone along the CLT. Second, Stein discrepancy is bounded from below by the skewness of a random vector.

These results lead to optimal rates of convergence in the multidimensional Central Limit Theorem in Wasserstein distance W_2, as well as entropic CLTs, with suboptimal rate. Our main estimate can also be reformulated as a quantitative improvement of the fact that among all isotropic measures, the standard Gaussian measure has the best Poincaré constant.

2 On existence of Stein kernels

Let ν be a probability measure on $\mathbb{R}^d$. Henceforth, we make the following assumption:

Assumption. *The measure ν is absolutely continuous with respect to the Lebesgue measure, and has finite second moment, i.e. $\int |x|^2 d\nu < \infty$.*

We shall work in the Sobolev space $W_\nu^{1,2}$ of vector valued functions, which we define as as the closure of the set of all smooth vector-valued functions $f : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ in $L^2(\nu)$, with respect to the usual Sobolev norm $\int (|f|^2 + \|\nabla f\|_{\mathrm{HS}}^2) d\nu$. We also define its restriction to the set of (vector-valued) functions with average zero $W_{\nu,0}^{1,2} := W_\nu^{1,2} \bigcap \{ f : \int f d\nu = 0 \}$.

Definition 2.1. *A function $\tau_\nu : \mathbb{R}^d \longrightarrow \mathcal{M}_d(\mathbb{R})$ is a Stein kernel for ν if for any $\varphi \in W_\nu^{1,2}$ equation (1) holds. The Stein discrepancy is defined as*

$$S(\nu|\gamma)^2 := \inf \int \|\tau_\nu - \mathrm{Id}\|_{\mathrm{HS}}^2 d\nu,$$

where the infimum is taken over all Stein kernels of ν, and takes value $+\infty$ if no Stein kernel exists.

One of the main applications of Stein kernels is that bounds on the Stein discrepancy can be used to obtain rates of convergence in the central limit theorem, as discussed in Section 4.

Definition 2.2. *A probability measure ν is said to satisfy a Poincaré inequality with constant C_p if for any locally Lipschitz function $f \in L^2(\nu)$ we have*

$$\mathrm{Var}_\nu(f) \leq C_p \int |\nabla f|^2 d\nu.$$

A measure satisfying a Poincaré inequality is also said to have spectral gap. The terminology comes from the fact that C_p^{-1} is a lower bound on the smallest positive eigenvalue of the operator $-\Delta + \nabla H \cdot \nabla$ in $L^2(\nu)$, where $H = -\log \frac{d\nu}{dx}$.

There is a vast literature on Poincaré inequalities, with many examples and abstract results giving sufficient conditions for one to hold. In particular, the class of measures satisfying a Poincaré inequality is stable under bounded perturbations and tensor products, and it contains the set of all log-concave probability measures. A more general sufficient condition for a measure with density e^{-V} to have spectral gap is

$$\exists a \in (0,1), R \geq 0, c > 0 \text{ such that } a|\nabla V(x)|^2 - \Delta V(x) \geq c \quad \forall |x| \geq R,$$

which was obtained in [6]. We refer to [5] for more background on Poincaré inequalities.

2.1 Finite Poincaré constant ensures existence of Stein kernel

Our main result of this section is that a Poincaré inequality ensures existence of a Stein kernel, and moreover that the Poincaré constant controls the Stein discrepancy. Stated more precisely,

Theorem 2.3. *Assume that ν is centered and satisfies a Poincaré inequality with constant C_p. Then there exists a unique function $g \in W_{\nu,0}^{1,2}$ such that $\tau_\nu = \nabla g$ is a Stein kernel for ν. Moreover,*

$$\int \|\tau_\nu\|_{\mathrm{HS}}^2 d\nu \leq C_p \int |x|^2 d\nu \tag{5}$$

so that the Stein discrepancy satisfies

$$S(\nu|\gamma)^2 \leq (C_p - 2) \int |x|^2 d\nu + d$$

The centering assumption on ν is necessary for the theorem to hold. Indeed, a necessary condition for existence of a Stein kernel is that ν is centered, seen by taking $\varphi = 1$ in the defining equation (1).

In most situations, we shall be using the above bounds for measures with second moment normalized with respect to dimension (e.g., as is the case for isotropic measures):

Corollary 2.4. *Let ν be a centered probability measure on $\mathbb{R}^d$ satisfying a Poincaré inequality with constant C_p, normalized so that $\int |x|^2 d\nu = d$. Then*

$$S(\nu|\gamma)^2 \leq d(C_p - 1).$$

A few remarks are in order:

- The standard Gaussian measure γ has Poincaré constant $C_p = 1$, so the above estimates dictate $S(\gamma|\gamma) = 0$, as desired.

- Stein discrepancy is additive on product measures (i.e., $S(\nu^{\otimes k}|\gamma^{\otimes k})^2 = kS(\nu|\gamma)^2$), whereas the Poincaré constant is independent of dimension (i.e., $C_p(\nu^{\otimes k}) = C_p(\nu)$). Thus, our estimates are dimension-free in nature.

- A finite Poincaré constant is by no means necessary for existence of a Stein kernel. In dimension one, the formula (4) works in more general situations. We will see other multi-dimensional examples further on.

Proof of Theorem 2.3. The result follows from an application of the Lax-Milgram theorem [32]. Indeed, $\int \langle \nabla f, \nabla h \rangle_{\mathrm{HS}} d\nu$ is a continuous bi-linear functional on $W_{\nu,0}^{1,2} \times W_{\nu,0}^{1,2}$, and is coercive by the assumption that ν satisfies a Poincaré inequality. Hence, there exists a unique $g \in W_{\nu,0}^{1,2}$ such that

$$\int \langle \nabla g, \nabla f \rangle_{\mathrm{HS}} d\nu = \int x \cdot f d\nu$$

for any $f \in W_{\nu,0}^{1,2}$, and in particular ∇g is a Stein kernel.

In addition, g minimizes the functional $J(f) := \frac{1}{2} \int \|\nabla f\|_{\mathrm{HS}}^2 d\nu - \int x \cdot f d\nu$. Indeed,

$$
\begin{aligned}
J(f) &= \frac{1}{2} \int \|\nabla f\|_{\mathrm{HS}}^2 d\nu - \int x \cdot f d\nu \\
&= \frac{1}{2} \int \|\nabla f\|_{\mathrm{HS}}^2 d\nu - \int \langle \nabla g, \nabla f \rangle_{\mathrm{HS}} d\nu \\
&\geq -\frac{1}{2} \int \|\nabla g\|_{\mathrm{HS}}^2 d\nu = \frac{1}{2} \int \|\nabla g\|_{\mathrm{HS}}^2 d\nu - \int x \cdot g d\nu = J(g).
\end{aligned}
$$

Noting from above that $J(g) = -\frac{1}{2} \int \|\nabla g\|_{\mathrm{HS}}^2 d\nu$, the Cauchy-Schwarz inequality and the Poincaré inequality for ν give, after a simple computation,

$$-\frac{1}{2} \int \|\nabla g\|_{\mathrm{HS}}^2 d\nu = J(g) \geq \frac{1}{2C_p} \int |g|^2 d\nu - \left(\int |g|^2 d\nu \right)^{1/2} \left(\int |x|^2 d\nu \right)^{1/2} \quad (6)$$

$$\geq -\frac{C_p}{2} \int |x|^2 d\nu, \quad (7)$$

establishing (5). $\qquad \square$

Remark 2.1. Even when ν does not satisfy a Poincaré inequality, if $g \in W_{\nu,0}^{1,2}$ minimizes the functional $J : f \mapsto \frac{1}{2} \int \|\nabla f\|_{\mathrm{HS}}^2 - \int x \cdot f d\nu$, then ∇g is a Stein Kernel for ν. To see this, consider a perturbation in the direction $h \in W_{\nu,0}^{1,2}$, which gives:

$$
\begin{aligned}
0 &\leq J(g + \varepsilon h) - J(g) \\
&= \varepsilon \left(\int \langle \nabla g, \nabla h \rangle d\nu - \int x \cdot h d\nu \right) + \frac{\varepsilon^2}{2} \int |\nabla h|^2 d\nu.
\end{aligned}
$$

Letting $\varepsilon \downarrow 0$ shows that $\int \langle \nabla g, \nabla h \rangle d\nu \geq \int x \cdot h d\nu$. Replacing h by $-h$ gives the reverse inequality.

Hence, a sufficient condition for existence of a Stein kernel is that the functional J has a minimum. Stated another way, there exists a finite constant $c > 0$ such that

$$\left(\int x \cdot f d\nu \right)^2 \leq c \int \|\nabla f\|_{\mathrm{HS}}^2 d\nu \quad \forall f \in W_{\nu,0}^{1,2}, \tag{8}$$

and moreover, equality is attained for some nonzero function g. This should be compared against the definition of the Poincaré inequality.

Of course, there exist probability measures that satisfy the (8) condition without satisfying a (converse weighted) Poincaré inequality. For example, if we consider two disjoint closed annuli C_1 and C_2 that are centered around the origin, then the uniform probability measure on $C_1 \cup C_2$ does not satisfy a Poincaré inequality, yet it does satisfy (8).

We conclude this section by noting that, as pointed out in [33], for log-concave measures there is a reverse version of our inequality:

Proposition 2.5. *Let ν be a centered log-concave measure. Then for some numerical constant C,*
$$C_p \leq C(1 + S(\nu|\gamma)^2).$$

This statement, combined with our main result, tells us that for log-concave measures, controlling the Stein discrepancy and controlling the Poincaré constant are equivalent. At first glance, the above estimate does not capture the dimension-free nature of the Poincaré constant. This may be a necessary downside of such bounds, since if we consider a measure of the form $\nu = \gamma_{d-1} \otimes \tilde{\nu}$ it behaves in the correct way, since the Poincaré constant is at least as bad as that of the projection along the worst direction.

As mentioned in [33], this estimate is obtained by combining the moment bound of Theorem 2.8 in [33] and Milman's results on obtaining estimates on Poincaré constants of log-concave measures by the worst variance of 1-Lipschitz functions [38].

2.2 Converse weighted Poincaré inequalities

A generalized form of the Poincaré inequality is still enough to ensure existence of a Stein kernels, up to an additional assumption on moment bounds.

Definition 2.6. *A probability measure ν is said to satisfy a converse weighted Poincaré inequality with weight $\omega : \mathbb{R} \longrightarrow \mathbb{R}_+^*$ if for any locally lipschitz $f \in L^2(\nu)$ we have*
$$\inf_{c \in \mathbb{R}} \int (f - c)^2 \omega d\nu \leq \int |\nabla f|^2 d\nu.$$

This definition originates from [15], and was further studied in [21]. Such inequalities are related to measure concentration for heavy-tailed distributions. We could incorporate a constant in front of the Dirichlet form in the definition, but we have chosen to absorb it into the weight to reduce notations.

Theorem 2.7. *Assume that ν is centered and satisfies a converse weighted Poincaré inequality with weight ω, and that $\int |x|^2 \omega^{-1} d\nu < \infty$. Then there exists a Stein kernel τ_ν for ν, that moreover satisfies*

$$\int \|\tau_\nu\|_{\mathrm{HS}}^2 d\nu \leq \int |x|^2 \omega^{-1} d\nu.$$

The proof is exactly the same as for the Poincaré case, we just use ω as an extra weight when using the Cauchy-Schwarz inequality

$$\left(\int x \cdot f d\nu\right)^2 \leq \inf_{c_1,..,c_d} \left(\int |x|^2 \omega^{-1} d\nu\right)\left(\int \sum (f_i - c_i)^2 \omega d\nu\right).$$

Converse Poincaré inequalities have been established for a large class of heavy-tailed probability distributions via Lyapunov function techniques in [21]. Here are some examples from [15, 21]:

Corollary 2.8. *Stein kernels exist for the following probability distributions on $\mathbb{R}^d$:*
   (i) Generalized Cauchy distributions $\nu_\beta(dx) := Z^{-1}(1+|x|^2)^{-\beta}$ for $\beta > \max((d+4)/2, d)$;
   (ii) Probability measures of the form $\nu(dx) = Z^{-1}e^{-V(x)^p}$ with V convex and $p > 0$, as soon as $\int |x|^{2+2(1-p)} d\nu < \infty$. In particular, subexponential distributions with density proportional to $e^{-|x|^p}$ with $p \in (0,1)$.

Note that these examples typically do not satisfy a classical Poincaré inequality.

2.3   Extension to non-Gaussian reference measures

Theorem 2.3 also generalizes to Stein kernels with non-Gaussian reference measures. Such an extension is natural in the framework of the generator approach to Stein's method, where an integration by parts formula for a given measure is obtained by finding a Markov generator that leaves the considered measure invariant. This approach was pioneered in [9, 28]. Stein's method for the approximation of non-Gaussian reference measures has had some successful applications in the study of convergence of Markov Chain Monte Carlo algorithms [27] and for generalizations of the fourth moment theorem [3]. The Gaussian functional inequalities of [33] were also extended to a class of non-Gaussian measures, using arguments from Bakry-Émery calculus.
   We can extend Theorem 2.3 to the situation where the Gaussian measure is replaced by a general reference measure $\mu = e^{-V}dx$, where $V : \mathbb{R}^d \to \mathbb{R}$ is a

smooth function. In this situation, a Stein kernel of a measure ν with respect to μ is defined by the relation

$$\int \nabla V \cdot \varphi d\nu = \int \langle \tau_\nu, \nabla \varphi \rangle_{\mathrm{HS}} d\nu \quad \forall \varphi \in W_\nu^{1,2}. \tag{9}$$

Applying the same arguments as for the Gaussian case, we obtain

Theorem 2.9. *Let* $\mu = e^{-V} dx$, *where* $V : \mathbb{R}^d \to \mathbb{R}$ *is smooth. Assume that* ν *satisfies a Poincaré inequality with constant* C_p, *that* $\int \nabla V d\nu = 0$ *and that* $\int |\nabla V|^2 d\nu < \infty$. *Then there exists a Stein kernel for* ν, *relative to* μ, *of the form* $\tau_\nu = \nabla g$ *for some* $g \in W_\nu^{1,2}$. *Moreover, we have the bound*

$$\int \|\tau_\nu\|_{\mathrm{HS}}^2 d\nu \le C_p \int |\nabla V|^2 d\nu. \tag{10}$$

Note that for polynomial potentials V, the finiteness of $\int |\nabla V|^2 d\nu$ automatically follows from the Poincaré inequality. As in the previous section, this result can easily be generalized to cover measures satisfying a converse weighted Poincaré inequality.

3   General bounds on the Stein discrepancy

3.1   Stein discrepancy controls skewness

Above, it was shown that in presence of a suitable Poincaré inequality, the Stein discrepancy is controlled from above by second moments. Here, we establish a complementary lower bound on the Stein discrepancy in terms of skewness:

Theorem 3.1. *Let* $X = (X_1, X_2, \ldots, X_d)$ *have law* ν. *If* ν *is isotropic with finite fourth moment, then*

$$S(\nu|\gamma)^2 \ge \frac{1}{9} \sum_{i=1}^d |\mathbb{E}[X_i^3]|^2.$$

Proof. First, we shall reduce the problem to the one-dimensional case. Let τ_ν be a Stein factor for ν. Then $\tau^i(x) := \mathbb{E}[\tau_\nu(X)|X_i = x]$ is a Stein kernel for X_i. Moreover, we have

$$S(\nu|\gamma)^2 = \int \sum_{1 \le i,j \le d} |\tau_{i,j} - \delta_{i,j}|^2 d\nu$$

$$\ge \int \sum_{i=1}^d |\tau_{i,i} - 1|^2 d\nu \ge \sum_{i=1}^d \mathbb{E}[|\tau^i(X_i) - 1|^2] \ge \sum_{i=1}^d S(\nu_i|\gamma_i)^2,$$

where ν_i is the law of X_i and γ_i is the ith marginal of γ. Hence it is enough to prove the theorem when $d = 1$.

Now let X be a real-valued random variable with mean zero and unit variance, and let ν_n be the law of the standardized sum $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$, where the X_i's are independent copies of X. Then

$$S(\nu|\gamma)^2 \geq nS(\nu_n|\gamma)^2 \tag{11}$$

for any $n \geq 1$ (see for example Section 2.5 in [33] and Theorem 3.2 in the next section). Moreover, it was established in [33] that the Stein discrepancy is always larger than the Wasserstein distance W_2 to the standard Gaussian measure. Hence $S(\nu_n|\gamma)^2 \geq W_2(\nu_n, \gamma)^2$ for all $n \geq 1$. Finally, Rio established in [44] that under our assumptions, $\sqrt{n} W_2(\nu_n, \gamma) \longrightarrow \frac{1}{3}|\mathbb{E}[X^3]|$, which concludes the proof. $\qquad\square$

3.2 Strict Monotonicity of the Stein Discrepancy in the CLT

Monotonicity of information measures along the CLT have a long history, going back to Shannon's conjecture on the monotonicity of the entropy, which was eventually resolved in [2]. More specifically, if $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$, where $X_1, \dots, X_n$ are i.i.d. isotropic random vectors, then both the entropy and Fisher information of S_n with respect to the standard gaussian measure are non-increasing in n. Following Artstein, Ball, Barthe and Naor's proof of this fact, several generalizations and alternative proofs have been discovered [50, 46, 35, 36, 37, 23].

Since the Stein discrepancy relates to both Fisher information and entropy in various ways [33], it is natural to conjecture that it also is non-increasing along the CLT. It turns out that this is indeed the case and, in fact, it is strictly decreasing. The following generalizes (11) along these lines:

Theorem 3.2. *Let ν be an isotropic probability measure on $\mathbb{R}^d$, and let ν_n denote the law of $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$, where $X_1, \dots, X_n$ are i.i.d. with law ν. Then*

$$S(\nu_n|\gamma)^2 \leq \frac{m}{n} S(\nu_m|\gamma)^2 \qquad 1 \leq m \leq n.$$

Proof. For $m \geq 1$, let τ_m denote a Stein kernel associated to S_m. Assuming such a τ_m exists (if it does not, the claim is vacuous), then for all $n \geq m$, the function

$$\tau_n(s_n) = \mathbb{E}[\tau_m(S_m)|S_n = s_n] \tag{12}$$

is a valid Stein kernel for S_n. Indeed, for $s_n = \sqrt{\frac{m}{n}} s_m + \tilde{s}$, any smooth function φ evaluated on s_n may also be considered as a smooth function of s_m for each fixed $\tilde{s}$. In particular, the chain rule directly yields

$$\nabla_{s_m} \varphi(s_n) = \sqrt{\frac{m}{n}} \nabla_{s_n} \varphi(s_n).$$

Therefore, starting with linearity of expectation and defining $\tilde{S} := S_n - \sqrt{\frac{m}{n}}S_m$, we may write

$$
\begin{aligned}
\mathbb{E}\langle S_n, \varphi(S_n)\rangle &= \sqrt{\frac{n}{m}}\mathbb{E}\langle S_m, \varphi(S_n)\rangle \\
&= \sqrt{\frac{n}{m}}\mathbb{E}[\mathbb{E}[\langle S_m, \varphi(S_n)\rangle | \tilde{S}]] \\
&= \sqrt{\frac{n}{m}}\mathbb{E}[\mathbb{E}[\langle \tau_m(S_m), \nabla_{s_m}\varphi(S_n)\rangle_{\mathrm{HS}} | \tilde{S}]] \\
&= \mathbb{E}[\langle \tau_m(S_m), \nabla_{s_n}\varphi(S_n)\rangle_{\mathrm{HS}}] \\
&= \mathbb{E}\langle \mathbb{E}[\tau_m(S_m)|S_n], \nabla\varphi(S_n)\rangle_{\mathrm{HS}},
\end{aligned}
$$

establishing (12).

Following [23], if a function $\vartheta : \mathbb{R}^d \to \mathbb{R}$ satisfies $\mathbb{E}\vartheta(S_m) = 0$, then

$$
\mathbb{E}[|\mathbb{E}[\vartheta(S_m)|S_n]|^2] \leq \frac{m}{n}\mathbb{E}[|\vartheta(S_m)|^2] \qquad\qquad 1 \leq m \leq n. \qquad (13)
$$

This inequality is due to Dembo, Kagan and Shepp [26]; see also Kamath and Nair [31]. Now, $\mathbb{E}[\tau_m(S_m)] = \mathrm{Id}$, so a direct application of (13) yields

$$
\mathbb{E}\|\tau_n(S_n) - \mathrm{Id}\|_{\mathrm{HS}}^2 = \mathbb{E}[\|\mathbb{E}[\tau_m(S_m) - \mathrm{Id}\,|S_n]\|_{\mathrm{HS}}^2] \leq \frac{m}{n}\mathbb{E}\|\tau_m(S_m) - \mathrm{Id}\|_{\mathrm{HS}}^2.
$$

Taking the infimum over all valid Stein kernels τ_m, τ_n finishes the proof.

$\square$

Remark 3.1. *The same result holds if (2) is adopted as the definition of a Stein kernel.*

4   Applications

4.1   Quantitative central limit theorems

We shall now discuss some applications of the bounds to quantitative central limit theorems in Wasserstein distance W_2, which is defined as

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

where the infimum is taken over all couplings π of the probability measures μ and ν. We refer the reader to the textbook [51] for more information about Wasserstein distances and optimal transport.

Let X be a centered isotropic random vector on $\mathbb{R}^d$ with law ν, and let ν_n be the law of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$, where the X_i are independent copies of X. As was already

discussed in the proof of Theorem 3.1 (or as a particular case of Theorem 3.2), we have the inequalities

$$W_2(\nu_n, \gamma)^2 \leq S(\nu_n|\gamma)^2 \leq \frac{1}{n} S(\nu|\gamma)^2. \tag{14}$$

From this remark, we immediately deduce the following result:

Theorem 4.1. *Assume that ν is centered, isotropic and satisfies a Poincaré inequality with constant C_p. Then*

$$W_2(\nu_n, \gamma)^2 \leq \frac{d(C_p - 1)}{n}.$$

We remark that the rate $W_2(\nu_n, \gamma) = O(1/\sqrt{n})$ is known to be optimal in general. Moreover, the dependence on the dimension is sharp, since it cannot be improved for product measures. To our knowledge, this seems to be the first result with sharp dependence on both the dimension and on n for W_2 and with assumptions satisfied by a large class of probability measures. A similar result can be obtained with converse weighted Poincaré inequalities, with the same sharp rate but a less explicit prefactor.

There are several similar results already present in the literature. In dimension one, a more general result has been obtained by Rio in [43, 44], where a finite fourth moment suffices. He also obtained convergence in stronger transport distances when the random variable has a finite exponential moment, which is a weaker assumption than our use of a Poincaré inequality. The proofs rely on an explicit representation of transport maps involving the repartition function of ν, which is unavailable in higher dimensions. Subsequently, Bobkov [11] combined optimal rates in the entropic CLT [14] with Talagrand's inequality to conclude $O(1/\sqrt{n})$ convergence of $W_2(\nu_n, \gamma)$ in dimension one, but left open the problem in higher dimensions.

In the multidimensional setting, Zhai [53] has recently claimed that for random variables in dimension d, we have $W_2(\nu_n, \gamma) \leq \frac{5\sqrt{d}\beta(1+\log n)}{\sqrt{n}}$, under the boundedness assumption $|X| \leq \beta$ almost surely. His assumptions are not directly comparable with ours, since bounded random variables do not necessarily satisfy a Poincaré inequality, while there are many examples of unbounded random variables that do satisfy one. However, it is true that every bounded random variable regularized via convolution with a Gaussian measure of arbitrarily small variance does satisfy a Poincaré inequality [10], which suggests that Zhai's result may potentially be improved to have optimal dependence on both dimension and n. Unfortunately, the bounds on the Poincaré constant obtained in [10] are exponential in β, so it is not clear whether Zhai's result may be recovered from our own via this route. In situations where both estimates apply, the bound in the present work will typically be smaller. For example, for high-dimensional product measures, the Poincaré constant is independent of the dimension, while β would be of order of $\sqrt{d}$. Moreover, we eliminate the extra $\log n$ factor. It may be relevant to point

out that both our assumptions and those of [53] fit in the framework of random variables with a finite exponential moment.

Also in higher dimensions, Bonis showed in [16] that, under the moment constraint $\mathbb{E}\|X\|^{2+m} < \infty$ for $m \in [0, 2]$, we have the asymptotic rate $W_2(\nu_n, \gamma) = O(n^{-1/2+(2-m)/4})$. However, in the case $m = 2$, the prefactor (which does not appear explicitly in Bonis' work) seems to have a suboptimal dependence on the dimension d [17].

In dimension 1, and for log-concave measures in higher dimension, the works [7, 8, 30] can be used to obtain a sharp rate of convergence in relative entropy when a Poincaré inequality holds, which implies convergence in W_2. These results however would rely on a bound on the relative entropy of ν, which in general would give a worse prefactor in the bound.

While the arguments we used rely on the independence of the X_i, we do not actually need them to be identically distributed. In particular, we have the following result:

Theorem 4.2. *Assume that X_i is isotropic and satisfies a Poincaré inequality with constant C_i. Let ν_n be the law of $\frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} X_i$. Then*

$$W_2(\nu_n, \gamma)^2 \leq \frac{d}{n^2} \sum_{i=1}^{n} (C_i - 1).$$

It is also possible to extend the method to *weakly* dependent random variables, using a standard splitting trick:

Theorem 4.3. *Let (X_i) be a sequence of centered isotropic random variables, and assume that there exists a k such that as soon as $|i - j| > k$ then X_i and X_j are independent. Assume moreover that $\mathbb{E}[X_i \cdot X_j] = 0$ for all distinct i, j, and that the law of each random variable satisfies a Poincaré inequality with uniform constant C_p. Then*

$$W_2(\nu_n, \gamma)^2 \leq \frac{d(C_p - 1) + 2dk}{\lfloor n/k \rfloor}$$

where ν_n is the law of $S_n := \frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} X_i$.

As an example, this theorem applies to $X_i := \prod\limits_{j=0}^{k-1} f_j(Y_{i+j})$ with the Y_i i.i.d. random variables satisfying some Poincaré inequality and the f_j bounded lipschitz functions with mean zero and unit variance.

Proof. We can define the partial sums

$$S_n^j := \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} X_{i(k+1)+j}$$

for $j = 1, .., k$. Then each S_n^j is a sum of independent random variables, so we can apply Theorem 4.2 to obtain convergence in W_2 distance to the Gaussian, i.e.

$$W_2(\nu_{j,n}, \gamma)^2 \le \frac{(C_p - 1)d}{n}$$

where $\nu_{j,n}$ is the law of S_n^j. Since $S_{kn} = \frac{1}{\sqrt{k}} \sum_{j=1}^{k} S_n^j$, it is easy to check that $W_2(\nu_{kn}, \gamma)^2 \le \frac{1}{k} \sum W_2(\nu_{j,n}, \gamma)^2 \le \frac{(C_p-1)d}{n}$. Moreover,

$$W_2(\nu_{kn+j}, \gamma)^2 \le \frac{kn}{kn+j} W_2(\nu_{kn}, \gamma)^2 + \frac{2dj}{kn+j},$$

where we split the sum and used the fact that the Wasserstein distance is bounded by the second moment to control the contribution of $X_{nk+1}, .., X_{nk+j}$. $\qquad\square$

4.2 Entropy bounds

Let $H_\gamma(\nu) := \int d\nu \log \frac{d\nu}{d\gamma}$ denote the entropy of ν relative to γ. As a direct consequence of the HSI inequality of [33], we also have the following rate of convergence in the entropic CLT:

Proposition 4.4. *Assume that ν satisfies a Poincaré inequality with constant C_p, and satisfies the Fisher information bound $\int |\nabla \log f|^2 d\gamma \le \alpha d$, where $f = \frac{d\nu}{d\gamma}$. Then we have*

$$H_\gamma(\nu_n) \le \frac{d(C_p - 1)}{2n} \log \left(1 + \frac{\alpha n}{(C_p - 1)} \right)$$

Convergence to the Gaussian measure in entropy is strictly stronger than convergence in W_2, due to Talagrand's inequality [49]. The choice of scaling in the dimension for the upper bound on the Fisher information reflects the fact that for product measures, it is of order d. In dimension one, the actual rate of convergence in the entropic CLT is asymptotically $1/n$ under a fourth moment condition [13], and non-asymptotically $1/n$ if the entropy of a single random variable is bounded [14] (with a prefactor that is exponential in the entropy). When the Poincaré inequality holds a non-asymptotic rate was obtained in [8, 1], and extended to multi-dimensional random vectors having log-concave density in [7]. Related results in dimension one were obtained in [30].

4.3 Fisher information bounds

In this section, we shall combine our main estimate with results of [42] to obtain bounds on the Fisher information of a sum of independent random variables, to which we add a small Gaussian noise. To this end, recall that the Fisher information of ν relative to γ is defined as $I_\gamma(\nu) := \int |\nabla \log f|^2 d\gamma$, where $f = \frac{d\nu}{d\gamma}$. After applying Theorem V.3 in [42], we get

Proposition 4.5. *Let (X_i) be a collection of independent centered isotropic random variables in $\mathbb{R}^d$ with Poincaré constants C_i. Let $W_n := \sqrt{1-t}Z + \frac{\sqrt{t}}{\sqrt{n}}\sum\limits_{i=1}^{n} X_i$, where Z is a standard Gaussian random variable independent of the X_i. If ν_n^t denotes the law of W_n, then*

$$I_\gamma(\nu_n^t) \leq \frac{t^2}{n^2(1-t)} \sum_{i=1}^{n} (C_i - 1)d.$$

In particular, if the X_i satisfy a Poincaré inequality with same constant C_p, then

$$I_\gamma(\nu_n^t) \leq \frac{t^2(C_p - 1)d}{n(1-t)}.$$

Due to Cramer's law [24], weak convergence of ν_n^t to γ is equivalent to convergence of $\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} X_i$. Unfortunately, Cramer's law is unstable in general [12], so we cannot directly deduce quantitative closeness of X to a Gaussian if $X + Z$ is for any random variable (although the counterexamples of [12] do not seem to satisfy a Poincaré inequality, so it may be that under such an extra assumption Cramer's law would be stable).

Rates of convergence in Fisher information in dimension one when the information of a single variable is finite have been obtained in [30]. In higher dimension, a quantitative bound on the difference between Fisher informations of ν_1 and ν_2 was obtained in [29], but does not readily lead to a quantitative central limit theorem.

Remark 4.1. *Instead of using the results of [42], it is possible to derive upper bounds on $I_\gamma(\nu_n^t)$ by $W_2(\nu_n, \gamma)^2$ using the gradient flow structure of the Ornstein-Uhlenbeck flow, as done for example in Theorem 24.16 of [52], and apply our bounds on the rate of convergence in W_2 distance to conclude.*

4.4 Stability of the Poincaré constant under a second moment constraint

Combined with the previously mentioned fact that Stein discrepancy controls W_2 distance to γ, our bound on the Stein discrepancy implies the following estimate:

Theorem 4.6. *Let ν be a centered probability measure on $\mathbb{R}^d$, normalized so that $\int |x|^2 d\nu = d$. Then its Poincaré constant C_p satisfies*

$$C_p \geq 1 + \frac{W_2(\nu, \gamma)^2}{d}.$$

This estimate is a quantitative reinforcement of the fact that among all probability measures with the same second moment, the Gaussian has the best Poincaré constant. More generally, it is a reinforcement of the fact that, given a sequence of centered measures (ν_n) with $\int |x|^2 d\nu_n = d$, if their Poincaré constants converge

to 1, then ν_n weakly converges to the standard Gaussian [18]. Once again, we note that this estimate depends optimally in the dimension.

In a different direction, De Philippis and Figalli [25] recently showed a similar quantitative stability result among a different class of measures: for densities that are of the form $e^{-V}\gamma$ with V convex, the Gaussian has the worst Poincaré constant, and we have a deficit of the form $1 - C_p \leq \epsilon \Longrightarrow W_1(\nu, \gamma) \leq C(d, \alpha)|\log \epsilon|^{-1/4+\alpha}$ for any $\alpha > 0$, for ϵ small enough. Our results are not directly comparable, since they concern completely different classes of measures. We just note that the dependence in ϵ in the result of [25] is not expected to be sharp. Indeed, in dimension one they show that $W_1(\nu, \gamma) \leq C\epsilon$. In spirit, this question is also similar to the stability problem for the Szegö-Weinberger inequality, that was solved in [19].

Finally, we observe that Theorem 2.7 leads to the more general analogous result for measures satisfying a converse weighted Poincaré inequality. Although it is easily seen that such inequalities are stable under log-bounded transformations of the measure [21], the following appears to be the first quantitative stability result along these lines:

Theorem 4.7. *Let ν be a centered probability measure on $\mathbb{R}^d$, normalized so that $\int |x|^2 d\nu = d$. If ν satisfies a converse weighted Poincaré inequality with weight function ω, then*

$$\frac{1}{d} \int |x|^2 \omega^{-1} d\nu \geq 1 + \frac{W_2(\nu, \gamma)^2}{d}.$$

By Hölder's inequality, the following corollary is immediate:

Corollary 4.8. *Let ν be a centered probability measure on $\mathbb{R}^d$, normalized so that $\int |x|^2 d\nu = d$. If ν satisfies a converse weighted Poincaré inequality with weight function ω, then for Hölder conjugate exponents $1 \leq p, q \leq \infty$,*

$$\|\tfrac{1}{d}|x|^2\|_{L^p(\nu)} \|\omega^{-1}\|_{L^q(\nu)} \geq 1 + \frac{W_2(\nu, \gamma)^2}{d}.$$

Of course, Theorem 4.6 coincides with the special case where $p = 1$ and $q = \infty$.

References

[1] 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).

[2] S. Artstein, K. Ball, F. Barthe and A. Naor, Solution of Shannon's Problem on the Monotonicity of Entropy, *Journal of the AMS* 17, 975-982 (2004).

[3] E.Azmoodeh, S.Campese and G.Poly, Fourth moment theorems for Markov diffusion generators. *Journal of Functional Analysis*, Volume 266, Issue 4 (2013).

[4] H. Airault, P. Malliavin, and F. Viens. Stokes formula on the Wiener space and n-dimensional Nourdin-Peccati analysis. *Journal of Functional Analysis*, 258(5):1763–1783, 2010.

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

[6] 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.

[7] K. Ball and V. H. Nguyen, Entropy jumps for isotropic log-concave random vectors and spectral gap, *Studia Mathematica*, vol. 213, no. 1, pp. 81–96, 2012.

[8] K. Ball, F. Barthe, and A. Naor, Entropy jumps in the presence of a spectral gap, *Duke Mathematical Journal*, vol. 119, no. 1, pp. 41–63, 2003.

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

[10] Bardet, J.B., Gozlan, N., Malrieu, F. and Zitt, P.A. Functional inequalities for Gaussian convolutions of compactly supported measures: explicit bounds and dimension dependence. *Bernoulli*, to appear (2016).

[11] Sergey G. Bobkov. Entropic approach to E. Rio's central limit theorem for W_2 transport distance. *Statistics and Probability Letters*, 83(7):1644–1648, 2013.

[12] Bobkov, S. G., Chistyakov, G. P., Gotze, F. Entropic instability of Cramer's characterization of the normal law. Selected Works of Willem van Zwet, pp. 231-242, Sel. Works Probab. Stat., Springer, New York, 2012.

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

[14] Bobkov, S. G.; Chistyakov, G. P.; Götze, F. Berry-Esseen bounds in the entropic central limit theorem. *Probab. Theory Related Fields* 2014 (159), pp. 435–478.

[15] S. Bobkov and M. Ledoux, Weighted Poincar-type inequalities for Cauchy and other convex measures *Ann. Probability* 37, 403-427 (2009)

[16] T. Bonis, Rates in the Central Limit Theorem and diffusion approximation via Steins Method. *Arxiv preprint* (2016).

[17] T. Bonis, *personal communication.*

[18] A.A. Borovkov and S.A. Utev. On an inequality and a related characterisation of the normal distribution. *Theory of Probability and Its Applications*, 28:219–228, 1984.

[19] L. Brasco and A. Pratelli, Sharp stability of some spectral inequalities, *Geometric and Functional Analysis*, 2012, Vol. 22, Issue 1, pp 107–135.

[20] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* 44, no.4, pp. 375-417 (1991).

[21] P. Cattiaux , N. Gozlan , A. Guillin, C. Roberto. Functional inequalities for heavy tails distributions and application to isoperimetry. *Electronic J. Probab.* 15 : 346–385, 2010

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

[23] T. A. Courtade, Monotonicity of Entropy and Fisher Information: A quick proof via maximal correlation. *Communications on Information and Systems*, vol. 16, no. 2, pp. 111–115, 2016.

[24] Cramér, H. Ueber eine Eigenschaft der Normalen Verteilungsfunktion. *Math. Zeitschrift*, 41 (1936), 405–414.

[25] G. De Philippis and A. Figalli, Rigidity and stability of Caffarelli's log-concave perturbation theorem. *preprint* (2016).

[26] A. Dembo, A. Kagan, and L. A. Shepp, "Remarks on the maximum correlation coefficient," *Bernoulli*, vol. 7, no. 2, pp. 343–350, 2001.

[27] P. Diaconis, Steins method for Markov chains: first examples. Lecture Notes–Monograph Series Volume 46, 2004, 26-41.

[28] F. Gotze, On the rate of convergence in the multivariate CLT. *Ann. Probab* 19, 724–739, 1991.

[29] O. Johnson, *Information theory and the central limit theorem.* Vol. 8. London: Imperial College Press, 2004.

[30] O. Johnson and A. Barron, Fisher information inequalities and the central limit theorem, *Probab. Theory Related Fields* 129 (2004), no. 3, 391–409.

[31] S. Kamath and C. Nair, "The strong data processing constant for sums of iid random variables," *Proceedings of the 2015 IEEE International Symposium on Information Theory*, Hong Kong, June 2015.

[32] Lax, P. D.; Milgram, A. N., "Parabolic equations", Contributions to the theory of partial differential equations, *Annals of Mathematics Studies*, 33, Princeton, N. J.: Princeton University Press, pp. 167–190 (1954).

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

[34] Ley, C., Reinert, G., and Swan, Y. Approximate computation of expectations: a canonical Stein operator. *to appear in Probability Surveys* (2017).

[35] M. Madiman and A. R. Barron, "The monotonicity of information in the central limit theorem and entropy power inequalities," *Proceedings of the 2006 IEEE International Symposium on Information Theory*, Seattle, Washington, July 2006.

[36] M. Madiman and A. R. Barron, "Generalized entropy power inequalities and monotonicity properties of information," *IEEE Transactions on Information Theory*, vol. 53, no. 7, pp. 2317–2329, 2007.

[37] M. Madiman and F. Ghassemi, "The entropy power of a sum is fractionally superadditive," *Proceedings of the 2009 IEEE International Symposium on Information Theory*, Seoul, Korea, July 2009.

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

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

[40] I. Nourdin, G. Peccati and A. Réveillac. Multivariate normal approximation using Steins method and Malliavin calculus. *Ann. I.H.P. Proba. Stat.*, 46(1):45–58, 2010

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

[42] I. Nourdin, G. Peccati and Y. Swan, Integration by parts and representation of information functionals. *Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT)*, Honolulu, HI. Pages 2217–2221 (2014).

[43] E. Rio, Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. H. Poincar Probab. Statist.* Vol. 45, 3, 802–817 (2009).

[44] E. Rio, Asymptotic constants for minimal distance in the central limit theorem *Electron. Commun. Probab.* Vol. 16 (2011), no. 9, 96–103.

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

[46] D. Shlyakhtenko, "Shannon's monotonicity problem for free and classical entropy," *Proc. Nat. Acad. Sci.*, vol. 104, no. 39, pp. 15254–15258, 2007.

[47] 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.

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

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

[50] A. M. Tulino and S. Verdú, "Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4295–4297, 2006.

[51] C. Villani, *Topics in optimal transportation.* Graduate Studies in Mathematics Vol. 58; 2003.

[52] C. Villani, *Optimal Transport, old and new.* Grundlehren der mathematischen Wissenschaften, 338, 2009.

[53] A. Zhai, A multivariate CLT in Wasserstein distance with near optimal convergence rate. *Arxiv preprint* (2016).