The strong data processing inequality under the heat flow

B. Klartag, O. Ordentlich [*]

Abstract

Let ν and μ be probability distributions on $\mathbb{R}^n$, and ν_s, μ_s be their evolution under the heat flow, that is, the probability distributions resulting from convolving their density with the density of an isotropic Gaussian random vector with variance s in each entry. This paper studies the rate of decay of $s \mapsto D(\nu_s \| \mu_s)$ for various divergences, including the χ^2 and Kullback-Leibler (KL) divergences. We prove upper and lower bounds on the strong data-processing inequality (SDPI) coefficients corresponding to the source μ and the Gaussian channel. We also prove generalizations of de Brujin's identity, and Costa's result on the concavity in s of the differential entropy of ν_s. As a byproduct of our analysis, we obtain new lower bounds on the mutual information between X and $Y = X + \sqrt{s}Z$, where Z is a standard Gaussian vector in $\mathbb{R}^n$, independent of X, and on the minimum mean-square error (MMSE) in estimating X from Y, in terms of the Poincaré constant of X.

1 Introduction

For two probability distributions ν and μ on $\mathbb{R}^n$, where ν is absolutely continuous with respect to μ, and a smooth, convex function $\varphi : (0, \infty) \to \mathbb{R}$ with $\varphi(1) = 0$, the φ-divergence [17] is defined as

$$D_\varphi(\nu \| \mu) = \mathbb{E}_\mu \left[\varphi \left(\frac{d\nu}{d\mu} \right) \right], \tag{1}$$

where $\varphi(0) = \lim_{t \to 0^+} \varphi(t)$. Let $Z \sim \mathcal{N}(0, \text{Id})$ be a standard Gaussian random variable in $\mathbb{R}^n$. For $s \geq 0$, denote by ν_s the probability distribution of the random variable $Y = X + \sqrt{s}Z$ when $X \sim \nu$ is independent of Z. Similarly, write μ_s for the probability distribution of Y when $X \sim \mu$ is independent of Z. By the data-processing inequality [17, 37] for φ-divergences, we have that

$$s \mapsto D_\varphi(\nu_s \| \mu_s)$$

is non-increasing in $s \geq 0$. The goal of this paper is to provide estimates on the rate of decay of $s \mapsto D_\varphi(\nu_s \| \mu_s)$ as a function of μ and uniformly over the measure ν. We stress that here *both* measures evolve according to the heat flow. This is in contrast to prior work that analyzed the evolution of the entropy and similar functionals of ν_s according to the heat flow. This corresponds to taking μ as the Lebesgue measure whose heat flow evolution satisfies $\mu_s = \mu$. See more details below.

To that end, we study the *strong data-processing inequality (SDPI) constant/contraction coefficient* for the probability distribution μ and the Gaussian channel [1], [37, chapter 33], which is defined as

$$\eta_\varphi(\mu, s) \triangleq \sup_{\nu: 0 < D_\varphi(\nu \| \mu) < \infty} \frac{D_\varphi(\nu_s \| \mu_s)}{D_\varphi(\nu \| \mu)}. \tag{2}$$

Recalling that $D_\varphi(\nu_s \| \mu_s) \geq 0$ (with equality if and only if $\nu_s = \mu_s$), by the data-processing inequality we have that $0 \leq \eta_\varphi(\mu, s) \leq 1$. We mostly analyze two choices of φ. The first is $\varphi(x) = \varphi_{\mathrm{KL}}(x) = x \log x$,[1] which results in the Kullback-Leibler (KL) divergence

$$D(\nu \| \mu) = D_{\mathrm{KL}}(\nu \| \mu) = \int_{\mathbb{R}^n} d\nu \log \frac{d\nu}{d\mu}, \tag{3}$$

and the second is $\varphi(x) = \varphi_{\chi^2}(x) = (x-1)^2$, which results in the χ^2-divergence

$$\chi^2(\nu \| \mu) = D_{\chi^2}(\nu \| \mu) = \int_{\mathbb{R}^n} \left(\frac{d\nu}{d\mu} - 1 \right)^2 d\mu. \tag{4}$$

It is common to rewrite the integrand in (4), by abuse of notation, as $(d\nu - d\mu)^2 / d\mu$, where $d\mu$ refers to the "density of μ" with respect to some ambient measure. We denote the corresponding contraction coefficients by $\eta_{\mathrm{KL}}(\mu, s)$ and $\eta_{\chi^2}(\mu, s)$. By [36, Theorem 2] we have that

$$\eta_{\chi^2}(\mu, s) \leq \eta_\varphi(\mu, s), \tag{5}$$

for any φ with $\varphi''(1) > 0$. In particular, $\eta_{\chi^2}(\mu, s) \leq \eta_{\mathrm{KL}}(\mu, s)$. Intuitively, inequality (5) follows by considering measures ν that are very close to μ, and using Taylor approximation.

Some of our results hold for further choices of φ beyond φ_{χ^2} and φ_{KL}. In particular, the crucial requirement for our results is that $1/\varphi''$ is concave, as in Chafaï [11]. This requirement is fulfilled for φ_{χ^2} and φ_{KL} (in fact for those functions, $1/\varphi''$ is linear). Another important class of functions for which $1/\varphi''$ is concave is $\varphi_\lambda(x) = x^\lambda - 1$, for $\lambda \in (1, 2)$. The corresponding divergence $D_{\varphi_\lambda}(\nu \| \mu)$, defined by (1), is related to the Rényi divergence of order λ, via the monotone transformation

$$D_\lambda(\nu \| \mu) = \frac{1}{\lambda - 1} \log(1 + D_{\varphi_\lambda}(\nu \| \mu)). \tag{6}$$

---

[1]In this paper, logarithms are always taken to the natural basis.

Thus, although $D_\lambda(\nu\|\mu)$ is not a φ-divergence, it does satisfy the data processing inequality, and bounds on the decay rate of $s \mapsto D_{\varphi_\lambda}(\nu_s\|\mu_s)$ immediately imply bounds on the decay rate of the Rényi divergence $s \mapsto D_\lambda(\nu_s\|\mu_s)$. We note that the concavity requirement on $1/\varphi''$ fails to hold for many other important φ-divergences, such as the Jensen-Shannon divergence ($\varphi(x) = x \log \frac{2x}{x+1} + \log \frac{2}{x+1}$), the Le Cam divergence ($\varphi(x) = (1-x)/(2x+2)$), the total variation divergence ($\varphi(x) = |x-1|/2$) and the Hellinger divergence ($\varphi(x) = (1-\sqrt{x})^2$).

It is well-known, see. e.g., [37, Theorem 33.12] that

$$\sqrt{\eta_{\chi^2}(\mu, s)} = S(X, X + \sqrt{s}Z) \tag{7}$$

where $X \sim \mu$ is independent of the standard Gaussian Z. Here $S(X, Y)$ is the *Hirschfeld-Gebelein-Rényi maximal correlation* [39] between the random variables X and Y defined as

$$S(X, Y) = \sup_{f, g : \mathbb{R}^n \to \mathbb{R}} \frac{\mathbb{E}f(X)g(Y) - \mathbb{E}(f(X))\mathbb{E}(g(Y))}{\sqrt{\mathrm{Var}(f(X))\mathrm{Var}(g(Y))}} \tag{8}$$

$$= \sup_f \sqrt{\mathrm{Var}(\mathbb{E}[f(X)|Y])}. \tag{9}$$

The supremum in (9) runs over all measurable functions $f : \mathbb{R}^n \to \mathbb{R}$ with $\mathbb{E}f(X) = 0$ and $\mathrm{Var}(f(X)) = 1$. In the passage from (8) to (9) we use the Cauchy-Schwartz inequality.

The quantities $\eta_{\mathrm{KL}}(\mu, s)$ and $\eta_{\chi^2}(\mu, s)$ can be thought of as (normalized) measures of statistical dependence between $X \sim \mu$ and $Y = X + \sqrt{s}Z$. Two other (un-normalized) measures of dependence we will consider are the minimum mean-square error (MMSE)[2]

$$\mathrm{mmse}(\mu, s) = \mathbb{E}|X - \mathbb{E}[X|X + \sqrt{s}Z]|^2, \tag{10}$$

and the mutual information

$$I(\mu, s) = I(X; X + \sqrt{s}Z) = h(X + \sqrt{s}Z) - h(X + \sqrt{s}Z|X) = h(\mu_s) - \frac{n}{2}\log(2\pi es),$$

where for a probability distribution μ with density ρ in $\mathbb{R}^n$, the differential entropy of $X \sim \mu$ is

$$h(X) = h(\mu) = -\int_{\mathbb{R}^n} \rho(x) \log \rho(x) dx. \tag{11}$$

Our results will be expressed in terms of the Poincaré constant $C_\mathrm{P}(X) = C_\mathrm{P}(\mu)$, of $X \sim \mu$, and the log-Sobolev constant $C_\mathrm{LS}(X) = C_\mathrm{LS}(\mu)$. The *Poincaré constant* of a random vector X in $\mathbb{R}^n$, denoted by $C_P(X)$, is the infimum over all $C \geq 0$ such that for any locally-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying $\mathbb{E}|\nabla f(X)|^2 < \infty$,

$$\mathrm{Var}(f(X)) \leq C \cdot \mathbb{E}|\nabla f(X)|^2. \tag{12}$$

---

[2]For a vector $z = (z_1, \ldots, z_n) \in \mathbb{R}^n$ we denote by $|z| = \|z\|_2 = \sqrt{\sum_j z_j^2}$ the ℓ_2 norm of z.

The *log-Sobolev constant* of a random vector X in $\mathbb{R}^n$, denoted by $C_{\mathrm{LS}}(X)$, is the infimum over all $C \geq 0$ such that for any locally-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying $\mathbb{E}f^2(X) = 1$ and $\mathbb{E}|\nabla f(X)|^2 < \infty$,

$$\mathbb{E}f^2(X) \log f^2(X) \leq 2C \cdot \mathbb{E}|\nabla f(X)|^2. \tag{13}$$

We list a few important properties of the Poincaré and log-Sobolev constants, see [5] for more explanations, proofs and references.

- The Poincaré constant and the log-Sobolev constant of the standard Gaussian random vector $Z \in \mathbb{R}^n$ satisfy $C_{\mathrm{P}}(Z) = C_{\mathrm{LS}}(Z) = 1$. Furthermore, for any random variable we have that

$$\frac{C_{\mathrm{P}}(X)}{\frac{1}{n}\mathbb{E}|X - \mathbb{E}X|^2} \geq 1 \qquad \text{and} \qquad \frac{C_{\mathrm{LS}}(X)}{\frac{1}{n}\mathbb{E}|X - \mathbb{E}X|^2} \geq 1.$$

and in both cases the lower bound is attained if and only if X is an isotropic Gaussian random variable.

- $C_{\mathrm{P}}(\mu) \leq C_{\mathrm{LS}}(\mu)$.

- For $p \geq 1$, let $\bar{\mathcal{B}}_p = \{x \in \mathbb{R}^n \; : \; \|x\|_p < \alpha\}$, where α is chosen such that $\mathrm{vol}(\bar{\mathcal{B}}_p) = 1$, be the unit-volume ℓ_p-ball. For $U_p \sim \mathrm{Uniform}(\bar{\mathcal{B}}_p)$, we have that

$$c < C_{\mathrm{P}}(U_p) < C \tag{14}$$

for certain explicit universal constants $c, C > 0$. For $p \in [2, +\infty]$ we furthermore have

$$c < C_{\mathrm{LS}}(U_p) < C \tag{15}$$

for universal constants $c, C > 0$. The log-Sobolev constant is not bounded by a universal constant when $p < 2$. These results may be extracted from the literature in the following way: First, up to a universal constant, Poincaré and log-Sobolev inequalities follows from a corresponding isoperimetric inequality (see Ledoux [32] for the log-Sobolev case and Cheeger [12] for the Poincaré case). Second, the corresponding isoperimetric inequalities were proven in Sodin [42] ($1 \leq p \leq 2$) and Latała and Wojtaszczyk [31].

- $C_{\mathrm{P}}(X)$ is finite for any log-concave random vector X in $\mathbb{R}^n$, see Bobkov [9].

- When X is a log-concave random vector in $\mathbb{R}^n$, the Kanann-Lovász-Simonovits (KLS) conjecture from [27] suggests that

$$\|\mathrm{Cov}(X)\|_{op} \leq C_{\mathrm{P}}(X) \leq C \cdot \|\mathrm{Cov}(X)\|_{op}, \tag{16}$$

where $C > 0$ is a universal constant. Here, $\mathrm{Cov}(X) \in \mathbb{R}^{n \times n}$ is the covariance matrix of X, whose i, j entry is $\mathbb{E}X_i X_j - \mathbb{E}X_i \mathbb{E}X_j$. We write $\|\cdot\|_{op}$ for the operator norm, thus[3]

$$\|\mathrm{Cov}(X)\|_{op} = \sup_{|\theta|=1} \mathrm{Var}(X \cdot \theta) \tag{17}$$

---

[3]We denote the standard inner product of two vectors $x, y \in \mathbb{R}^n$ by $x \cdot y = \langle x, y \rangle = x^T y$.

where the supremum runs over all unit vectors $\theta \in \mathbb{R}^n$. The left-hand side inequality in (16) follows from the definitions (12) and (17). The current state of the art regarding the conjectural right-hand side inequality in (16) is the bound from [28]

$$C_{\mathrm{P}}(X) \leq C \log n \cdot \|\mathrm{Cov}(X)\|_{op}, \tag{18}$$

where $C > 0$ is a universal constant and X is an arbitrary log-concave random variable in $\mathbb{R}^n$.

We recall that an absolutely continuous probability distribution μ on $\mathbb{R}^n$, or a random variable $X \sim \mu$ is log-concave if it is supported in an open, convex set $K \subseteq \mathbb{R}^n$ (which could be $\mathbb{R}^n$ itself) and has a density of the form e^{-H} in K where H is a convex function. Thus a Gaussian measure is log-concave, as well as the uniform measure on an arbitrary convex body.

1.1 Main results on χ^2 divergence

For $\eta_{\chi^2}(\mu, s)$, we prove the following results.

Theorem 1.1. *For any probability distribution μ and $s > 0$*

$$1 - \frac{\mathrm{mmse}(\mu, s)}{\mathbb{E}_\mu |X - \mathbb{E}X|^2} \leq \eta_{\chi^2}(\mu, s) = S^2(X, X + \sqrt{s}Z) \leq \frac{1}{1 + \frac{s}{C_{\mathrm{P}}(\mu)}}. \tag{19}$$

If μ is additionally log-concave, then,

$$\eta_{\chi^2}(\mu, s) \geq e^{-s/C_P(\mu)}. \tag{20}$$

Theorem 1.2. *For any log-concave probability distribution μ and function $f : \mathbb{R}^n \to \mathbb{R}$, the function $s \mapsto \log \mathrm{Var}(\mathbb{E}[f(X)|X + \sqrt{s}Z])$ is convex. In particular, whenever μ is log-concave, $s \mapsto \log S(X, X + \sqrt{s}Z)$ is convex, and consequently, $s \mapsto \eta_{\chi^2}(\mu, s)$ is also convex.*

Inequality (20) as well as Theorem 1.2 are mostly an interpretation of the results of [29]. Theorem 1.1 implies that for log-concave μ the rate at which $s \mapsto \eta_{\chi^2}(\mu, s)$ decreases with s is largely determined by $C_{\mathrm{P}}(\mu)$. In particular, if $s_\mu^{\chi^2}(\alpha) = \min\{s : \eta_{\chi^2}(\mu, s) \leq \alpha\}$ is the required time for $\eta_{\chi^2}(\mu, s)$ to drop below $\alpha \in (0, 1)$, then

$$C_{\mathrm{P}}(\mu) \cdot \log \frac{1}{\alpha} \leq s_\mu^{\chi^2}(\alpha) \leq C_{\mathrm{P}}(\mu) \cdot \left(\frac{1}{\alpha} - 1\right). \tag{21}$$

For α close to 1, the upper bound and lower bound nearly coincide. We refer to the quantity $s_\mu^{\chi^2}(\alpha)$ for $\alpha = 1/2$ as the "χ^2 half-blurring time" of the measure μ. In the log-concave case, the χ^2 half-blurring time of μ has the order of magnitude of $C_P(\mu)$, up to an explicit multiplicative universal constant. A Corollary of Theorem 1.1 is the following:

Corollary 1.3.

$$\mathrm{mmse}(\mu, s) \geq \frac{\mathbb{E}_\mu |X - \mathbb{E}X|^2}{1 + \frac{C_\mathrm{P}(\mu)}{s}} = \frac{nP}{1 + \frac{P}{s} \cdot \bar{C}_\mathrm{P}(\mu)} \tag{22}$$

where

$$P = \frac{1}{n} \mathbb{E}_\mu |X - \mathbb{E}X|^2, \tag{23}$$

and

$$\bar{C}_\mathrm{P}(\mu) = \frac{C_\mathrm{P}(\mu)}{P}. \tag{24}$$

Let us compare Corollary 1.3 with known bounds in Estimation Theory. Assume that μ has smooth density ρ and recall that the Bayesian Cramér-Rao lower bound (CRLB)/multivariate van Trees inequality [22, eq.10] for the additive isotropic Gaussian noise case, gives

$$\mathrm{mmse}(\mu, s) \geq \frac{n^2 s}{n + s\mathcal{J}(\mu)} = \frac{nP}{\bar{\mathcal{J}}(\mu) + \frac{P}{s}}, \tag{25}$$

where

$$\mathcal{J}(\mu) = \int_{x \in \mathbb{R}^n} \frac{|\nabla \rho(x)|^2}{\rho(x)} dx. \tag{26}$$

and

$$\bar{\mathcal{J}}(\mu) = \frac{\mathcal{J}(\mu) \cdot P}{n}. \tag{27}$$

We see that the lower bound from Corollary 1.3 is tighter than the Bayesian CRLB whenever $\frac{P}{s}(\bar{C}_\mathrm{P}(\mu) - 1) < \bar{\mathcal{J}}(\mu) - 1$. In particular, whenever $\bar{C}_\mathrm{P}(\mu)$ is finite, the bound from Corollary 1.3 is tighter than the Bayesian CRLB for low enough signal-to-noise ratio (SNR) $\frac{P}{s}$. There are also cases where $\bar{\mathcal{J}}(\mu)$ is infinite, rendering the Bayesian CRLB useless for all $s > 0$, whereas $\bar{C}_\mathrm{P}(\mu)$ is finite. An example of such probability distribution is the uniform distribution over a convex set. We remark that there are also examples where $\mathcal{J}(\mu)$ is finite and $C_\mathrm{P}(\mu)$ infinite, such as the density $t \mapsto e^{-(t^2+1)^\alpha}$ on the real line for $0 < \alpha < 1/2$.

Using the I-MMSE identity of Guo, Shamai and Verdù [24, Theorem 2], we can deduce the following.

Corollary 1.4.

$$I(\mu, s) = I(X; X + \sqrt{s}Z) \geq \frac{n}{2} \cdot \frac{1}{\bar{C}_\mathrm{P}(\mu)} \log\left(1 + \frac{P}{s} \cdot \bar{C}_\mathrm{P}(\mu)\right), \tag{28}$$

where P and $\bar{C}_\mathrm{P}(\mu)$ are as defined in (23) *and* (24)*, respectively.*

It is easy to verify that the function $t \mapsto \frac{1}{t} \log(1 + \frac{P}{s} \cdot t)$ is decreasing. Thus, the lower bound above is monotonically decreasing in $\bar{C}_{\mathrm{P}}(\mu)$, which in turn is minimized for the isotropic Gaussian distribution, for which its value is 1. Thus, the lower bound is tight for the isotropic Gaussian distribution, and controls the mutual information loss of the probability distribution μ from the upper bound $\frac{n}{2} \log(1 + \frac{P}{s})$ via the distance from Gaussianity measure $\bar{C}_{\mathrm{P}}(\mu) \geq 1$.

It is insightful to compare the lower bound from Corollary 1.4 to the standard entropy power inequality (EPI)-based [16, Theorem 17.7.3] lower bound

$$I(\mu, s) \geq \frac{n}{2} \log \left(1 + \frac{P}{s} \cdot \bar{N}(\mu) \right). \tag{29}$$

where

$$\bar{N}(\mu) = \frac{e^{\frac{2}{n} h(\mu)}}{2 \pi e \cdot P} \tag{30}$$

is the normalized entropy power (recall that $\bar{N}(\mu) \leq 1$ with equality iff μ is Gaussian isotropic). It is easy to see that whenever $\bar{C}_{\mathrm{P}}(\mu) < \infty$, the new lower bound from Corollary 1.4 is tighter than the EPI-based bound (29) for s large enough (low SNR).

Note that using the Bayesian CRLB (25) and the I-MMSE identity, one easily obtains the lower bound

$$I(\mu, s) \geq \frac{n}{2} \log \left(1 + \frac{P}{s} \cdot \frac{1}{\bar{\mathcal{J}}(\mu)} \right). \tag{31}$$

However, since for any probability distribution μ we have that $\bar{\mathcal{J}}(\mu) \cdot \bar{N}(\mu) \geq 1$, with equality iff μ is the isotropic Gaussian distribution [10], [19, Theorem 16], this bound is subsumed by the EPI-based bound (29).

1.2 Main results on KL-divergence

We now move on to present our results concerning the KL divergence. Our first result is a generalization of de Bruijn's identity.

Proposition 1.5. *Let μ and ν be probability measures on $\mathbb{R}^n$ with $D(\nu \,||\, \mu) < \infty$, and assume $\mathbb{E}_\mu |X|^4 < \infty$. Then for any $s > 0$,*

$$\frac{d}{ds} D(\nu_s \,||\, \mu_s) = -\frac{1}{2} J(\nu_s \,||\, \mu_s), \tag{32}$$

where the Fisher information of a probability distribution ν relative to μ is, with $f = d\nu/d\mu$,

$$J(\nu \,||\, \mu) = \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} d\mu. \tag{33}$$

Proposition 1.5 easily implies de Bruijn's identity.

Some of the proceeding results rely on the correctness of a certain conjecture, namely Conjecture 4.2, stated in Section 4. Calling it a conjecture is perhaps too modest: we provide an almost full of proof for Conjecture 4.2. The only gap between our "proof" and a fully rigorous proof is the justification of order exchange between integration and differentiation, as well as justifying several integration by parts (that is, showing that the involved functions vanish at infinity). While it is not trivial to rigorously justify these steps, we are quite certain they are correct under some minimal regularity assumptions. In fact, the justification of such steps is often ignored in the literature. For example, to the best of our knowledge, the first time a fully rigorous proof was given for de Bruijn's identity was by Barron in 1984 [6], while prior proofs, such as that of Blachman [8] did not justify the exchange of differentiation and integration. Whenever a statement relies on the correctness of Conjecture 4.2, we explicitly emphasize it.

We prove the following bounds for the KL contraction coefficient:

Theorem 1.6. *For any probability measure μ and $s > 0$,*

$$1 - \frac{\mathrm{mmse}(\mu, s)}{\mathbb{E}_\mu |X - \mathbb{E}X|^2} \le \eta_{KL}(\mu, s) \le \frac{1}{1 + \frac{s}{C_{\mathrm{LS}}(\mu)}}. \tag{34}$$

If additionally μ is log-concave, then, assuming the validity of Conjecture 4.2, we also have

$$1 - \frac{s}{C_{\mathrm{LS}}(\mu)} \le \eta_{KL}(\mu, s). \tag{35}$$

The left-hand side inequality in (34) immediately follows from (5) and (19).

Theorem 1.7. *Assuming validity of Conjecture 4.2, for any probability distribution ν and any log-concave probability distribution μ, the mapping $s \mapsto D(\nu_s \| \mu_s)$ is convex. In particular, assuming validity of Conjecture 4.2, whenever μ is log-concave, $s \mapsto \eta_{KL}(\mu, s)$ is convex*

Theorem 1.7 implies that $s \mapsto h(\nu_s)$ is concave, which is [13, Corollary 1]. Thus, Theorem 1.7 is stronger than Costa'a Corollary about concavity of $s \mapsto h(\nu_s)$. On the other hand, Theorem 1.7 does not imply the concavity of $s \mapsto \exp\left(\frac{2}{n} h(\nu_s)\right)$, which is Costa's main result in [13].

As in our analysis for $\eta_{\chi^2}(s, \mu)$, we see that Theorem 1.6 implies that for log-concave μ the rate at which $s \mapsto \eta_{\mathrm{KL}}(\mu, s)$ decreases with s is largely determined by $C_{\mathrm{LS}}(\mu)$. In particular, if $s_\mu^{\mathrm{KL}}(\alpha) = \min\{s : \eta_{\mathrm{KL}}(\mu, s) \le \alpha\}$ is the required time for $\eta_{\mathrm{KL}}(\mu, s)$ to drop below $\alpha \in (0, 1)$, then

$$(1 - \alpha) \cdot C_{\mathrm{LS}}(\mu) \le s_\mu^{\mathrm{KL}}(\alpha) \le \left(\frac{1 - \alpha}{\alpha}\right) C_{\mathrm{LS}}(\mu). \tag{36}$$

For α close to 1, the upper bound and lower bound nearly coincide. We refer to the quantity $s_\mu^{\mathrm{KL}}(\alpha)$ for $\alpha = 1/2$ as the "KL half-blurring time" of the measure μ. In the log-concave case, the KL half-blurring time of μ has the order of magnitude of $C_{\mathrm{LS}}(\mu)$, up to an explicit multiplicative universal constant.

8

1.3 Main results on general divergences

Fix a smooth, convex function $\varphi : (0, \infty) \to \mathbb{R}$ with $\varphi(1) = 0$ and set $\varphi(0) = \lim_{t \to 0^+} \varphi(t)$. In [11], Chafaï introduced the concepts of φ-Sobolev inequalities and φ-entropy, which turn out to fit nicely with our study of the evolution of the φ-divergence $D_\varphi(\nu_s \, || \, \mu_s)$.

The φ-Sobolev constant of a random vector X in $\mathbb{R}^n$, denoted by $C_\varphi(X)$, is the infimum over all $C \geq 0$ such that for any locally-Lipschitz function $f : \mathbb{R}^n \to (0, \infty)$ with $\mathbb{E}\varphi(X) = 1$,

$$\mathbb{E}\varphi(f(X)) \leq \frac{C}{2} \cdot \mathbb{E}\varphi''(f(X))|\nabla f(X)|^2. \tag{37}$$

This generalized the definition of the Poincaré constant and the log-Sobolev constant. Indeed, in the case where $\varphi(x) = x \log x$ is the KL-divergence, we have $C_\varphi(X) = C_{\mathrm{LS}}(X)$, and a simple argument reveals that when $\varphi(x) = (x-1)^2$ we have $C_\varphi(X) = C_{\mathrm{P}}(X)$. When $X \sim \mu$ we set $C_\varphi(\mu) = C_\varphi(X)$. In the general case, our definition (37) is slightly different from the one in Chafaï in that we impose the additional requirement that $\mathbb{E}f(X) = 1$, which implies that $\varphi(\mathbb{E}f(X)) = 0$. In Chafaï the left-hand side of (37) is replaced by $\mathbb{E}\varphi(f(X)) - \varphi(\mathbb{E}f(X))$. In the cases of interest just described, the two definitions coincide by homogeneity.

Given two probability distributions μ and ν on $\mathbb{R}^n$, with ν absolutely continuous with a smooth density with respect to μ, we defined the φ-Fisher information via

$$J_\varphi(\nu \, || \, \mu) = \int_{\mathbb{R}^n} \varphi''(f)|\nabla f|^2 d\mu,$$

where $f = d\nu/d\mu$. Thus we have the defining inequality of the φ-Sobolev constant,

$$D_\varphi(\nu \, || \, \mu) \leq \frac{C_\varphi(\mu)}{2} \cdot J_\varphi(\nu \, || \, \mu). \tag{38}$$

Some of the properties of the Poincaré and log-Sobolev constants generalize to the more general context of the φ-Sobolev constant. The φ-Sobolev constant is 2-homogeneous: For any fixed number $\lambda > 0$,

$$C_\varphi(\lambda X) = \lambda^2 C_\varphi(X).$$

This follows from the definition (37), by substituting $g(x) = f(\lambda x)$. The Poincaré constant is the smallest of all φ-Sobolev constants:

Proposition 1.8. *Assume that $\varphi''(1) > 0$. Then for any random vector X for which $C_\varphi(X) < \infty$,*

$$C_P(X) \leq C_\varphi(X).$$

Proposition 1.8 must be known, yet for lack of concise reference we provide its proof below. The function $1/\varphi''$ plays an important role in the sequel. When it is a concave function on $(0, \infty)$ – this is equivalent to condition (H1) from Chafaï [11] – it is possible to compute the φ-Sobolev constant of the Gaussian measure. The following proposition follows from the results of Chafaï [11] and Proposition 1.8. For convenience we provide a proof.

Proposition 1.9. *Let $\varphi : (0, \infty) \to \mathbb{R}$ be a smooth, convex function with $\varphi(1) = 0$ such that $1/\varphi''$ is concave. Let X be a standard Gaussian random vector in $\mathbb{R}^n$. Then,*

$$C_\varphi(X) = 1.$$

Proposition 1.5, the de Bruijn identity, may be generalized even further:

Proposition 1.10. *Let $\varphi(x) = \varphi_\lambda(x) = x^\lambda - 1$, for $\lambda > 1$. Let μ and ν be probability distributions on $\mathbb{R}^n$ with $D_\varphi(\nu \parallel \mu) < \infty$, and assume $\mathbb{E}_\mu |X|^4 < \infty$. Then for any $s > 0$,*

$$\frac{d}{ds} D_\varphi(\nu_s \parallel \mu_s) = -\frac{1}{2} J_\varphi(\nu_s \parallel \mu_s). \tag{39}$$

This proposition formally implies Proposition 1.5 since $D_{\mathrm{KL}}(\nu \| \mu) = \lim_{\lambda \to 1} \frac{1}{\lambda - 1} \log(1 + D_{\varphi_\lambda}(\nu \parallel \mu))$. While we have only proved Proposition 1.10 for $\varphi(x) = \varphi_\lambda(x)$, $\lambda > 1$, we believe it should be valid for any smooth, convex function $\varphi : (0, \infty) \to \mathbb{R}$ with $\varphi(1) = 0$. We have only used the assumption that $\varphi(x) = \varphi_\lambda(x)$, $\lambda > 1$, for justifying integration under the integral sign, and for justifying integration by parts. The following is a generalization of Theorem 1.6.

Theorem 1.11. *Let $\varphi(x) = \varphi_\lambda(x) = x^\lambda - 1$, for $1 < \lambda \leq 2$, such that $1/\varphi''$ is concave. Then, for any probability measure μ and $s > 0$,*

$$1 - \frac{\mathrm{mmse}(\mu, s)}{\mathbb{E}_\mu |X - \mathbb{E}X|^2} \leq \eta_\varphi(\mu, s) \leq \frac{1}{1 + \frac{s}{C_\varphi(\mu)}}. \tag{40}$$

If additionally μ is log-concave, then, assuming the validity of Conjecture 4.2, we also have that

$$1 - \frac{s}{C_\varphi(\mu)} \leq \eta_\varphi(\mu, s). \tag{41}$$

The left-hand side inequality in (40) immediately follows from (5) and (19). We proceed with a generalization of Theorem 1.7.

Theorem 1.12. *Let $\varphi(x) = \varphi_\lambda(x) = x^\lambda - 1$, for $1 < \lambda \leq 2$, such that $1/\varphi''$ is concave, such that $1/\varphi''$ is concave. Then, assuming the validity of Conjecture 4.2, for any probability measure ν and any log-concave probability measure μ, the mapping $s \mapsto D_\varphi(\nu_s \| \mu_s)$ is convex. In particular, whenever μ is log-concave, $s \mapsto \eta_\varphi(\mu, s)$ is convex.*

For the geometric meaning of the φ-Sobolev constant, and for the relations to concentration of measure, we refer the reader to [11, Section 3.4].

1.4 Related Work

Contraction coefficients for the Gaussian channel were studied in [35] (see also [21]), and it was shown that there exist probability distributions μ with bounded second moment $\mathbb{E}_\mu |X - \mathbb{E}X|^2$ for which $\eta_{\chi^2}(\mu, s) = \eta_{\mathrm{KL}}(\mu, s) = 1$. Our results show that while a bounded second moment does not imply non-trivial $\eta_{\chi^2}(\mu, s)$ or $\eta_{\mathrm{KL}}(\mu, s)$, bounded Poincarè constant $C_{\mathrm{P}}(\mu)$ or log-Sobolev constant $C_{\mathrm{LS}}(\mu)$, respectively, do imply non-trivial contraction coefficient. Moreover, in the log-concave case the Poincarè constant or log-Sobolev constant are essentially equivalent to the corresponding contraction coefficient.

For discrete channels, Raginsky [38] has estimated the SDPI constants $\eta_{\chi^2}(\mu, K_{Y|X})$ and $\eta_{\mathrm{KL}}(\mu, K_{Y|X})$, as well as for other choices of φ, of a source μ and a general discrete channel $K_{Y|X}$ as a function of a Poincarè constant $C_{\mathrm{P}}(\mu, K_{Y|X})$ or a log-Sobolev constant $C_{\mathrm{LS}}(\mu, K_{Y|X})$, respectively, corresponding to *both* the source μ and *the channel* $K_{Y|X}$. See [38] for the precise definitions of those constants and the relations to the contraction coefficients. We stress that the Poincarè constant $C_{\mathrm{P}}(\mu)$ and log-Sobolev constant $C_{\mathrm{LS}}(\mu)$ used in our bounds depend only on the source μ. Furthermore, our bounds only address the Gaussian channels, while the estimates in [38] hold for any discrete channel (though this class does not include the Gaussian channel).

The general problem of developing lower bounds, amenable to evaluation, on the MMSE in estimating a random variable X from a dependent random variable Y is a classic topic in information theory, signal processing, and statistics, and so is the special case of estimation in Gaussian noise. Classic references include [46, 44, 48, 50]. See also [37, Chapter 30]. Our lower bound from Corollary 1.3 varies smoothly with the noise level s, similarly to the Bayesian CRLB, and therefore, cannot capture threshold phenomena/phase transitions of the MMSE in s. However, to the best of our knowledge, no previous bounds have estimated $\mathrm{mmse}(\mu, s)$ in terms of the Poincarè constant $C_{\mathrm{P}}(\mu)$. A recent work by Zieder, Dytso and Cardone [49] develops a lower bound on the MMSE for a class of additive channels from X to Y, including the Gaussian one. Their bound is expressed in terms of the random variable $\kappa(X)$, where $\kappa(x) = (C_{\mathrm{P}}(W_{Y|X=x}))^{-1}$ is the inverse of the Poincarè constant corresponding to the conditional distributions of the channel's output given the input x, as well as the variance of the conditional information density between X and Y. Our bound from Corollary 1.3 (which holds only for the Gaussian channel), on the other hand, depends *only* on Poincarè constant of the *source* $C_{\mathrm{P}}(\mu)$.

The function $s \mapsto h(X + \sqrt{s}Z)$ for isotropic Gaussian Z statistically independent of $X \sim \mu$, has been studied intensively, from the first days of information theory. de Bruijn's identity [43, 13] (see also [16, Theorem 17.7.2]) shows that the derivative of this function with respect to s is $\frac{1}{2}\mathcal{J}(\mu_s)$. As explained here, our Proposition 1.5 is a stronger and more general identity. de Bruijn's identity was instrumental for proving the entropy power inequality [43, 8]. Decades later, Costa used de Bruijn's identity for showing that $s \mapsto e^{\frac{2}{n}h(X+\sqrt{s}Z)}$ is concave, which immediately also implies the concavity of $s \mapsto h(X + \sqrt{s}Z)$. Several alternative proofs for Costa's EPI were given in [20, 47, 25, 40, 14]. An important point of view on the evolution of curvature under the heat flow, involving curvature and dimension was developed by Bakry, Émery and the Toulouse school, see the book [5].

In Theorem 1.7 we establish the convexity of $s \mapsto D(\nu_s \| \mu_s)$ for log-concave μ, which gen-

eralizes the latter, and weaker, statement of Costa. Unfortunately, we were not able to prove convexity of $s \mapsto \log D(\nu_s \| \mu_s)$. Further improvements of the entropy power inequality for the random variable $X + \sqrt{s}Z$ were established by Courtade [14]. The relation between $I(\mu, s)$ and $\mathrm{mmse}(\mu, s)$, called the I-MMSE relation, which is intimately related to de Bruijn's identity, was discovered by Guo, Shamai and Verdú [24]. To the best of our knowledge, our Corollary 1.4 is the first lower bound on $I(\mu, s)$ in terms of $C_{\mathrm{P}}(\mu)$. An upper bound in a somewhat similar spirit was derived in [3]. Inequalities interpolating between the Poincaré and the log-Sobolev inequality appear in Latała and Oleszkiewicz [30]. This is a family of inequalities parameterized by a parameter $1 \leq p \leq 2$.

In the remainder of this paper, we prove the theorems stated above.

2 Minimum-squared error and mutual information

2.1 Proof of the lower bound in Theorem 1.1

We show that for any pair of random variables X, Y in $\mathbb{R}^n \times \mathcal{A}$, where $\mathcal{A}$ is some abstract alphabet, it holds that

$$S^2(X, Y) \geq 1 - \frac{\mathrm{mmse}(X|Y)}{\mathbb{E}|X - \mathbb{E}X|^2}, \tag{42}$$

where

$$\mathrm{mmse}(X|Y) = \mathbb{E}|X - \mathbb{E}[X|Y]|^2. \tag{43}$$

The idea is to use linear test functions in the definition (9) of the maximal correlation $S(X, Y)$. For the case $n = 1$, the lower bound was already observed by Reǹyi [39], see also [4, eq. 17]. Here we give a proof for the general case. Let

$$i^* = \underset{i \in [n]}{\mathrm{argmax}} \, \frac{\mathrm{Var}(\mathbb{E}[X_i|Y])}{\mathrm{Var}(X_i)}, \tag{44}$$

and set $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ as

$$\tilde{f}(X) = \frac{X_{i^*} - \mathbb{E}X_{i^*}}{\sqrt{\mathrm{Var}(X_{i^*})}},$$

such that $\mathbb{E}\tilde{f}(X) = 0$ and $\mathrm{Var}(\tilde{f}(X)) = 1$. Since for non-negative numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$ it holds that $\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \leq \max_{i \in [n]} \frac{a_i}{b_i}$, we have that

$$
\mathrm{Var}(\mathbb{E}[\tilde{f}(X)|Y]) = \frac{\mathrm{Var}(\mathbb{E}[X_{i^*}|Y])}{\mathrm{Var}(X_{i^*})}
$$

$$
\geq \frac{\sum_{i=1}^n \mathrm{Var}(\mathbb{E}[X_i|Y]}{\sum_{i=1}^n \mathrm{Var}(X_i)} \tag{45}
$$

$$
= \frac{\sum_{i=1}^n \mathrm{Var}(X_i) - \mathbb{E}(\mathrm{Var}[X_i|Y])}{\sum_{i=1}^n \mathrm{Var}(X_i)} \tag{46}
$$

$$
= 1 - \frac{\mathrm{mmse}(X|Y)}{\mathbb{E}_\mu|X - \mathbb{E}[X]|^2}. \tag{47}
$$

The statement follows by invoking (9)

$$
S^2(X;Y) = \sup_f \mathrm{Var}(\mathbb{E}[f(X)|Y]) \geq \mathrm{Var}(\mathbb{E}[\tilde{f}(X)|Y]). \tag{48}
$$

2.2   Proof of Corollary 1.4

Let

$$
\tilde{I}(\mu, \rho) = I(X; \sqrt{\rho}X + Z), \tag{49}
$$

$$
\widetilde{\mathrm{mmse}}(\mu, \rho) = \mathbb{E}|X - \mathbb{E}[X|\sqrt{\rho}X + Z]|^2. \tag{50}
$$

Clearly, $I(\mu, s) = \tilde{I}(\mu, \frac{1}{s})$ and $\mathrm{mmse}(\mu, s) = \widetilde{\mathrm{mmse}}(\mu, \frac{1}{s})$. By Corollary 1.3, we therefore have that

$$
\widetilde{\mathrm{mmse}}(\mu, \rho) = \mathrm{mmse}\left(\mu, \frac{1}{\rho}\right) \geq \frac{nP}{1 + \rho\bar{C}_\mathrm{P}(\mu)P}. \tag{51}
$$

By the I-MMSE identity [24, Theorem 2],

$$
\frac{d}{d\rho}\tilde{I}(\mu, \rho) = \frac{1}{2}\widetilde{\mathrm{mmse}}(\mu, \rho). \tag{52}
$$

Thus,

$$
\tilde{I}(\mu, \rho) = \frac{1}{2}\int_0^\rho \widetilde{\mathrm{mmse}}(\mu, t)dt \geq \frac{1}{2}\int_0^\rho \frac{nP}{1 + t\bar{C}_\mathrm{P}(\mu)P}dt = \frac{n}{2}\frac{1}{\bar{C}_\mathrm{P}(\mu)}\log\left(1 + \bar{C}_\mathrm{P}(\mu)P \cdot \rho\right). \tag{53}
$$

where in the last equality we have used the identity $\int_0^\rho \frac{a}{1+bt}dt = \frac{a}{b}\log(1 + b\rho)$ for $a, b > 0$. The claimed result follows by recalling that $I(\mu, s) = \tilde{I}(\mu, \frac{1}{s})$.

13

3 χ^2-divergence and proofs for blurring time bounds

In this section we complete the proof of Theorem 1.1. Let X be a random vector attaining values in $\mathbb{R}^n$. When proving Theorem 1.1 we may assume that $C_P(X) < \infty$ as otherwise the conclusion is vacuous. Recall from above that a log-concave random vector has a finite Poincaré constant, and that for any additional random vector Y,

$$S^2(X, Y) = \sup_f \mathrm{Var}(\mathbb{E}[f(X)|Y]) \tag{54}$$

where the supreumum runs over all measurable functions $f : \mathbb{R}^n \to \mathbb{R}$ with $\mathbb{E}f(X) = 0$ and $\mathrm{Var}(f(X)) = 1$. The requirement that $\mathbb{E}f(X) = 0$ is actually not necessary. Let Z be a standard Gaussian random vector in $\mathbb{R}^n$, independent of X. Recall that we are interested in the χ^2 contraction coefficient

$$\eta_{\chi^2}(\mu, s) = S^2(X, X + \sqrt{s}Z).$$

Following Klartag and Putterman [29], for a function $f : \mathbb{R}^n \to \mathbb{R}$ with $\mathbb{E}|f(X)| < \infty$ and for $s > 0$ we write

$$Q_s f(X + \sqrt{s}Z) = \mathbb{E}[f(X)|X + \sqrt{s}Z].$$

Let μ be the probability distribution on $\mathbb{R}^n$ that is the distribution law of the random vector X, and let μ_s be the distribution law of $X + \sqrt{s}Z$. The operator

$$Q_s : L^2(\mu) \to L^2(\mu_s)$$

is of norm at most one, since

$$\int_{\mathbb{R}^n} |Q_s f|^2 d\mu_s = \mathbb{E}\left|Q_s f(X + \sqrt{s}Z)\right|^2 = \mathbb{E}\left|\mathbb{E}[f(X)|X + \sqrt{s}Z]\right|^2 \leq \mathbb{E}|f(X)|^2.$$

Lemma 3.1. *For any $s > 0$, the quantity $S^2(X, X + \sqrt{s}Z)$ is the square of the operator norm of $Q_s : L^2(\mu) \to L^2(\mu_s)$ restricted to the subspace of functions of μ-average zero. In other words, $S^2(X, X + \sqrt{s}Z)$ is the minimal number $M \geq 0$ such that for any $f \in L^2(\mu)$ with $\int f d\mu = 0$,*

$$\|Q_s f\|_{L^2(\mu_s)}^2 \leq M \cdot \|f\|_{L^2(\mu)}^2. \tag{55}$$

Proof. By (54), for any $s > 0$,

$$S^2(X, X + \sqrt{s}Z) = \sup_f \mathrm{Var}(\mathbb{E}[f(X)|X + \sqrt{s}Z]) = \sup_f \mathrm{Var}(Q_s f(X + \sqrt{s}Z)),$$

where the supremum runs over all f with $\mathbb{E}f(X) = 0$ and $\mathrm{Var}f(X) = 1$. Therefore,

$$S^2(X, X + \sqrt{s}Z) = \sup_f \int_{\mathbb{R}^n} |Q_s f|^2 d\mu_s = \sup_f \|Q_s f\|_{L^2(\mu_s)}^2,$$

where the supremum runs over all f with $\int f d\mu = 0$ and $\int f^2 d\mu = 1$. $\qquad\square$

The operator Q_s may be expressed as an integral operator. Write

$$\gamma_s(x) = (2\pi s)^{-n/2} \exp(-|x|^2/(2s)) \tag{56}$$

for the density in $\mathbb{R}^n$ of a Gaussian random vector of mean zero and covariance $s \cdot \mathrm{Id}$. By considering the joint distribution of $(X, X + \sqrt{s}Z)$ and writing conditional expectation as an integral, we see that

$$Q_s f(y) = \frac{\int_{\mathbb{R}^n} \gamma_s(x - y) f(x) d\mu(x)}{(\rho * \gamma_s)(y)}. \tag{57}$$

The proof of Theorem 1.1 requires two differentiations of the expression on the left-hand side of (55) with respect to s. It will be convenient to consider a subclass of well-behaved functions of $L^2(\mu)$ in order to justify the differentiations under the integral sign. Write ρ for the density of μ.

As in Klartag and Putterman [29], we say that a function $f : \mathbb{R}^n \to \mathbb{R}$ has subexponential decay relative to ρ if there exist $C, a > 0$ such that

$$|f(x)| \leq \frac{C}{\sqrt{\rho(x)}} e^{-a|x|} \qquad (x \in \mathbb{R}^n). \tag{58}$$

If ρ decays exponentially at infinity – for instance, if ρ is log-concave – then all polynomials have subexponential decay relative to ρ. We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is μ-*tempered* if it is smooth and if all of its partial derivatives of all orders have subexponential decay relative to ρ. Write $\mathcal{F}_\mu$ for the collection of all μ-tempered functions on $\mathbb{R}^n$. Since $\mathcal{F}_\mu$ contains all compactly-supported smooth functions, it is a dense subspace of $L^2(\mu)$.

Lemma 3.2. *If f is μ-tempered, then $Q_s f$ is μ_s-tempered for any $s > 0$.*

Lemma 3.2 is proven in [29, Lemma 2.2] under the additional assumption that μ is log-concave. The log-concavity assumption is only used in the proof of Lemma 2.2 in [29] in order to show that for any $p > 0$,

$$\int_{\mathbb{R}^n} |x|^p d\mu(x) < \infty. \tag{59}$$

However, a measure μ with $C_P(\mu) < \infty$ clearly satisfies (59). In fact, since $f(x) = |x|$ is a 1-Lipschitz function, we even know that $\int_{\mathbb{R}^n} e^{\alpha f} d\mu < \infty$ for some positive $\alpha > 0$, this goes back to Gromov and Milman [23]. Hence Lemma 3.2 applies for any probability measure μ with a finite Poincaré constant.

We write $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$ for the Hessian matrix of f at the point x. We abbreviate $X_s = X + \sqrt{s}Z$. The following lemma is proven in [29] as well:

Lemma 3.3. *For any $s > 0$ and $f \in \mathcal{F}_\mu$, setting $f_s = Q_s f$,*

$$\frac{\partial}{\partial s} \mathrm{Var}(f_s(X_s)) = -\mathbb{E}|\nabla f_s(X_s)|^2, \tag{60}$$

and

$$\frac{\partial}{\partial s} \mathbb{E}|\nabla f_s(X_s)|^2 = -\mathbb{E}\|\nabla^2 f_s(X_s)\|_{HS}^2 - 2\mathbb{E}(\nabla^2 \psi_s)(X_s) \nabla f_s(X_s) \cdot \nabla f_s(X_s), \tag{61}$$

where $\rho_s = e^{-\psi_s}$ is the density of X_s and $\|A\|_{HS}$ is the Hilbert-Schmidt norm (also known as the Frobenius norm) of the matrix $A \in \mathbb{R}^{n \times n}$.

Recall that the density of the random vector $X = X_0$ is the function $\rho = e^{-\psi}$. The Laplace operator associated with μ is defined for $u \in \mathcal{F}_\mu$ via

$$Lu = L_\mu u = \Delta u - \nabla \psi \cdot \nabla u. \tag{62}$$

For any $u, v \in \mathcal{F}_\mu$ we have the integration by parts formula

$$\int_{\mathbb{R}^n} (Lu)v\, d\mu = -\int_{\mathbb{R}^n} \langle \nabla u, \nabla v \rangle d\mu \tag{63}$$

and the (integrated) Bochner formula

$$\int_{\mathbb{R}^n} (Lu)^2 d\mu = \int_{\mathbb{R}^n} \|\nabla^2 u\|_{HS}^2 d\mu + \int_{\mathbb{R}^n} \langle (\nabla^2 \psi) \nabla u, \nabla u \rangle d\mu, \tag{64}$$

where $\|\nabla^2 u\|_{HS}$ is the Hilbert-Schmidt norm of the Hessian matrix $\nabla^2 u$. Formulae (63) and (64) are proven by intergation by parts, see e.g., Ledoux [33, Section 2.3]. The μ-temperedness of u, v are used in order to discard the boundary terms. The operator $-L$ is symmetric and positive semi-definite in $\mathcal{F}_\mu \subseteq L^2(\mu)$. Recall the well-known subadditivity property of the Poincaré constant (e.g. [15] or Theorem 5.1),

$$C_{\mathrm{P}}(X_s) = C_{\mathrm{P}}(X + \sqrt{s}Z) \leq C_{\mathrm{P}}(X) + C_{\mathrm{P}}(\sqrt{s}Z) = C_{\mathrm{P}}(X) + s.$$

Proof of Theorem 1.1. We begin with the proof of the right-hand side inequality in (19), as the left-hand side inequality was already proven above. If $C_P(X) = +\infty$ then this inequality is vacuously true, hence we may assume that $C_P(X) < \infty$. Recall that $X_s = X + \sqrt{s}Z$. By Lemma 3.1, we need to show that for any function f with $\mathbb{E}|f(X)|^2 < \infty$,

$$\mathrm{Var}\left(\mathbb{E}\left[f(X)|X_s\right]\right) = \mathrm{Var}(Q_s f(X_s)) \leq \frac{1}{1 + s/C_P(\mu)} \cdot \mathrm{Var} f(X). \tag{65}$$

Recall that μ is the distribution law of X. Since $Q_s : L^2(\mu) \to L^2(\mu_s)$ is a bounded operator and $\mathcal{F}_\mu$ is dense in $L^2(\mu)$, it suffices to prove (65) under the additional assumption that $f \in \mathcal{F}_\mu$. Abbreviate $f_s = Q_s(f)$. By using (60) and the Poincaré inequality, we get

$$\frac{\partial}{\partial s} \mathrm{Var}(f_s(X_s)) \leq -\frac{1}{C_{\mathrm{P}}(X_s)} \mathrm{Var}(f_s(X_s)) \leq -\frac{1}{s + C_{\mathrm{P}}(X)} \mathrm{Var}(f_s(X_s)),$$

where we used subadditivity in the last passage. Therefore

$$\frac{\partial}{\partial s} \log \mathrm{Var}(f_s(X_s)) \leq -\frac{1}{s + C_{\mathrm{P}}(X)}.$$

By integrating from 0 to s we conclude that

$$\log \frac{\mathrm{Var}(f_s(X_s))}{\mathrm{Var} f(X)} \leq -\int_0^s \frac{dx}{x + C_{\mathrm{P}}(X)} = \log \frac{1}{1 + s/C_{\mathrm{P}}(X)},$$

16

proving (65). This proves the right-hand side inequality in (19). Let us now assume that X is log-concave and prove (20). This is in fact proven in [29], but let us briefly repeat the argument here for convenience. Let $0 < \varepsilon < C_{\mathrm{P}}(\mu)$. By the definition of $C_{\mathrm{P}}(\mu)$, there exists a non-constant function f such that

$$\mathrm{Var} f(X) \geq (C_{\mathrm{P}}(\mu) - \varepsilon) \cdot \mathbb{E} |\nabla f(X)|^2. \tag{66}$$

Thanks to the appendix of [7], we may assume that f is smooth and compactly-supported in $\mathbb{R}^n$, and in particular $f \in \mathcal{F}_\mu$. According to (60) and (61) we may differentiate the Rayleigh quotient

$$\frac{\partial}{\partial s} \frac{\mathbb{E}|\nabla f_s(X_s)|^2}{\mathrm{Var}(f_s(X_s))} = -\mathbb{E}|\nabla^2 \tilde{f}_s(X_s)|^2 - 2\mathbb{E}(\nabla^2 \psi_s)(X_s)\nabla \tilde{f}_s(X_s) \cdot \nabla \tilde{f}_s(X_s) - \mathbb{E}|\nabla \tilde{f}_s(X_s)|^4$$

where we normalize $\tilde{f}_s = f_s / \sqrt{\mathrm{Var}(f_s(X_s))}$. As explained in the proof of [29, Theorem 2.4], it follows from the spectral theorem that

$$\langle L^2_{\mu_s} \tilde{f}_s, \tilde{f}_s \rangle_{L^2(\mu_s)} \geq \langle L\tilde{f}_s, \tilde{f}_s \rangle^2_{L^2(\mu_s)} = \left(\int_{\mathbb{R}^n} |\nabla \tilde{f}_s|^2 d\mu_s \right)^2.$$

Therefore, from the Bochner formula (64),

$$\frac{\partial}{\partial s} \frac{\mathbb{E}|\nabla f_s(X_s)|^2}{\mathrm{Var}(f_s(X_s))} \leq -\mathbb{E}(\nabla^2 \psi_s)(X_s)\nabla \tilde{f}_s(X_s) \cdot \nabla \tilde{f}_s(X_s) \leq 0. \tag{67}$$

where the last inequality is the only place where we use log-concavity; indeed, since X is log-concave, so is the random vector $X + \sqrt{s}Z$ by the Prékopa-Leindler inequality (see e.g. the first pages of Pisier [34], or Davidovič, Korenbljum and Hacet [18]). Thus $e^{-\psi_s}$ is a log-concave density, which amounts to the fact that the symmetric matrix $\nabla^2 \psi_s$ is positive semi-definite. Hence the term involving the Hessian $\nabla^2 \psi_s$ in (67) is non-positive. Consequently, $\mathbb{E}|\nabla f_s(X_s)|^2/\mathrm{Var}(f_s(X_s))$ is non-increasing in s and

$$\log \frac{\mathrm{Var}(f_s(X_s))}{\mathrm{Var} f(X)} = -\int_0^s \frac{\mathbb{E}|\nabla f_t(X_t)|^2}{\mathrm{Var}(f_t(X_t))} dt \geq -s \cdot \frac{\mathbb{E}|\nabla f(X)|^2}{\mathrm{Var}(f(X))} \geq -\frac{s}{C_{\mathrm{P}}(\mu) - \varepsilon}.$$

Hence for any $\varepsilon > 0$ we found f such that

$$\mathrm{Var}\left(\mathbb{E}[f(X)|X_s]\right) = \mathrm{Var}(f_s(X_s)) \geq \exp\left(-\frac{s}{C_{\mathrm{P}}(\mu) - \varepsilon}\right) \mathrm{Var} f(X).$$

Since $\varepsilon > 0$ is arbitrary, and in view of Lemma 3.1 this proves that $S^2(X, X + \sqrt{s}Z) \geq \exp(-s/C_{\mathrm{P}}(\mu))$. Inequality (20) is thus proven. $\square$

Remark 3.4. Thomas Courtade and Joseph Lehec communicated to us an alternative proof of the right-hand side inequality in (19). Their proof does not require differentiation with respect to the parameter s.

4 Identities for KL-divergence and general divergences

In this section we prove Propositions 1.5 and 1.10, the generalized de Bruijn identity. We keep the notation from the previous section. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a μ-integrable function.

As in the previous section, some care is needed when differentiating under the integral sign and when integrating by parts and neglecting the boundary terms. As opposed to the previous section analyzing the case of the χ^2-divergence, here we will be rather brief in the justification. As in [29] and in Section 3, the basic idea is to introduce a suitable class $\mathcal{F}_{\mu,\varphi}$ of smooth functions, which allow for integrations by parts without boundary terms, and is preserved by the Q_s-dynamics. Then one approximates the density of an arbitrary measure ν with $J_\varphi(\nu \mid\mid \mu) < \infty$ with a function from this class.

From Klartag and Putterman [29], we know that for any $f \in \mathcal{F}_\mu$ and $s > 0$,

$$\frac{d}{ds}Q_s f = \square_s Q_s f \tag{68}$$

where

$$\square_s = L_s - \frac{\Delta}{2}$$

and $L_s = L_{\mu_s}$ is the Laplace operator associated with the measure μ_s. We write ρ_s for the density of μ_s.

Lemma 4.1. *Let $\varphi(x) = \varphi_\lambda(x) = x^\lambda - 1$, for $\lambda > 1$. Let μ and ν be probability distributions on $\mathbb{R}^n$ with $D_\varphi(\nu \mid\mid \mu) < \infty$, and assume $\mathbb{E}_\mu|X|^4 < \infty$. Denote $f = d\nu/d\mu$ such that $f_s = Q_s f = d\mu_s/d\nu_s$, for $s > 0$. we have*

$$\frac{d}{ds}\mathbb{E}\varphi(f_s(X_s)) = -\frac{1}{2} \cdot \mathbb{E}\varphi''(f_s(X_s))|\nabla f_s(X_s)|^2 = -\frac{1}{2}\int_{\mathbb{R}^n} \varphi''(f_s)|\nabla f_s|^2 d\mu_s. \tag{69}$$

In particular, if $D_{KL}(\nu \mid\mid \mu) < \infty$

$$\frac{d}{ds}\mathbb{E}f_s(X_s) \log f_s(X_s) = -\frac{1}{2} \cdot \mathbb{E}\frac{|\nabla f(X_s)|^2}{f_s(X_s)} \tag{70}$$

Proof. We prove (69) below for any smooth, convex function $\varphi : (0, \infty) \to \mathbb{R}$ with $\varphi(1) = 0$, without justifying differentiation under the integral sign, and without justifying integration by parts. The justifications for these steps for $\varphi(x) = \varphi_\lambda(x) = x^\lambda - 1$, for $\lambda > 1$, under the assumptions that $D_\varphi(\nu \mid\mid \mu) < \infty$, and $\mathbb{E}_\mu|X|^4 < \infty$, are given in the appendix. We also prove in the appendix that (69) implies (70).

Recall that $\partial \rho_s/\partial s = \Delta \rho_s/2$ by the heat equation, as $\mu_s = \mu * \gamma_s$ and ρ_s is the density of μ_s. This and (68) are the main places where we use the heat equation in our arguments. Let us compute:

$$\frac{d}{ds}\int_{\mathbb{R}^n} \varphi(f_s)\rho_s = \int_{\mathbb{R}^n} \varphi'(f_s)\square_s f_s \cdot \rho_s + \int_{\mathbb{R}^n} \varphi(f_s)\frac{\Delta \rho_s}{2}.$$

18

Integrating by parts we continue with

$$
= \int_{\mathbb{R}^n} \left[\varphi'(f_s) \left(L_s f_s - \frac{\Delta f_s}{2} \right) + \frac{\Delta(\varphi(f_s))}{2} \right] \rho_s
$$

$$
= \int_{\mathbb{R}^n} \left[\varphi'(f_s) L_s f_s - \varphi'(f_s) \frac{\Delta f_s}{2} + \frac{\varphi'(f_s) \Delta f_s + \varphi''(f_s) |\nabla f_s|^2}{2} \right) \right] \rho_s
$$

$$
= \int_{\mathbb{R}^n} \left[-\varphi''(f_s) |\nabla f_s|^2 + \frac{\varphi''(f_s) |\nabla f_s|^2)}{2} \right] \rho_s,
$$

where we used the integration by parts (63) in the last passage. $\qquad\square$

Proof of Proposition 1.5 and Proposition 1.10. Follow directly from Lemma 4.1, and the definition of D_φ and J_φ. $\qquad\square$

We move on to computing the second derivative of $\mathbb{E}\varphi(f(X_s))$.

Conjecture 4.2. *Denoting*

$$
\kappa = -\frac{1}{2\varphi''} \left(\frac{1}{\varphi''} \right)'',
$$

and

$$
g_s = \varphi'(f_s)
$$

we have, under mild regularity assumptions, the "Bochner-type" formula

$$
\frac{d}{ds} \int_{\mathbb{R}^n} \varphi''(f_s) |\nabla f_s|^2 d\mu_s = -\int_{\mathbb{R}^n} \left[2(\nabla^2 \psi_s) \nabla g_s \cdot \nabla g_s + \left| \nabla^2 g_s \right|^2 + \kappa(f_s) |\nabla g_s|^4 \right] \frac{d\mu_s}{\varphi''(f_s)},
$$

where we recall that $\rho_s = e^{-\psi_s}$.

Sketch of proof. We assume below that differentiation under the integral sign is valid, and that all functions involved in integration in parts vanish. Justifying these assumptions is the missing step in proving this conjecture. We now prove the "Bochner-typ" formula above, under the assumption that those steps are indeed valid (i.e., ignoring all regularity issues).

Our approach follows the M-function of Ivanishvili and Volberg [26] and the dynamic approach to the Γ-calculus is presented in [29]. We consider a function $M(x, y)$ of two real variables of the form

$$
M(x, y) = y\varphi''(x). \tag{71}
$$

Thus $M(x, y) = y/x$ in the case where $\varphi(x) = x \log x$ and $M(x, y) = 2y$ if $\varphi(x) = (x - 1)^2$. In view of Lemma 4.1, we would need to compute

$$
\frac{d}{ds} \mathbb{E} M(f_s(X_s), |\nabla f_s(X_s)|^2).
$$

19

By the chain rule, for any functions f, g,

$$\Delta M(f, g) = M_x \Delta f + M_y \Delta g + M_{xx} |\nabla f|^2 + 2M_{xy} \nabla f \cdot \nabla g + M_{yy} |\nabla g|^2, \qquad (72)$$

where we abbreviate $M_x = \partial M / \partial x, M_y = \partial M / \partial y$ etc. Let us also introduce the dynamic Γ-calculus notation from [29]. We set

$$\Gamma_1(u, v) = \nabla u \cdot \nabla v$$

and

$$\Gamma_2(u, v) := \Box_s \Gamma_1(u, v) - \Gamma_1(\Box_s u, v) - \Gamma_1(u, \Box_s v)$$

which satisfies

$$\Gamma_2(u, u) = \|\nabla^2 u\|_{HS}^2 - 2\nabla^2(\log \rho_s) \nabla u \cdot \nabla u. \qquad (73)$$

We abbreviate $\Gamma_1(u) = \Gamma_1(u, u)$ and $\Gamma_2(u) = \Gamma_2(u, u)$.

We first prove that for a general function M of two variables,

$$\frac{d}{ds} \mathbb{E} M(f_s(X_s), |\nabla f_s(X_s)|^2) \qquad (74)$$

$$= -\int_{\mathbb{R}^n} \left[M_y \Gamma_2(f_s) + \frac{M_{xx}}{2} \Gamma_1(f_s) + M_{xy} \Gamma_1(f_s, \Gamma_1(f_s)) + \frac{M_{yy}}{2} \Gamma_1(\Gamma_1(f_s)) \right] d\mu_s.$$

To this end, abbreviate $M = M(f_s, |\nabla f_s|^2), M_x = M_x(f_s, |\nabla f_s|^2)$ etc. Then,

$$\frac{d}{ds} \int_{\mathbb{R}^n} M(f_s, |\nabla f_s|^2) \rho_s = \int_{\mathbb{R}^n} \left[\frac{\Delta M}{2} + M_x \Box_s f_s + 2M_y \Gamma_1(\Box_s f_s, f_s) \right] \rho_s \qquad (75)$$

$$= \int_{\mathbb{R}^n} \left[\frac{\Delta M}{2} + M_x \Box_s f_s + M_y \left(\Box_s \Gamma_1(f_s, f_s) - \Gamma_2(f_s, f_s) \right) \right] \rho_s$$

$$= \int_{\mathbb{R}^n} \left[\frac{\Delta M}{2} + M_x \left(L_s - \frac{\Delta}{2} \right) f_s + M_y \left(\left(L_s - \frac{\Delta}{2} \right) \Gamma_1(f_s) - \Gamma_2(f_s) \right) \right] \rho_s.$$

By (72) with $f = f_s, g = \Gamma_1(f_s)$,

$$\Delta M = M_x \Delta f_s + M_y \Delta \Gamma_1(f_s) + M_{xx} \Gamma_1(f_s) + 2M_{xy} \Gamma_1(f_s, \Gamma_1(f_s)) + M_{yy} \Gamma_1(\Gamma_1(f_s)).$$

Therefore,

$$\frac{1}{2} (\Delta M - M_x \Delta f_s - M_y \Delta \Gamma_1(f_s)) = \frac{M_{xx}}{2} \Gamma_1(f_s) + M_{xy} \Gamma_1(f_s, \Gamma_1(f_s)) + \frac{M_{yy}}{2} \Gamma_1(\Gamma_1(f_s)).$$

Consequently, the integral in (75) equals

$$\int_{\mathbb{R}^n} \left[M_x L_s f_s + M_y L_s \Gamma_1(f_s) - M_y \Gamma_2(f_s) + \frac{M_{xx}}{2} \Gamma_1(f_s) + M_{xy} \Gamma_1(f_s, \Gamma_1(f_s)) + \frac{M_{yy}}{2} \Gamma_1(\Gamma_1(f_s)) \right] \rho_s$$

20

Integrate by parts the two terms involving L_s using (63) to obtain

$$\int_{\mathbb{R}^n} \left[M_x L_s f_s + M_y L_s \Gamma_1(f_s)\right] \rho_s = \int_{\mathbb{R}^n} \left[-\nabla M_x \cdot \nabla f_s - \nabla M_y \cdot \nabla \Gamma_1(f_s)\right] \rho_s$$

$$= \int_{\mathbb{R}^n} \left[-(M_{xx}\nabla f_s + M_{xy}\nabla\Gamma_1(f_s)) \cdot \nabla f_s - (M_{xy}\nabla f_s + M_{yy}\nabla\Gamma_1(f_s)) \cdot \nabla\Gamma_1(f_s)\right] \rho_s$$

$$= \int_{\mathbb{R}^n} \left[-M_{xx}\Gamma_1(f_s) - 2M_{xy}\Gamma_1(\Gamma_1(f_s), f_s) - M_{yy}\Gamma_1(\Gamma_1(f_s))\right] \rho_s.$$

Hence, the integral in (75) equals

$$-\int_{\mathbb{R}^n} \left[M_y\Gamma_2(f_s) + \frac{M_{xx}}{2}\Gamma_1(f_s) + M_{xy}\Gamma_1(f_s, \Gamma_1(f_s)) + \frac{M_{yy}}{2}\Gamma_1(\Gamma_1(f_s))\right] \rho_s$$

and (74) is proven. In the specific case where M is the function given by (71) we conclude from (74) that

$$\frac{d}{ds}\int_{\mathbb{R}^n} \varphi''(f_s)|\nabla f_s|^2 d\mu_s \tag{76}$$

$$= -\int_{\mathbb{R}^n} \left[\varphi''(f_s)\Gamma_2(f_s) + \frac{\varphi^{(4)}(f_s)}{2}\Gamma_1(f_s)^2 + \varphi^{(3)}(f_s)\Gamma_1(f_s, \Gamma_1(f_s))\right] d\mu_s$$

$$= -\int_{\mathbb{R}^n} \left[2\varphi''(\nabla^2\psi_s)\nabla f_s \cdot \nabla f_s + \varphi''|\nabla^2 f_s|^2 + \frac{\varphi^{(4)}}{2}|\nabla f_s|^4 + 2\varphi^{(3)}(\nabla^2 f_s)\nabla f_s \cdot \nabla f_s\right] d\mu_s$$

$$= -\int_{\mathbb{R}^n} \left[2\varphi''(\nabla^2\psi_s)\nabla f_s \cdot \nabla f_s + \varphi''\left|\nabla^2 f_s + \frac{\varphi^{(3)}}{\varphi''}\nabla f_s \otimes \nabla f_s\right|^2 + \left\{\frac{\varphi^{(4)}}{2} - \frac{\left(\varphi^{(3)}\right)^2}{\varphi''}\right\}|\nabla f_s|^4\right] d\mu_s$$

Note that

$$\frac{\varphi^{(4)}}{2} - \frac{\left(\varphi^{(3)}\right)^2}{\varphi''} = -\frac{(\varphi'')^2}{2}\left(\frac{1}{\varphi''}\right)'' = (\varphi'')^3\kappa.$$

Consequently,

$$\frac{d}{ds}\int_{\mathbb{R}^n} \varphi''(f_s)|\nabla f_s|^2 d\mu_s$$

$$= -\int_{\mathbb{R}^n} \left[2\varphi''(\nabla^2\psi_s)\nabla f_s \cdot \nabla f_s + \varphi''\left|\nabla^2 f_s + \frac{\varphi^{(3)}}{\varphi''}\nabla f_s \otimes \nabla f_s\right|^2 + \kappa(\varphi')^3|\nabla f_s|^4\right] d\mu_s.$$

Since $g_s = \varphi'(f_s)$ we have $\nabla g_s = \varphi''(f_s)\nabla f_s$ and $\nabla^2 g_s = \varphi''(f_s)\nabla^2 f_s + \varphi^{(3)}(f_s)\nabla f_s \otimes \nabla f_s$, and the lemma is proven. $\qquad\square$

Specializing Conjecture 4.2 to the case where $\varphi(x) = x\log x$, we obtain the following:

Proposition 4.3. *Assuming the validity of Conjecture 4.2, With $g_s = \log Q_s f$ we have*

$$\frac{d}{ds}\mathbb{E}\frac{|\nabla f_s(X_s)|^2}{f_s(X_s)} = -\int_{\mathbb{R}^n} \left[|\nabla^2 g_s|^2 + 2\langle(\nabla^2\psi_s)\nabla g_s, \nabla g_s\rangle\right] f_s d\mu_s.$$

21

Proof. This follows from Conjecture 4.2 with $\varphi(x) = x \log x$, since in this case $1/\varphi''(x) = x$ and $\kappa \equiv 0$. $\qquad\blacksquare$

5   Proofs of inequalities related to φ-Sobolev constants

We keep the notation and assumptions of the previous section. We need to use the subadditivity property of the φ-Sobolev constants, proven by Chafaï [11, Corollary 3.1] in greater generality. For the convenience of the reader we include here a statement and a proof of the following proposition, as well as of Proposition 1.9, which also goes back to Chafaï [11].

Theorem 5.1 ("Subadditivity of the φ-Sobolev constants", Chafaï [11]). *Let X and Y be independent random vectors in $\mathbb{R}^n$. Let $\varphi : (0, \infty) \to \mathbb{R}$ be a smooth, convex function with $\varphi(1) = 0$ such that $1/\varphi''$ is concave. Then,*

$$C_\varphi(X + Y) \leq C_\varphi(X) + C_\varphi(Y).$$

Proof. for any locally-Lipschitz function $f : \mathbb{R}^n \to [0, \infty)$ with $\mathbb{E}|f(X + Y)| < \infty$, denoting $g(x) = \mathbb{E}f(x + Y)$ we have

$$\begin{aligned}
&\mathbb{E}\varphi(f(X + Y)) - \varphi(\mathbb{E}(f(X + Y))) \\
&\quad = \mathbb{E}_X \left[\mathbb{E}_Y \varphi(f(X + Y)) - \varphi(\mathbb{E}_Y \varphi(f(X + Y))) \right] + \mathbb{E}_X \varphi(g(X)) - \varphi(\mathbb{E}_X g(X)) \\
&\quad \leq \frac{C_\varphi(Y)}{2} \mathbb{E}_X \mathbb{E}_Y \varphi''(f(X + Y))|\nabla f(X + Y)|^2 + \frac{C_\varphi(X)}{2} \mathbb{E}_X \varphi''(g(X))|\nabla g(X)|^2.
\end{aligned}$$

To conclude the proof it remains to show that for any fixed $x \in \mathbb{R}^n$,

$$\varphi''(g(x))|\nabla g(x)|^2 \leq \mathbb{E}_Y \varphi''(f(x + Y))|\nabla f(x + Y)|^2.$$

Indeed, this would imply that

$$\mathbb{E}\varphi(f(X + Y)) - \varphi(\mathbb{E}(\varphi(X + Y))) \leq \frac{C_\varphi(Y) + C_\varphi(X)}{2} \mathbb{E}\varphi''(f(X + Y))|\nabla f(X + Y)|^2,$$

and hence $C_\varphi(X + Y) \leq C_\varphi(Y) + C_\varphi(X)$. Since $\nabla g(x) = \mathbb{E}\nabla f(x + Y)$, all that remains is to show that

$$\varphi''(\mathbb{E}f(x + Y))|\mathbb{E}\nabla f(x + Y)|^2 \leq \mathbb{E}_Y \varphi''(f(x + Y))|\nabla f(x + Y)|^2.$$

By Jensen's inequality, this would follow once we show that the function

$$(x, y) \mapsto \varphi''(x)|y|^2 \qquad\qquad (x > 0, y \in \mathbb{R}^n) \tag{77}$$

is jointly convex in x and y. Since $\varphi''(x)|y|^2 = \sum_j \varphi''(x)y_j^2$, it suffices to show that the function

$$(x, y) \mapsto \varphi''(x)y^2 \qquad\qquad (x > 0, y \in \mathbb{R}) \qquad\qquad (78)$$

is jointly convex in (x, y). The Hessian of this function is the matrix

$$\begin{pmatrix} \varphi^{(4)}(x)y^2 & 2\varphi^{(3)}(x)y \\ 2\varphi^{(3)}(x)y & 2\varphi''(x) \end{pmatrix}$$

Since this symmetric 2×2 matrix has a non-negative entry on the diagonal, namely $2\varphi''(x)$, by the Sylvester criterion it is positive semi-definite if and only if its determinant is non-negative. The determinant equals

$$2y^2 \left[\varphi''\varphi^{(4)} - 2(\varphi^{(3)})^2 \right] = -2y^2(\varphi'')^3 \left(\frac{1}{\varphi''} \right)'',$$

where we used the formula $(1/\varphi'')'' = -(\varphi''\varphi^{(4)} - 2(\varphi^{(3)})^2)/(\varphi'')^3$. Since $\varphi'' \geq 0$ and $1/\varphi''$ is concave, the determinant is non-negative, and the function in (78) is convex. This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proof of Proposition 1.8. For any locally-Lipschitz function f,

$$\mathbb{E}\varphi(f(X)) - \varphi(\mathbb{E}f(X)) \leq \frac{C_\varphi(X)}{2} \cdot \mathbb{E}\varphi''(f(X))|\nabla f(X)|^2.$$

Let $g : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz, bounded function. Then there exists $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$, the function $g_0 = 1 + \varepsilon g$ attains values at I. In fact, since φ is smooth, there exists $M > 0$ and $\varepsilon_1 < \varepsilon_0$ such that for $\varepsilon < \varepsilon_1$,

$$\left| \varphi(g_0(x)) - \varepsilon\varphi'(1)g(x) - \frac{\varepsilon^2}{2}\varphi''(1)g^2(x) \right| \leq \varepsilon^3 M$$

and

$$\varphi''(g_0(x)) \geq (1 - \varepsilon M) \cdot \varphi''(1).$$

Consequently, for any $0 < \varepsilon < \varepsilon_1$,

$$\frac{\varepsilon^2}{2}\varphi''(1)Var(g(X)) - 2M\varepsilon^3 \leq \mathbb{E}\varphi(f(X)) - \varphi(\mathbb{E}f(X))$$

and

$$\mathbb{E}\varphi''(f(X))|\nabla f(X)|^2 \geq (1 - \varepsilon M)\varepsilon^2\varphi''(1) \cdot \mathbb{E}|\nabla g(X)|^2.$$

By using the φ-Sobolev inequality and dividing by $\varepsilon^2\varphi''(1) > 0$, and then letting ε tends to zero we obtain

$$\frac{1}{2}Var(g(X)) \leq \frac{C_\varphi(X)}{2}\mathbb{E}|\nabla g(X)|^2. \qquad\qquad (79)$$

23

This inequality holds for a Lipschitz, bounded function g. A standard truncation argument shows that (79) holds true for an arbitrary locally-Lipschitz function g. Indeed, It suffices to replace $g(x)$ by $g_M(x) = \theta(x/M)g(x)$ where $\theta : \mathbb{R}^n \to [0,1]$ is a compactly supported smooth function that equals one in a neighborhood of the origin. The function g_M is a Lipschitz, bounded function, and a short argument detailed e.g. in the proof of Proposition 27 in [7] allows to take the limit as $M \to \infty$ in the Poincaré inequality. This shows that $C_P(X) \leq C_\varphi(X)$. $\qquad\square$

Proof of Proposition 1.9. Let Z be a standard Gaussian random vector, independent of X. When X is a standard Gaussian, the operator Q_s has a pleasant form given by the Mehler formula. Recall that in general,

$$Q_s f(X + \sqrt{s}Z) = \mathbb{E}\left[f(X)|X + \sqrt{s}Z\right]$$

and hence

$$\mathbb{E}Q_s f(X + \sqrt{s}Z) = \mathbb{E}f(X).$$

In the case where X is Gaussian, the density of X conditioned on $X + \sqrt{s}Z = y$ is

$$\frac{\gamma_1(x)\gamma_s(y - x)}{\gamma_{1+s}(y)} = \gamma_{\frac{s}{s+1}}\left(x - \frac{y}{s+1}\right)$$

where γ_s is the Gaussian density, as in (56). In other words, the distribution of X conditioned on $X + \sqrt{s}Z$ is that of a Gaussian random vector of mean $\frac{X+\sqrt{s}Z}{s+1}$ and covariance $\frac{s}{s+1}\mathrm{Id}$. Therefore, for any $y \in \mathbb{R}^n$,

$$Q_s f(y) = \mathbb{E}f\left(\frac{y}{s+1} + \sqrt{\frac{s}{s+1}}Z\right) \tag{80}$$

and

$$\nabla Q_s f(y) = \frac{1}{s+1}\mathbb{E}\nabla f\left(\frac{y}{s+1} + \sqrt{\frac{s}{s+1}}Z\right).$$

Consequently, in all of $\mathbb{R}^n$ we have the identity,

$$\varphi''(Q_s f)|\nabla Q_s f|^2 = \frac{1}{(s+1)^2} \cdot \varphi''(Q_s f)|Q_s(\nabla f)|^2 \tag{81}$$

where Q_s is acting on a vector field entry by entry, i.e., $Q_s(\nabla f) = (Q_s(\partial_1 f), \ldots Q_s(\partial_n f))$. Since the function in (77) is convex in x and y, and since Q_s is an averaging operator as we see from (80), we may use Jensen's inequality. We obtain from (81) that

$$\varphi''(Q_s f)|\nabla Q_s f|^2 \leq \frac{1}{(s+1)^2} \cdot Q_s\left[\varphi''(f)|\nabla f|^2\right] \tag{82}$$

Abbreviate $f_s = Q_s f$ and $X_s = X + \sqrt{s}Z$, and observe that

$$\mathbb{E}\varphi(f(X)) - \varphi(\mathbb{E}f(X)) = -\int_0^\infty \frac{\partial}{\partial s}\mathbb{E}\varphi(\mathbb{E}\left[f(X)|X + \sqrt{s}Z\right])ds = -\int_0^\infty \frac{\partial}{\partial s}\mathbb{E}\varphi(f_s(X_s))ds.$$

From the generalized de Bruijn identity in the form of Lemma 4.1 above, and from (82),

$$\mathbb{E}\varphi(f(X)) - \varphi(\mathbb{E}f(X)) = \frac{1}{2}\int_0^\infty \mathbb{E}\varphi''(f_s(X_s))|\nabla f_s(X_s)|^2 ds$$

$$\leq \int_0^\infty \frac{1}{2(s+1)^2}\mathbb{E}Q_s\varphi''(f(X_s))|\nabla f(X_s)|^2 ds$$

$$= \int_0^\infty \frac{1}{2(s+1)^2}\mathbb{E}\left[\varphi''(f)|\nabla f|^2\right](X))ds$$

$$= \frac{1}{2}\mathbb{E}\left[\varphi''(f)|\nabla f|^2\right](X).$$

This shows that $C_\varphi(X) \leq 1$. The inequality $C_\varphi(X) \geq C_P(X) = 1$ follows from Proposition 1.8. $\qquad\square$

Question 5.2. *Is there an extremality property of the log-Sobolev constant $C_{\mathrm{LS}}(X)$ among all φ-Sobolev constants $C_\varphi(X)$ where φ ranges over the class of convex function with $\varphi(1) = 0$ such that $1/\varphi''$ is concave?*

Remark 5.3. In the proof of Proposition 1.9 we used the fact that when X is a standard Gaussian random vector, for any f,

$$|\nabla Q_s f|^2 \leq \frac{1}{(s+1)^2}Q_s(|\nabla f|^2). \tag{83}$$

Since $\varphi'' \geq 0$, inequality (83) and the convexity of the function $(x,y) \mapsto \varphi''(x)y^2$, suffices for concluding (82). Inequality (83) is the only property of a Gaussian random vector that was used in the proof of Proposition 1.9. Arguing as in the proof of [29, Lemma 2.5], one can show that (83) holds true in the case where the law μ of X is more log-concave than the Gaussian measure, i.e. when its density $e^{-\psi}$ satisfies $\nabla^2\psi \geq \mathrm{Id}$ everywhere in $\mathbb{R}^n$. We omit the details. This yields another proof of a result from Chafaï [11]: If $1/\varphi''$ is concave and X is more log-concave than the Gaussian, then,

$$C_\varphi(X) \leq 1.$$

Proof of Theorem 1.11 and Theorem 1.12. We mimick the proof of Theorem 1.1. Begin with the proof of the right-hand side inequality in (40), as the left-hand side inequality was already proven above. If $C_\varphi(X) = +\infty$ then this inequality is vacuously true, hence we may assume that $C_\varphi(X) < \infty$. Assume that $D_\varphi(\nu \,||\, \mu) < \infty$, and write $f = d\nu/d\mu$, so that $f \geq 0$ satisfies $\int f d\mu = 1$. Clearly $f_s = Q_s f$ is a probability density with respect to μ_s, and in fact $d\nu_s/d\mu_s = Q_s f$. We need to show that

$$D_\varphi(\nu_s \,||\, \mu_s) = \int_{\mathbb{R}^n} \varphi(Q_s f)d\mu_s \leq \frac{1}{1+s/C_\varphi(\mu)} \cdot \int_{\mathbb{R}^n} \varphi(f)d\mu = \frac{1}{1+s/C_\varphi(\mu)} \cdot D_\varphi(\nu \,||\, \mu). \tag{84}$$

The two equalities in (84) follow from the definition of D_φ, and it remains to prove the inequality. Let $X \sim \mu$ and recall that $X_s = X + \sqrt{s}Z$. By the generalized de Bruijn identity, i.e. Proposition

1.10, and the definition of the φ-Sobolev constant,

$$\frac{\partial}{\partial s}D_\varphi(\nu_s \parallel \mu_s) = -\frac{1}{2}J_\varphi(\nu_s \parallel \mu_s) \leq -\frac{1}{C_\varphi(X_s)}D_\varphi(\nu_s \parallel \mu_s) \leq -\frac{1}{C_\varphi(X) + C_\varphi(\sqrt{s}Z)}D_\varphi(\nu_s \parallel \mu_s),$$

where in the last passage we used subadditivity, i.e., Theorem 5.1. By homogeneity and Proposition 1.9,

$$\frac{\partial}{\partial s}D_\varphi(\nu_s \parallel \mu_s) \leq -\frac{1}{C_\varphi(X) + s}D_\varphi(\nu_s \parallel \mu_s).$$

Therefore

$$\frac{\partial}{\partial s}\log D_\varphi(\nu_s \parallel \mu_s) \leq -\frac{1}{s + C_\varphi(X)}.$$

By integrating from 0 to s we conclude that

$$\log \frac{D_\varphi(\nu_s \parallel \mu_s)}{D_\varphi(\nu \parallel \mu)} \leq -\int_0^s \frac{dx}{x + C_\varphi(X)} = \log \frac{1}{1 + s/C_\varphi(X)}.$$

This proves the right-hand side inequality in (40).

Let us now assume that X is log-concave and prove (41). Let $0 < \varepsilon < C_\varphi(\mu)$. There exists non-negative function f with $\mathbb{E}f(X) = 1$ such that

$$\mathbb{E}\varphi(f(X)) \geq (C_\varphi(\mu) - \varepsilon) \cdot \mathbb{E}\varphi''(f(X))|\nabla f(X)|^2. \tag{85}$$

By an argument similar to the one from the appendix of [7], we may assume that f is smooth. Let ν satisfy $d\nu/d\mu = f$. By Conjecture 4.2,

$$\frac{d}{ds}J_\varphi(\nu_s \parallel \mu_s) = \frac{d}{ds}\int_{\mathbb{R}^n} \varphi''(f_s)|\nabla f_s|^2 d\mu_s$$

$$= -\int_{\mathbb{R}^n} \left[2(\nabla^2\psi_s)\nabla g_s \cdot \nabla g_s + \left|\nabla^2 g_s\right|^2 + \kappa(f_s)|\nabla g_s|^4\right]\frac{d\mu_s}{\varphi''(f_s)} \leq 0, \tag{86}$$

where $f_s = Q_s f$. Indeed, each of the three summands in the integral is non-negative; the first is non-negative since $\nabla^2\psi_s \geq 0$ by log-concavity, the second is always non-negative, and the third is non-negative since $\kappa \geq 0$ as $1/\varphi''$ is concave. We thus proved that $J_\varphi(\nu_s \parallel \mu_s)$ is decreasing with s. Since $J_\varphi(\nu_s \parallel \mu_s)$ is the derivative of $D_\varphi(\nu_s \parallel \mu_s)$, we conclude that $D_\varphi(\nu_s \parallel \mu_s)$ is a concave function of s. Moreover, by the generalized de Brujin identity,

$$D_\varphi(\nu_s \parallel \mu_s) - D_\varphi(\nu \parallel \mu) = -\int_0^s J_\varphi(\nu_t \parallel \mu_t)dt \geq -s \cdot J_\varphi(\nu \parallel \mu) \geq -\frac{s}{C_\varphi(\mu) - \varepsilon} \cdot D_\varphi(\nu \parallel \mu).$$

and hence

$$D_\varphi(\nu_s \parallel \mu_s) \geq \left(1 - \frac{s}{C_\varphi(\mu) - \varepsilon}\right)D_\varphi(\nu \parallel \mu).$$

Since $\varepsilon > 0$ is arbitrary, this proves that

$$\eta_\varphi(\mu, s) \geq 1 - \frac{s}{C_\varphi(\mu)}.$$

$\blacksquare$

26

Proof of Theorem 1.6 and Theorem 1.7. These follow from Theorem 1.11 and Theorem 1.12, respectively. □

Appendix A Technical justifications for Lemma 4.1

We assume $\varphi(x) = \varphi_\lambda(x) = x^\lambda - 1$, for $\lambda > 1$. Let ν and μ be two probability distributions on $\mathbb{R}^n$. We make the following regularity assumptions:

1. $D_\varphi(\nu\|\mu) < \infty$

2. $\mathbb{E}_\mu |X|^4 < \infty$

Let $f = d\nu/d\mu$, such that $f_s = d\nu_s/d\mu_s$. The following proposition will be useful in the derivation below.

Proposition A.1. *Let $\varphi(x) = \varphi_\lambda(x) = x^\lambda - 1$, $\lambda > 1$, and assume $D_\varphi(\nu\|\mu) < \infty$. Then, for any $s > 0$ we also have $D_{\varphi^2}(\nu_s\|\mu_s) < \infty$, and in particular, $\chi^2(\nu\|\mu) < \infty$. Furthermore, $D_{\varphi_{4\lambda}}(\nu_s\|\mu_s) < \infty$.*

Proof. Recall that the Rènyi entropy of order λ is defined as

$$D_\lambda(\nu_s\|\mu_s) = \frac{1}{\lambda - 1} \log\left(1 + \mathbb{E}\left[\varphi_\lambda\left(\frac{d\nu_s}{d\mu_s}(X_s)\right)\right]\right) = \frac{1}{\lambda - 1} \log\left(1 + \mathbb{E}\left[\varphi_\lambda\left(f_s(X_s)\right)\right]\right), \quad (87)$$

where $\varphi_\lambda(x) = x^\lambda - 1$, and $f = d\nu/d\mu$, such that $f_s = d\nu_s/d\mu_s$. It therefore follows that

$$\varphi_\lambda^2(x) = (x^\lambda - 1)^2 = (x^{2\lambda} - 1) - 2(x^\lambda - 1) = \varphi_{2\lambda}(x) - 2\varphi_\lambda(x), \quad (88)$$

so that

$$D_\varphi(\nu_s\|\mu_s) = \mathbb{E}\varphi_\lambda(f_s(X_s)) = e^{(\lambda - 1)D_\lambda(\nu_s\|\mu_s)} - 1 \quad (89)$$

$$D_{\varphi^2}(\nu_s\|\mu_s) = \mathbb{E}\varphi_{2\lambda}(f_s(X_s)) - 2\mathbb{E}\varphi_\lambda(f_s(X_s)) = e^{(2\lambda - 1)D_{2\lambda}(\nu_s\|\mu_s)} - 2e^{(\lambda - 1)D_\lambda(\nu_s\|\mu_s)} + 1. \quad (90)$$

By the data-processing inequality, and the assumption that $D_\varphi(\nu\|\mu)$ is finite, we have

$$D_\varphi(\nu_s\|\mu_s) \leq D_\varphi(\nu\|\mu) < \infty. \quad (91)$$

Therefore, by (87), it follows that $D_\lambda(\nu_s\|\mu_s)$ is also finite. From [2, Theorem 1] (see also [41] and [45, Theorem 30]) it follows that finiteness of $D_\lambda(\nu_s\|\mu_s)$ for $\lambda > 1$ also implies the finiteness of $D_{2\lambda}(\nu_s\|\mu_s)$ (and also of $D_{4\lambda}(\nu_s\|\mu_s)$). To see this, note that for $\lambda > 1$ [2, Theorem 1] reads

$$D_\lambda(\nu_s\|\mu_s) = \sup_{\xi \ll \nu_s}\left[D_{\mathrm{KL}}(\xi\|\mu_s) - \frac{\lambda}{\lambda - 1}D_{\mathrm{KL}}(\xi\|\nu_s)\right].$$

Thus, finiteness of $D_\lambda(\nu_s\|\mu_s)$ implies that for any distribution $\xi \ll \nu_s$ we have $D_{\mathrm{KL}}(\xi\|\mu_s) < \infty$. Using the same equation again, with 2λ (or 4λ), we see that $D_{2\lambda}(\nu_s\|\mu_s)$ must be finite as well. Thus

$$D_\varphi(\nu_s\|\mu_s) < \infty \Rightarrow D_\lambda(\nu_s\|\mu_s) < \infty \Rightarrow D_{2\lambda}(\nu_s\|\mu_s) < \infty \Rightarrow D_{\varphi^2}(\nu_s\|\mu_s) < \infty. \tag{92}$$

To prove that $\chi^2(\nu\|\mu)$ is finite, we note that $\chi^2(\nu\|\mu) = D_{\varphi_{\lambda'}}(\nu\|\mu)$ with $\lambda' = 2$. Since $\lambda > 1$, we have that $\lambda' < 2\lambda$. We have already shown that $D_{2\lambda}(\nu\|\mu)$ is finite. This together with the fact that $\lambda \mapsto D_\lambda(\nu\|\mu)$ is non-decreasing in $\lambda > 1$, shows that $D_{\lambda'}(\nu\|\mu)$ is finite, which in turn implies that $D_{\varphi_{\lambda'}}(\nu\|\mu)$ is finite. $\qquad\square$

A.1 Justification of first integration in parts

We prove that

$$\sum_{i=1}^n \int_{\mathbb{R}^n} |\varphi(f_s)\frac{\partial}{\partial x_i}\rho_s| < \infty. \tag{93}$$

We have

$$\varphi(f_s)\nabla\rho_s = \varphi(f_s)\sqrt{\rho_s} \cdot \frac{\nabla\rho_s}{\sqrt{\rho_s}}. \tag{94}$$

Applying the Cauchy-Schwartz inequality, we therefore have that

$$\sum_{i=1}^n \int_{\mathbb{R}^n} |\varphi(f_s)\frac{\partial}{\partial x_i}\rho_s| = n\sum_{i=1}^n \frac{1}{n}\int_{\mathbb{R}^n} |\varphi(f_s)\sqrt{\rho_s}| \cdot |\frac{\frac{\partial}{\partial x_i}\rho_s}{\sqrt{\rho_s}}| \tag{95}$$

$$\leq n\sum_{i=1}^n \frac{1}{n}\sqrt{\int_{\mathbb{R}^n} \varphi^2(f_s)\rho_s \cdot \int_{\mathbb{R}^n} \frac{\left(\frac{\partial}{\partial x_i}\rho_s\right)^2}{\sqrt{\rho_s}}} \tag{96}$$

$$\leq \sqrt{n}\sqrt{\int_{\mathbb{R}^n} \varphi^2(f_s)\rho_s \cdot \int_{\mathbb{R}^n} \frac{|\nabla\rho_s|^2}{\rho_s}} \tag{97}$$

$$= \sqrt{\mathbb{E}\varphi^2(f_s(X_s))}\sqrt{n \cdot \mathcal{J}(\mu_s)}. \tag{98}$$

We have that $\mathcal{J}(\mu_s) < \infty$ for any $s > 0$. Finiteness of $\mathbb{E}\varphi^2(f_s(X_s)) = D_{\varphi^2}(\nu_s\|\mu_s)$ follows from Proposition A.1. We therefore conclude that (93) holds.

A.2 Justification of second integration in parts

We prove that

$$\sum_{i=1}^n \int_{\mathbb{R}^n} |\rho_s\frac{\partial}{\partial x_i}\varphi(f_s)| < \infty. \tag{99}$$

28

Recalling that $\varphi(x) = x^\lambda - 1$ we have

$$\varphi'(x) = \lambda x^{\lambda-1}, \quad \varphi''(x) = \lambda(\lambda-1)x^{\lambda-2}, \tag{100}$$

and therefore

$$\frac{\varphi'(x)}{\sqrt{\varphi''(x)}} = \left(\frac{\lambda}{\lambda-1}x^\lambda\right)^{1/2} = \left(\frac{\lambda}{\lambda-1}(\varphi(x)+1)\right)^{1/2}. \tag{101}$$

Consequently, (recall that $\varphi'' > 0$)

$$\rho_s \nabla \varphi(f_s) = \rho_s \varphi'(f_s) \nabla f_s = \left(\sqrt{\varphi''(f_s)\rho_s} \nabla f_s\right) \left(\frac{\varphi'(f_s)}{\sqrt{\varphi''(f_s)}}\sqrt{\rho_s}\right) \tag{102}$$

$$= \sqrt{\frac{\lambda}{\lambda-1}} \left(\sqrt{\varphi''(f_s)\rho_s} \nabla f_s\right) \left(\sqrt{(\varphi(f_s)+1)\rho_s}\right). \tag{103}$$

Applying the Cauchy-Schwartz inequality, we therefore have that

$$\sum_{i=1}^{n} \int_{\mathbb{R}^n} |\rho_s \frac{\partial}{\partial x_i}\varphi(f_s)| = \sqrt{\frac{\lambda}{\lambda-1}} n \sum_{i=1}^{n} \frac{1}{n} \int_{\mathbb{R}^n} \left(\sqrt{\varphi''(f_s)\rho_s} \frac{\partial}{\partial x_i} f_s\right) \left(\sqrt{(\varphi(f_s)+1)\rho_s}\right) \tag{104}$$

$$\leq \sqrt{\frac{\lambda}{\lambda-1}} n \sum_{i=1}^{n} \frac{1}{n} \sqrt{\int_{\mathbb{R}^n} \varphi''(f_s)\rho_s \left(\frac{\partial}{\partial x_i} f_s\right)^2 \int_{\mathbb{R}^n} (\varphi(f_s)+1)\rho_s} \tag{105}$$

$$= \sqrt{\frac{\lambda}{\lambda-1}} \sqrt{n(1 + \mathbb{E}\varphi(f_s(X_s))) \cdot \int_{\mathbb{R}^n} \varphi''(f_s)|\nabla f_s|^2 d\mu_s}. \tag{106}$$

Consequently, the assumption that $D_\varphi(\nu_s\|\mu_s)$ is finite and that $\int_{\mathbb{R}^n} \varphi''(f_s)|\nabla f_s|^2 d\mu_s$ is finite (otherwise the statement is void), implies that (99) holds.

A.3 Justification of taking derivative under the integral sign

We show that $\frac{d}{ds}\int_{\mathbb{R}^n} \varphi(f_s)\rho_s = \int_{\mathbb{R}^n} \frac{d}{ds}\varphi(f_s)\rho_s$. To that end, we will show that for any $0 < a < b$ close enough to each other, it holds that $\sup_{s\in[a,b]} |\frac{d}{ds}\varphi(f_s)\rho_s|$ is integrable, and use the dominated convergence theorem. In particular, we may assume $\frac{b}{a} \leq \frac{4\lambda-1}{4\lambda-2}$.

Let β be the density corresponding to ν, and $\beta_s = \beta * \gamma_s$ the density corresponding to ν_s. Recalling that $\varphi(x) = x^\lambda - 1$, we have that $\varphi(f_s) = \left(\frac{\beta_s}{\rho_s}\right)^\lambda - 1$, so that

$$\frac{d}{ds}\varphi(f_s)\rho_s = \frac{d}{ds}\left(\beta_s^\lambda \rho_s^{1-\lambda} - \rho_s\right) \tag{107}$$

$$= \lambda\beta_s' \left(\frac{\beta_s}{\rho_s}\right)^{\lambda-1} + (1-\lambda)\rho_s' \left(\frac{\beta_s}{\rho_s}\right)^\lambda - \rho_s' \tag{108}$$

$$= \rho_s \left(\frac{\beta_s}{\rho_s}\right)^\lambda \left(\lambda \cdot \frac{\beta_s'}{\beta_s} + (1-\lambda) \cdot \frac{\rho_s'}{\rho_s}\right) - \rho_s'. \tag{109}$$

29

Thus,

$$\int_{\mathbb{R}^n} \sup_{s \in [a,b]} |\frac{d}{ds}\varphi(f_s)\rho_s| \tag{110}$$

$$\leq \lambda \underbrace{\int_{\mathbb{R}^n} \sup_{s \in [a,b]} \left| \rho_s \left(\frac{\beta_s}{\rho_s}\right)^\lambda \frac{\beta_s'}{\beta_s} \right|}_{I} + |1-\lambda| \underbrace{\int_{\mathbb{R}^n} \sup_{s \in [a,b]} \left| \rho_s \left(\frac{\beta_s}{\rho_s}\right)^\lambda \frac{\rho_s'}{\rho_s} \right|}_{II} + \underbrace{\int_{\mathbb{R}^n} \sup_{s \in [a,b]} |\rho_s'|}_{III}. \tag{111}$$

Before bounding integrals I, II, and III, we develop explicit expressions for ρ_s, β_s, ρ_s' and β_s'. Recall that

$$\rho_s(y) = \mathbb{E}_\mu[\gamma_s(y - X)], \tag{112}$$

$$\frac{d}{ds}\gamma_s(y - X) = \frac{1}{s}\gamma_s(y - X)\left[\frac{|y - X|^2}{2s} - \frac{n}{2}\right]. \tag{113}$$

Differentiation under the integral sign is valid, see e.g. [6], and therefore

$$\rho_s'(y) = \frac{d}{ds}\rho_s(y) = \mathbb{E}_\mu\left[\frac{d}{ds}\gamma_s(y - X)\right] = \frac{1}{s}\left[\mathbb{E}_\mu\left[\frac{|y - X|^2}{2s}\gamma_s(y - X)\right] - \frac{n}{2}\rho_s(y)\right]. \tag{114}$$

Similarly,

$$\beta_s(y) = \mathbb{E}_\nu[\gamma_s(y - X)], \quad \beta_s'(y) = \frac{d}{ds}\beta_s(y) = \frac{1}{s}\left[\mathbb{E}_\nu\left[\frac{|y - X|^2}{2s}\gamma_s(y - X)\right] - \frac{n}{2}\beta_s(y)\right]. \tag{115}$$

We begin with showing that the integral I is finite. Using the Cauchy-Schwartz inequality, we have

$$\int_{\mathbb{R}^n} \sup_{s \in [a,b]} \left| \rho_s \left(\frac{\beta_s}{\rho_s}\right)^\lambda \frac{\beta_s'}{\beta_s} \right| \leq \left(\int_{\mathbb{R}^n} \sup_{s \in [a,b]} \left| \rho_s \left(\frac{\beta_s}{\rho_s}\right)^{2\lambda} \right| \cdot \int_{\mathbb{R}^n} \sup_{s \in [a,b]} \left| \rho_s \left(\frac{\beta_s'}{\beta_s}\right)^2 \right| \right)^{1/2}. \tag{116}$$

We bound each of the two integrals in the product above. We have that

$$\sup_{s \in [a,b]} \left| \rho_s \left(\frac{\beta_s}{\rho_s}\right)^{2\lambda}(y) \right| \leq \sup_{s \in [a,b]} (2\pi s)^{-n/2} \sup_{s \in [a,b]} \left(\mathbb{E}_\nu\left[\exp\left(-\frac{1}{2s}|y - X|^2\right)\right] \right)^{2\lambda} \tag{117}$$

$$\cdot \sup_{s \in [a,b]} \left(\mathbb{E}_\mu\left[\exp\left(-\frac{1}{2s}|y - X|^2\right)\right] \right)^{1-2\lambda} \tag{118}$$

$$= (2\pi a)^{-n/2} \left(\mathbb{E}_\nu\left[\exp\left(-\frac{1}{2b}|y - X|^2\right)\right] \right)^{2\lambda} \left(\mathbb{E}_\mu\left[\exp\left(-\frac{1}{2a}|y - X|^2\right)\right] \right)^{1-2\lambda} \tag{119}$$

$$= C\rho_a^{1-2\lambda}\beta_b^{2\lambda} = C\rho_a\left(\frac{\beta_b}{\rho_a}\right)^{2\lambda} = C\rho_a\left(\frac{\rho_b}{\rho_a}\frac{\beta_b}{\rho_b}\right)^{2\lambda} \tag{120}$$

$$= C\rho_a\left[\left(\frac{\rho_b}{\rho_a}\right)^{2\lambda-1/2}\right]\left[\left(\frac{\rho_b}{\rho_a}\right)^{1/2}\left(\frac{\beta_b}{\rho_b}\right)^{2\lambda}\right] \tag{121}$$

30

where $C = (b/a)^{n\lambda}$. With this, we may apply the Cauchy-Schwartz inequality and obtain

$$\int_{\mathbb{R}^n} \sup_{s \in [a,b]} \left| \rho_s \left(\frac{\beta_s}{\rho_s} \right)^{2\lambda} \right| \le C \left(\mathbb{E}_{\mu_a} \left[\left(\frac{\rho_b}{\rho_a} \right)^{4\lambda-1} \right] \mathbb{E}_{\mu_b} \left[\left(\frac{\beta_b}{\rho_b} \right)^{4\lambda} \right] \right)^{1/2}. \qquad (122)$$

We have that $\mathbb{E}_{\mu_b} \left[\left(\frac{\beta_b}{\rho_b} \right)^{4\lambda} \right] = 1 + D_{\varphi_{4\lambda}}(\nu_b \| \mu_b)$ is finite due to Proposition A.1. Furthermore, by the data-processing inequality

$$\mathbb{E}_{\mu_a} \left[\left(\frac{\rho_b}{\rho_a} \right)^{4\lambda-1} \right] = 1 + D_{\varphi_{4\lambda-1}}(\mu_b \| \mu_a) \le 1 + D_{\varphi_{4\lambda-1}}(\mathcal{N}(0, b \cdot \mathrm{Id}) \| \mathcal{N}(0, a \cdot \mathrm{Id})). \qquad (123)$$

It is easy to see that $D_{\varphi_\lambda}(\mathcal{N}(0, \sigma_0^2 \cdot \mathrm{Id}) \| \mathcal{N}(0, \sigma_1^2 \cdot \mathrm{Id}) < \infty$ provided that $\frac{\sigma_0^2}{\sigma_1^2} < \frac{\lambda}{\lambda-1}$, see [45, eq. (10)]. Thus, $\mathbb{E}_{\mu_a} \left[\left(\frac{\rho_b}{\rho_a} \right)^{4\lambda-1} \right] < \infty$, since we assumed $\frac{b}{a} \le \frac{4\lambda-1}{4\lambda-2}$.

We move on to upper bounding the second integral in (116). Using (115) we have that

$$\left| \frac{\beta_s'}{\beta_s} \right| = \frac{1}{s} \left| \frac{\mathbb{E}_\nu \left[\frac{|y-X|^2}{2s} \gamma_s(y-X) \right] - \frac{n}{2} \beta_s}{\beta_s} \right| \le \frac{1}{s} \left[\frac{n}{2} + \frac{\mathbb{E}_\nu \left[\frac{|y-X|^2}{2s} \gamma_s(y-X) \right]}{\beta_s} \right] \qquad (124)$$

The function $|y - X| \mapsto \frac{|y-X|^2}{2s}$ is increasing, while the function $|y - X| \mapsto \gamma_s(|y - X|)$ is decreasing. Consequently,

$$\mathbb{E}_\nu \left[\frac{|y-X|^2}{2s} \gamma_s(y-X) \right] \le \mathbb{E}_\nu \left[\frac{|y-X|^2}{2s} \right] \mathbb{E}_\nu \left[\gamma_s(y-X) \right] = \beta_s \cdot \mathbb{E}_\nu \left[\frac{|y-X|^2}{2s} \right]. \qquad (125)$$

Using the Cauchy Schwartz inequality, we can further bound

$$\mathbb{E}_\nu \left[\frac{|y-X|^2}{2s} \right] = \mathbb{E}_\mu \left[\frac{d\nu}{d\mu}(X) \frac{|y-X|^2}{2s} \right] \le \left(\mathbb{E}_\mu \left[\left(\frac{d\nu}{d\mu} \right)^2 \right] \cdot \mathbb{E}_\mu \left[\frac{|y-X|^4}{4s^2} \right] \right)^{1/2} \qquad (126)$$

$$= \left((\chi^2(\nu \| \mu) + 1) \cdot \mathbb{E}_\mu \left[\frac{|y-X|^4}{4s^2} \right] \right)^{1/2} \qquad (127)$$

$$\le c \cdot \left(\mathbb{E}_\mu \left[\frac{|y-X|^4}{4s^2} \right] \right)^{1/2}, \qquad (128)$$

where the fact that $\chi^2(\nu \| \mu)$ is bounded follows from Proposition A.1. Thus,

$$\left| \frac{\beta_s'}{\beta_s} \right| \le \frac{1}{s} \left[\frac{n}{2} + c \cdot \left(\mathbb{E}_\mu \left[\frac{|y-X|^4}{4s^2} \right] \right)^{1/2} \right]. \qquad (129)$$

In particular, for $s \in [a, b]$, we have

$$\left| \frac{\beta_s'}{\beta_s} \right|^2 \leq \frac{1}{a^2} \left[\frac{n^2}{4} + cn \left(\mathbb{E}_\mu \left[\frac{|y - X|^4}{4a^2} \right] \right)^{1/2} + c^2 \cdot \mathbb{E}_\mu \left[\frac{|y - X|^4}{4a^2} \right] \right] \quad (130)$$

$$= c_1 + c_2 \left(\mathbb{E}_\mu \left[\frac{|y - X|^4}{4b^2} \right] \right)^{1/2} + c_3 \mathbb{E}_\mu \left[\frac{|y - X|^4}{4b^2} \right]. \quad (131)$$

Noting further that for all $s \in [a, b]$ we have that $\rho_s \leq (b/a)^{n/2} \rho_b$, it holds that

$$\int_{\mathbb{R}^n} \sup_{s \in [a,b]} \left| \rho_s \left(\frac{\beta_s'}{\beta_s} \right)^2 \right| \leq \int_{\mathbb{R}^n} c_1' \rho_b(y) + c_2' \rho_b(y) \left(\mathbb{E}_\mu \left[\frac{|y - X|^4}{4b^2} \right] \right)^{1/2} + c_3' \rho_b(y) \mathbb{E}_\mu \left[\frac{|y - X|^4}{4b^2} \right] dy \quad (132)$$

$$\leq c_1' + c_2' \left(\mathbb{E}_{Y \sim \rho_b} \mathbb{E}_{X \sim \mu} \left[\frac{|Y - X|^4}{4b^2} \right] \right)^{1/2} + c_3' \mathbb{E}_{Y \sim \rho_b} \mathbb{E}_{X \sim \mu} \left[\frac{|Y - X|^4}{4b^2} \right], \quad (133)$$

where we have used Jensen's inequality above. The expression above is finite by our assumption that $\mathbb{E}_\mu |X|^4 < \infty$. Thus, integral I is finite.

The proof that integral II is finite is nearly identical, where the only difference is that we do not need the change of measure from ν to μ "trick" we have used in (126)-(128), and instead we have the trivial bound $\mathbb{E}_\mu \left[\frac{|y-X|^2}{2s} \right] \leq \left(\mathbb{E}_\mu \left[\frac{|y-X|^4}{4s^2} \right] \right)^{1/2}$.

We are left with showing that integral III converges. This follows since

$$\int_{\mathbb{R}^n} \sup_{s \in [a,b]} |\rho_s'| = \int_{\mathbb{R}^n} \sup_{s \in [a,b]} \rho_s \left| \frac{\rho_s'}{\rho_s} \right| \leq \left(\int_{\mathbb{R}^n} \sup_{s \in [a,b]} \rho_s \cdot \int_{\mathbb{R}^n} \sup_{s \in [a,b]} \rho_s \left| \frac{\rho_s'}{\rho_s} \right|^2 \right)^{1/2} \quad (134)$$

$$\leq (b/a)^{n/2} \left(\int_{\mathbb{R}^n} \sup_{s \in [a,b]} \rho_s \left| \frac{\rho_s'}{\rho_s} \right|^2 \right)^{1/2}, \quad (135)$$

and we have already shown that this integral converges.

A.4 Proof of that (69) implies (70)

Let $\lambda > 1$ and $\varphi_\lambda(x) = x^\lambda - 1$. By (69), we have that for any $s > 0$,

$$\frac{d}{ds} D_{\varphi_\lambda}(\nu_s \parallel \mu_s) = -\frac{1}{2} J_{\varphi_\lambda}(\nu_s \parallel \mu_s). \quad (136)$$

where

$$J_{\varphi_\lambda}(\nu \parallel \mu) = \int_{\mathbb{R}^n} \varphi_\lambda''(f) |\nabla f|^2 d\mu = \lambda(\lambda - 1) \int_{\mathbb{R}^n} f^{\lambda-1} \frac{|\nabla f|^2}{f} d\mu,$$

where $f = d\nu/d\mu$. Recalling the definition of Rényi divergence (87), we have that for any $\lambda > 1$

$$\frac{d}{ds} D_\lambda(\nu_s \parallel \mu_s) = -\frac{1}{2} \frac{\lambda \int_{\mathbb{R}^n} f_s^{\lambda-1} \frac{|\nabla f_s|^2}{f_s} d\mu_s}{1 + D_{\varphi_\lambda}(\nu_s \parallel \mu_s)} \triangleq G(\lambda, s). \quad (137)$$

32

We have that $G(\lambda, s)$ is continuous in both $\lambda > 1$ and $s > 0$, and that $\lim_{\lambda \to 1} G(\lambda, s) = -\frac{1}{2} J(\nu_s \| \mu_s)$. Recall that

$$D_{\mathrm{KL}}(\nu_s \| \mu_s) = \lim_{\lambda \to 1} D_\lambda(\nu_s \| \mu_s). \tag{138}$$

The theorem then follows by exchanging of limits in

$$\frac{d}{ds} \lim_{\lambda \to 1} D_\lambda(\nu_s \| \mu_s) = \lim_{\lambda \to 1} \frac{d}{ds} D_\lambda(\nu_s \| \mu_s) = \lim_{\lambda \to 1} G(\lambda, s) = -\frac{1}{2} J(\nu_s \| \mu_s), \tag{139}$$

which is valid because $G(\lambda, s)$ is continuous in $(\lambda, s) \in [1, \infty) \times (0, \infty)$.

References

[1] Rudolf Ahlswede and Peter Gács. Spreading of sets in product spaces and hypercontraction of the markov operator. *The annals of probability*, pages 925–939, 1976.

[2] Venkat Anantharam. A variational characterization of rényi divergences. *IEEE Transactions on Information Theory*, 64(11):6979–6989, 2018.

[3] Efe Aras, Kuan-Yun Lee, Ashwin Pananjady, and Thomas A Courtade. A family of Bayesian Cramér-Rao bounds, and consequences for log-concave priors. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 2699–2703. IEEE, 2019.

[4] Shahab Asoodeh, Mario Diaz, Fady Alajaji, and Tamás Linder. Estimation efficiency under privacy constraints. *IEEE Transactions on Information Theory*, 65(3):1512–1534, 2018.

[5] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*. Springer, 2014.

[6] Andrew R Barron. Monotonic central limit theorem for densities. *Department of Statistics, Stanford University, California, Tech. Rep*, 50, 1984.

[7] Franck Barthe and Bo'az Klartag. Spectral gaps, symmetries and log-concave perturbations. *Bulletin of the Hellenic Mathematical Society*, 64:1–31, 2020.

[8] N. Blachman. The convolution inequality for entropy powers. *IEEE Transactions on Information Theory*, 11(2):267–271, 1965.

[9] Sergey G Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *The Annals of Probability*, 27(4):1903–1921, 1999.

[10] Eric A Carlen. Superadditivity of Fisher's information and logarithmic sobolev inequalities. *Journal of Functional Analysis*, 101(1):194–211, 1991.

[11] Djalil Chafaï. Entropies, convexity, and functional inequalities, On Φ-entropies and Φ-Sobolev inequalities. *Journal of Mathematics of Kyoto University*, 44(2):325 – 363, 2004.

[12] J. Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems in analysis, Princeton Univ. Press*, pages 195–199, 1970.

[13] Max Costa. A new entropy power inequality. *IEEE Transactions on Information Theory*, 31(6):751–760, 1985.

[14] Thomas A Courtade. A strong entropy power inequality. *IEEE Transactions on Information Theory*, 64(4):2173–2192, 2017.

[15] Thomas A Courtade. Bounds on the Poincaré constant for convolution measures. *Ann. Inst. H. Poincaré Probab. Statist.*, 56(1):566–579, 2020.

[16] TM Cover and Joy A Thomas. *Elements of information theory*. Hoboken, NJ: Wiley-Interscience, 2006.

[17] Imre Csiszár. On information-type measure of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[18] Yu S Davidovich, Boris I Korenblum, and Boris Isakovich Khatset. A certain property of logarithmically concave functions. In *Doklady Akademii Nauk*, volume 185, pages 1215–1218. Russian Academy of Sciences, 1969.

[19] A. Dembo, T.M. Cover, and J.A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, 1991.

[20] Amir Dembo. Simple proof of the concavity of the entropy power with respect to added Gaussian noise. *IEEE Transactions on Information Theory*, 35(4):887–888, 1989.

[21] Flavio du Pin Calmon, Yury Polyanskiy, and Yihong Wu. Strong data processing inequalities for input constrained additive noise channels. *IEEE Transactions on Information Theory*, 64(3):1879–1892, 2017.

[22] Richard D Gill and Boris Y Levit. Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, pages 59–79, 1995.

[23] Mikhael Gromov and Vitali D Milman. A topological application of the isoperimetric inequality. *American Journal of Mathematics*, 105(4):843–854, 1983.

[24] Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE transactions on information theory*, 51(4):1261–1282, 2005.

[25] Dongning Guo, Shlomo Shamai, and Sergio Verdú. Proof of entropy power inequalities via mmse. In *2006 IEEE International Symposium on Information Theory*, pages 1011–1015. IEEE, 2006.

[26] P. Ivanisvili and A. Volberg. Isoperimetric functional inequalities via the maximum principle: the exterior differential systems approach. *Oper. Theory Adv. Appl.*, 261:281–305, 2018.

[27] Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13:541–559, 1995.

[28] Bo'az Klartag. Logarithmic bounds for isoperimetry and slices of convex sets. *Ars Inveniendi Analytica*, Paper No. 4:17pp, 2023.

[29] Bo'az Klartag and Eli Putterman. Spectral monotonicity under Gaussian convolution. *arXiv preprint arXiv:2107.09496*, 2021.

[30] R. Latała and K. Oleszkiewicz. Between Sobolev and Poincaré. *Geom. Aspects of Funct. Anal., Israel seminar, Lecture Notes in Math.*, Vol. 1745:147–168, 2000.

[31] R. Latała and J. O. Wojtaszczyk. On the infimum convolution inequality. *Studia Math.*, 189:147–187, 2008.

[32] M. Ledoux. A simple analytic proof of an inequality by P. Buser. *Proc. Amer. Math. Soc.*, 121:951–959, 1994.

[33] Michel Ledoux. *The concentration of measure phenomenon*. American Mathematical Soc., 2001.

[34] Gilles Pisier. *The volume of convex bodies and Banach space geometry*, volume 94. Cambridge University Press, 1999.

[35] Yury Polyanskiy and Yihong Wu. Dissipation of information in channels with input constraints. *IEEE Transactions on Information Theory*, 62(1):35–55, 2015.

[36] Yury Polyanskiy and Yihong Wu. Strong data-processing inequalities for channels and bayesian networks. In *Convexity and Concentration*, pages 211–249. Springer, 2017.

[37] Yury Polyanskiy and Yihong Wu. Information theory: From coding to learning. *Book draft*, 2022.

[38] Maxim Raginsky. Strong data processing inequalities and ϕ-sobolev inequalities for discrete channels. *IEEE Transactions on Information Theory*, 62(6):3355–3389, 2016.

[39] Alfréd Rényi. On measures of dependence. *Acta mathematica hungarica*, 10(3-4):441–451, 1959.

[40] Olivier Rioul. Information theoretic proofs of entropy power inequalities. *IEEE transactions on information theory*, 57(1):33–55, 2010.

[41] Ofer Shayevitz. On rényi measures and hypothesis testing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 894–898. IEEE, 2011.

[42] Sasha Sodin. An isoperimetric inequality on the ℓ_p balls. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44:362–373, 2008.

[43] Aart J Stam. Some inequalities satisfied by the quantities of information of fisher and shannon. *Information and Control*, 2(2):101–112, 1959.

[44] Harry L Van Trees and Kristine L Bell. *Bayesian bounds for parameter estimation and nonlinear filtering/tracking*. Wiley-IEEE press, 2007.

[45] Tim van Erven and Peter Harremos. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

[46] Harry L Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.

[47] Cédric Villani. A short proof of the concavity of entropy power". *IEEE Transactions on Information Theory*, 46(4):1695–1696, 2000.

[48] Ehud Weinstein and Anthony J Weiss. A general class of lower bounds in parameter estimation. *IEEE Transactions on Information Theory*, 34(2):338–342, 1988.

[49] Ian Zieder, Alex Dytso, and Martina Cardone. An mmse lower bound via poincaré inequality. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 957–962. IEEE, 2022.

[50] Jacob Ziv and Moshe Zakai. Some lower bounds on signal parameter estimation. *IEEE transactions on Information Theory*, 15(3):386–391, 1969.