Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2010, Vol. 46, No. 1, 72–96
DOI: 10.1214/08-AIHP309

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

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

Djalil Chafaï[a] and Florent Malrieu[b]

[a] *UMR181 INRA ENVT, 23 Chemin des Capelles F-31076 Toulouse Cedex 3, France. UMR5583 CNRS Institut de Mathématiques de Toulouse (IMT), Université de Toulouse III, 118 route de Narbonne, F-31062, Toulouse Cedex, France. E-mail: chafai@math.univ-toulouse.fr*
[b] *UMR 6625 CNRS Institut de Recherche Mathématique de Rennes (IRMAR), Université de Rennes I, Campus de Beaulieu, F-35042 Rennes* CEDEX*, France. E-mail: florent.malrieu@univ-rennes1.fr*

Received 7 May 2008; revised 10 November 2008; accepted 10 December 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 potential with multiple deep wells. We study in the present work fine properties of mixtures with respect to concentration of measure and Sobolev 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. Additionally, our analysis of Sobolev type inequalities for two-component 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-component mixture may remain bounded as the mixture proportion goes to 0 or 1 while the logarithmic Sobolev constant may surprisingly blow up. This counter-intuitive result is not reducible to support disconnections, and appears as a reminiscence of the variance-entropy comparison on the two-point space. As far as mixtures are concerned, the logarithmic Sobolev inequality is less stable than the Poincaré inequality and the sub-Gaussian concentration for Lipschitz functions. We illustrate our results on a gallery of concrete two-component mixtures. This work leads to many open questions.

Résumé. Les mélanges dont il est question ici sont des combinaisons convexes de lois de probabilité. Malgré cette défnition simple, un mélange peut étre beaucoup plus subtil que ses composants. Un mélange de lois gaussiennes par exemple peut donner lieu à des potentiels à profonds puits multiples. Dans ce travail, nous étudions les propriétés fines des mélanges vis à vis de la concentration de la mesure et des inégalités de type Sobolev. Nous proposons des bornes sur la transformée de Laplace faisant intervenir le diamètre de la famille mélangée pour une distance de transport. Notre analyse des inégalités de type Sobolev pour les mélanges à deux composants révèle des relations naturelles avec une forme d'isopérimétrie pour les bandes, ainsi qu'avec le transport optimal sous contrainte de support. Nous établissons que la constante de Poincaré peut rester bornée lorsque la proportion du mélange tend vers 0 tandis que la constante de Sobolev logarithmique peut exploser. Ce phénomène contre intuitif n'est pas réductible à un problème de support et peut être vu comme une trace de la comparaison variance-entropie sur l'espace à deux points. Pour les mélanges, la propriété de concentration de la mesure sous-gaussienne et l'inégalité de Poincaré sont plus stables que l'inégalité de Sobolev logarithmique. Nous illustrons nos résultats avec une collection d'exemples à deux composants concrets. Ce travail conduit à plusieurs questions ouvertes.

MSC: 60E15; 49Q20; 46E35; 62E99

Keywords: Transportation cost distances; Mallows or Wasserstein distance; Mixtures of distributions; Finite Gaussian mixtures; Concentration of measure; Gaussian bounds; Tails probabilities; Deviation inequalities; Functional inequalities; Poincaré inequalities; Gross logarithmic Sobolev inequalities; Band isoperimetry; Transportation of measure; Mass transportation

1. Introduction

Mixtures of distributions are ubiquitous in stochastic analysis, modelling, simulation, and statistics, see, for instance, the monographs [16,18,39,40,50]. 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-component mixture." More generally, a *finite mixture* takes the form $p_1\mu_1 + \cdots + p_n\mu_n$ where $\mu_1, \dots, \mu_n$ 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 $1 \le i \le n$. In that case, for certain choices of $m_1, \dots, m_n$ and $\sigma_1, \dots, \sigma_n$, 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. Another 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 class 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, [13].

At this point, it is enlightening to introduce a 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

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

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 $\Theta = \{1, \dots, n\}$ and $\nu = p_1\delta_1 + \cdots + p_n\delta_n$ and

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

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 a certain class of ν as a particular Choquet's integral, see [43] and [17]. 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, [7]. 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.

Our first aim is to investigate the fine behavior of concentration of measure for mixtures, for instance, for a two-component 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 these functional inequalities for mixtures, and in particular for two-component mixtures. Our work reveals striking unexpected phenomena. In particular, our results suggest that the logarithmic Sobolev 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 [20] and [3], 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. We denote by $\|\cdot\|_2$ the Euclidean norm of $\mathbb{R}^d$. A function $f : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz when

$$\|f\|_{\mathrm{Lip}} = \sup_{x \ne y} \frac{|f(x) - f(y)|}{\|x - y\|_2} < \infty.$$

Let μ be a law 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 transform $\alpha_\mu : \mathbb{R} \to [0, \infty]$ of μ by

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

The Cramér–Chernov–Chebyshev inequality gives, for every $r > 0$,

$$\beta_\mu(r) = \sup_{\|f\|_{\mathrm{Lip}} \le 1} \mu\big(|f - \mathbf{E}_\mu f| \ge r\big) \le 2 \exp\Big(-\sup_{\lambda > 0}\big(r\lambda - \alpha_\mu(\lambda)\big)\Big) \tag{2}$$

and the supremum in the right-hand side is a Fenchel–Legendre transform of α_μ. 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 a *sub-Gaussian concentration of measure* for Lipschitz functions when there exists a constant $C \in (0, \infty)$ such that for every real number λ,

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

The log-Laplace–Lipschitz quadratic bound (3) implies via (2) that for every $r > 0$,

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

Actually, it was shown (see [15] and [9]) 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

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

Linear or quadratic upper bounds for α_μ may be deduced from functional inequalities such as Poincaré and (Gross) logarithmic Sobolev inequalities [23,24]. We say that μ satisfies a Poincaré inequality of constant $C \in (0, \infty)$ when for every smooth $h : \mathbb{R}^d \to \mathbb{R}$,

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

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 constant C is called the *optimal Poincaré constant of* μ and is denoted $C_{\mathrm{PI}}(\mu)$ with the convention $\inf \varnothing = \infty$. Similarly, μ satisfies a (Gross) logarithmic Sobolev inequality of constant $C \in (0, \infty)$ when

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

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* or *free energy* of h^2 for μ, with the convention $0 \log(0) = 0$. As for the Poincaré inequality, the smallest possible C is the *optimal logarithmic Sobolev constant of* μ and is denoted $C_{\mathrm{GI}}(\mu)$ with $\inf \varnothing = \infty$. Standard linearization arguments give that

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

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 coordinate functions. More precisely, the first inequality in (8) follows from (6) by taking $h = \langle \cdot, u\rangle$ where u runs over the unit sphere while the second inequality in (8) follows by considering the directional derivative of both sides of (7) at the constant function 1.

A basic example is given by Gaussian laws for which equalities are achieved in (8). A wide class of laws satisfy Poincaré and logarithmic Sobolev inequalities. Beyond Gaussian laws, a criterion due to Bakry and Émery [1,2] (see also [12,42,46] and [10,11]) states that if μ has Lebesque 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 $C_{\mathrm{PI}}(\mu) \le \kappa$ and $C_{\mathrm{GI}}(\mu) \le 2\kappa$ with equality in both cases when μ is Gaussian. This log-concave criterion appears as a comparison with Gaussians. Note that in general, $C_{\mathrm{GI}}(\mu) < \infty$ implies $C_{\mathrm{PI}}(\mu) < \infty$ but the converse is false. For instance, the law with density proportional to $\exp(-|x|^a)$ on $\mathbb{R}$ satisfies a Poincaré inequality iff $a \ge 1$ and a logarithmic Sobolev inequality iff $a \ge 2$, see, for example, [1], Chapter 6. Note also that if μ has disconnected support, then necessarily $C_{\mathrm{PI}}(\mu) = C_{\mathrm{GI}}(\mu) = \infty$. To see it, consider a non-constant h which is constant on each connected component of the support of μ. This is for instance the case 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 logarithmic Sobolev inequality (7) implies 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 [34], Chapter 5, that for any reals λ and $r > 0$

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

The same method yields from (6) a sub-exponential upper bound for β_μ of the form $c_1 \exp(-c_2 r)$ for some constants $c_1, c_2 > 0$, see, for instance, [22] and [33], Section 2.5.

Both Poincaré and logarithmic Sobolev inequalities are invariant by the action of the translation group and the orthogonal group. More generally, let us denote by $f \cdot \mu$ the image measure of μ by the map f. Both (6) and (7) are stable by Lipschitz maps in the sense that $C_{\mathrm{PI}}(f \cdot \mu) \le \|f\|_{\mathrm{Lip}}^2 C_{PI}(\mu)$ and $C_{\mathrm{GI}}(f \cdot \mu) \le \|f\|_{\mathrm{Lip}}^2 C_{\mathrm{GI}}(\mu)$. On the real line, C_{PI} and C_{GI} can be controlled via "simple" variational bounds such as (18). Both (6) and (7) are also stable by bounded perturbations on the log-density of μ, see [25,27] and [3] for further details. 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 of (6) and (7), see, for example, [1], Chapters 1 and 3, [9] and [21].

The case of mixtures

The integral criterion (5) shows that if the components of a mixture uniformly satisfy a sub-Gaussian concentration of measure for Lipschitz functions, and if the mixing law has compact support, then the mixture also satisfies a sub-Gaussian concentration of measure for Lipschitz functions. Such bounds appear, for instance, in [6]. However, this observation does not give any fine quantitative estimate on the dependency over the weights for a finite mixture. Regarding Poincaré and logarithmic Sobolev inequalities, it is clear that a finite mixture of Gaussians will satisfy 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.

An upper bound for the Poincaré constant of univariate finite Gaussian mixture was provided by Johnson [29], 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 [35], 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 [30,31] in connection with compound Poisson processes and discrete modified logarithmic Sobolev inequalities.

Outline of the article

Recall that the aim of the present work is to study the fine properties of mixture of law with respect to concentration of measure and Sobolev type 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 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 (see Section 2 for a precise definition) of the mixed family. Section 3 is devoted to upper bounds on α_μ for two-component mixtures $\mu = \mu_p = p\mu_1 + q\mu_0$. Our result is mainly based on a Laplace–Lipschitz counterpart of the optimal logarithmic Sobolev inequality for asymmetric Bernoulli measures. In particular, we show that if μ_0 and μ_1 satisfy a 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 $\min(p, q)$ goes to 0. In Section 4, we study Poincaré and logarithmic Sobolev inequalities for two-component mixtures. A decomposition of variance and entropy allows us to reduce the problem to the Poincaré and logarithmic Sobolev 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 can remain bounded as $\min(p, q)$ goes to 0, while the logarithmic Sobolev constant may surprisingly blow up at speed $-\log(\min(p, q))$. This counter-intuitive result shows that as far as the mixture of laws is concerned, the logarithmic Sobolev inequality does not behave like the sub-Gaussian concentration for Lipschitz functions. We also illustrate our results on a gallery of concrete two-component mixtures. In particular, we show that the blow up of the logarithmic Sobolev constant as $\min(p, q)$ goes to 0 is not necessarily related to support problems.

Open problems

The study of Poincaré and logarithmic Sobolev 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 maybe a link with the decomposition approach used in [28] for Markov chains. One can also explore the tensor products of mixtures, which are again mixtures. Another interesting problem is the development of a direct approach for transportation cost and measure-capacities inequalities (see [5]) 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 not necessarily finite mixture. The notion of diameter that we shall use is related to coupling and transportation cost. Recall that for every $k \geq 1$, the Wasserstein (or transportation cost) distance of order k between two laws μ_1 and μ_2 on $\mathbb{R}^d$ is defined by (see [51,52] and [44,47])

$$W_k(\mu_1, \mu_2) = \inf_\pi \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^k \, \mathrm{d}\pi(x, y) \right)^{k^{-1}}, \tag{10}$$

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 $k \geq 1$,

$$W_k(\mu_1, \mu_2)^k = \int_0^1 \left| F_1^{-1}(x) - F_2^{-1}(x) \right|^k \mathrm{d}x \quad \text{and} \quad W_1(\mu_1, \mu_2) = \int_{\mathbb{R}} \left| F_1(x) - F_2(x) \right| \mathrm{d}x, \tag{11}$$

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

$$W_1(\mu_1, \mu_2) = \sup_{\|f\|_{\mathrm{Lip}} \leq 1} \left(\int_{\mathbb{R}^d} f \, \mathrm{d}\mu_1 - \int_{\mathbb{R}^d} f \, \mathrm{d}\mu_2 \right). \tag{12}$$

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 the so-called *transportation cost inequalities*, popularized by Marton [36,37], Talagrand [49], and Bobkov and Götze [8]. See, for instance, the books [34,51,52] for a review. The link with concentration of measure was recently deeply explored by Gozlan, see [21]. We will not use this interesting line of research in the present paper.

Theorem 2.1 (General Laplace–Lipschitz bound via diameter). *Let $\mu = \mathcal{M}(\nu, (\mu_\theta)_{\theta\in\Theta})$ be a general mixture. If this mixture satisfies the uniform bounds*

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

then for every $\lambda > 0$ we have

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

Proof of Theorem 2.1. The key point is that if $\|f\|_{\mathrm{Lip}} \le 1$ then for every $\lambda > 0$,

$$\frac{\mathbf{E}_\mu(\mathrm{e}^{\lambda f})}{\mathrm{e}^{\lambda \mathbf{E}_\mu f}} = \mathrm{e}^{-\lambda \mathbf{E}_\mu f}\int_\Theta \mathbf{E}_{\mu_\theta}\bigl(\mathrm{e}^{\lambda f}\bigr)\nu(\mathrm{d}\theta) \le \int_\Theta \mathrm{e}^{\alpha_\theta(\lambda)+\lambda(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f)}\nu(\mathrm{d}\theta). \tag{13}$$

As a consequence, we get

$$\alpha_\mu(\lambda) \le \overline{\alpha}(\lambda) + \sup_{\|f\|_{\mathrm{Lip}}\le 1} \log\int_\Theta \mathrm{e}^{\lambda(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f)}\nu(\mathrm{d}\theta). \tag{14}$$

Thanks to the relation (12), we obtain

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

This shows that the second term in the right-hand side of (14) is bounded by $\overline{W}\lambda$. Alternatively, one can use the Hoeffding bound [26] which says that if X is a centered bounded random variable with oscillation $c = \sup X - \inf X$ then

$$\mathbf{E}\bigl(\mathrm{e}^{\lambda X}\bigr) \le \mathrm{e}^{\lambda^2 c^2/8}.$$

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 $c \le \sup_{\theta,\theta'}(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_{\mu_{\theta'}} f) = \overline{W}$. $\square$

Example 2.2 (Finite mixtures). *For a finite mixture $\mu = p_1\mu_1 + \cdots + p_n\mu_n = \mathcal{M}(\nu, (\mu_i)_{1\le i\le n})$ 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 $(\mu_i)_{1\le i\le n}$ (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 non-negative $(d \times d)$-matrices. Note that $\Gamma(\theta)$ is allowed to be singular, that is, not of full rank. The spectrum of $\Gamma(\theta)$ is real and non-negative. If $\lambda_1(\theta) \ge \cdots \ge \lambda_d(\theta)$ are the eigenvalues of $\Gamma(\theta)$, we define $\rho = \sup_{\theta\in\Theta}\lambda_1(\theta) = \sup_{\theta\in\Theta}\|\Gamma(\theta)\|_{2\to 2}$. Now fix some mixing law ν on Θ and consider the mixture $\mu = \mathcal{M}(\nu, (\mu_\theta)_{\theta\in\Theta})$. Then for every $\lambda > 0$,*

$$\alpha_\mu(\lambda) \le \frac{\rho}{2}\lambda^2 + \frac{1}{8}\min\bigl(8\overline{W}\lambda, \overline{W}^2\lambda^2\bigr).$$

One can deduce an upper bound for $\overline{W}$ from the following lemma.

Lemma 2.4 (W_1-distance of two multivariate Gaussian laws). *Let $\mu_0 = \mathcal{N}(m(0), \Gamma(0))$ and $\mu_1 = \mathcal{N}(m(1), \Gamma(1))$ be two Gaussian laws on $\mathbb{R}^d$. For $\theta \in \{0, 1\}$, we denote by*

$$\lambda_1(\theta) \ge \cdots \ge \lambda_d(\theta)$$

the ordered spectrum of $\Gamma(\theta)$ and by $(v_i(\theta))_{1\le i\le d}$ an associated orthonormal basis of eigenvectors. Assume, without loss of generality, that $v_i(0) \cdot v_i(1) \ge 0$ for every $1 \le i \le d$ where "$\cdot$" stands for the Euclidean scalar product of $\mathbb{R}^d$. Then $W_1(\mu_0, \mu_1)$ is bounded above by

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

The reader may find in [48], Theorem 3.2, a formula in the same spirit for $W_2(\mu_0, \mu_1)$.

Proof of Lemma 2.4. The triangle inequality for the W_1 distance gives

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

Now, if $(Y_i)_{1\le i\le d}$ 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 (10) and Jensen's inequality, we get

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

At this step, we note that

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

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

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

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*

$$W_1(\mu_\theta, \mu_{\theta'}) = \left|\theta - \theta'\right|.$$

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

$$\left|g(\theta) - g(\theta')\right| \le \mathbf{E}\left(\left|f(X+\theta) - f(X+\theta')\right|\right) \le \left|\theta - \theta'\right|,$$

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

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

and for any $\lambda > 0$

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

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 $\forall \lambda > 0$,

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

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*

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

where $\gamma \ge 2$ is some fixed real number. Note that ν has a non-compact support and that μ does not 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 (11) for W_1. Starting from (13), one has by the Cauchy–Schwarz inequality

$$\left(\frac{\mathbf{E}_\mu(\mathrm{e}^{\lambda f})}{\mathrm{e}^{\lambda \mathbf{E}_\mu f}}\right)^2 \le \int_\Theta \mathrm{e}^{\theta^2 \lambda^2} \nu(\mathrm{d}\theta) \int_\Theta \mathrm{e}^{2\lambda(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f)} \nu(\mathrm{d}\theta). \tag{15}$$

Note that ν satisfies condition (5) and $\alpha_\nu(\lambda) \le C\lambda^2$ 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

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

where $X \sim \mathcal{N}(0,1)$. Also, for the second term in the right-hand side of (15) we have

$$\int_\Theta \mathrm{e}^{2\lambda(\mathbf{E}_{\mu_\theta} f - \mathbf{E}_\mu f)} \nu(\mathrm{d}\theta) \le \mathrm{e}^{\alpha_\nu(2\lambda)} \le \mathrm{e}^{4C\lambda^2}.$$

If $\gamma = 2$ then $\alpha_\mu(\lambda) \le 2C\lambda^2 - \frac{1}{4}\log(1-\lambda^2) \le 2C - \frac{1}{4}\log(1-\lambda)$ if $\lambda < 1$, which gives, after some computations, the deviation bound, for some other constants $C' > 0$ and $C'' > 0$,

$$\mu(F - \mathbf{E}_\mu f \ge r) \le C' \mathrm{e}^{-C'' r}.$$

Assume in contrast that $\gamma > 2$. *Since* $\theta^2\lambda^2 \le \gamma^{-1}\theta^\gamma + C_0\lambda^{2\gamma/(\gamma-2)}$ *for some constant* $C_0 > 0$ *which may depend on* γ *but not on* λ *and* θ, *we get, for some constants* $C_1 > 0$ *and* $C_2 > 0$,

$$\int_0^\infty \exp\left(\theta^2\lambda^2\right)\nu(\mathrm{d}\theta) \le C_1 \exp\left(C_2\lambda^{2\gamma/(\gamma-2)}\right).$$

This gives $\alpha_\mu(\lambda) \le C_3\lambda^{2\gamma/(\gamma-2)} + C_4$ *for some constants* $C_3 > 0$ *and* $C_4 > 0$, *which yields a deviation bound of the form* (*for some constants* $C_5 > 0$ *and* $C_6 > 0$)

$$\mu(f - \mathbf{E}_\mu f \ge r) \le C_5 \exp\left(-C_6 r^{2-4/(\gamma+2)}\right).$$

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-component 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-component 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 continuity conventions $c_{1/2} = 1/8$ *and* $c_0 = c_1 = 0$. *Then for any* $\lambda > 0$,

$$\alpha_\mu(\lambda) \le \max(\alpha_{\mu_0}, \alpha_{\mu_1})(\lambda) + \begin{cases} c_p\lambda^2 W_1(\mu_0, \mu_1)^2 & \text{if } \lambda W_1(\mu_0, \mu_1) \le x_p, \\ \max(p, q)\left(\lambda W_1(\mu_0, \mu_1) - \frac{1}{2}x_p\right) & \text{otherwise.} \end{cases}$$

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 $\alpha_\mu \le \max(\alpha_{\mu_1}, \alpha_{\mu_2})$ 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-component 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* $\lambda > 0$

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

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

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

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 finite mixture, we get, as in the general case, for any $f \in \mathrm{Lip}(\mathcal{X}, \mathbb{R})$ and $\lambda > 0$,

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

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

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

where $\cosh_p(x) := p\mathrm{e}^{qx} + q\mathrm{e}^{-px}$. Since $g(1) - g(0) = \mathbf{E}_{\mu_1} f - \mathbf{E}_{\mu_0} f$, we get by (12)

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

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

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

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

$$\alpha_\mu(\lambda) \le \max(\alpha_{\mu_0}, \alpha_{\mu_1})(\lambda) + \log\max(\cosh_p, \cosh_q)\bigl(\lambda W_1(\mu_0, \mu_1)\bigr).$$

Since $(\cosh_q - \cosh_p)'(x) = 2pq(\cosh(px) - \cosh(qx))$, one has, for every $x \ge 0$,

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

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

$$\log\max(\cosh_p, \cosh_q)(x) = \log\cosh_p(x) \le c_p x^2.$$

On the other hand,

$$\log\cosh_p(x) = qx + \log\bigl(p + q\mathrm{e}^{-x}\bigr) \le qx.$$

Now, for $x = x_p$, the slope of $x \mapsto c_p x^2$ is equal to q and the tangent is $y = q(x - x_p/2)$. On the other hand, the convexity of $x \mapsto \log\cosh_p(x)$ yields $\log\cosh_p(x) \le q(x - x_p)$ for $x \ge x_p$ (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 law. This lemma is an optimal version of the Hoeffding bound [26] in the case of a Bernoulli law.

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

$$\sup_{x>0} x^{-2}\log\bigl(p\mathrm{e}^{qx} + q\mathrm{e}^{-px}\bigr) = c_p = \frac{q - p}{4(\log(q) - \log(p))} \tag{16}$$

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 $1/(4c_p)$ is the optimal constant of the logarithmic Sobolev inequality for the asymmetric Bernoulli measure $q\delta_0 + p\delta_1$ (see Lemma 4.1).

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

$$\psi(x) = \log\bigl(p\mathrm{e}^{qx} + q\mathrm{e}^{-px}\bigr).$$

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 $(0,\widehat{x}_p)$ and negative on $(\widehat{x}_p,+\infty)$. Since $\gamma'(0)=0$ and γ' achieves its maximum for $x=\widehat{x}_p$ and γ' goes to $-q$ at infinity and there exists an unique $y_p>0$ (in fact $y_p>\widehat{x}_p$) 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 $x=z_p>y_p>\widehat{x}_p$. In fact, z_p is equal to $2\widehat{x}_p$. Indeed, we have

$$\psi'(x)=pq\frac{\mathrm{e}^{qx}-\mathrm{e}^{-px}}{p\mathrm{e}^{qx}+q\mathrm{e}^{-px}}.$$

Now, we compute

$$\psi'(2\widehat{x}_p)=pq\frac{(q/p)^{2q}-(p/q)^{2p}}{p(q/p)^{2q}+q(p/q)^{2p}}=\cdots=q^2-p^2=q-p$$

and

$$\begin{aligned}2\psi(2\widehat{x}_p)&=2\log\bigl(p(q/p)^{2q}+q(p/q)^{2p}\bigr)\\&=2\log\bigl((q+p)(q/p)^{q-p}\bigr)\\&=2\widehat{x}_p\psi'(2\widehat{x}_p).\end{aligned}$$

Thus, $2\widehat{x}_p$ is the unique positive solution of $2\psi(x)=x\psi'(x)$. As a conclusion, we get $c_p=\psi(2\widehat{x}_p)/(4\widehat{x}_p^2)$, which gives the desired formula after some algebra. $\square$

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 $r\geq 0$,*

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

This bound remains relevant as $\sigma\to 0$ since we recover the bound for the Bernoulli mixing law $\nu=p\delta_1+q\delta_0$. On the other hand, any concentration bound deduced from a logarithmic Sobolev 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 (resp. 1), the concentration bound of μ_0 (resp. μ_1) and not only the maximum of the bounds of the two components. It is possible to 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 $\theta_1>\theta_0>0$. Theorem 2.1 gives*

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

On the other hand, one has

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

where

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

Then, Lemma 3.3 ensures that for every $\varepsilon > 0$,

$$\log \int e^{H_\lambda(\theta)+\lambda g(\theta)} \nu(d\theta) \le c_p \big(H_\lambda(1) + \lambda g(1) - H_\lambda(0) - \lambda g(0)\big)^2$$
$$\le c_p \left(\frac{1}{\varepsilon} \big|H_\lambda(1) - H_\lambda(0)\big|^2 + \varepsilon \big|\lambda g(1) - \lambda g(0)\big|^2 \right).$$

Choosing $\varepsilon = \lambda$ leads to

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

As a conclusion α_μ can be controlled in (at least) these two ways:

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

The second one provides sharp bounds for $\lambda \le f(1/c_p)$ whereas the first one is useful for $\lambda \ge f(1/c_p)$ (where f is an increasing function which is computable).

4. Gross–Poincaré inequalities for two-component 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, [34]. It is quite natural to ask for such functional inequalities for mixtures. Before attacking the problem, some facts have to be emphasized.

As already mentioned in the introduction, a law μ with disconnected support cannot satisfy a Poincaré or a logarithmic Sobolev inequality. In particular, a mixture of laws with disjoint supports cannot satisfy such functional inequalities. 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 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-component mixtures, and try to get sharp bounds on the Poincaré and logarithmic Sobolev constants. The two-component case is fundamental. The extension of the results to more general finite mixtures is possible by following roughly the same scheme, see Remark 4.2 below.

For the logarithmic Sobolev inequality of two-component mixtures, we will make use of the following optimal two-point lemma, obtained years ago independently by Diaconis and Saloff-Coste and Higushi and Yoshida. An elementary proof due to Bobkov is given by Saloff-Coste in his Saint-Flour Lecture Notes [45].

Lemma 4.1 (Optimal logarithmic Sobolev 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*

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

Moreover, the function of p in front of the right-hand side cannot be improved.

Note that the constant in front of the right-hand side of the inequality provided by Lemma 4.1 is nothing else but $pq/(4c_p)$ where c_p is as in Theorem 3.1 and Lemma 3.3. At this stage, it is important to understand the deep difference between the Poincaré and the logarithmic Sobolev inequalities at the level of the two-point space. On the two-point space, the Poincaré inequality turns out to be a simple equality, and Lemma 4.1 is in fact an entropy-variance comparison. Namely, for every $p \in (0,1)$ and $f:\{0,1\} \to \mathbb{R}$,

$$\mathbf{Ent}_{p\delta_1+q\delta_0}\big(f^2\big) \le \frac{\log(q)-\log(p)}{q-p}\mathbf{Var}_{p\delta_1+q\delta_0}(f).$$

This inequality is optimal and $(\log(q)-\log(p))/(q-p)$ tends to $+\infty$ as $\min(p,q)$ goes to 0. Also, for strongly asymmetric Bernoulli measures, the entropy of the square can take extremely big values for a fixed prescribed variance. This elementary phenomenon helps to better understand the surprising difference in the behavior of the Poincaré and logarithmic Sobolev constants of certain two-component mixtures exhibited in the sequel. Moreover, this observation suggests that we use asymmetric test functions inspired from the two-point space in order to show that the logarithmic Sobolev constant may blow up when the mixing law is strongly asymmetric. We shall adopt however another (quantitative) route.

4.1. *Decomposition of the variance and entropy of the mixture*

Let μ_0 and μ_1 be two laws 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 and bound the variance of $f:\mathbb{R}^d \to \mathbb{R}$ with respect to μ_p as

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

For the entropy, let us write

$$\mathbf{Ent}_{\mu_p}\big(f^2\big) = \mathbf{E}_\nu\big(\theta \mapsto \mathbf{Ent}_{\mu_\theta}\big(f^2\big)\big) + \mathbf{Ent}_\nu\big(\theta \mapsto \mathbf{E}_{\mu_\theta}\big(f^2\big)\big).$$

Applying Lemma 4.1 to the function $\theta \mapsto \sqrt{\mathbf{E}_{\mu_\theta}(f^2)}$, one gets

$$\mathbf{Ent}_\nu\big(\theta \mapsto \mathbf{E}_{\mu_\theta}\big(f^2\big)\big) \le \frac{pq(\log q - \log p)}{q-p}\Big(\sqrt{\mathbf{E}_{\mu_0}\big(f^2\big)} - \sqrt{\mathbf{E}_{\mu_1}\big(f^2\big)}\Big)^2.$$

Since $\mathbf{E}_{\mu_0}(f)\mathbf{E}_{\mu_1}(f) \le \sqrt{\mathbf{E}_{\mu_0}(f^2)\mathbf{E}_{\mu_1}(f^2)}$, we have

$$\begin{aligned}\Big(\sqrt{\mathbf{E}_{\mu_0}\big(f^2\big)} - \sqrt{\mathbf{E}_{\mu_1}\big(f^2\big)}\Big)^2 &= \mathbf{E}_{\mu_0}\big(f^2\big) + \mathbf{E}_{\mu_1}\big(f^2\big) - 2\sqrt{\mathbf{E}_{\mu_0}\big(f^2\big)\mathbf{E}_{\mu_1}\big(f^2\big)}\\ &\le \mathbf{Var}_{\mu_0}(f) + \mathbf{Var}_{\mu_1}(f) + (\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2.\end{aligned}$$

(Note that the right-hand side is not equal to zero if $\mu_0 = \mu_1$.) Using the Poincaré inequalities for μ_0 and μ_1 provides the following control of the entropy:

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

(The worst term is the last one since it always explodes as $\min(p,q)$ goes to zero.) We thus see that in both cases (Poincaré and logarithmic Sobolev 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 $(\mu_i)_{1\le i\le n}$ be a family of probability measures on $\mathbb{R}^d$. Consider the finite mixture $\mu = \mathcal{M}(\nu, (\mu_i)_{1\le i\le n})$ 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*

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

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

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

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

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

As for the two-component case, this further reduces the Poincaré inequality for μ to the control of the mean-differences $(\mathbf{E}_{\mu_i} f - \mathbf{E}_{\mu_j} f)^2$ in terms of $\mathbf{E}_\mu(|\nabla f|^2)$. An analogous approach for the entropy and the logarithmic Sobolev inequality can be obtained by using [14], *Theorem* A1 *p.* 49, *for instance.*

4.2. *Control of the mean-difference in dimension one*

The following lemma 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). If* $\mathrm{co}(S)$ *denotes the convex envelope of the set $S = \mathrm{supp}(\mu_0) \cup \mathrm{supp}(\mu_1)$, then, for any $p \in (0, 1)$, with $\mu_p = p\mu_1 + q\mu_0$ and $q = 1 - p$, we have*

$$(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 \le I(p)\,\mathbf{E}_{\mu_p}\big(f'^2\big), \quad \text{where } I(p) = \int_{\mathrm{co}(S)} \frac{(F_1(x) - F_0(x))^2}{pf_1(x) + qf_0(x)}\,\mathrm{d}x,$$

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

$$\frac{1}{2\max(p,q)} I\left(\frac{1}{2}\right) \le I(p) \le \frac{1}{2\min(p,q)} I\left(\frac{1}{2}\right). \tag{17}$$

Furthermore, if S is not connected then I is constant and equal to ∞, while the convexity of I ensures that $\sup_{p\in(0,1)} I(p) = \max(I(0^+), I(1^-))$ *where*

$$I\big(0^+\big) = \lim_{p\to 0^+} I(p) \quad \text{and} \quad I\big(1^-\big) = \lim_{p\to 1^-} I(p),$$

and $I(p) < \infty$ for every p in $(0, 1)$ if and only if $\max(I(0^+), I(1^-)) < \infty$.

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

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

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

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

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

$$\begin{aligned}(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 &= \left(\int_{\mathrm{co}(S)} \frac{F_0(x) - F_1(x)}{\sqrt{pf_1(x) + qf_0(x)}} f'(x)\sqrt{pf_1(x) + qf_0(x)}\,\mathrm{d}x\right)^2 \\ &\le I(p)\int_{\mathrm{co}(S)} f'(x)^2\big(pf_1(x) + qf_0(x)\big)\,\mathrm{d}x = I(p)\mathbf{E}_{\mu_p}\big(f'^2\big).\end{aligned}$$

The equality case of the Cauchy–Schwarz inequality provides the optimality of $I(p)$. The bound (17) follows from $2\min(p,q)(f_0 + f_1)/2 \le pf_1 + qf_0 \le 2\max(p,q)(f_0 + f_1)/2$. The other claims of the lemma are immediate. □

4.3. *Control of the Poincaré and logarithmic Sobolev constants*

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

Theorem 4.4 (Poincaré and logarithmic Sobolev inequalities for two-component 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-component mixture $\mu_p = p\mu_1 + q\mu_0$ with $0 \le p \le 1$ and $q = 1 - p$. If $I(p)$ is as in Lemma 4.3 then for every $p \in (0,1)$,*

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

and

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

In particular, since $\sup_{p\in(0,1)} I(p) = \max(I(0^+), I(1^-))$ where $I(0^+)$ and $I(1^-)$ are as in Lemma 4.3, we get the following uniform bounds:

$$\sup_{p\in(0,1)} C_{\mathrm{PI}}(\mu_p) \le \max\big(C_{\mathrm{PI}}(\mu_0), C_{\mathrm{PI}}(\mu_1)\big) + \frac{1}{4}\max\big(I\big(0^+\big), I\big(1^-\big)\big).$$

Moreover, if $I(0^+) < \infty$ (respectively if $I(1^-) < \infty$) then

$$\limsup_{p\to 0^+ \text{ respectively } 1^-} C_{\mathrm{PI}}(\mu_p) \le \max\big(C_{\mathrm{PI}}(\mu_0), C_{\mathrm{PI}}(\mu_1)\big).$$

The upper bounds given by Theorem 4.4 must be understood in $[0,\infty]$ since the right-hand side can be infinite (in such a case the bound is of course useless). Additionally, by Lemma 4.3, the function $p \mapsto I(p)$ is convex, and it is possible that $I(1/2) < \infty$ while $\max(I(0^+), I(1^-)) = \infty$. The following corollary provides a uniform bound on the Poincaré constant of a two-component mixture in terms of $I(1/2)$ without using $\max(I(0^+), I(1^-))$. This corollary has no immediate logarithmic Sobolev counterpart, as explained in the remark below following the proof of the corollary.

Corollary 4.5 (Uniform Poincaré inequality for two-component 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 mixture $\mu_p = p\mu_1 + q\mu_0$ for every $p \in [0,1]$. We have then*

$$\max_{p\in[0,1]} C_{\mathrm{PI}}(\mu_p) \le \max\big(C_{\mathrm{PI}}(\mu_0), C_{\mathrm{PI}}(\mu_1)\big) + \frac{1}{2}I\left(\frac{1}{2}\right),$$

where $I(1/2)$ is as in Lemma 4.3.

Proof. We observe that thanks to (17), one has

$$pqI(p) = \max(p,q)\min(p,q)I(p) \le \frac{1}{2}I\left(\frac{1}{2}\right)$$

and Theorem 4.4 provides the desired result. □

Remark 4.6 (Blow-up of the logarithmic Sobolev constant). *With the notations of Corollary 4.5, we have, by using the same argument, that for every $p \in (0,1)$,*

$$C_{\mathrm{GI}}(\mu_p) \le \max\bigl(C_{\mathrm{GI}}(\mu_0), C_{\mathrm{GI}}(\mu_1)\bigr) + \frac{1}{2}\frac{\log(q)-\log(p)}{q-p}\left(I\left(\frac{1}{2}\right) + \max\bigl(C_{\mathrm{PI}}(\mu_0), C_{\mathrm{PI}}(\mu_1)\bigr)\right).$$

Since $(\log(q)-\log(p))/(q-p)$ goes to $+\infty$ at speed $-\log(\min(p,q))$ as $\min(p,q)$ goes to 0, we cannot derive a uniform logarithmic Sobolev inequality for two-component mixtures under the sole assumption that $I(1/2) < \infty$. Surprisingly, we shall see in the sequel that this behavior is sharp and cannot be improved in general for two-component mixtures.

4.4. *The fundamental example of two Gaussians with identical variance*

It was already observed by Johnson in [29], Theorem 1.1, p. 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

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

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 $\min_{1\le i\le n} p_i$ goes to 0. Madras and Randall have also obtained [35], 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 $\min_{1\le i\le n} p_i$ goes to 0. In the sequel, we show that the Poincaré constant can remain actually bounded as $\min_{1\le i\le n} p_i$ goes to 0. To fix ideas, we will consider the special case of a two-component 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-component mixture $\mu_p = p\mu_1 + q\mu_0$. Then*

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

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

$$C_{\mathrm{PI}}(\mu_{1/2}) \le 1 + a\frac{2\Phi(a)-1}{2\varphi(a)}.$$

Note that as a function of p, the obtained upper bounds on the constants are continuous on the whole interval $[0,1]$. 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 of Corollary 4.7. 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 here equal by symmetry. Since

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

one has

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

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

$$\begin{aligned}
I\left(\frac{1}{2}\right) &= 2\int_{\mathbb{R}}\frac{\Phi(x+a)-\Phi(x-a)}{\varphi(x+a)+\varphi(x-a)}\big(\Phi(x+a)-\Phi(x-a)\big)\,\mathrm{d}x\\
&\le 2\tau_a\int_{\mathbb{R}}\big(\Phi(x+a)-\Phi(x-a)\big)\,\mathrm{d}x\\
&= 4a\tau_a.
\end{aligned}$$

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

The following lemma shows that $I(1/2)$ is related to some kind of "band isoperimetry." Note that Lemma 4.3 provides a more general approach beyond the Gaussian case.

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

$$\frac{\Phi(x+a)-\Phi(x-a)}{\varphi(x+a)+\varphi(x-a)}\le\frac{\Phi(+a)-\Phi(-a)}{\varphi(+a)+\varphi(-a)}=\tau_a.$$

Moreover, this constant cannot be improved. As an example, one has $\tau_1\approx 1.410686134$.

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

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

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

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

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

$$\tau \ge \frac{\Phi(+1) - \Phi(-1)}{\varphi(+1) + \varphi(-1)},$$

whereas the condition $\alpha''(0) \ge 0$ holds for any τ. The case where $a \ne 1$ is similar. □

Remark 4.9 (Relation with isoperimetry). *If $A_x = [x-a, x+a]$ then $\partial A_x = \{x-a, x+a\}$. If $\gamma = \mathcal{N}(0,1)$ then $\gamma(A_x) = \Phi(x+a) - \Phi(x-a)$ while $\gamma_s(\partial A_x) = \varphi(x+a) + \varphi(x-a)$ where γ_s is the surface measure associated to γ, see* [32]. *Lemma* 4.8 *expresses that for any $A \in \mathcal{C}_a = \{A_x; x \in \mathbb{R}\}$, we have $\gamma(A) \le \tau_a \gamma_s(\partial A)$ and equality is achieved for $A = A_0$. Recall that the Gaussian isoperimetric inequality states that $(\varphi \circ \Phi^{-1})(\gamma(A)) \le \gamma_s(\partial A)$ for any regular $A \subset \mathbb{R}$ with equality when A is a half line, see* [32] *and references therein.*

4.5. *Gallery of examples of one-dimensional two-component mixtures*

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

$$\max(b_-, b_+) \le C_{\mathrm{GI}}(\mu) \le 16 \max(b_-, b_+), \tag{18}$$

where

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

and

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

These bounds appear in [5], *Remark* 7, *p.* 9, *as a refinement of a famous criterion by Bobkov and Götze based on previous works of Hardy and Muckenhoupt, see also* [41]. *More generally, the notion of measure capacities constitutes a powerful tool for the control of C_{PI} and C_{GI}, see* [38] *and* [4,5]. *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* [1], *Chapter* 6, *p.* 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 4.10 (Crude lower bound). *Let μ be some distribution on $\mathbb{R}$ with density $f > 0$ then for every median m of μ and every $x \le m$, by denoting $\Psi(u) = -u\log(u)$,*

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

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 $0 \le p \le 1$, we consider the two-component mixture $\mu_p = p\mu_1 + q\mu_0$. The sharp analysis of the logarithmic Sobolev 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 logarithmic Sobolev constant of μ_p. Some of them are quite surprising and reveal hidden subtleties of the logarithmic Sobolev inequality as $\min(p,q)$ goes to $0 \ldots$. The key tools used here are Theorem 4.4 and Lemma 4.10.

4.5.1. *One Gaussian and a sub-Gaussian*

Setting. *Here $\mu_1 = \mathcal{N}(0,1)$ while μ_0 is such that $f_0 \le \kappa f_1$ for some finite constant $\kappa \ge 1$.*

Claim. *For every $0 < p < 1$ we have $C_{\mathrm{PI}}(\mu_p) \le \max(1, C_{\mathrm{PI}}(\mu_0)) + Dq$. This upper bound goes to* $\max(1, C_{\mathrm{PI}}(\mu_0))$ *as $p \to 1$ and is additionally uniformly bounded when p runs over* $(0,1)$. *Similarly, $C_{\mathrm{GI}}(\mu_p) \le \alpha - \beta\log(p)$ for* *some constants $\alpha, \beta > 0$ which do not depend on p. This upper bound blows up at speed $-\log(p)$ as $p \to 0$. This is actually the real behavior of $C_{\mathrm{GI}}(\mu_p)$ in some situations as shown in 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 $F_0 \le \kappa F_1$ and $1 - F_0 \le \kappa(1 - F_1)$. Thus, for some $D > 0$ and every $0 < p < 1$,

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

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

4.5.2. *Two Gaussians with identical mean*

We have already considered the mixture of two Gaussians with identical variances and different means in Section 4.4. Here we consider a mixture of two Gaussians with identical means and different variances. It turns out that this Gaussian mixture is a simple Gaussian sub-case of Section 4.5.1, for which we are able to provide a more precise bound for C_{GI}.

Setting. $\mu_1 = \mathcal{N}(0, \sigma^2)$ *with* $\sigma > 1$ *and* $\mu_0 = \mathcal{N}(0,1)$.

Claim. *There exists $C > 0$ such that, for any $p < 1/2$,*

$$I(p) \le C\left(\frac{1}{p}\right)^{(\sigma^2-2)/(\sigma^2-1)} \quad \textit{and} \quad C_{\mathrm{PI}}(\mu_p) \le \sigma^2 + Cp^{1/(\sigma^2-1)}.$$

Moreover we have $\sup_{p\in(0,1)} C_{\mathrm{PI}}(\mu_p) < \infty$.

Proof. We have $f_0 \le \kappa f_1$ for some $\kappa > 1$, and we recover the setting of Section 4.5.1. Let us provide now an upper bound for $I(p)$ when p is close to 0. We have $pf_1(x) \ge qf_0(x)$ if and only if $|x| \ge \overline{x}_p$ where

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

We have, for some constant $C > 0$,

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

since $2q \ge 1$ and $F_1(x) \le f_1(x)/|x|$. If p is sufficiently small then $\overline{x}_p > 1$ and

$$\int_{-\infty}^{-1} \frac{1}{x^2}\frac{f_1(x)^2}{pf_1(x)+qf_0(x)}\,\mathrm{d}x \le 2\int_{-\overline{x}_p}^{-1} \frac{1}{x^2}\frac{f_1(x)^2}{f_0(x)}\,\mathrm{d}x + \frac{1}{p}F_1(-\overline{x}_p).$$

By the definition of $\overline{x}_p$, for some $C > 0$,

$$\frac{1}{p}F_1(-\overline{x}_p) \le \frac{C}{p}\mathrm{e}^{-\overline{x}_p^2/(2\sigma^2)} \le C\left(\frac{1}{p}\right)^{(\sigma^2-2)/(\sigma^2-1)}.$$

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

$$\int_{-\bar{x}_p}^{-1} \frac{1}{x^2} \frac{f_1(x)^2}{f_0(x)}\, \mathrm{d}x \le C \int_{-\bar{x}_p}^{-1} \frac{1}{x^2} \mathrm{e}^{(\sigma^2-2)x^2/(2\sigma^2)}\, \mathrm{d}x.$$

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

$$\int_{-\bar{x}_p}^{-1} \mathrm{e}^{(\sigma^2-2)x^2/(2\sigma^2)}\, \mathrm{d}x \le C \mathrm{e}^{(\sigma^2-2)\bar{x}_p^2/(2\sigma^2)} \le C\left(\frac{1}{p}\right)^{(\sigma^2-2)/(\sigma^2-1)}.$$

As a conclusion, if $\sigma^2 \le 2$, then $\sup_{p\in(0,1)} I(p) < \infty$, whereas if $\sigma^2 > 2$, then for some constant $C > 0$ and any $p < 1/2$,

$$I(p) \le C\left(\frac{1}{p}\right)^{(\sigma^2-2)/(\sigma^2-1)}.$$

The bound of C_{PI} follows from Theorem 4.4. For the logarithmic Sobolev inequality, one may use the Bobkov-Götze criterion. □

4.5.3. *Two uniforms with overlapping supports*
Setting. *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*

$$C_{\mathrm{PI}}(\mu_p) \le \pi^{-2} + \frac{a^2}{3}\big(3pq(1-a)+a\big)$$

and for $p \le 1/2$,

$$C_{\mathrm{GI}}(\mu_p) \ge \frac{a^2}{600}\log(1/p).$$

Proof. It is known (see [19]) 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 $p \in (0,1)$,

$$I(p) = \int_0^a \frac{x^2}{q}\, \mathrm{d}x + \int_a^1 \frac{a^2}{p+q}\, \mathrm{d}x + \int_1^{a+1} \frac{(1+a-x)^2}{p}\, \mathrm{d}x = \frac{a^2}{3pq}\big(3pq(1-a)+a\big). \qquad \square$$

The minoration of $C_{\mathrm{GI}}(\mu_p)$ follows from Lemma 4.10:

$$150 C_{\mathrm{GI}}(\mu_p) \ge \Psi\big(\mu_p(0, a/2]\big) \int_{a/2}^{a} \frac{1}{f_p(y)}\, \mathrm{d}y = \Psi\left(\frac{pa}{2}\right)\frac{a}{2p}.$$

4.5.4. *One Gaussian and a uniform*
Setting. *Here $\mu_1 = \mathcal{N}(0,1)$ and $\mu_0 = \mathcal{U}([-1,+1])$.*

Claim. *There exists a real constant $C > 0$ such that $C_{\mathrm{GI}}(\mu_p) \ge -C\log(p)$ for every $p \in (0,1)$. Also, $C_{\mathrm{GI}}(\mu_p)$ blows up at speed $-\log(p)$ as $p \to 0^+$. Moreover, μ_p satisfies a sub-Gaussian concentration of measure for Lipschitz functions, uniformly in $p \in (0,1)$. This similarity with the Bernoulli law $\mathcal{B}(p)$ 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 $f_0 \le \kappa f_1$ for some constant $\kappa \ge 1$. Also, for every $p \in (0,1)$, the result of Section 4.5.1 gives that $C_{\mathrm{GI}}(\mu_p) \le \alpha - \beta \log(p)$ for some constants $\alpha > 0$ and $\beta > 0$ independent of p. Now, by Lemma 4.10,

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

4.5.5. *Surprising blow up*

Setting. *Here $f_1(x) = Z_1^{-1}\mathrm{e}^{-x^2}$ and $f_0(x) = Z_0^{-1}\mathrm{e}^{-|x|^a}$ for some fixed real number $a > 2$, with $Z_1 = \pi^{-1/2}$ and $Z_0 = 2\Gamma(a^{-1})a^{-1}$. Note that μ_0 has lighter tails than μ_p with $p > 0$.*

Claim. *There exists a real constant $C > 0$ which may depend on a such that*

$$C_{\mathrm{GI}}(\mu_p) \ge C\big(-\log(p)\big)^{1-2a^{-1}}$$

for small enough p. In particular, $C_{\mathrm{GI}}(\mu_p)$ blows up as $p \to 0^+$.

Comments. *As mentioned in the Introduction, we have* $\max(C_{\mathrm{GI}}(\mu_0), C_{\mathrm{GI}}(\mu_1)) < \infty$. *We have seen in Section 4.5.2 that $C_{\mathrm{GI}}(\mu_p)$ does not blow up as $p \to 0^+$ if $a = 2$. Here $a > 2$, and μ_0 has strictly lighter tails than μ_p for every $p \in (0,1)$, and moreover, this difference is at the level of the log-power of the tails, not only at the level of the constants in front of the log-power. The potential ($-\log$-density) of μ_p has multiple wells, see Fig. 1. This example shows also that the blow up speed of $C_{\mathrm{GI}}(\mu_p)$ as $p \to 0^+$ cannot be improved by considering a mixture of fully supported laws. Note that $\mu_0 \to \mathcal{U}([-1,+1])$ as $a \to \infty$, and the result is thus compatible with Section 4.5.4.*

Proof of Claim. Since $f_0 \le \kappa f_1$ for some constant $\kappa \ge 1$, Section 4.5.1 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

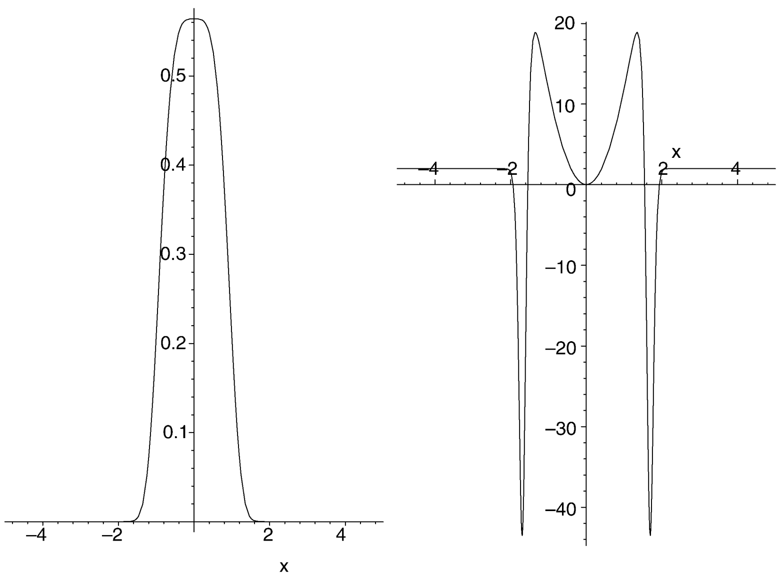

Fig. 1. Density and second derivative of $-\log$-density of μ_p for Section 4.5.5 with $p = 1/100$ and $a = 4$. The second plot reveals a deep multiple wells potential.

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 $q f_0(x) \le p f_1(x)$ for large enough x. Let us find some explicit bound on x. The inequality $q f_0(x) \le p f_1(x)$ writes $|x|^a - x^2 \ge \log(q Z_1) - \log(p Z_0)$. Now, $|x|^a - x^2 \ge \frac{1}{2}|x|^a$ for $|x|^{a-2} \ge 2$. The non-negative solution of $\frac{1}{2}|x|^a = \log(q Z_1) - \log(p Z_0)$ is

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

If p is small enough, then $|\overline{x}_p|^{a-2} \ge 2$ and therefore, $q f_0(x) \le p f_1(x)$ for any $|x| \ge \overline{x}_p$. Now, by Lemma 4.10, for small enough p,

$$150\, C_{\mathrm{GI}}(\mu_p) \ge \Psi\big(p F_1(-2\overline{x}_p) + q F_0(-2\overline{x}_p)\big) \int_{-2\overline{x}_p}^{0} \frac{1}{p f_1(u) + q f_0(u)} \,\mathrm{d}u.$$

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

$$\Psi\big(p F_1(-2\overline{x}_p) + q F_0(-2\overline{x}_p)\big) \ge \Psi\big(p F_1(-2\overline{x}_p)\big) \ge -p F_1(-2\overline{x}_p) \log(p) \ge C \frac{\mathrm{e}^{-4\overline{x}_p^2}}{\overline{x}_p} \Psi(p).$$

On the other hand, since $q f_0(x) \le p f_1(x)$ for $|x| \ge \overline{x}_p$, we get for some constant $C > 0$,

$$\int_{-2\overline{x}_p}^{0} \frac{1}{p f_1(u) + q f_0(u)} \,\mathrm{d}u \ge \int_{-2\overline{x}_p}^{-\overline{x}_p} \frac{\mathrm{d}u}{2 p f_1(u)} \ge \frac{C \mathrm{e}^{4\overline{x}_p^2}}{p \overline{x}_p}.$$

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

$$150\, C_{\mathrm{GI}}(\mu_p) \ge -C \frac{\log(p)}{\overline{x}_p^2}.$$

Now, by using the explicit expression of $\overline{x}_p$, we finally obtain for some real constant $C > 0$,

$$C_{\mathrm{GI}}(\mu_p) \ge C\big(-\log(p)\big)^{1-2a^{-1}}.$$

□

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 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, [51]) 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

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

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

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

By the Cauchy–Schwarz inequality, we get

$$(\mathbf{E}_{\mu_1} f - \mathbf{E}_{\mu_0} f)^2 \le \left(\int_{\mathbb{R}^d} \big|T(x) - x\big|^2 \,\mathrm{d}\mu_0(x)\right)\left(\int_0^1 \int_{\mathbb{R}^d} \big|\nabla f(x)\big|^2 \,\mathrm{d}\mu_{(s)}(x) \,\mathrm{d}s\right)$$

and therefore

$$(\mathbf{E}_{\mu_1} f - \mathbf{E}_{\mu_0} f)^2 \le W_2(\mu_1, \mu_0)^2 \int_{\mathbb{R}^d} \int_0^1 \big|\nabla f(x)\big|^2 \,\mathrm{d}\mu_{(s)}(x) \,\mathrm{d}s.$$

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 $\overline{\mu} \leq C_p \mu_p$ where

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

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 pathwise 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 $f : \mathbb{R}^d \to \mathbb{R}$,

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

As a consequence, we have

$$(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 \leq \bigg(\int_{S_0}\int_0^1 \big|\dot{\gamma}_x(s)\big|^2\,\mathrm{d}s\,\mu_0(\mathrm{d}x)\bigg)\bigg(\int_{S_0}\int_0^1 |\nabla f|^2\big(\gamma_x(s)\big)\,\mathrm{d}s\,\mu_0(\mathrm{d}x)\bigg).$$

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 $\overline{\mu}(A) = \int_0^1 \mu_{(s)}(A)\,\mathrm{d}s$. With this notation, we have

$$(\mathbf{E}_{\mu_0} f - \mathbf{E}_{\mu_1} f)^2 \leq \bigg(\int_{S_0}\int_0^1 \big|\dot{\gamma}_x(s)\big|^2\,\mathrm{d}s\,\mu_0(\mathrm{d}x)\bigg)\bigg(\int_{\mathbb{R}^d} |\nabla f|^2(x)\,\overline{\mu}(\mathrm{d}x)\bigg).$$

Note that

$$\bigg(\int_{S_0}\int_0^1 \big|\dot{\gamma}_x(s)\big|^2\,\mathrm{d}s\,\mu_0(\mathrm{d}x)\bigg) \geq W_2(\mu_0, \mu_1)^2$$

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 follows then immediately if there exists a real constant $C_p > 0$ such that $\overline{\mu} \leq C_p \mu_p$. The problem is thus reduced to 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 4.11 (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*

$$C_p = \frac{1}{\min(p,q)} \quad \textit{and} \quad W_2(\mu_0, \mu_1)^2 = 1.$$

Therefore, for every $0 < p < 1$ *and any smooth* $f : \mathbb{R}^2 \to \mathbb{R}$,

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

Example 4.12 (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* $\overline{\mu} \leq C_p \mu_p$.

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 $x \in \operatorname{supp}(\mu_0)$ *and every* $0 \le s \le 1$, *with* $\mathbf{1} = (e_1, e_1)$,

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

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* $0 \le s \le 1$,

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

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

$$\int_{S_0} \int_0^1 \left|\dot{\gamma}_x(s)\right|^2 \mathrm{d}s\,\mu_0(\mathrm{d}x) = \frac{8}{3} \quad \text{and} \quad \overline{\mu} \le \frac{4}{\min(p,q)} \mu_p.$$

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

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

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.

Acknowledgments

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. The final form of the paper has benefited from the fine comments of two anonymous reviewers.

References

[1] 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.) MR1845806

[2] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités (XIX, 1983/84)* 177–206. *Lecture Notes in Math.* **1123**. Springer, Berlin, 1985. MR0889476

[3] D. Bakry, M. Ledoux and F.-Y. Wang. Perturbations of functional inequalities using growth conditions. *J. Math. Pures Appl. (9)* **87** (2007) 394–407. MR2317340

[4] F. Barthe, P. Cattiaux and C. Roberto. Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry. *Rev. Mat. Iberoamericana* **22** (2006) 993–1067. MR2320410

[5] F. Barthe and C. Roberto. Sobolev inequalities for probability measures on the real line. *Studia Math.* **159** (2003) 481–497. (Dedicated to Professor Aleksander Pełczyński on the occasion of his 70th birthday (Polish).) MR2052235

[6] S. G. Bobkov. Concentration of normalized sums and a central limit theorem for noncorrelated random variables. *Ann. Probab.* **32** (2004) 2884–2907. MR2094433

[7] S. G. Bobkov. Generalized symmetric polynomials and an approximate de Finetti representation. *J. Theoret. Probab.* **18** (2005) 399–412. MR2137450

[8] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** (1999) 1–28. MR1682772

[9] F. Bolley and C. Villani. Weighted Csiszár–Kullback–Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse Math.* **14** (2005) 331–352. MR2172583

[10] L. A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.* **214** (2000) 547–563. MR1800860

[11] L. A. Caffarelli. Erratum: [*Comm. Math. Phys.* **214** (2000) 547–563]. *Comm. Math. Phys.* **225** (2002) 449–450.

[12] E. A. Carlen. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *J. Funct. Anal.* **101** (1991) 194–211. MR1132315

[13] S. Chatterjee. Spin glasses and Stein's method. Preprint, 2007. Available at arXiv:0706.3500v2 [math.PR].
[14] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** (1996) 695–750. MR1410112
[15] H. Djellout, A. Guillin and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32** (2004) 2702–2732. MR2078555
[16] B. S. Everitt and D. J. Hand. *Finite Mixture Distributions*. Chapman & Hall, London, 1981. MR0624267
[17] V. P. Fonf, J. Lindenstrauss and R. R. Phelps. Infinite dimensional convexity. In *Handbook of the Geometry of Banach Spaces* 599–670. North-Holland, Amsterdam, 2001. MR1863703
[18] S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models.* Springer, New York, 2006.
[19] I. Gentil. Inégalités de Sobolev logarithmique et de Poincaré pour la loi uniforme. Unpublished 2004. Available on the author's web page.
[20] I. Gentil and C. Roberto. Spectral gaps for spin systems: Some non-convex phase examples. *J. Funct. Anal.* **180** (2001) 66–84. MR1814423
[21] N. Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. Preprint, 2008. Available at arXiv: 0804.3089 [math.PR].
[22] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *Amer. J. Math.* **105** (1983) 843–854. MR0708367
[23] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** (1975) 1061–1083. MR0420249
[24] L. Gross. Hypercontractivity, logarithmic Sobolev inequalities, and applications: A survey of surveys. In *Diffusion, Quantum Theory, and Radically Elementary Mathematics* 45–73. *Math. Notes* **47**. Princeton Univ. Press, Princeton, NJ, 2006. MR2325763
[25] B. Helffer. *Semiclassical Analysis, Witten Laplacians, and Statistical Mechanics. Series in Partial Differential Equations and Applications* **1**. World Scientific Publishing, River Edge, NJ, 2002. MR1936110
[26] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** (1963) 13–30. MR0144363
[27] R. Holley and D. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Statist. Phys.* **46** (1987) 1159–1194. MR0893137
[28] 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) 1741–1765. MR2099650
[29] O. Johnson. Convergence of the Poincaré constant. *Teor. Veroyatnost. i Primenen.* **48** (2003) 615–620. MR2141356
[30] I. Kontoyiannis and M. Madiman. Measure concentration for compound Poisson distributions. *Electron. Comm. Probab.* **11** (2006) 45–57. MR2219345
[31] L. Kontoyiannis and M. Madiman. Entropy, compound Poisson approximation, log-Sobolev inequalities and measure concentration. In *Information Theory Workshop, 24–29 Oct. 2004. IEEE* 71–75, 2004.
[32] R. Latała. On some inequalities for Gaussian measures. In *Proceedings of the International Congress of Mathematicians, Vol. II Beijing, 2002* 813–822. Higher Ed. Press, Beijing, 2002. MR1957087
[33] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXIII* 120–216. Springer, Berlin, 1999. MR1767995
[34] M. Ledoux. *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI, 2001. MR1849347
[35] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Ann. Appl. Probab.* **12** (2002) 581–606. MR1910641
[36] K. Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory* **32** (1986) 445–446. MR0838213
[37] K. Marton. Bounding $\overline{d}$-distance by informational divergence: A method to prove measure concentration. *Ann. Probab.* **24** (1996) 857–866. MR1404531
[38] V. G. Maz'ja. *Sobolev Spaces*. Springer, Berlin, 1985. (Translated from the Russian by T. O. Shaposhnikova.) MR0817985
[39] G. McLachlan and K. Basford. *Mixture Models. Statistics: Textbooks and Monographs* **84**. Marcel Dekker Inc., New York, 1988. MR0926484
[40] G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley, New York, 2000. MR1789474
[41] L Miclo. Quand est-ce que des bornes de Hardy permettent de calculer une constante de Poincaré exacte sur la droite?. Preprint, 2005. Available at http://hal.archives-ouvertes.fr/hal-00017875/en/.
[42] F. Otto and M. G. Reznikoff. A new criterion for the logarithmic Sobolev inequality and two applications. *J. Funct. Anal.* **243** (2007) 121–157. MR2291434
[43] R. R. Phelps, *Lectures on Choquet's Theorem*, 2nd edition. *Lecture Notes in Math.* **1757**. Springer, Berlin, 2001. MR1835574
[44] S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models*. Wiley, Chichester, 1991. MR1105086
[45] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1996)* 301–413. *Lecture Notes in Math.* **1665**. Springer, Berlin, 1997. MR1490046
[46] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control* **2** (1959) 101–112. MR0109101
[47] V. N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.* (1979) 1–178. (Cover to cover translation of *Trudy Mat. Inst. Steklov* **141** (1976).) MR0530375
[48] A. Takatsu. On Wasserstein geometry of the space of Gaussian measures, 2008. Available at arXiv:0801.2250 [math.DG].
[49] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **6** (1996) 587–600. MR1392331
[50] D. M. Titterington, A. F. M. Smith and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions*. Wiley, Chichester, 1985. MR0838090
[51] C. Villani. *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI, 2003. MR1964483
[52] C. Villani. *Optimal Transport, Old and New*. Springer, Berlin, 2009. MR2459454