hal-00277936, version 1 - 7 May 2008

On fine properties of mixtures with respect to concentration of measure and Sobolev type inequalities

Djalil Chafaï and Florent Malrieu

Preprint – May 2008

Abstract

Mixtures are convex combinations of laws. Despite this simple definition, a mixture can be far more subtle than its mixed components. For instance, mixing Gaussian laws may produce a wild potential with multiple wells. We study in the present work fine properties of mixtures with respect to concentration of measure and Gross type functional inequalities. We provide sharp Laplace bounds for Lipschitz functions in the case of generic mixtures, involving a transportation cost diameter of the mixed family. We also provide precise upper bounds for two-components mixtures. Additionally, our analysis of Gross type inequalities for two-components mixtures reveals natural relations with some kind of band isoperimetry and support constrained interpolation via mass transportation. We show that the Poincaré constant of a two-components mixture may remain bounded as the mixture proportion goes to 0 or 1 while the Gross constant may surprisingly blow up. Additionally, this counter-intuitive result is not reducible to support disconnections. As far as mixture of distributions are concerned, the Gross inequality is less stable than the sub-Gaussian concentration for Lipschitz functions. We illustrate our results on a gallery of concrete two-components mixtures.

Keywords. Mixtures of distributions; finite Gaussian mixtures; functional inequalities; concentration of measure; transportation of measure; Sobolev type inequalities; Gaussian bounds; tails probabilities; deviation inequalities.

AMS-MSC. 60E15; 49Q20; 46E35; 62E99

Contents

1 Introduction **2**

2 General Laplace bounds for Lipschitz functions **7**

3 Concentration bounds for two-components mixtures **10**

4 Gross-Poincaré inequalities for two-components mixtures **14**
4.1 Decomposition of the variance and entropy of the mixture 14
4.2 Control of the mean-difference in dimension one 15
4.3 Control of the Poincaré and Gross constants 16
4.4 The fundamental example of two Gaussians with identical variance 17
4.5 Gallery of examples of one-dimensional two-components mixtures 19
4.5.1 Two Gaussians with identical mean 20
4.5.2 Two Uniforms with overlapping supports 21
4.5.3 One Gaussian and a sub-Gaussian 21
4.5.4 One Gaussian and a uniform 22

4.5.5 Surprising blowing mixture 22
4.6 Multivariate mean-difference bound 23

References **26**

1 Introduction

Mixtures of distributions are ubiquitous in Stochastic Analysis, Modelling, Simulation, and Statistics, see for instance the monographs [19, 21, 43, 44, 54]. Recall that a mixture of distributions is nothing else but a *convex combination* of these distributions. For instance, if μ_0 and μ_1 are two laws on the same space, and if $p \in [0,1]$ and $q = 1-p$, then the law $p\mu_1 + q\mu_0$ is a "two-components mixture". More generally, a *finite mixture* takes the form $p_1\mu_1 + \cdots + p_n\mu_n$ where $\mu_1, \ldots, \mu_n$ are probability measures on a common measurable space and $p_1\delta_1 + \cdots + p_n\delta_n$ is a finite discrete probability measure. A widely used example is given by finite mixtures of Gaussians for which $\mu_i = \mathcal{N}(m_i, \sigma_i^2)$ for every $1 \leqslant i \leqslant n$. In that case, for certain choices of $m_1, \ldots, m_n$ and $\sigma_1, \ldots, \sigma_n$, the mixture

$$p_1\mathcal{N}(m_1, \sigma_1^2) + \cdots + p_n\mathcal{N}(m_n, \sigma_n^2)$$

is multi-modal and its log-density is a multiple wells potential. For instance, each component μ_i may correspond typically in Statistics to a sub-population, in Information Theory to a channel, and in Statistical Physics to an equilibrium. A very natural example is given by the invariant measures of finite Markov chains, which are mixtures of the invariant measures uniquely associated to each recurrent classes of the chain. A more subtle example is the local field of the Sherrington-Kirkpatrick model of spin glasses which gives rise to a mixture of two univariate Gaussians with equal variances, see for instance [16].

At this point, it is enlightening to introduce a bit more abstract point of view. Let ν be a probability measure on some measurable space Θ and $(\mu_\theta)_{\theta\in\Theta}$ be a collection of probability measures on some common fixed measurable space $\mathcal{X}$, such that the map $\theta \mapsto \mathbf{E}_{\mu_\theta} f$ is measurable for any fixed bounded continuous $f : \mathcal{X} \to \mathbb{R}$. The mixture $\mathcal{M}(\nu, \mu_{\theta\in\Theta})$ is the law on $\mathcal{X}$ defined for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$ by

$$\mathbf{E}_{\mathcal{M}(\nu,\mu_{\theta\in\Theta})} f = \int_\Theta \int_{\mathcal{X}} f(x)\, d\mu_\theta(x)\, d\nu(\theta) = \mathbf{E}_\nu(\theta \mapsto \mathbf{E}_{\mu_\theta} f).$$

Here ν is the *mixing law* whereas $(\mu_\theta)_{\theta\in\Theta}$ are the *mixed laws* or the *mixture components* or even the *mixed family*. With these new notations, and for the finite mixture example mentioned earlier we have $\Theta = \{1, \ldots, n\}$ and $\nu = p_1\delta_1 + \cdots + p_n\delta_n$ and

$$\mathcal{M}(\nu, (\mu_\theta)_{\theta\in\Theta}) = \mathcal{M}(p_1\delta_1 + \cdots + p_n\delta_n, \{\mu_1, \ldots, \mu_n\}) = p_1\mu_1 + \cdots + p_n\mu_n.$$

The mixture $\mathcal{M}(\nu, \mu_{\theta\in\Theta})$ can be seen as a sort of general convex combination in the convex set of probability measures on $\mathcal{X}$. It appears for certain class of ν as a particular Choquet's Integral, see [48] and [20]. On the other hand, the case where the mixture components are product measures is also related to exchangeability and De Finetti's Theorem, see for instance [9]. In terms of random variables, if (X, Y) is a couple of random variables then the law $\mathcal{L}(X)$ of X is a mixture of the family of conditional laws $\mathcal{L}(X|Y = y)$ with the mixing law $\mathcal{L}(Y)$. By this way, mixing appears as the dual of the so-called disintegration

of measure. Here and in the whole sequel, the term "mixing" refers to the mixture of distributions as defined above and has a priori nothing to do with weak dependence.

The first aim of the present work is to investigate the fine behavior of concentration of measure for mixtures, for instance the behavior of the concentration of measure of a two-components mixture $p\mu_1 + q\mu_0$ as $\min(p, q)$ goes to 0. It is well known that Poincaré and Gross (logarithmic Sobolev) functional inequalities are powerful tools in order to obtain concentration of measure. Also, our second aim is to investigate the fine behavior of the Poincaré and Gross inequalities for mixtures, and in particular for two-components mixtures. Our work reveals striking unexpected phenomena. In particular, our results suggest that the Gross inequality, which implies sub-Gaussian concentration, is very sensitive to mixing, in contrast with the sub-Gaussian concentration itself which is far more stable. As in [23] and [4], our work is connected to the more general problem of the behavior of optimal constants for sequences of probability measures.

Let us start with the notion of *concentration of measure* for Lipschitz functions. Recall that a function $f : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz when

$$\|f\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} < \infty.$$

Let μ be a probability measure on $\mathbb{R}^d$ such that $\mathbf{E}_\mu |f| < \infty$ for every Lipschitz function f. This holds true for instance when μ has a finite first moment. We always make implicitly this assumption in the sequel. We define now the log-Laplace $\alpha_\mu : \mathbb{R} \to [0, \infty]$ of μ by

$$\alpha_\mu(\lambda) = \log \sup_{\|f\|_{\mathrm{Lip}} \leqslant 1} \mathbf{E}_\mu\Big(e^{\lambda(f - \mathbf{E}_\mu f)}\Big). \tag{1}$$

The Cramér-Chernov-Chebychev inequality gives immediately for every real number $r > 0$

$$\beta_\mu(r) = \sup_{\|f\|_{\mathrm{Lip}} \leqslant 1} \mu(|f - \mathbf{E}_\mu f| \geqslant r) \leqslant 2 \exp\left(\sup_{\lambda > 0} (r\lambda - \alpha_\mu(\lambda))\right). \tag{2}$$

Note that β_μ is a uniform upper bound on the tails probabilities of Lipschitz images of μ. We are interested in the control of β_μ via α_μ in the case where $\mu = \mathcal{M}(\nu, (\mu_\theta)_{\theta \in \Theta})$, in terms of the mixing law ν and of the log-Laplace bounds $(\alpha_{\mu_\theta})_{\theta \in \Theta}$ for the mixed family.

We say that μ satisfies to *sub-Gaussian concentration of measure* for Lipschitz functions when there exists a constant $C \in (0, \infty)$ such that for every real number λ

$$\alpha_\mu(\lambda) \leqslant \frac{1}{4} C \lambda^2. \tag{3}$$

A direct consequence of this log-Laplace-Lipschitz quadratic bound is that for every $r > 0$

$$\beta_\mu(r) \leqslant 2 \exp\left(-\frac{r^2}{C}\right). \tag{4}$$

Actually, it was shown (see [18] and [12]) that up to constants, (3) and (4) are equivalent, and are also equivalent to the existence of a constant $\varsigma \in (0, \infty)$ and $x_0 \in \mathbb{R}^d$ such that

$$\int_{\mathbb{R}^d} e^{\varsigma |x - x_0|^2} \, \mu(dx) < \infty. \tag{5}$$

Linear or quadratic upper bounds for α_μ may be deduced from functional inequalities such as Poincaré inequalities and Gross (logarithmic Sobolev) inequalities [26, 27]. More precisely, let us focus on the case where μ is a law on $\mathbb{R}^d$. We say that μ satisfies to a Poincaré inequality of constant $C \in (0,\infty)$ when for every smooth $h : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbf{Var}_\mu(h) \leqslant C\,\mathbf{E}(|\nabla h|^2) \tag{6}$$

where $\mathbf{Var}_\mu(h) = \mathbf{E}_\mu(h^2) - (\mathbf{E}_\mu h)^2$ is the variance of h for μ. The smallest possible C is called the *optimal Poincaré constant of* μ and is denoted $C_{\mathrm{PI}}(\mu)$ with the convention $\inf \emptyset = \infty$. Similarly, μ satisfies to a Gross inequality of constant $C \in (0,\infty)$ when

$$\mathbf{Ent}_\mu(h^2) \leqslant C\,\mathbf{E}(|\nabla h|^2) \tag{7}$$

for every smooth function $f : \mathbb{R}^d \to \mathbb{R}$, where $\mathbf{Ent}_\mu(h^2) = \mathbf{E}_\mu(h^2 \log h^2) - \mathbf{E}_\mu(h^2) \log \mathbf{E}_\mu(h^2)$ is the "entropy" of h^2 for μ (with the convention $0 \log(0) = 0$). As for the Poincaré inequality, the smallest possible C is the *optimal Gross constant of* μ and is denoted $C_{\mathrm{GI}}(\mu)$ with $\inf \emptyset = \infty$. Standard linearization arguments give that

$$\rho(K_\mu) \leqslant C_{\mathrm{PI}}(\mu) \leqslant \frac{1}{2}\, C_{\mathrm{GI}}(\mu) \tag{8}$$

where $\rho(K_\mu)$ stands for the spectral radius of the covariance matrix K_μ of μ defined by $(K_\mu)_{i,j} = \mathbf{E}_\mu(x_i x_j) - \mathbf{E}_\mu(x_i)\mathbf{E}_\mu(x_j)$ where x_i and x_j are the coordinates functions. A wide class of laws satisfy to Poincaré and Gross inequalities. A basic example is given by Gaussian laws for which equalities are achieved in (8). Note that $C_{\mathrm{GI}}(\mu) < \infty$ implies $C_{\mathrm{PI}}(\mu) < \infty$ but (10) shows that the converse is false. Note also that if μ has disconnected support, then necessarily $C_{\mathrm{PI}}(\mu) = C_{\mathrm{GI}}(\mu) = \infty$. To see it, just consider a non constant function h which is constant on each connected component of the support of μ. This is for instance the case for for the two-component mixture $\mu = p\mu_1 + q\mu_0 = \mathcal{M}(p\delta_1 + q\delta_0, \{\mu_0, \mu_1\})$ with $p \in (0,1)$ and $q = 1-p$ where μ_0 and μ_1 have disjoint supports.

The Gross inequality (7) encloses a sub-Gaussian concentration of measure for Lipschitz images of μ. Namely, using (7) with $h = \exp(\frac{1}{2}\lambda f)$ for a real number λ and a smooth Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$ gives via Rademacher's Theorem and a standard argument attributed to Herbst [38, Chapter 5] that for any reals λ and $r > 0$

$$\alpha_\mu(\lambda) \leqslant \frac{1}{4} C_{\mathrm{GI}} \lambda^2 \quad \text{and} \quad \beta_\mu(r) \leqslant 2 \exp\left(-\frac{r^2}{C_{\mathrm{GI}}}\right). \tag{9}$$

The same method yields from (6) a sub-exponential upper bound for β_μ of the form $c_1 \exp(-c_2 r)$, see [25] and [37, Section 2.5]. Beyond Gaussian laws, a criterion due to Bakry and Émery [3, 2] (see also [46], and the ideas of Stam [51, 15]) states that if μ has density e^{-V} on $\mathbb{R}^d$ such that $x \mapsto V(x) - \frac{1}{2\kappa}|x|^2$ is convex for some fixed real $\kappa > 0$ then

$$C_{\mathrm{PI}}(\mu) \leqslant \kappa \quad \text{and} \quad C_{\mathrm{GI}}(\mu) \leqslant 2\kappa$$

with equality when μ is Gaussian. A striking result due to Caffarelli [13, 14] and based on the Brenier-McCann Theorem and the Monge-Ampère equation states that such uniformly log-concave laws are Lipschitz images of a Gaussian. It is immediate via Rademacher's Theorem that Poincaré and Gross inequalities are stable by Lipschitz maps, i.e.

$$C_{\mathrm{PI}}(f \cdot \mu) \leqslant \|f\|^2_{\mathrm{LIP}} C_{\mathrm{PI}}(\mu) \quad \text{and} \quad C_{\mathrm{GI}}(f \cdot \mu) \leqslant \|f\|^2_{\mathrm{LIP}} C_{\mathrm{GI}}(\mu)$$

where $f \cdot \mu$ denotes the image measure of μ by f. The equality is achieved if $f(x) = ax + b$ (translation and dilation). If Φ is the cumulative distribution function of the standard Gaussian measure $\mathcal{N}(0,1)$ on $\mathbb{R}$, then $\Phi \cdot \mathcal{N}(0,1) = \mathcal{U}([0,1])$ and since $\|\Phi\|_{\mathrm{Lip}} \leqslant (2\pi)^{-1}$ we get $C_{\mathrm{PI}}(\mathcal{U}([0,1])) \leqslant (2\pi)^{-1}$ and $C_{\mathrm{GI}}(\mathcal{U}([0,1])) \leqslant \pi^{-1}$. Actually, it is shown in [22] that

$$2C_{\mathrm{PI}}(\mathcal{U}([0,1])) = C_{\mathrm{GI}}(\mathcal{U}([0,1])) = \frac{2}{\pi^2}.$$

On the real line, C_{PI} and C_{GI} can be controlled via "simple" variational bounds (see (20) page 19). This allows to analyze the finiteness of these constants for laws lying between the exponential and the Gaussian. Namely, if μ_a is the law on $\mathbb{R}$ with density proportional to $x \mapsto e^{-|x|^a}$ for some fixed real $a > 0$, then (see [2, Chapter 6], [33], and [11])

$$C_{\mathrm{PI}}(\mu_a) < \infty \text{ iff } a \geqslant 1 \quad \text{while} \quad C_{\mathrm{GI}}(\mu_a) < \infty \text{ iff } a \geqslant 2. \tag{10}$$

Both Poincaré and Gross inequalities are also stable by bounded perturbations, in the sense that if μ_B has density e^B with respect to μ then, with $\mathrm{osc}(B) = \sup B - \inf B$,

$$C_{\mathrm{PI}}(\mu_B) \leqslant e^{2\mathrm{osc}(B)} C_{\mathrm{PI}}(\mu) \quad \text{and} \quad C_{\mathrm{GI}}(\mu_B) \leqslant e^{2\mathrm{osc}(B)} C_{\mathrm{GI}}(\mu).$$

These bounds, due to Holley and Stroock [30], follow from variational formulas for variance and entropy, see also [28] and [4] for further developments. In view of sub-exponential or sub-Gaussian concentration bounds, the main advantage of (6) and (7) over a direct approach based on α_μ or β_μ lies in the stability by tensor products, namely

$$C_{\mathrm{PI}}(\mu_1 \otimes \mu_2) \leqslant \max(C_{\mathrm{PI}}(\mu_1), C_{\mathrm{PI}}(\mu_2)) \quad \text{and} \quad C_{\mathrm{GI}}(\mu_1 \otimes \mu_2) \leqslant \max(C_{\mathrm{GI}}(\mu_1), C_{\mathrm{GI}}(\mu_2))$$

which leads to dimension-free sub-exponential and sub-Gaussian upper bounds for $\beta_{\mu^{\otimes n}}$ via $\alpha_{\mu^{\otimes n}}$. It is important to realize that in general, α_μ is not stable by tensor product in the sense that the upper bound on $\alpha_{\mu^{\otimes n}}$ deduced from an upper bound for α_μ may depend on n. We refer to [12] and [24] and references therein for further details.

Recall that for every $k \geqslant 1$, the Wasserstein (or transportation cost) distance of order k between two probability measures μ_1 and μ_2 on $\mathbb{R}^d$ is defined by (see [55, 56] and [49, 52])

$$W_k(\mu_1, \mu_2) = \inf_\pi \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^k \, d\pi(x,y) \right)^{k^{-1}} \tag{11}$$

where π runs over the set of laws on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ_1 and μ_2. The W_k-convergence is equivalent to the weak convergence together with the convergence of moments up to order k. In dimension $d = 1$, we have, by denoting F_1 and F_2 the cumulative distribution functions of μ_1 and μ_2, with generalized inverses F_1^{-1} and F_2^{-1}, for every $k \geqslant 1$,

$$W_k(\mu_1, \mu_2)^k = \int_0^1 \left| F_1^{-1}(x) - F_2^{-1}(x) \right|^k dx \quad \text{and} \quad W_1(\mu_1, \mu_2) = \int_{\mathbb{R}} |F_1(x) - F_2(x)| \, dx \tag{12}$$

where the last expression of W_1 follows from the Kantorovich-Rubinstein dual formulation

$$W_1(\mu_1, \mu_2) = \sup_{\|f\|_{\mathrm{Lip}} \leqslant 1} \left(\int_{\mathbb{R}^d} f \, d\mu_1 - \int_{\mathbb{R}^d} f \, d\mu_2 \right). \tag{13}$$

Note that if μ_1 does not give mass to points then $\mu_2 = (F_2^{-1} \circ F_1) \cdot \mu_1$. The transportation cost distances lead to another type of functional inequalities popularized by Marton [40, 41], Talagrand [53], and Bobkov & Götze [10]. Namely, we say that a probability measure μ on $\mathbb{R}^d$ satisfies to a W_k inequality [38, 55, 56] when there exists a constant $C_k \in (0, \infty)$ such that for every density f with respect to μ,

$$W_k(\mu, f\mu)^2 \leqslant C_k \mathbf{Ent}_\mu(f). \tag{14}$$

It is known [12, 18] that the W_1 inequality is equivalent to a sub-Gaussian concentration of measure for Lipschitz functions (3) and to a square exponential integrability (5). In particular, the W_1 inequality is not stable by tensor product (the constant may depend on the tensor power). In contrast, the W_2 inequality, which is stronger, is stable by tensor products and is actually equivalent to a sub-Gaussian concentration of measure with a constant independent of the dimension (see the recent work of Gozlan [24]).

It was shown by Otto & Villani [47] that the Gross inequality (7) implies the W_2 inequality, and that the converse is true under some log-concavity assumption. Under the same assumption, the square exponential integrability (5) implies the Gross inequality (7) (see [1, 57]). We refer to [24, 6, 56] for recent accounts on these links. Back to mixtures, we provide in Section 4.5.5 a simple example of a two-component mixture $p\mu_1 + q\mu_0$ for which the log-concave curvature and the Gross constant blow up as p goes to 0 while the sub-Gaussian concentration of measure for Lipschitz functions remains bounded.

The case of mixtures. The integral criterion (5) shows that if each component of a mixture satisfies to a sub-Gaussian concentration of measure for Lipschitz functions, and if the mixing law has compact support, then the mixture also satisfies to sub-Gaussian concentration of measure for Lipschitz functions. Such bounds appear for instance in [8]. However, this observation does not give any fine quantitative estimate on the dependency over the weights for a finite mixture. Regarding Poincaré and Gross inequalities, it is clear that a finite mixture of Gaussians will satisfies to such inequalities since its log-density is a bounded perturbation of a uniformly concave function. Here again, this does not give any fine control on the constants. Notice also that if the union of the supports of the mixture components is not connected, then the mixture cannot satisfy to a Poincaré or Gross inequality, even if each component does.

An upper bound for the Poincaré constant of univariate finite Gaussian mixture was provided by Johnson [32, Theorem 1.1 and Section 2]. Unfortunately, this upper bound blows up when the minimum weight of the mixing law goes to 0. A more general upper bound for finite mixtures of overlapping densities was obtained by Madras and Randall [39, Theorem 1.2 and Section 5]. Here again, the bound blows up when the minimum weight of the mixing law goes to 0. Some aspects of Poisson mixtures are considered by Kontoyannis and Madiman [34, 35] in connection with compound Poisson processes and discrete modified Gross inequalities.

Outline of the article. Recall that the aim of the present work is to study fine properties of mixture of distributions with respect to concentration of measure for Lipschitz functions, and also with respect to Poincaré and Gross functional inequalities. The analysis of various elementary examples shows actually that such a general objective is very ambitious. Also, we decided to focus in the present work on more tractable situations. Section 2 provides general Laplace bounds for Lipschitz functions in the case of generic mixtures. These upper bounds on α_μ (and thus β_μ) for a mixture μ involve

the W_1-diameter of the mixed family. Section 3 is devoted to upper bounds on α_μ for two-components mixtures $\mu = \mu_p = p\mu_1 + q\mu_0$. Our result is mainly based on a Laplace-Lipschitz counterpart of the optimal Gross inequality for asymmetric Bernoulli measures. In particular, we show that if μ_0 and μ_1 have sub-Gaussian concentration for Lipschitz functions, then it is also the case for the mixture μ_p, with a quite satisfactory and intuitive behavior as p goes to 0 or 1. In section 4, we study Poincaré and Gross inequalities for two components mixtures. A decomposition of variance and entropy allows to reduce the problem to the Poincaré and Gross inequalities for each component, to discrete inequalities for the Bernoulli mixing law $p\delta_1 + q\delta_0$, and to the control of a mean-difference term. This last term can be controlled in turn by using some support-constrained transportation, leading to very interesting open questions in dimension > 1. The Poincaré constant of the two-component mixture remains bounded as $\min(p, q)$ goes to 0, while the Gross constant may surprisingly blow up at speed $-\log(\min(p, q))$. This counter-intuitive result shows that as far as mixture of distributions are concerned, the Gross inequality does not behave like the sub-Gaussian concentration for Lipschitz functions. We also illustrate our results on a gallery of concrete two-components mixtures. In particular, we show that the blowing of the Gross constant as $\min(p, q)$ goes to 0 is not necessarily related to support problems!

Open problems. The study of Poincaré and Gross inequalities for multivariate or non-finite mixtures is an interesting open problem, for which we give some clues at the end of Section 4 in terms of support-constrained transportation interpolation. There is also a probably interesting link with the decomposition approach used in [31] for Markov chains. Another interesting open problem is the development of a direct approach for W_k inequalities (14) and for measure capacities inequalities [7] for mixtures, even in the finite univariate case.

2 General Laplace bounds for Lipschitz functions

Intuitively, the concentration of measure of a finite mixture may be controlled by the worst concentration of the components and some sort of diameter of the mixed family. We shall confirm, extend, and illustrate this intuition for a (non necessarily finite) mixture $\mu = \mathcal{M}(\nu, (\mu_\theta)_{\theta\in\Theta})$ by providing upper bounds for α_μ. The key point is that if $\|f\|_{\mathrm{Lip}} < \infty$ and $\lambda > 0$ then

$$\frac{\mathbf{E}_\mu\big(e^{\lambda f}\big)}{e^{\lambda \mathbf{E}_\mu f}} = e^{-\lambda \mathbf{E}_\mu f} \int_\Theta \mathbf{E}_{\mu_\theta}\Big(e^{\lambda f}\Big)\, \nu(d\theta) \leqslant \int_\Theta e^{\alpha_\theta(\lambda) + \lambda(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f)}\, \nu(d\theta). \tag{15}$$

Theorem 2.1 (Laplace bound via diameter). *If*

$$\overline{\alpha} = \sup_{\theta\in\Theta} \alpha_\theta < \infty \quad \textit{and} \quad \overline{W} = \sup_{\theta,\theta'\in\Theta} W_1(\mu_\theta, \mu_{\theta'}) < \infty$$

then for every $\lambda > 0$ we have

$$\alpha_\mu(\lambda) \leqslant \overline{\alpha}(\lambda) + \frac{1}{8} \min\Big(8\overline{W}\lambda, \overline{W}^2\lambda^2\Big).$$

Proof. The bound (15) implies that for every $\lambda > 0$,

$$\alpha_\mu(\lambda) \leqslant \overline{\alpha}(\lambda) + \sup_F \log \int_\Theta e^{\lambda(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f)}\, \nu(d\theta). \tag{16}$$

Thanks to the relation (13), we obtain

$$\begin{aligned}\mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f &= \int_\Theta \left(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_{\mu_{\theta'}} f\right) \nu(d\theta') \\ &\leqslant \int_\Theta W_1(\mu_\theta, \mu_{\theta'})\, \nu(d\theta') \leqslant \overline{W}.\end{aligned}$$

This shows that the second term in the right hand side of (16) is bounded by $\overline{W}\lambda$. Alternatively, one can use the Hoeffding bound [29] which says that if X is a centered bounded random variable with oscillation $c = \sup X - \inf X$ then

$$\mathbf{E}\Big(e^{\lambda X}\Big) \leqslant e^{\frac{1}{8}\lambda^2 c^2}.$$

The desired bound in terms of $\overline{W}^2\lambda^2$ follows by taking $X = \mathbf{E}_{\mu_Y} f - \mathbf{E}_\mu f$ where $Y \sim \nu$ and noticing that $c \leqslant \sup_{\theta,\theta'} \left(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_{\mu_{\theta'}} f\right) = \overline{W}$. □

Example 2.2 (Finite mixtures). *For a finite mixture $\mu = p_1\mu_1 + \cdots + p_n\mu_n = \mathcal{M}(\nu, (\mu_i)_{1\leqslant i\leqslant n})$ where $\nu = p_1\delta_1 + \cdots + p_n\delta_n$, the mixing measure ν is supported by a finite set. In that case, Theorem 2.1 gives an immediate Laplace bound, involving the worst bound for the mixture components $(\mu_i)_{1\leqslant i\leqslant n}$ (this cannot be improved in general). However, in Section 3, we provide sharper bounds by improving the dependency over ν in the case where $n = 2$.*

Example 2.3 (Bounded mixtures of multivariate Gaussians). *Here $\mu_\theta = \mathcal{N}(m(\theta), \Gamma(\theta))$ where $m : \Theta \to \mathbb{R}^d$ and $\Gamma : \mathbb{R}^d \to \mathcal{S}_d^+$ are two measurable bounded functions and $\mathcal{S}_d^+$ is the cone of symmetric nonnegative $d \times d$ matrices. Note that Γ is allowed to be singular (not of full rank). We denote by $\lambda_1(\theta) \geqslant \cdots \geqslant \lambda_d(\theta)$ the spectrum of $\Gamma(\theta)$ and $C = \sup_{\theta\in\Theta} \lambda_1(\theta)$. Now fix some mixing distribution ν on Θ and consider the mixture $\mu = \mathcal{M}(\nu, (\mu_\theta)_{\theta\in\Theta})$. Then for every $\lambda > 0$,*

$$\alpha_\mu(\lambda) \leqslant \frac{C}{2}\lambda^2 + \frac{1}{8}\min(8\overline{W}\lambda, \overline{W}^2\lambda^2).$$

One can deduce a bound of $\overline{W}$ from the following lemma.

Lemma 2.4. *Let $\mu_0 = \mathcal{N}(m(0), \Gamma(0))$ and $\mu_1 = \mathcal{N}(m(1), \Gamma(1))$ be two Gaussian measures on $\mathbb{R}^d$. For $\theta \in \{0, 1\}$, we denote by $\lambda_1(\theta) \geqslant \cdots \geqslant \lambda_d(\theta)$ the ordered spectrum of $\Gamma(\theta)$ and by $(v_i(\theta))_{1\leqslant i\leqslant d}$ an associated orthonormal basis of eigenvectors. Assume (without loss of generality) that $v_i(0) \cdot v_i(1) \geqslant 0$ for every $1 \leqslant i \leqslant d$ where "$\cdot$" stands for the standard Euclidean scalar product of $\mathbb{R}^d$. Then $W_1(\mu_0, \mu_1)$ is bounded above by*

$$|m(1) - m(0)| + \sqrt{\sum_{i=1}^d \left\{ \left(\sqrt{\lambda_i(1)} - \sqrt{\lambda_i(0)}\right)^2 + 2\sqrt{\lambda_i(1)\lambda_i(0)}(1 - v_i(1) \cdot v_i(0)) \right\}}.$$

Proof. The triangle inequality for the W_1 distance gives

$$\begin{aligned}W_1(\mu_0, \mu_1) &\leqslant W_1(\mu_0, \mathcal{N}(m(1), \Gamma(0))) + W_1(\mathcal{N}(m(1), \Gamma(0)), \mu_1) \\ &\leqslant |m(1) - m(0)| + W_1(\mathcal{N}(0, \Gamma(0)), \mathcal{N}(0, \Gamma(1))).\end{aligned}$$

Now, if $(Y_i)_{1\leqslant i\leqslant d}$ are i.i.d. real random variables of law $\mathcal{N}(0,1)$ then the law of

$$X_\theta = \sum_{i=1}^{d} Y_i \sqrt{\lambda_i(\theta)} v_i(\theta)$$

is $\mathcal{N}(0,\Gamma(\theta))$ for $\theta \in \{0,1\}$. Moreover, from the expression (11) we get

$$W_1(\mathcal{N}(0,\Gamma(0)),\mathcal{N}(0,\Gamma(1)))^2 \leqslant (\mathbb{E}|X_1 - X_0|)^2 \leqslant \mathbb{E}(|X_1 - X_0|^2).$$

At this step, we note that

$$\begin{aligned}|X_1 - X_0|^2 =& \sum_{i=1}^{d} Y_i^2 \Big|\sqrt{\lambda_i(1)} v_i(1) - \sqrt{\lambda_i(0)} v_i(0)\Big|^2 \\ &+ 2\sum_{i<j} Y_i Y_j \Big(\sqrt{\lambda_i(1)} v_i(1) - \sqrt{\lambda_i(0)} v_i(0)\Big) \cdot \Big(\sqrt{\lambda_i(1)} v_i(1) - \sqrt{\lambda_i(0)} v_i(0)\Big).\end{aligned}$$

Since (Y_i) are i.i.d. $\mathcal{N}(0,1)$ and $(v_i(\theta))_{1\leqslant i\leqslant d}$ is orthonormal for $\theta \in \{0,1\}$, one has

$$\begin{aligned}\mathbb{E}(|X_1 - X_0|^2) &= \sum_{i=1}^{d} \Big|\sqrt{\lambda_i(1)} v_i(1) - \sqrt{\lambda_i(0)} v_i(0)\Big|^2 \\ &= \sum_{i=1}^{d} \Big\{ \Big(\sqrt{\lambda_i(1)} - \sqrt{\lambda_i(0)}\Big)^2 + 2\sqrt{\lambda_i(1)\lambda_i(0)}(1 - v_i(1)\cdot v_i(0)) \Big\}.\end{aligned}$$

□

Of course the assumptions of Theorem 2.1 may be relaxed. Instead of trying to deal with generic abstract results, let us provide some highlighting examples.

Example 2.5 (Gaussian mixture of translated Gaussians). *Here $\Theta = \mathbb{R}$ and $\mu_\theta = \mathcal{N}(\theta,\sigma^2)$ for some fixed $\sigma > 0$, and the mixing law is also Gaussian $\nu = \mathcal{N}(0,\tau^2)$ for some fixed $\tau > 0$. In this case, $\overline{\alpha}(\lambda) = \frac{1}{2}\sigma^2\lambda^2$ but $\overline{W}$ is infinite since $W_1(\mu_\theta,\mu_{\theta'}) = |\theta - \theta'|$. In particular, Theorem 2.1 is useless. Nevertheless, the function $\theta \mapsto g(\theta) = \mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f$ is Lipschitz since*

$$\big|g(\theta) - g(\theta')\big| \leqslant \mathbf{E}\big(\big|f(X+\theta) - f(X+\theta')\big|\big) \leqslant \big|\theta - \theta'\big|,$$

where $X \sim \mathcal{N}(0,1)$. As a consequence,

$$\sup_{\|f\|_{\mathrm{Lip}}\leqslant 1} \log \int_\Theta e^{\lambda(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f)}\, \nu(d\theta) \leqslant \frac{\tau^2\lambda^2}{2},$$

and for any $\lambda > 0$

$$\alpha_\mu(\lambda) \leqslant \frac{\sigma^2 + \tau^2}{2}\lambda^2.$$

The same argument may be used more generally for position mixtures. For instance if η is some fixed probability measure on $\mathbb{R}^d$ and $\mu_\theta = \eta * \delta_\theta$ for $\theta \in \mathbb{R}^d$ then $\forall \lambda > 0$,

$$\alpha_\mu(\lambda) \leqslant \alpha_\eta(\lambda) + \alpha_\mu(\lambda).$$

In this particular case, $\mu = \nu * \eta$ and the bound above follows also by tensorization!

Example 2.6 (Mixture of scaled Gaussians: from exponential to Gaussian tails). *Here we take* $\Theta = [0, \infty)$ *and* $\mu_\theta = \mathcal{N}(0, \theta^2)$ *with a mixing measure* ν *of density*

$$\theta \mapsto \frac{\gamma}{\Gamma(\gamma^{-1})} \exp\left(-\theta^\gamma\right) \mathrm{I}_{[0,\infty)}(\theta)$$

where $\gamma \geqslant 2$ *is some fixed real number. Since* ν *has a non-compact support, the mixture* μ *cannot satisfy the integral criterion* (5)*. This means that* μ *cannot have sub-Gaussian tails. Note also that both* $\overline{\alpha}(\lambda)$ *and* $\overline{W}$ *are infinite since*

$$\alpha_\theta(\lambda) = \frac{\theta^2\lambda^2}{2} \quad \textit{and} \quad W_1(\mu_\theta, \mu_{\theta'}) = \sqrt{\frac{2}{\pi}}\left|\theta - \theta'\right|$$

where we used (12) *for* W_1*. Starting from* (15)*, one has by Cauchy-Schwarz's inequality*

$$\left(\frac{\mathbf{E}_\mu\left(e^{\lambda f}\right)}{e^{\lambda \mathbf{E}_\mu f}}\right)^2 \leqslant \int_{\mathbb{R}} e^{\theta^2\lambda^2}\, \nu(d\theta) \int_{\mathbb{R}} e^{2\lambda(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f)}\, \nu(d\theta). \tag{17}$$

Note that ν *satisfies condition* (5) *and* $\alpha_\nu(\lambda) \leqslant C\lambda^2$ *for some real constant* $C > 0$*. Here and in the sequel, the constant* C *may vary from line to line and may be chosen independent of* γ*. On the other hand, the centered function* $g(\theta) = \mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f$ *is 1-Lipschitz since*

$$\left|g(\theta) - g(\theta')\right| = \left|\mathbf{E} f(\theta X) - \mathbf{E} f(\theta' X)\right| \leqslant \left|\theta - \theta'\right| \mathbf{E}(|X|)$$

where $X \sim \mathcal{N}(0, 1)$*. This allows to bound the second term in the right hand side of* (17) *as*

$$\int_{\mathbb{R}} e^{2\lambda(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f)}\, \nu(d\theta) \leqslant e^{\alpha_\nu(2\lambda)} \leqslant e^{4C\lambda^2}.$$

If $\gamma = 2$ *then* $\alpha_\mu(\lambda) \leqslant (1 - 2\lambda^2)^{-1/2} + 2C\lambda^2$ *if* $\sqrt{2}\lambda < 1$*, which gives the deviation bound*

$$\mu(F - \mathbf{E}_\mu f \geqslant r) \leqslant e^{-Cr}.$$

Assume in contrast that $\gamma > 2$*. Since* $\theta^2\lambda^2 \leqslant \gamma^{-1}\theta^\gamma + C\lambda^{\frac{2\gamma}{\gamma-2}}$*, we have*

$$\int_0^\infty \exp\left(\theta^2\lambda^2\right) \nu(d\theta) \leqslant C \exp\left(C\lambda^{\frac{2\gamma}{\gamma-2}}\right).$$

This gives $\alpha_\mu(\lambda) \leqslant C\lambda^{\frac{2\gamma}{\gamma-2}} + C$ *which yields the deviation bound*

$$\mu(f - \mathbf{E}_\mu f \geqslant r) \leqslant C \exp\left(-Cr^{2-\frac{4}{\gamma+2}}\right).$$

Note that ν *goes to the uniform law on* $[0, 1]$ *as* $\gamma \to \infty$ *and the Gaussian tail reappears.*

3 Concentration bounds for two-components mixtures

In this section, we investigate the special case where the mixing measure ν is the Bernoulli measure $\mathcal{B}(p) = p\delta_1 + q\delta_0$ where $q = 1 - p$. We are interested in the study of the sharp dependence of the concentration bounds on p, especially when p is close to 0 or 1.

Theorem 3.1 (Two-components mixture). *Let μ_0 and μ_1 be two probability measures on $\mathcal{X}$ and $\mu = p\mu_1 + q\mu_0$ with $p \in [0,1]$ and $q = 1-p$. Define $x_p = \max(p,q)/(2c_p)$ where*

$$c_p = \frac{q-p}{4(\log(q) - \log(p))}$$

with the conventions $c_{1/2} = 1/8$ and $c_0 = c_1 = 0$. Then for any $\lambda > 0$,

$$\alpha_\mu(\lambda) \leqslant \max(\alpha_{\mu_0}, \alpha_{\mu_1})(\lambda) + \begin{cases} c_p \lambda^2 W_1(\mu_0,\mu_1)^2 & \text{if } \lambda W_1(\mu_0,\mu_1) \leqslant x_p \\ \max(p,q)(\lambda W_1(\mu_0,\mu_1) - x_p) & \text{otherwise.} \end{cases}$$

Note that if $\min(p,q) \to 0$, then $c_p \sim -(4\log(p))^{-1} \to 0$ and $x_p \to \infty$, and we thus recover an upper bound of the form $\alpha_\mu \leqslant \max(\alpha_{\mu_1}, \alpha_{\mu_2})$ as $\min(p,q) \to 0$, which is satisfactory. The two different upper bounds given by Theorem 3.1 provide two different upper bounds for the concentration of measure of the mixture μ, illustrated by the following Corollary (the proof of the Corollary is immediate and is left to the reader).

Corollary 3.2 (Two-components mixtures with sub-Gaussian tails). *Let μ_0 and μ_1 be two probability measures on $\mathcal{X}$ and $\mu = p\mu_1 + q\mu_0$ for some $p \in [0,1]$ with $q = 1-p$. If there exists a real constant $C > 0$ such that for any $\lambda > 0$*

$$\max(\alpha_{\mu_0}, \alpha_{\mu_1})(\lambda) \leqslant \frac{1}{2} C \lambda^2$$

then for every $r \geqslant 0$, with $\overline{W} = W_1(\mu_0,\mu_1)$,

$$\beta_\mu(r) \leqslant 2 \begin{cases} \exp\left(-\dfrac{r^2}{2C + 4c_p \overline{W}^2}\right) & \text{if } r \leqslant \max(p,q)\Big(\frac{C}{2c_p\overline{W}} + \overline{W}\Big), \\ \exp\left(-\dfrac{1}{2C}(r - \max(p,q)\overline{W})^2 - \dfrac{\max(p,q)^2}{4c_p}\right) & \text{otherwise.} \end{cases}$$

Proof of Theorem 3.1. We have $\mu = q\mu_0 + p\mu_1 = \mathcal{M}(\nu, \{\mu_0, \mu_1\})$ where $\nu := q\delta_0 + p\delta_1$. For this special finite mixture, we get, as in the general case, for any $F \in \mathrm{Lip}(\mathcal{X}, \mathbb{R})$ and any $\lambda > 0$,

$$\log\left(\frac{\mathbf{E}_\mu\big(e^{\lambda f}\big)}{e^{\lambda \mathbf{E}_\mu f}}\right) \leqslant \max(\alpha_{\mu_0}, \alpha_{\mu_1})(\lambda) + \log\left(\frac{\mathbf{E}_\nu\big(e^{\lambda g}\big)}{e^{\lambda \mathbf{E}_\nu g}}\right),$$

where $g(i) := \mathbf{E}_{\mu_i} f$. At this step, we use the particular nature of ν, which gives the identity

$$\frac{\mathbf{E}_\nu\big(e^{\lambda g}\big)}{e^{\lambda \mathbf{E}_\nu g}} = \cosh_p(\lambda(g(1) - g(0))),$$

where $\cosh_p(x) := pe^{qx} + qe^{-px}$. Since $g(1) - g(0) = \mathbf{E}_{\mu_1} f - \mathbf{E}_{\mu_0} f$, we get by (13)

$$-W_1(\mu_0,\mu_1) \leqslant g(1) - g(0) \leqslant W_1(\mu_0,\mu_1).$$

Since $\cosh_p(-x) = \cosh_q(x)$ for any $x \in \mathbb{R}$, we get for any $\lambda > 0$,

$$\sup_{\|F\|_{\mathrm{Lip}} \leqslant 1} \left(\frac{\mathbf{E}_\nu\big(e^{\lambda g}\big)}{e^{\lambda \mathbf{E}_\nu g}}\right) = \max\left(\cosh_p, \cosh_q\right)(\lambda W_1(\mu_0,\mu_1)).$$

Putting all together, we obtain, for any $\lambda > 0$,

$$\alpha_\mu(\lambda) \leqslant \max(\alpha_{\mu_0}, \alpha_{\mu_1})(\lambda) + \log \max\left(\cosh_p, \cosh_q\right)(\lambda W_1(\mu_0, \mu_1)),$$

Since the derivative of $\cosh_q - \cosh_p$ is equal to $pq(\sinh(q\cdot) - \sinh(q\cdot))$, one has, for every $x \geqslant 0$,

$$\max\left(\cosh_p, \cosh_q\right)(x) = \cosh_{\min(p,q)}(x).$$

Let us assume that $p \leqslant q$. Lemma 3.3 ensures that, for every $x \geqslant 0$,

$$\log \max\left(\cosh_p, \cosh_q\right)(x) = \log \cosh_p(x) \leqslant c_p x^2.$$

On the other hand,

$$\log \cosh_p(x) = qx + \log\left(p + qe^{-x}\right) \leqslant qx.$$

Now, for $x = x_p$, the slope of $x \mapsto c_p x^2$ is equal to q and the tangent is $y = q(x - x_p)$. On the other hand, the convexity of $x \mapsto \log \cosh_p(x)$ yields $\log \cosh_p(x) \leqslant q(x - x_p)$ for $x \geqslant x_p$ (drawing a picture may help the reader). The desired conclusion follows immediately. □

The proof of Theorem 3.1 relies on lemma 3.3 below which provides a Gaussian bound for the Laplace transform of a Lipschitz function with respect to a Bernoulli measure. This lemma can be seen as an improvement of the Hoeffding bound [29] in the case of a Bernoulli measure.

Lemma 3.3 (Two points lemma). *For any $0 \leqslant p \leqslant 1/2$, we have*

$$\sup_{x>0} x^{-2} \log(pe^{qx} + qe^{-px}) = c_p = \frac{q - p}{4(\log(q) - \log(p))} \tag{18}$$

with the natural conventions $c_0 = 0$ and $c_{1/2} = 1/8$ as in Theorem 3.1. Moreover, the supremum in x is achieved for $x = 2(\log(q) - \log(p))$.

The constant c_p is also equal, as it will appear in the proof, to $\sup_{\lambda>0} \alpha_{\mathcal{B}(p)}(\lambda)/\lambda^2$. The classical Hoeffding bound for this supremum is $c_{1/2} = 1/8$ which is the maximum of c_p over p. Additionally, the quantity $pq/(4c_p)$ is the optimal constant of the Gross inequality for the asymmetric Bernoulli measure $q\delta_0 + p\delta_1$.

Proof of Lemma 3.3. Let us define $x_p = \log(q/p)$ and $\beta(x) = x^{-2}\psi(x)$ where

$$\psi(x) = \log(pe^{qx} + qe^{-px}).$$

The function ψ is "strongly convex" at the origin ($\psi(0) = \psi'(0) = 0$ and $\psi''(0) = pq$ and $\psi'''(0) > 0$) and linear at infinity ($\psi'(\infty) = q$). Therefore, the supremum of β is achieved for some $x > 0$. The derivative of β has the sign of $\gamma(x) := x\psi'(x) - 2\psi(x)$. Furthermore,

$$\gamma'(x) = x\psi''(x) - \psi'(x) \quad \text{and} \quad \gamma''(x) = x\psi'''(x).$$

As a consequence, γ'' has the sign of ψ''' which is positive on $(0, x_p)$ and negative on $(x_p, +\infty)$. Since $\gamma'(0) = 0$ and γ' achieves its maximum for $x = x_p$ and γ' goes to $-q$ at infinity and there exists an unique $y_p > 0$ (in fact $y_p > x_p$) such that $\gamma'(y_p) = 0$. As a conclusion, since $\gamma(0) = 0$ and γ is increasing on $(0, y_p)$ and $\gamma(x)$ goes to $-\infty$ as x goes to

infinity, $\gamma(x)$ is equal to zero exactly two times: for $x = 0$ and $x = z_p > y_p > x_p$ In fact, z_p is equal to $2x_p$. Indeed, we have

$$\psi'(x) = pq\frac{e^{qx} - e^{-px}}{pe^{qx} + qe^{-px}}.$$

Now, we compute

$$\psi'(2x_p) = pq\frac{(q/p)^{2q} - (p/q)^{2p}}{p(q/p)^{2q} + q(p/q)^{2p}} = \cdots = q^2 - p^2 = q - p,$$

and

$$\begin{aligned} 2\psi(2x_p) &= 2\log(p(q/p)^{2q} + q(p/q)^{2p}) \\ &= 2\log((q+p)(q/p)^{q-p}) \\ &= 2x_p\psi'(2x_p). \end{aligned}$$

Thus, $2x_p$ is (the unique positive) solution of $2\psi(x) = x\psi'(x)$. As a conclusion, $c_p = \psi(2x_p)/(4x_p^2)$ which gives the desired formula after some algebra. □

Remark 3.4 (Advantage of direct Laplace bounds). *Consider a mixture $\mu = p\mu_1 + q\mu_0$ of two Gaussian laws μ_0 and μ_1 on $\mathbb{R}$ with same variance σ^2 and different means. Corollary 3.2 ensures that for every $r \geqslant 0$,*

$$\beta_\mu(r) \leqslant 2\exp\left(-\frac{r^2}{2\sigma^2 + 4c_pW_1(\mu_0,\mu_1)^2}\right).$$

This bound remains relevant as σ goes to zero (we recover the bound for the Bernoulli mixing measure $\nu = p\delta_1 + q\delta_0$). On the other hand, any concentration bound deduced from a Gross inequality would blow up as σ goes to zero, as we shall see in Section 4.

Remark 3.5 (Inhomogeneous tails). *It is satisfactory to recover, when p goes to 0 or 1, the concentration bound of one of the component of the mixture. Nevertheless, one could expect to recover the one of μ_0 (and not only the maximum of the bounds of the two-components). It is possible exhibit two regimes, corresponding to small and big values of λ. Assume that $\mu_i = \mathcal{N}(0,\theta_i^2)$ for $i \in \{0,1\}$ with $\theta_1 > \theta_0 > 0$. We already know that*

$$\alpha_\mu(\lambda) \leqslant \frac{\theta_1^2\lambda^2}{2} + (\theta_1 - \theta_0)\lambda.$$

On the other hand, one has

$$\frac{\mathbf{E}_\mu\big(e^{\lambda f}\big)}{e^{\mathbf{E}_\mu(\lambda f)}} \leqslant \int \alpha_{\mu_\theta}(\lambda)\,\nu(d\theta) + \log\int e^{H_\lambda(\theta)+\lambda g(\theta)}\,\nu(d\theta),$$

where

$$H_\lambda(\theta) = \alpha_{\mu_\theta}(\lambda) + \int \alpha_{\mu_{\theta'}}(\lambda)\,\nu(d\theta') \quad \textit{and} \quad g(\theta) = \mathbf{E}_{\mu_\theta}f - \mathbf{E}_\mu f.$$

Then, Lemma 3.3 ensures that

$$\begin{aligned} \log\int e^{H_\lambda(\theta)+\lambda g(\theta)}\,\nu(d\theta) &\leqslant c_p(H_\lambda(1) + \lambda g(1) - H_\lambda(0) - \lambda g(0))^2 \\ &\leqslant c_p\left(\frac{1}{\varepsilon}|H_\lambda(1) - H_\lambda(0)|^2 + \varepsilon|\lambda g(1) - \lambda g(0)|^2\right). \end{aligned}$$

Choosing $\varepsilon = \lambda$ leads to

$$\log \int e^{H_\lambda(\theta)+\lambda g(\theta)}\, \nu(d\theta) \leqslant c_p \left(\frac{(\theta_1^2-\theta_0^2)^2}{4} + (\theta_1-\theta_0)^2 \right) \lambda^3.$$

As a conclusion α_μ can be control by (at least) these two ways:

$$\alpha_\mu(\lambda) \leqslant \begin{cases} \frac{\theta_1^2\lambda^2}{2} + (\theta_1-\theta_0)\lambda, \\ \frac{p\theta_1^2+q\theta_0^2\lambda^2}{2} + c_p\left(\frac{(\theta_1^2-\theta_0^2)^2}{4} + (\theta_1-\theta_0)^2 \right)\lambda^3. \end{cases}$$

The second one provide sharp bounds for $\lambda \leqslant f(1/c_p)$ whereas the second one is useful for $\lambda \geqslant f(1/c_p)$ (where f is an increasing function which is computable).

4 Gross-Poincaré inequalities for two-components mixtures

It is known that functional inequalities such as Poincaré and Gross (logarithmic Sobolev) inequalities provide, via Laplace exponential bounds, dimension free concentration bounds, see for instance [38]. It is quite natural to ask for such functional inequalities for mixtures. Before attacking the problem, some facts have to be emphasized.

Note that a probability measure μ with non-connected support cannot satisfy to a Poincaré or to a Gross inequality (just consider a non-constant function which is constant on each connected component). In particular, a mixture of distributions with disjoint supports cannot satisfy to such functional inequalities. For instance, one can think about the two-components mixture of uniforms $p\mu_1 + q\mu_0$ where $\mu_0 = \mathcal{U}(I)$ and $\mu_1 = \mathcal{U}(J)$ with $0 < p < 1$ and $\overline{I} \cap \overline{J} = \emptyset$. This observation suggests that in order to obtain a functional inequality for a mixture, one has probably to control the considered functional inequality for each component of the mixture and to ensure that the support of the mixture is connected. It is important to realize that such a connectivity problem is due to the peculiarities of the Poincaré and Gross functional inequalities, but does not pose a real problem for the concentration of measure properties, as suggested by Theorem 3.1 and Remark 3.4 for instance. In the sequel, we will focus on the case of two-components mixtures, and try to get sharp bounds on the Poincaré and Gross constants for the mixture.

For the Gross inequality of two-components mixtures, we will make use of the following two-points Lemma, obtained years ago independently by Diaconis & Saloff-Coste and Higushi & Yoshida, see [50] and references therein.

Lemma 4.1 (Optimal Gross inequality for Bernoulli measures). *For every $p \in (0,1)$ and every $f : \{0,1\} \to \mathbb{R}$, and with the convention $(\log(q) - \log(p))/(q-p) = 2$ if $p = q = 1/2$, we have*

$$\mathbf{Ent}_{p\delta_1+(1-p)\delta_0}\left(f^2\right) \leqslant \frac{\log(q)-\log(p)}{q-p} pq (f(0)-f(1))^2.$$

4.1 Decomposition of the variance and entropy of the mixture

Let μ_0 and μ_1 be two probability measures on $\mathbb{R}^d$, $p \in [0,1]$, $q = 1-p$, $\nu = p\delta_1 + q\delta_0$, and $\mu_p = p\mu_1 + q\mu_0$. Then, one can decompose the variance of $f : \mathbb{R}^d \to \mathbb{R}$ with respect to μ_p

as

$$\begin{aligned}\mathbf{Var}_{\mu_p}(f) &= \mathbf{E}_\nu(\theta \mapsto \mathbf{Var}_{\mu_\theta}(f)) + \mathbf{Var}_\nu(\theta \mapsto \mathbf{E}_{\mu_\theta} f) \\ &= \mathbf{E}_\nu(\theta \mapsto \mathbf{Var}_{\mu_\theta}(f)) + pq(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 \\ &\leqslant \max(C_{\mathrm{PI}}(\mu_0), C_{\mathrm{PI}}(\mu_1))\mathbf{E}_\mu(|\nabla f|^2) + pq(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2.\end{aligned}$$

For the entropy, by using Lemma 4.1 for ν we can write

$$\begin{aligned}\mathbf{Ent}_{\mu_p}\big(f^2\big) &= \mathbf{E}_\nu\big(\theta \mapsto \mathbf{Ent}_{\mu_\theta}\big(f^2\big)\big) + \mathbf{Ent}_\nu\big((\theta \mapsto \mathbf{E}_{\mu_\theta} f)^2\big) \\ &\leqslant \mathbf{E}_\nu\big(\theta \mapsto \mathbf{Ent}_{\mu_\theta}\big(f^2\big)\big) + \frac{pq(\log q - \log p)}{q-p}(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 \\ &\leqslant \max(C_{\mathrm{GI}}(\mu_0), C_{\mathrm{GI}}(\mu_1))\mathbf{E}_\mu(|\nabla f|^2) + \frac{pq(\log q - \log p)}{q-p}(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2.\end{aligned}$$

We thus see that in both cases (Poincaré and Gross inequalities), the problem can be reduced to the control of the mean-difference term $(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2$ in terms of $\mathbf{E}_\mu(|\nabla f|^2)$ for every smooth function f. Note that this task is impossible if μ_0 and μ_1 have disjoint supports.

Remark 4.2 (Finite mixtures). *Let $(\mu_i)_{1\leqslant i\leqslant n}$ be a family of probability measures on $\mathbb{R}^d$. Consider the finite mixture $\mu = \mathcal{M}(\nu, (\mu_i)_{1\leqslant i\leqslant n})$ with mixing measure $\nu = p_1\delta_1 + \cdots + p_n\delta_n$. The decomposition of variance is a general fact valid in particular for μ, and writes*

$$\mathbf{Var}_\mu(f) = \mathbf{E}_\nu(\theta \mapsto \mathbf{Var}_{\mu_\theta}(f)) + \mathbf{Var}_\nu(\theta \mapsto \mathbf{E}_{\mu_\theta} f)\,.$$

Here again, the first term in the right hand side may be controlled with the Poincaré inequality for each of the components $(\mu_i)_{1\leqslant i\leqslant n}$. For the second term of the right hand side, it remains to notice that for every $g : \Theta = \{1, \ldots, n\} \to \mathbb{R}$,

$$\mathbf{Var}_\nu(g) = \frac{1}{2}\sum_{i,j} p_i p_j (g(i) - g(j))^2 = \sum_{i<j} p_i p_j (g(i) - g(j))^2$$

which gives for $g = \mathbf{E}_{\mu_\theta}(f)$ the identity

$$\mathbf{Var}_\nu(\mathbf{E}_{\mu_\theta} f) = \sum_{i<j} p_i p_j \big(\mathbf{E}_{\mu_i} f - \mathbf{E}_{\mu_j} f\big)^2.$$

As for the two-components case, this further reduces the Poincaré inequality for μ to the control of the mean-differences $\big(\mathbf{E}_{\mu_i} f - \mathbf{E}_{\mu_j} f\big)^2$ in terms of $\mathbf{E}_\mu(|\nabla f|^2)$. An analogous approach for the entropy and the Gross inequality can be obtained by using [17, th. A1 p. 49] for instance.

4.2 Control of the mean-difference in dimension one

The following lemma below provides the control of the mean-difference term $(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2$ in the case where μ_0 and μ_1 are probability measures on $\mathbb{R}$ (i.e. $d = 1$).

Lemma 4.3 (Control of the mean-difference term in dimension one). *Let μ_0 and μ_1 be two probability distributions on $\mathbb{R}$ absolutely continuous with respect to the Lebesgue measure. Let us denote by F_0 (respectively F_1) the cumulative distribution function and f_0 (respectively f_1) the probability density function of μ_0 (respectively μ_1). Let $S = \mathrm{supp}(\mu_0) \cup \mathrm{supp}(\mu_1)$ and $\mathrm{Co}(S)$ its convex envelope. Then, for any $p \in (0,1)$, with $\mu_p = p\mu_1 + q\mu_0$ and $q = 1-p$,*

$$(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 \leqslant I(p)\, \mathbf{E}_{\mu_p}(f'^2) \quad \textit{where} \quad I(p) = \int_{\mathrm{Co}(S)} \frac{(F_1(x) - F_0(x))^2}{p f_1(x) + q f_0(x)}\, dx,$$

and the constant $I(p)$ cannot be improved. The function $p \mapsto I(p)$ is convex, and

$$\frac{1}{2\max(p,q)} I\left(\frac{1}{2}\right) \leqslant I(p) \leqslant \frac{1}{2\min(p,q)} I\left(\frac{1}{2}\right). \tag{19}$$

Additionally, if S is not connected then I is constant and equal to ∞. Furthermore, the convexity of I ensure that $I(p)$ is finite for every p in $(0,1)$ if and only if

$$I(0^+) = \lim_{p \to 0^+} I(p) < \infty \quad \textit{and} \quad I(1^-) = \lim_{p \to 1^-} I(p) < \infty.$$

Proof. For a compactly supported f, an integration by parts gives for every $\theta \in \{0,1\}$,

$$\mathbf{E}_{\mu_\theta} f = \int_{\mathbb{R}} f(x) f_\theta(x)\, dx = -\int_{\mathbb{R}} f'(x) F_\theta(x)\, dx.$$

Since $F_1 - F_0 = 0$ outside $\mathrm{Co}(S)$ we have

$$\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f = \int_{\mathrm{Co}(S)} (F_1(x) - F_0(x)) f'(x)\, dx.$$

It remains to use the Cauchy-Schwarz inequality, which gives

$$\begin{aligned}(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 &= \left(\int_{\mathrm{Co}(S)} \frac{F_0(x) - F_1(x)}{\sqrt{p f_1(x) + q f_0(x)}} f'(x) \sqrt{p f_1(x) + q f_0(x)}\, dx \right)^2 \\ &\leqslant I(p) \int_{\mathrm{Co}(S)} f'(x)^2 (p f_1(x) + q f_0(x)) dx = I(p) \mathbf{E}_{\mu_p}(f'^2).\end{aligned}$$

The equality case of the Cauchy-Schwarz inequality provides the optimality of $I(p)$. The other claims of the lemma are immediate. □

4.3 Control of the Poincaré and Gross constants

By combining the decomposition of the variance and of the entropy given at the beginning of the current section with Lemma 4.3 and Lemma 4.1, we obtain the following Theorem.

Theorem 4.4 (Poincaré and Gross inequalities for two-components mixtures). *Let μ_0 and μ_1 be two probability distributions on $\mathbb{R}$ absolutely continuous with respect to the Lebesgue measure, and consider the two-components mixture $\mu_p = p\mu_1 + q\mu_0$ with $0 \leqslant p \leqslant 1$ and $q = 1-p$. If $I(p)$ is as in Lemma 4.3 then for every $p \in (0,1)$*

$$C_{\mathrm{PI}}(\mu_p) \leqslant \max(C_{\mathrm{PI}}(\mu_0), C_{\mathrm{PI}}(\mu_1)) + pqI(p)$$

and

$$C_{\mathrm{GI}}(\mu_p) \leqslant \max(C_{\mathrm{GI}}(\mu_0), C_{\mathrm{GI}}(\mu_1)) + \frac{\log q - \log p}{q-p} pqI(p).$$

In particular, if $\sup_{p\in(0,1)} I(p) < \infty$ *then the optimal Poincaré and Gross constant for* μ_p *goes to the maximum constants of* μ_0 *and* μ_1 *as* $pq \to 0$.

Corollary 4.5 (Uniform Poincaré for two-components mixtures). *Let* μ_0 *and* μ_1 *be two probability distributions on* $\mathbb{R}$ *absolutely continuous with respect to the Lebesgue measure. If* μ_0 *and* μ_1 *satisfy to a Poincaré inequality with respective constants* C_0 *and* C_1 *and if* $I(1/2) < \infty$ *then the mixture* $\mu_p = p\mu_1 + q\mu_0$ *satisfies to a Poincaré inequality with constant* $\max(C_0, C_1) + I(1/2)/2$ *which does not depend on* $p \in [0,1]$.

Proof. Thanks to (19), one has $pqI(p) = \max(p,q)\min(p,q)I(p) \leqslant I(1/2)/2$ and Theorem 4.4 provides the result. □

Remark 4.6 (Blow-up of the Gross constant). *Corollary 4.5 does not work for the Gross inequality, since the upper bound on the constant blows up at speed* $-\log(\min(p,q))$ *as* $pq \to 0$. *Surprisingly, we shall see in the sequel that this behavior is sharp and cannot be improved in general for two-components mixtures.*

4.4 The fundamental example of two Gaussians with identical variance

It was already observed by Johnson in [32, Theorem 1.1 page 536] that for the finite univariate Gaussian mixture $\mu = p_1\mathcal{N}(m_1,\tau^2) + \cdots + p_n\mathcal{N}(m_n,\tau^2)$, we have

$$C_{\mathrm{PI}}(\mu) \leqslant \tau\left(1 + \frac{\sigma^2}{\tau \min_{1\leqslant i\leqslant n} p_i} \exp\left(\frac{\sigma^2}{\tau \min_{1\leqslant i\leqslant n} p_i}\right)\right)$$

where $\sigma^2 = (p_1m_1^2+\cdots+p_nm_n^2)-(p_1m_1+\cdots+p_nm_n)^2$ is the variance of $p_1\delta_{m_1}+\cdots+p_n\delta_{m_n}$. This upper bound on the Poincaré constant blows up as $\min_{1\leqslant i\leqslant n} p_i$ goes to 0. Madras and Randall have also obtained [39, Theorem 1.2 and Section 5] upper bounds for the Poincaré constant of non-Gaussian finite mixtures under an overlapping condition on the supports of the components. As for the result of Johnson mentioned earlier, their upper bound blows up when the minimum weight of the mixing law $\min_{1\leqslant i\leqslant n} p_i$ goes to 0. In the sequel, we show that the Poincaré constant may remain actually bounded as $\min_{1\leqslant i\leqslant n} p_i$ goes to 0. To fix ideas, we will consider the special case of a two-components mixture of two Gaussian distributions $\mathcal{N}(-a,1)$ and $\mathcal{N}(+a,1)$. As usual, we denote by Φ (respectively φ) the cumulative distribution function (respectively probability density function) of the standard Gaussian measure $\mathcal{N}(0,1)$.

Corollary 4.7 (Mixture of two Gaussians with identical variance). *For any* $a > 0$ *and* $0 < p < 1$, *let* $\mu_0 = \mathcal{N}(-a,1)$ *and* $\mu_1 = \mathcal{N}(+a,1)$, *and define the two-components mixture* $\mu_p = p\mu_1 + q\mu_0$. *Then*

$$C_{\mathrm{PI}}(\mu_p) \leqslant 1 + pq4a^2\left(\Phi(2a)e^{4a^2} + \frac{2a}{\sqrt{2\pi}}e^{2a^2} + \frac{1}{2}\right)$$

and

$$C_{\mathrm{GI}}(\mu_p) \leqslant 2 + \frac{\log(q) - \log(p)}{q-p} pq4a^2\left(\Phi(2a)e^{4a^2} + \frac{2a}{\sqrt{2\pi}}e^{2a^2} + \frac{1}{2}\right).$$

Additionally, a sharper upper bound for $p = 1/2$ is given by

$$C_{\mathrm{PI}}(\mu_{1/2}) \leqslant 1 + a\frac{2\Phi(a) - 1}{2\varphi(a)} \quad \textit{and} \quad C_{\mathrm{GI}}(\mu_{1/2}) \leqslant 2 + a\frac{2\Phi(a) - 1}{\varphi(a)}.$$

Note that as a function of $p \in (0, 1)$, the optimal constants are bounded and continuous at $p = 0$ and $p = 1$. The bound (8) expressed in the univariate situation implies that C_{PI} is always greater than or equal to the variance of the probability measure. Here, the variance of μ_p is equal to $1 + 4apq$. Then the upper bound on the Poincaré constant given above is sharp for any $p \in (0, 1)$ as a goes to 0.

Proof. Lemma 4.3 ensures that $p \mapsto I(p)$ is a convex function: let us have a look at $I(0^+)$ and $I(1^-)$ which are equal by symmetry. Since

$$\Phi(x + a) - \Phi(x - a) = \int_{-a}^{+a} \varphi(x + u)\, du \leqslant 2a \begin{cases} \varphi(x + a) & \text{if } x < -a, \\ \varphi(0) & \text{if } -a \leqslant x \leqslant a, \\ \varphi(x - a) & \text{if } a < x, \end{cases}$$

one has

$$\begin{aligned} I(1^-) &= \int_{\mathbb{R}} \frac{(\Phi(x + a) - \Phi(x - a))^2}{\varphi(x - a)}\, dx \\ &\leqslant 4a^2 \left(\int_{-\infty}^{-a} \frac{\varphi(x + a)^2}{\varphi(x - a)}\, dx + \varphi(0)^2 \int_{-a}^{+a} \frac{1}{\varphi(x - a)}\, dx + \int_{+a}^{+\infty} \varphi(x - a)\, dx \right) \\ &\leqslant 4a^2 \left(e^{4a^2} \int_{-\infty}^{-a} e^{-\frac{(x+3a)^2}{2}} \frac{1}{\sqrt{2\pi}}\, dx + \frac{1}{\sqrt{2\pi}} \int_0^{2a} e^{\frac{x^2}{2}}\, dx + \int_0^{+\infty} \varphi(x)\, dx \right) \\ &\leqslant 4a^2 \left(\Phi(2a) e^{4a^2} + \frac{2a}{\sqrt{2\pi}} e^{2a^2} + \frac{1}{2} \right). \end{aligned}$$

Then, the first statement follows from Theorem 4.4. For the second one, by Lemma 4.8 given at the end of the section, we have

$$\begin{aligned} I\left(\frac{1}{2}\right) &= 2 \int_{\mathbb{R}} \frac{\Phi(x + a) - \Phi(x - a)}{\varphi(x + a) + \varphi(x - a)} (\Phi(x + a) - \Phi(x - a))\, dx \\ &\leqslant 2c_a \int_{\mathbb{R}} (\Phi(x + a) - \Phi(x - a))\, dx \\ &= 4ac_a. \end{aligned}$$

This gives as expected $I(1/2) \leqslant 2a(2\Phi(a) - 1)/\varphi(a)$. □

The following lemma shows that $I(1/2)$ is related to some kind of "band isoperimetry".

Lemma 4.8 (Band bound). *For any $x \in \mathbb{R}$ and any $a > 0$,*

$$\frac{\Phi(x + a) - \Phi(x - a)}{\varphi(x + a) + \varphi(x - a)} \leqslant \frac{\Phi(+a) - \Phi(-a)}{\varphi(+a) + \varphi(-a)} = c_a$$

Moreover, this constant cannot be improved. As an example, one has $c_1 \approx 1.410686134$.

Proof. Assume that $a = 1$. Let $c > 0$ and define for any $x \in \mathbb{R}$

$$\alpha(x) = \Phi(x+1) - \Phi(x-1) - c(\varphi(x+1) + \varphi(x-1)).$$

One has $\alpha'(x) = 0$ iff $c(1 + x + (x-1)e^{2x}) = e^{2x} - 1$. Thus, either $x = 0$, or

$$c^{-1} = \beta(x) = -1 + x\coth(x).$$

The function β is even, convex, and achieves its global minimum 0 at $x = 0$. Therefore, the equation $\alpha'(x) = 0$ has three solutions $\{-x_c, 0, +x_c\}$, where $x_c > 0$ satisfies $c\beta(x_c) = 1$. Since $\lim_{x\to\pm\infty} \alpha(x) = 0$, one has $\alpha \leqslant 0$ on $\mathbb{R}$ if and only if $\alpha(0) \leqslant 0$ and $\alpha''(0) \leqslant 0$. The condition $\alpha(0) \leqslant 0$ is fulfilled as soon as

$$c \geqslant \frac{\Phi(+1) - \Phi(-1)}{\varphi(+1) + \varphi(-1)}$$

whereas the condition $\alpha''(0) \geqslant 0$ holds for any c. □

If $A_x = [x-a, x+a]$ then $\partial A_x = \{x-a, x+a\}$. If $\gamma = \mathcal{N}(0,1)$ then

$$\gamma(A_x) = \Phi(x+a) - \Phi(x-a) \quad \text{and} \quad \gamma_s(\partial A_x) = \varphi(x+a) + \varphi(x-a)$$

where γ_s is the surface measure associated to γ (see [36] and references therein). Lemma 4.8 expresses that for any $A \in \mathcal{C}_a = \{A_x; x \in \mathbb{R}\}$

$$\gamma(A) \leqslant c_a \gamma_s(\partial A),$$

and equality is achieved for $A = A_0$. The Gaussian isoperimetric inequality states that $(\varphi \circ \Phi^{-1})(\gamma(A)) \leqslant \gamma_s(\partial A)$ for any regular $A \subset \mathbb{R}$ with equality when A is a half line.

Note also that from the definition of $I(p)$ in Lemma 4.3, we get for every $p \in (0,1)$

$$I(p) \leqslant \left(\sup_{x \in \mathrm{Co}(S)} \frac{|F_1(x) - F_0(x)|}{p f_1(x) + q f_0(x)} \right)^2.$$

In particular, if the right hand side is finite then $I(p)$ is finite.

4.5 Gallery of examples of one-dimensional two-components mixtures

Recall that if μ is a probability measure on $\mathbb{R}$ with density $f > 0$ and median m then

$$\max(b_-, b_+) \leqslant C_{\mathrm{GI}}(\mu) \leqslant 16 \max(b_-, b_+) \tag{20}$$

where

$$b_+ = \sup_{x > m} \mu([x, +\infty)) \log\left(1 + \frac{1}{2\mu([x,+\infty))}\right) \int_m^x \frac{1}{f(y)}\, dy,$$

and

$$b_- = \sup_{x < m} \mu((-\infty, x]) \log\left(1 + \frac{1}{2\mu((-\infty,x])}\right) \int_x^m \frac{1}{f(y)}\, dy.$$

These bounds appear in [7, Remark 7 page 9] as a refinement of a famous criterion by Bobkov and Götze based on previous works of Hardy and Muckenhoupt, see also [45]. More generally, the notion of measure capacities constitutes a powerful tool for the control of C_{PI} and C_{GI}, see [42] and [5, 7]. In the present article, we only use a weak version of such criteria, stated in the following lemma, and which can be found for instance in [2, Chapter 6 page 107]. We will typically use it in order to show that $C_{\mathrm{GI}}(p_1\mu + q\mu_0)$ blows up as p goes to 0 or 1 for certain choices of μ_0 and μ_1.

Lemma 4.9 (Crude lower bound). *Let μ be some distribution on $\mathbb{R}$ with density $f > 0$ then for every median m of μ and every $x \in \mathbb{R}$, by denoting $\Psi(u) = -u\log(u)$,*

$$150\, C_{\mathrm{GI}}(\mu) \geqslant \Psi(\mu(-\infty, x]) \int_x^m \frac{1}{f(y)}\, dy.$$

In this whole section, μ_0 and μ_1 are absolutely continuous probability measures on $\mathbb{R}$ with cumulative distribution functions F_0 and F_1 and probability density functions f_0 and f_1. For every $0 \leqslant p \leqslant 1$, we consider the two-components mixture $\mu_p = p\mu_1 + q\mu_0$. The sharp analysis of the Gross constant for finite mixtures is a difficult problem. Also, we decided to focus on some enlightening examples, by providing a gallery of special cases of μ_0 and μ_1 for which we are able to control the dependence over p of the Poincaré and Gross constant of μ_p. Some of them are quite surprising and reveal hidden subtleties of the Gross inequality as $\min(p, q)$ goes to $0\ldots$

We have already considered the mixture of two Gaussians with identical variance in section 4.4. In the next example, we consider a mixture of two Gaussians with identical mean and different variances.

4.5.1 Two Gaussians with identical mean

Settings. $\mu_1 = \mathcal{N}(0, \sigma^2)$ with $\sigma > 1$ and $\mu_0 = \mathcal{N}(0, 1)$. Also, $\mu_p \to \mathcal{N}(0, 1)$ as $p \to 0$.
Claim. For every $p \in (0, 1)$, we have $C_{\mathrm{GI}}(\mu_p) < \infty$. Moreover, $p \mapsto C_{\mathrm{GI}}(\mu_p)$ is bounded on $(0, 1)$ and goes to $2\sigma^2 = C_{\mathrm{GI}}(\mu_1)$ as $p \to 0$.
Proof. We assume (without loss of generality) that $\sigma^2 > 1$. We have $F_0 \leqslant \kappa F_1$ for some $\kappa > 1$. Once again, $I(p)$ is bounded above by $(2/p)I(1/2) < +\infty$ for every $0 < p < 1$. One can ask now if $I(0^+) < \infty$ or not. Let us define $x_p > 0$ as follows:

$$x_p = \sqrt{\frac{2\sigma^2}{\sigma^2 - 1} \log\left(\frac{q\sigma}{p}\right)}.$$

Then $pf_1(x) \geqslant qf_0(x)$ if and only if $|x| \geqslant x_p$. We have, for some $C > 0$,

$$\begin{aligned} I(p) &\leqslant 2\int_{-\infty}^{-1} \frac{F_1(x)^2}{pf_1(x) + qf_0(x)}\, dx + 2\int_{-1}^{0} \frac{F_1(x)^2}{f_0(x)} dx \\ &\leqslant 2\int_{-\infty}^{-1} \frac{1}{x^2} \frac{f_1(x)^2}{pf_1(x) + qf_0(x)}\, dx + C, \end{aligned}$$

since $2q \geqslant 1$ and $F_1(x) \leqslant f_1(x)/|x|$. If p is sufficiently small then $x_p > 1$ and

$$\int_{-\infty}^{-1} \frac{1}{x^2} \frac{f_1(x)^2}{pf_1(x) + qf_0(x)}\, dx \leqslant 2\int_{-x_p}^{-1} \frac{1}{x^2} \frac{f_1(x)^2}{f_0(x)}\, dx + \frac{1}{p} F_1(-x_p).$$

By the definition of x_p, for some $C > 0$,

$$pF_1(-x_p) \leqslant Cpe^{-x_p^2/(2\sigma^2)} \leqslant C\Big(\frac{1}{p}\Big)^{\frac{\sigma^2-2}{\sigma^2-1}}.$$

If $\sigma^2 \leqslant 2$, then this quantity is bounded as a function of p. On the other hand, for some $C > 0$,

$$\int_{-x_p}^{-1} \frac{1}{x^2}\frac{f_1(x)^2}{f_0(x)}\,dx \leqslant C\int_{-x_p}^{-1} \frac{1}{x^2} e^{\frac{\sigma^2-2}{2\sigma^2}x^2}\,dx.$$

Once again, if $\sigma^2 \leqslant 2$, then this function of p is bounded. If $\sigma^2 > 2$, then, for some $C > 0$,

$$\int_{-x_p}^{-1} e^{\frac{\sigma^2-2}{2\sigma^2}x^2}\,dx \leqslant Ce^{\frac{\sigma^2-2}{2\sigma^2}x_p^2} \leqslant C\Big(\frac{1}{p}\Big)^{\frac{\sigma^2-2}{\sigma^2-1}}.$$

Finally, if $\sigma^2 \leqslant 2$, then $\sup_{p\in(0,1)} I(p) < \infty$, whereas if $\sigma^2 > 2$, then for $C > 0$ and any $p < 1/2$,

$$I(p) \leqslant C\Big(\frac{1}{p}\Big)^{\frac{\sigma^2-2}{\sigma^2-1}}.$$

Since $pI(p) \to 0$ as $p \to 0$, the desired result follows from Theorem 4.4.

4.5.2 Two Uniforms with overlapping supports

Settings. Here $\mu_0 = \mathcal{U}([0,1])$ and $\mu_1 = \mathcal{U}([a, a+1])$ for some $a \in [0,1]$.
Claim. For every $p \in (0,1)$, we have

$$C_{\mathrm{PI}}(\mu_p) \leqslant \pi^{-2} + a^2\Big(1 + \frac{a}{3} - pq\Big) \quad \text{and} \quad C_{\mathrm{GI}}(\mu_p) \leqslant 2\pi^{-2} + \frac{\log(q) - \log(p)}{q-p} a^2\Big(1 + \frac{a}{3} - pq\Big).$$

Proof. Recall that $C_{\mathrm{PI}}(\mathcal{U}([0,1]) = \pi^{-2}$ while $C_{\mathrm{GI}}(\mathcal{U}([0,1]) = 2\pi^{-2}$. By translation invariance, we also have $C_{\mathrm{PI}}(\mathcal{U}([1,1+a]) = \pi^{-2}$ and $C_{\mathrm{GI}}(\mathcal{U}([1,1+a]) = 2\pi^{-2}$. The desired result follows from Theorem 4.4 since for $p \in (0,1)$,

$$I(p) = \int_0^a \frac{x^2}{p}\,dx + \int_a^1 \frac{a^2}{p+q}\,dx + \int_1^{a+1} \frac{(1+a-x)^2}{q}\,dx = \frac{a^2}{pq}\Big(1 + \frac{a}{3} - pq\Big).$$

4.5.3 One Gaussian and a sub-Gaussian

Settings. Here $\mu_1 = \mathcal{N}(0,1)$ with density f_1 while μ_0 is absolutely continuous on $\mathbb{R}$ with density f_0 such that $f_0 \leqslant \kappa f_1$ for some finite real constant $\kappa \geqslant 1$. Note that $\mu_p \to \mu_1$ as $p \to 1$.
Claim. For every $0 < p < 1$ we have $C_{\mathrm{PI}}(\mu_p) \leqslant \max(1, C_{\mathrm{PI}}(\mu_0)) + Dq$. This constant goes to $\max(1, C_{\mathrm{PI}})$ as $p \to 1$ and is additionally uniformly bounded when p runs over $(0,1)$. Similarly, $C_{\mathrm{GI}}(\mu_p) \leqslant \alpha - \beta\log(p)$ for some constants $\alpha > 0$ and $\beta > 0$ which do not depend on p. As in the case of two-uniforms mixture, this upper bound blows up at speed $-\log(p)$ as p goes to 0. This is actually the real behavior of the Gross constant in certain situations as shown by section 4.5.4!

Proof. Since $\mu_1 = \mathcal{N}(0,1)$, we have $C_{\mathrm{PI}}(\mu_1) = 1$ and $C_{\mathrm{GI}}(\mu_1) = 2$. By hypothesis, we have $F_0 \leqslant \kappa F_1$ and $1 - F_0 \leqslant \kappa(1 - F_1)$. Thus, for some $D > 0$ and every $0 < p < 1$,

$$I(p) \leqslant \frac{2(1+\kappa^2)}{p}\left(\int_{-\infty}^{0} \frac{F_1^2(x)}{f_1(x)}\,dx + \int_{0}^{+\infty} \frac{(1-F_1(x))^2}{f_1(x)}\,dx\right) = \frac{D}{p} < \infty.$$

Now Theorem 4.4 shows that $C_{\mathrm{PI}}(\mu_p) \leqslant \max(1, C_{\mathrm{PI}}(\mu_0)) + Dq$. The desired upper bound for $C_{\mathrm{GI}}(\mu_p)$ follows by the same way and we leave the details to the reader.

4.5.4 One Gaussian and a uniform

Settings. Here $\mu_1 = \mathcal{N}(0,1)$ and $\mu_0 = \mathcal{U}([-1,+1])$. Note that $\mu_p \to \mathcal{U}([-1,+1])$ as $p \to 0$.
Claim. There exists a real constant $C \in (0,\infty)$ such that $C_{\mathrm{GI}}(\mu_p) \geqslant -C\log(p)$ for every $p \in (0,1)$. Also, $C_{\mathrm{GI}}(\mu_p)$ blows up at speed $-\log(p)$ as p tends to 0, as for the asymmetric Bernoulli measure $\mathcal{B}(p)$ (see Lemma 4.1). Moreover, $\mathcal{B}(p)$ and the mixture μ_p satisfy to a sub-Gaussian concentration of measure for Lipschitz functions, uniformly in p. This similarity suggests that the blow up phenomenon of $C_{\mathrm{GI}}(\mu_p)$ is due to the asymptotic support reduction from $\mathbb{R}$ to $[-1,+1]$ when p goes to 0. Actually, section 4.5.5 shows that this intuition is false.
Proof. We have $f_0 \leqslant \kappa f_1$ for some constant $\kappa \geqslant 1$. Also, for every $p \in (0,1)$, the result of Section 4.5.3 gives that $C_{\mathrm{GI}}(\mu_p) \leqslant \alpha - \beta\log(p)$ for some constants $\alpha > 0$ and $\beta > 0$ independent of p. Now, by Lemma 4.9,

$$\begin{aligned}
150\,C_{\mathrm{GI}}(p) &\geqslant \Psi(pF_1(-2) + qF_0(-2)) \int_{-2}^{0} \frac{1}{pf_1(u) + qf_0(u)}\,du \\
&= \Psi(pF_1(-2)) \int_{-2}^{0} \frac{1}{pf_1(u) + qf_0(u)}\,du \\
&\geqslant -\left(F_1(-2)\int_{-2}^{-1} \frac{1}{f_1(u)}\,du\right)\log(p).
\end{aligned}$$

4.5.5 Surprising blowing mixture

Settings. Here $f_1(x) = Z_1^{-1}e^{-x^2}$ and $f_0(x) = Z_0^{-1}e^{-|x|^a}$ for some fixed real number $a > 2$, with $Z_1 = \pi^{-1/2}$ and $Z_0 = 2\Gamma(a^{-1})a^{-1}$. Note that $\mu_p \to \mu_0$ as $p \to 0$. The limit has smaller tails.
Claim. There exists a real constant $C > 0$ which may depend on a such that

$$C_{\mathrm{GI}}(\mu_p) \geqslant C(-\log(p))^{1-a^{-1}}$$

for small enough p. In particular, $C_{\mathrm{GI}}(\mu_p)$ blows up as $p \to 0$. Also, the blow up speed of C_{GI} as $p \to 0$ cannot be improved by considering a mixture of fully supported probability measures! Note that $\mu_0 \to \mathcal{U}([-1,+1])$ as $a \to \infty$, and the result is compatible with section 4.5.4.
Proof. Since $f_0 \leqslant \kappa f_1$ for some constant $\kappa \geqslant 1$, the results of Section 4.5.3 gives $C_{\mathrm{GI}}(\mu_p) < \infty$ for every $p \in (0,1)$. Moreover, $p \mapsto C_{\mathrm{GI}}(\mu_p)$ is uniformly bounded on $(p_0,1)$ for every $p_0 > 0$. Let us study the behavior of this function as $p \to 0$. In the sequel we assume that $p < p_0$ where p_0 satisfies $p_0 Z_0 = q_0 Z_1$. The immediate tails comparison

gives $qf_0(x) \leqslant pf_1(x)$ for large enough x. Let us find some explicit bound on x. The inequality $qf_0(x) \leqslant pf_1(x)$ writes $|x|^a - x^2 \geqslant \log(qZ_1) - \log(pZ_0)$. Now, $|x|^a - x^2 \geqslant \frac{1}{2}|x|^a$ for $|x|^{a-2} \geqslant 2$. The non-negative solution of $\frac{1}{2}|x|^a = \log(qZ_1) - \log(pZ_0)$ is

$$x_p = \left(2\log\left(\frac{q}{p}\frac{Z_1}{Z_0}\right)\right)^{1/a}.$$

If p is small enough, then $|x_p|^{a-2} \geqslant 2$ and therefore, $qf_0(x) \leqslant pf_1(x)$ for any $|x| \geqslant x_p$. Now, by Lemma 4.9, for small enough p,

$$150\, C_{\mathrm{GI}}(\mu_p) \geqslant \Psi(pF_1(-2x_p) + qF_0(-2x_p)) \int_{-2x_p}^{0} \frac{1}{pf_1(u) + qf_0(u)}\, du.$$

For small enough p, we have $\max(F_0, F_1)(-2x_p) < e^{-1}$ and thus, for some real constant $C > 0$,

$$\Psi(pF_1(-2x_p) + qF_0(-2x_p)) \geqslant \Psi(pF_1(-2x_p)) \geqslant -pF_1(-2x_p)\log(p) \geqslant C\frac{e^{-4x_p^2}}{x_p}\Psi(p).$$

On the other hand, since $qf_0(x) \leqslant pf_1(x)$ for $|x| \geqslant x_p$, we have for some real constant $C > 0$,

$$\int_{-2x_p}^{0} \frac{1}{pf_1(u) + qf_0(u)}\, du \geqslant \int_{-2x_p}^{-x_p} \frac{du}{2pf_1(u)} \geqslant \frac{Ce^{4x_p^2}}{px_p}.$$

Consequently, for some real constant $C > 0$,

$$150\, C_{\mathrm{GI}}(\mu_p) \geqslant -C\frac{\log(p)}{x_p^2}.$$

Now, by using the explicit expression of x_p, we finally obtain for some real constant $C > 0$,

$$C_{\mathrm{GI}}(\mu_p) \geqslant C\,(-\log(p))^{1-a^{-1}}.$$

4.6 Multivariate mean-difference bound

It is quite natural to ask for a multidimensional counterpart of the mean-difference Lemma 4.3. Let us give some informal ideas to attack this quite delicate problem. Let μ_0 and μ_1 be two probability measures on $\mathbb{R}^d$, and consider as usual the mixture $\mu_p = p\mu_1 + q\mu_0$ with $p \in (0,1)$ and $q = 1-p$. It is well known (see for instance [55]) that if μ_0 and μ_1 are regular enough, then there exists a map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that the image measure $T \cdot \mu_0$ of μ_0 by T is μ_1 and

$$W_2(\mu_0, \mu_1)^2 = \int_{\mathbb{R}^d} |T(x) - x|^2\, \mu_0(dx).$$

If we denote by $\mu_{(s)}$ the image measure of μ_0 by $x \mapsto sT(x) + (1-s)x$ for every $0 < s < 1$, then

$$(\mathbf{E}_{\mu_1} f - \mathbf{E}_{\mu_0} f)^2 = \left(\int_0^1 \int_{\mathbb{R}^d} (T(x) - x) \cdot \nabla f(sT(x) + (1-s)x)\, d\mu_0(x)\, ds\right)^2.$$

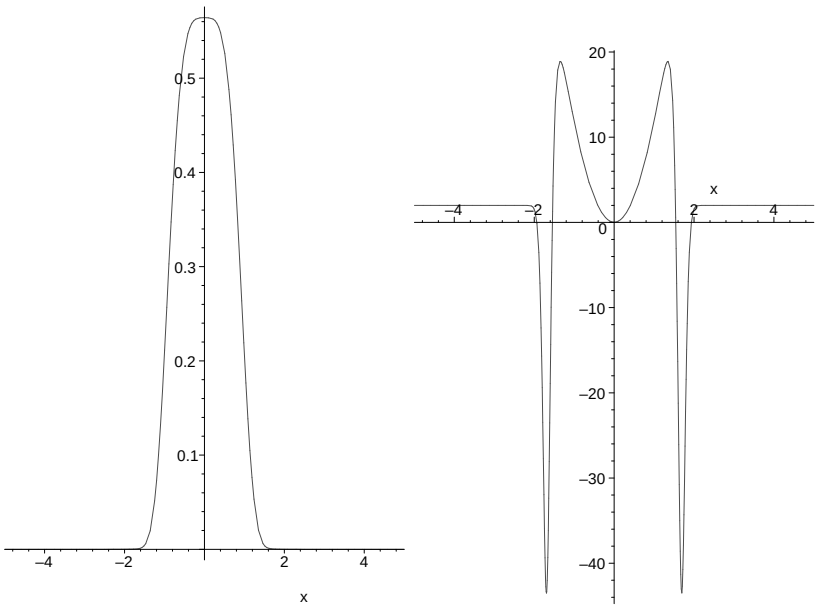


Figure 1: Density and $-\log$-density of μ_p for Example 4.5.5 with $p = 1/100$ and $a = 4$.

By Cauchy-Schwarz's inequality, we get

$$(\mathbf{E}_{\mu_1} f - \mathbf{E}_{\mu_0} f)^2 \leqslant \left(\int_{\mathbb{R}^d} |T(x) - x|^2 \, d\mu_0(x) \right) \left(\int_0^1 \int_{\mathbb{R}^d} |\nabla f(x)|^2 \, d\mu_s(x) \, ds \right)$$

and therefore

$$(\mathbf{E}_{\mu_1} f - \mathbf{E}_{\mu_0} f)^2 \leqslant W_2(\mu_1, \mu_0)^2 \int_{\mathbb{R}^d} \int_0^1 |\nabla f(x)|^2 \, d\mu_s(x) \, ds.$$

This shows that in order to control the mean-difference term $(\mathbf{E}_{\mu_1} f - \mathbf{E}_{\mu_0} f)^2$ by $\mathbf{E}_{\mu_p}(|\nabla f|^2)$, it is enough to find a real constant $C_p > 0$ such that $\overline{\mu} \leqslant C_p \mu_p$ where

$$\overline{\mu}(A) = \int_0^1 \mu_{(s)}(A) \, ds.$$

Unfortunately, this is not feasible if for some $s \in (0,1)$, the support of $\mu_{(s)}$ is not included in the support of μ_p (union of the supports of μ_0 and μ_1 if $p \in (0,1)$). This problem is due to the linear interpolation used to define $\mu_{(s)}$ via T. The linear interpolation will fail if the support of μ_p is a non-convex connected set. Let us adopt an alternative path-wise interpolation scheme. For each $x \in S_0 = \mathrm{supp}(\mu_0)$, let us pick a continuous and piecewise smooth interpolating path $\gamma_x : [0,1] \to \mathbb{R}^d$ such that $\gamma_x(0) = x$ and $\gamma_x(1) = T(x)$. Then for every smooth $f : \mathbb{R}^d \to \mathbb{R}$

$$f(x) - f(T(x)) = \int_0^1 \dot{\gamma}_x(s) \, \nabla f(\gamma_x(s)) \, ds \leqslant \sqrt{\int_0^1 |\dot{\gamma}_x(s)|^2 \, ds} \, \sqrt{\int_0^1 |\nabla f|^2(\gamma_x(s)) \, ds}.$$

As a consequence, we have

$$(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 \leqslant \left(\int_{S_0} \int_0^1 |\dot{\gamma}_x(s)|^2 \, ds \, \mu_0(dx) \right) \left(\int_{S_0} \int_0^1 |\nabla f|^2(\gamma_x(s)) \, ds \, \mu_0(dx) \right).$$

Now, let $\mu_{(s)}$ be the image measure of μ_0 by the map $x \mapsto \gamma_x(s)$, where here again $\overline{\mu}$ is the measure defined by $\overline{\mu}(A) = \int_0^1 \mu_{(s)}(A) \, ds$. With this notation, we have

$$(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 \leqslant \left(\int_{S_0} \int_0^1 |\dot{\gamma}_x(s)|^2 \, ds \, \mu_0(dx) \right) \left(\int_{\mathbb{R}^d} |\nabla f|^2(x) \, \overline{\mu}(dx) \right).$$

Note that

$$\left(\int_{S_0} \int_0^1 |\dot{\gamma}_x(s)|^2 \, ds \, \mu_0(dx) \right) \geqslant W_2(\mu_0, \mu_1)^2$$

with equality when γ_x is the linear interpolation map between x and $T(x)$ for every $x \in S_0$. The mean-difference control that we seek for follows then immediately if there exists a real constant $C_p > 0$ such that $\overline{\mu} \leqslant C_p \mu_p$. The problem is thus reduced in the choice of an interpolation scheme γ such that the support of $\overline{\mu}$ is included in the support of μ_p (which is the union of the supports of μ_0 and μ_1 as soon as $0 < p < 1$). Let us give now two enlightening examples.

Example 4.10 (When the linear interpolation map is optimal). *Consider the two-dimensional example where $\mu_0 = \mathcal{U}([0,2] \times [0,2])$ and $\mu_1 = \mathcal{U}([1,3] \times [0,2])$. If γ is the natural linear interpolation map given by $\gamma_x(s) = x + s e_1$ then $\mu_{(s)} = \mathcal{U}([s, s+2] \times [0,2])$ is supported inside $\mathrm{supp}(\mu_0) \cup \mathrm{supp}(\mu_1)$. This is due to the convexity of this union. Also, the linear interpolation map is here optimal. Moreover, elementary computations reveal that $C_p = 1/\min(p,q)$ and $W_2(\mu_0, \mu_1)^2 = 1$. Therefore, for every $0 < p < 1$ and any smooth $f : \mathbb{R}^2 \to \mathbb{R}$,*

$$(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 \leqslant \frac{1}{\min(p,q)} \mathbf{E}_{\mu_p}(|\nabla f|^2).$$

Example 4.11 (When the linear interpolation map fails). *In contrast, for the example where $\mu_0 = \mathcal{U}([0,2] \times [0,2])$ and $\mu_1 = \mathcal{U}([1,3] \times [1,3])$ and if γ is the natural linear interpolation map given by $\gamma_x(s) = x + s(e_1 + e_2)$ then $\mu_{(s)}$ is not supported in $\mathrm{supp}(\mu_0) \cup \mathrm{supp}(\mu_1)$ and this union is not convex. If $A = [0,1] \times [2,3]$ then $\mu_{(s)}(A) > 0$ for every $0 < s < 1$ while $\mu_p(A) = 0$ for every $0 < p < 1$ and hence there is no finite constant $C_p > 0$ such that $\overline{\mu} \leqslant C_p \mu_p$. This shows that the linear interpolation map fails here. Let us give an alternative interpolation map which leads to the desired result. We set for every $x \in \mathrm{supp}(\mu_0)$ and every $0 \leqslant s \leqslant 1$, with $\mathbf{1} = (e_1, e_1)$,*

$$\gamma_x(s) = \begin{cases} (1-s)x + 2s\mathbf{1} & \text{if } 0 \leqslant s \leqslant \frac{1}{2} \\ sx + \mathbf{1} & \text{otherwise.} \end{cases}$$

This corresponds to a two-steps linear interpolation between the squares $[0,2]^2$ and $[1,3]^2$ with intermediate square $[1,2]^2$. For every $0 \leqslant s \leqslant 1$,

$$\mu_{(s)} = \begin{cases} \mathcal{U}([2s,2]^2) & \text{if } 0 \leqslant s \leqslant \frac{1}{2} \\ \mathcal{U}([1,1+2s]^2) & \text{otherwise.} \end{cases}$$

Note that we constructed γ in such a way that $\mu_{(s)}$ is always supported in $\mathrm{supp}(\mu_0) \cup \mathrm{supp}(\mu_1)$. *Elementary computations reveal that for every* $0 < p < 1$

$$\int_{S_0} \int_0^1 |\dot{\gamma}_x(s)|^2 \, ds \, \mu_0(dx) = \frac{8}{3} \quad \textit{and} \quad \overline{\mu} \leqslant \frac{4}{\min(p,q)} \mu_p.$$

Finally, putting all together, we obtain for every $0 < p < 1$ *and smooth* $f : \mathbb{R}^2 \to \mathbb{R}$,

$$(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 \leqslant \frac{32}{3 \min(p,q)} \mathbf{E}_{\mu_p}(|\nabla f|^2).$$

As a conclusion, one can retain that the natural interpolation problem associated to the control of the mean-difference involves a kind of support-constrained interpolation for mass transportation.

Acknowledgements. The authors would like to warmly thank Arnaud Guillin, Sébastien Gouëzel, and all the participants of the Working Seminar *Inégalités fonctionnelles* held on March 2008 at the Institut Henri Poincaré of Paris for fruitful discussions.

References

[1] Sh. Aida, *Uniform positivity improving property, Sobolev inequalities, and spectral gaps*, J. Funct. Anal. **158** (1998), no. 1, 152–185. 1

[2] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Société Mathématique de France, Paris, 2000, Preface by D. Bakry and M. Ledoux. 1, 4.5

[3] D. Bakry and M. Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math., vol. 1123, Springer, Berlin, 1985, pp. 177–206. 1

[4] D. Bakry, M. Ledoux, and F.-Y. Wang, *Perturbations of functional inequalities using growth conditions*, J. Math. Pures Appl. (9) **87** (2007), no. 4, 394–407. 1, 1

[5] F. Barthe, P. Cattiaux, and C. Roberto, *Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry*, Rev. Mat. Iberoam. **22** (2006), no. 3, 993–1067. 4.5

[6] F. Barthe and A. V. Kolesnikov, *Mass transport and variants of the logarithmic Sobolev inequality*, preprint `arXiv:0709.3890v1 [math.PR]`, 2007. 1

[7] F. Barthe and C. Roberto, *Sobolev inequalities for probability measures on the real line*, Studia Math. **159** (2003), no. 3, 481–497, Dedicated to Professor Aleksander Pełczyński on the occasion of his 70th birthday (Polish). 1, 4.5

[8] S. G. Bobkov, *Concentration of normalized sums and a central limit theorem for noncorrelated random variables*, Ann. Probab. **32** (2004), no. 4, 2884–2907. 1

[9] ———, *Generalized symmetric polynomials and an approximate de Finetti representation*, J. Theoret. Probab. **18** (2005), no. 2, 399–412. 1

[10] S. G. Bobkov and F. Götze, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. **163** (1999), no. 1, 1–28. 1

[11] S. G. Bobkov and B. Zegarlinski, *Entropy bounds and isoperimetry*, Mem. Amer. Math. Soc. **176** (2005), no. 829, x+69. 1

[12] F. Bolley and C. Villani, *Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities*, Ann. Fac. Sci. Toulouse (2005), 331–352. 1, 1, 1

[13] L. A. Caffarelli, *Monotonicity properties of optimal transportation and the FKG and related inequalities*, Comm. Math. Phys. **214** (2000), no. 3, 547–563. 1

[14] ———, *Erratum: [Comm. Math. Phys.* **214** *(2000), no. 3, 547–563]*, Comm. Math. Phys. **225** (2002), no. 2, 449–450. 1

[15] E. A. Carlen, *Superadditivity of Fisher's information and logarithmic Sobolev inequalities*, J. Funct. Anal. **101** (1991), no. 1, 194–211. 1

[16] S. Chatterjee, *Spin glasses and Stein's method*, preprint `arXiv:0706.3500v2 [math.PR]`, 2007. 1

[17] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. **6** (1996), no. 3, 695–750. 4.2

[18] H. Djellout, A. Guillin, and L. Wu, *Transportation cost-information inequalities and applications to random dynamical systems and diffusions*, Ann. Probab. **32** (2004), no. 3B, 2702–2732. 1, 1

[19] B. S. Everitt and D. J. Hand, *Finite mixture distributions*, Chapman & Hall, London, 1981, Monographs on Applied Probability and Statistics. 1

[20] V. P. Fonf, J. Lindenstrauss, and R. R. Phelps, *Infinite dimensional convexity*, Handbook of the geometry of Banach spaces, Vol. I, North-Holland, Amsterdam, 2001, pp. 599–670. 1

[21] S. Frühwirth-Schnatter, *Finite mixture and Markov switching models*, Springer Series in Statistics, Springer, New York, 2006. 1

[22] I. Gentil, *Inégalités de sobolev logarithmique et de poincaré pour la loi uniforme*, unpublished note, 2004. 1

[23] I. Gentil and C. Roberto, *Spectral gaps for spin systems: some non-convex phase examples*, J. Funct. Anal. **180** (2001), no. 1, 66–84. 1

[24] N. Gozlan, *A characterization of dimension free concentration in terms of transportation inequalities*, preprint `arXiv: 0804.3089 [math.PR]`, 2008. 1, 1

[25] M. Gromov and V. D. Milman, *A topological application of the isoperimetric inequality*, Amer. J. Math. **105** (1983), no. 4, 843–854. 1

[26] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083. 1

[27] ———, *Hypercontractivity, logarithmic Sobolev inequalities, and applications: a survey of surveys*, Diffusion, quantum theory, and radically elementary mathematics, Math. Notes, vol. 47, Princeton Univ. Press, Princeton, NJ, 2006, pp. 45–73. 1

[28] B. Helffer, *Semiclassical analysis, Witten Laplacians, and statistical mechanics*, Series in Partial Differential Equations and Applications, vol. 1, World Scientific Publishing Co. Inc., River Edge, NJ, 2002. 1

[29] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30. 2, 3

[30] R. Holley and D. Stroock, *Logarithmic Sobolev inequalities and stochastic Ising models*, J. Statist. Phys. **46** (1987), no. 5-6, 1159–1194. 1

[31] M. Jerrum, J.-B. Son, P. Tetali, and E. Vigoda, *Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains*, Ann. Appl. Probab. **14** (2004), no. 4, 1741–1765. 1

[32] O. Johnson, *Convergence of the Poincaré constant*, Teor. Veroyatnost. i Primenen. **48** (2003), no. 3, 615–620, preprint version available at `arXiv.org:math.PR/0206227`. 1, 4.4

[33] O. Kavian, G. Kerkyacharian, and B. Roynette, *Quelques remarques sur l'ultracontractivité*, J. Funct. Anal. **111** (1993), no. 1, 155–196. 1

[34] I. Kontoyiannis and M. Madiman, *Measure concentration for compound Poisson distributions*, Electron. Comm. Probab. **11** (2006), 45–57 (electronic). 1

[35] L. Kontoyiannis and M. Madiman, *Entropy, compound Poisson approximation, log-Sobolev inequalities and measure concentration*, Information Theory Workshop,24–29 Oct. 2004. IEEE, 2004, pp. 71–75. 1

[36] R. Latała, *On some inequalities for Gaussian measures*, Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002) (Beijing), Higher Ed. Press, 2002, pp. 813–822. 4.4

[37] M. Ledoux, *Concentration of measure and logarithmic Sobolev inequalities*, Séminaire de Probabilités, XXXIII, Springer, Berlin, 1999, pp. 120–216. 1

[38] ———, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. 1, 1, 4

[39] N. Madras and D. Randall, *Markov chain decomposition for convergence rate analysis*, Ann. Appl. Probab. **12** (2002), no. 2, 581–606. 1, 4.4

[40] K. Marton, *A simple proof of the blowing-up lemma*, IEEE Trans. Inform. Theory **32** (1986), no. 3, 445–446. 1

[41] ———, *Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration*, Ann. Probab. **24** (1996), no. 2, 857–866. 1

[42] V. G. Maz'ja, *Sobolev spaces*, Springer Series in Soviet Mathematics, Springer-Verlag, Berlin, 1985, Translated from the Russian by T. O. Shaposhnikova. 4.5

[43] G. McLachlan and D. Peel, *Finite mixture models*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York, 2000. 1

[44] G. J. McLachlan and K. E. Basford, *Mixture models*, Statistics: Textbooks and Monographs, vol. 84, Marcel Dekker Inc., New York, 1988, Inference and applications to clustering. 1

[45] L Miclo, *Quand est-ce que des bornes de Hardy permettent de calculer une constante de Poincaré exacte sur la droite ?*, preprint, `http://hal.archives-ouvertes.fr/hal-00017875/en/`, 2005. 4.5

[46] F. Otto and M. G. Reznikoff, *A new criterion for the logarithmic Sobolev inequality and two applications*, J. Funct. Anal. **243** (2007), no. 1, 121–157. 1

[47] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), no. 2, 361–400. 1

[48] R. R. Phelps, *Lectures on Choquet's theorem*, second ed., Lecture Notes in Mathematics, vol. 1757, Springer-Verlag, Berlin, 2001. 1

[49] S. T. Rachev, *Probability metrics and the stability of stochastic models*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Ltd., Chichester, 1991. 1

[50] L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413. 4

[51] A. J. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Information and Control **2** (1959), 101–112. 1

[52] V. N. Sudakov, *Geometric problems in the theory of infinite-dimensional probability distributions*, Proc. Steklov Inst. Math. (1979), no. 2, i–v, 1–178, Cover to cover translation of Trudy Mat. Inst. Steklov **141** (1976). 1

[53] M. Talagrand, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal. **6** (1996), no. 3, 587–600. 1

[54] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical analysis of finite mixture distributions*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Ltd., Chichester, 1985. 1

[55] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. 1, 1, 4.6

[56] ———, *Optimal transport, old and new*, lecture notes, École d'été de Saint-Flour 2005, preprint, 2007. 1, 1, 1

[57] F.-Y. Wang, *Logarithmic Sobolev inequalities on noncompact Riemannian manifolds*, Probab. Theory Related Fields **109** (1997), no. 3, 417–424. 1

Compiled May 7, 2008

Djalil Chafaï mailto:chafai(AT)math.univ-toulouse.fr

UMR 181 INRA ENVT, 23 Chemin des Capelles F-31076 Toulouse Cedex 3, France.

UMR 5583 CNRS Institut de Mathématiques de Toulouse (IMT)
Université de Toulouse III, 118 route de Narbonne, F-31062, Toulouse Cedex, France.

Florent Malrieu mailto:florent.malrieu(AT)univ-rennes1.fr

UMR 6625 CNRS Institut de Recherche Mathématique de Rennes (IRMAR)
Université de Rennes I, Campus de Beaulieu, F-35042 Rennes Cedex, France.