Supplement to *Log-Concave Sampling*

Sinho Chewi

September 26, 2024

Abstract

This document contains supplementary material to the book [Che24] which was omitted for space.

Contents

1 Supplement to Chapter 2

1.1 Proof of Marton's tensorization

We recall the statement and then give a proof.

Theorem 1.1 (Marton's tensorization [Mar96]). Let $\mathcal{X}_1, \dots, \mathcal{X}_N$ be Polish spaces equipped with probability measures $\pi_1, \dots, \pi_N$ respectively. Let $\mathcal{X} \coloneqq \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$ be equipped with the product measure $\pi \coloneqq \pi_1 \otimes \cdots \otimes \pi_N$.

Let $\varphi : [0, \infty) \to [0, \infty)$ be convex and for $i \in [N]$, let $c_i : \mathcal{X}_i \times \mathcal{X}_i \to [0, \infty)$ be a lower semicontinuous cost function. Suppose that

$$\inf_{\gamma_i \in \mathcal{C}(\pi_i, \nu_i)} \varphi\Bigl(\int c_i \, \mathrm{d}\gamma_i\Bigr) \le 2\sigma^2 \, \mathsf{KL}(\nu_i \parallel \pi_i)\,, \qquad \forall \nu_i \in \mathcal{P}(\mathcal{X}_i)\,,\ \forall i \in [N]\,.$$

Then, it holds that

$$\inf_{\gamma \in \mathcal{C}(\pi, \nu)} \sum_{i=1}^N \varphi\Bigl(\int c_i(x_i, y_i) \, \gamma(\mathrm{d}x_{1:N}, \mathrm{d}y_{1:N})\Bigr) \le 2\sigma^2 \, \mathsf{KL}(\nu \parallel \pi)\,, \qquad \forall \nu \in \mathcal{P}(\mathcal{X})\,.$$

Proof. The proof goes by induction on N, with $N = 1$ being trivial. So, assume that the result is true in dimension N, and let us prove it for dimension $N + 1$.

Let $\nu \in \mathcal{P}(\mathcal{X}) = \mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_{N+1})$, let $\nu_{1:N}$ denote its $\mathcal{X}_1 \times \cdots \times \mathcal{X}_N$ marginal, and let $\nu_{N+1|1:N}$ denote the corresponding conditional kernel (and similarly for π). Let K denote the set of conditional kernels $y_{1:N} \mapsto \gamma_{N+1|1:N}(\cdot \mid y_{1:N})$ such that for $\nu_{1:N}$-a.e. $y_{1:N} \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$, it holds that $\gamma_{N+1|1:N}(\cdot \mid y_{1:N}) \in \mathcal{C}(\pi_{N+1}, \nu_{N+1|1:N}(\cdot \mid y_{1:N}))$. Instead of minimizing over all $\gamma \in \mathcal{C}(\pi, \nu)$, we can minimize over couplings γ such that for all bounded $f \in C(\mathcal{X} \times \mathcal{X})$,

$$\int f \, \mathrm{d}\gamma = \int \Bigl(\int f(x_{1:N+1}, y_{1:N+1}) \, \gamma_{N+1|1:N}(\mathrm{d}x_{N+1}, \mathrm{d}y_{N+1} \mid y_{1:N})\Bigr) \, \gamma_{1:N}(\mathrm{d}x_{1:N}, \mathrm{d}y_{1:N})\,,$$

for some $\gamma_{1:N} \in \mathcal{C}(\pi_{1:N}, \nu_{1:N})$ and $\gamma_{N+1|1:N} \in \mathsf{K}$.[1] Thus,

$$\inf_{\gamma \in \mathcal{C}(\pi, \nu)} \sum_{i=1}^{N+1} \varphi\Bigl(\int c_i(x_i, y_i) \, \gamma(\mathrm{d}x_{1:N+1}, \mathrm{d}y_{1:N+1})\Bigr)$$

[1] Suppose $N = 2$ and $(X_1, X_2) \sim \pi$ and $(Y_1, Y_2) \sim \nu$. Observe that a general coupling $p \in \mathcal{C}(\pi, \nu)$ factorizes as $p(x_1, x_2, y_1, y_2) = p_{X_1}(x_1) \, p_{X_2}(x_2) \, p_{Y_1, Y_2 | X_1, X_2}(y_1, y_2 \mid x_1, x_2)$. In contrast, we are restricting to couplings of the form $p(x_1, x_2, y_1, y_2) = p_{X_1}(x_1) \, p_{Y_1|X_1}(y_1 \mid x_1) \, p_{X_2}(x_2) \, p_{Y_2|X_2, Y_1}(y_2 \mid x_2, y_1)$.

$$
\begin{aligned}
&\le \inf_{\gamma_{1:N}\in\mathcal{C}(\pi_{1:N},\nu_{1:N})}\Big\{\sum_{i=1}^N \varphi\Big(\int c_i(x_i,y_i)\,\gamma_{1:N}(\mathrm{d}x_{1:N},\mathrm{d}y_{1:N})\Big) \\
&\qquad\qquad + \inf_{\gamma_{N+1|1:N}\in\mathsf{K}} \varphi\Big(\int \big(\int c_{N+1}\,\mathrm{d}\gamma_{N+1|1:N}(\cdot\mid y_{1:N})\big)\,\gamma_{1:N}(\mathrm{d}x_{1:N},\mathrm{d}y_{1:N})\Big)\Big\} \\
&\le \inf_{\gamma_{1:N}\in\mathcal{C}(\pi_{1:N},\nu_{1:N})}\Big\{\sum_{i=1}^N \varphi\Big(\int c_i(x_i,y_i)\,\gamma_{1:N}(\mathrm{d}x_{1:N},\mathrm{d}y_{1:N})\Big) \\
&\qquad\qquad + \inf_{\gamma_{N+1|1:N}\in\mathsf{K}} \int \varphi\Big(\int c_{N+1}\,\mathrm{d}\gamma_{N+1|1:N}(\cdot\mid y_{1:N})\Big)\,\gamma_{1:N}(\mathrm{d}x_{1:N},\mathrm{d}y_{1:N})\Big\}\,.
\end{aligned}
$$

Then, after checking that the integrands are indeed measurable,

$$
\begin{aligned}
&\inf_{\gamma_{N+1|1:N}\in\mathsf{K}} \int \varphi\Big(\int c_{N+1}\,\mathrm{d}\gamma_{N+1|1:N}(\cdot\mid y_{1:N})\Big)\,\gamma_{1:N}(\mathrm{d}x_{1:N},\mathrm{d}y_{1:N}) \\
&\quad = \int \inf_{\gamma_{N+1|1:N}\in\mathcal{C}(\pi_{N+1},\nu_{N+1|1:N}(\cdot|y_{1:N}))} \varphi\Big(\int c_{N+1}\,\mathrm{d}\gamma_{N+1|1:N}(\cdot\mid y_{1:N})\Big)\,\gamma_{1:N}(\mathrm{d}x_{1:N},\mathrm{d}y_{1:N}) \\
&\quad \le 2\sigma^2 \int \mathsf{KL}\big(\nu_{N+1|1:N}(\cdot\mid y_{1:N}) \,\big\|\, \pi_{N+1}\big)\,\gamma_{1:N}(\mathrm{d}x_{1:N},\mathrm{d}y_{1:N}) \\
&\quad = 2\sigma^2 \int \mathsf{KL}\big(\nu_{N+1|1:N}(\cdot\mid y_{1:N}) \,\big\|\, \pi_{N+1}\big)\,\nu_{1:N}(\mathrm{d}y_{1:N})\,,
\end{aligned}
$$

where we used the assumption. On the other hand, the inductive hypothesis is

$$
\inf_{\gamma_{1:N}\in\mathcal{C}(\pi_{1:N},\nu_{1:N})} \sum_{i=1}^N \varphi\Big(\int c_i(x_i,y_i)\,\gamma_{1:N}(\mathrm{d}x_{1:N},\mathrm{d}y_{1:N})\Big) \le 2\sigma^2\,\mathsf{KL}(\nu_{1:N}\,\|\,\pi_{1:N})\,.
$$

The chain rule for the KL divergence yields

$$
\mathsf{KL}(\nu\,\|\,\pi) = \mathsf{KL}(\nu_{1:N}\,\|\,\pi_{1:N}) + \int \mathsf{KL}\big(\nu_{N+1|1:N}(\cdot\mid y_{1:N}) \,\big\|\, \pi_{N+1}\big)\,\nu_{1:N}(\mathrm{d}y_{1:N})\,.
$$

Therefore, we have proven

$$
\inf_{\gamma\in\mathcal{C}(\pi,\nu)} \sum_{i=1}^{N+1} \varphi\Big(\int c_i(x_i,y_i)\,\gamma(\mathrm{d}x_{1:N+1},\mathrm{d}y_{1:N+1})\Big) \le 2\sigma^2\,\mathsf{KL}(\nu\,\|\,\pi)\,. \qquad \square
$$

The preceding proof is supposed to be a straightforward proof by induction, but it is rather cumbersome to write out precisely.

As an application of the tensorization principle, we will examine the tensorization properties of the T_1 inequality.

Example 1.1 (tensorization of T_1)**.** We will use the cost $c_i = \mathsf{d}_i$, where d_i is a lower semicontinuous metric on $\mathcal{X}_i$, and we take the convex function $\varphi(x) := x^2$. Suppose that for each $i \in [N]$, the measure $\pi_i \in \mathcal{P}(\mathcal{X})$ satisfies the T_1 inequality

$$W_1^2(\nu_i, \pi_i) \le 2\sigma^2 \,\mathsf{KL}(\nu_i \,\|\, \pi_i), \qquad \forall \nu_i \in \mathcal{P}(\mathcal{X}_i)\,.$$

Let $\pi := \pi_1 \otimes \cdots \otimes \pi_N$ be the product measure and let $\nu \in \mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)$. Suppose also that $\alpha_1, \dots, \alpha_N > 0$ are numbers with $\sum_{i=1}^N \alpha_i^2 = 1$. Then, Marton's tensorization (Theorem 1.1) yields

$$\begin{aligned}
2\sigma^2 \,\mathsf{KL}(\nu \,\|\, \pi) &\ge \Big(\sum_{i=1}^N \alpha_i^2\Big) \inf_{\gamma \in \mathcal{C}(\pi,\nu)} \sum_{i=1}^N \Big(\int \mathsf{d}_i(x_i, y_i)\, \gamma(\mathrm{d}x_{1:N}, \mathrm{d}y_{1:N})\Big)^2 \\
&\ge \inf_{\gamma \in \mathcal{C}(\pi,\nu)} \Big(\int \sum_{i=1}^N \alpha_i \mathsf{d}_i(x_i, y_i)\, \gamma(\mathrm{d}x_{1:N}, \mathrm{d}y_{1:N})\Big)^2,
\end{aligned}$$

where we used the Cauchy–Schwarz inequality. This is a T_1 inequality for the weighted distance $\mathsf{d}_\alpha(x_{1:N}, y_{1:N}) := \sum_{i=1}^N \alpha_i \mathsf{d}_i(x_i, y_i)$.

Together with results from §1.2.3, this tensorization result is already powerful enough to recover the bounded differences concentration inequality (see Exercise 1.3), but it is not fully satisfactory as it yields a transport inequality for a weighted metric. On the other hand, we recall that Marton's argument shows that the T_2 inequality *does* tensorize. This is explored further in §1.3.

1.2 Concentration of measure

Here, we expand on the relationship between functional inequalities and concentration of measure. Recall that we work on a Polish space (that is, a complete separable metric space) $(\mathcal{X}, \mathsf{d})$ unless otherwise stated.

1.2.1 Equivalence between the mean and the median

Some statements regarding concentration are more easily phrased in terms of concentration around the median rather than around the mean. The following result shows that, up to numerical constants, the mean and the median are equivalent. To state the result in generality, we introduce the idea of an Orlicz norm.

Definition 1.1 (Orlicz norm)**.** If $\psi : [0,\infty) \to [0,\infty)$ is a convex strictly increasing function with $\psi(0) = 0$ and $\psi(x) \to \infty$ as $x \to \infty$, then it is an **Orlicz function**.

For a real-valued random variable X, its **Orlicz norm** is defined to be

$$\|X\|_\psi := \inf\Big\{t > 0 \,\Big|\, \mathbb{E}\,\psi\big(\frac{|X|}{t}\big) \le 1\Big\}.$$

Examples of Orlicz functions include $\psi(x) = x^p$ for $p \ge 1$, for which the corresponding Orlicz norm is the $L^p(\mathbb{P})$ norm, and $\psi_2(x) := \exp(x^2) - 1$ for which the Orlicz norm $\|X\|_{\psi_2}$ captures the sub-Gaussianity of X.

Lemma 1.1 (mean and median [Mil09])**.** Let ψ be an Orlicz function and let X be a real-valued random variable. Then,

$$\frac{1}{2}\,\|X - \mathbb{E}\,X\|_\psi \le \|X - \operatorname{med} X\|_\psi \le 3\,\|X - \mathbb{E}\,X\|_\psi\,.$$

Proof. We can assume that X is not constant; from the properties of Orlicz functions, $\psi^{-1}(t)$ is well-defined for any $t > 0$. Then,

$$\begin{aligned}\|X - \mathbb{E}\,X\|_\psi &\le \|X - \operatorname{med} X\|_\psi + \|\operatorname{med} X - \mathbb{E}\,X\|_\psi \\ &= \|X - \operatorname{med} X\|_\psi + |\operatorname{med} X - \mathbb{E}\,X|\,\|1\|_\psi \\ &\le \|X - \operatorname{med} X\|_\psi + \mathbb{E}|X - \operatorname{med} X|\,\|1\|_\psi\,.\end{aligned}$$

Since

$$\mathbb{E}\,\psi\Big(\frac{|X - \operatorname{med} X|}{\mathbb{E}|X - \operatorname{med} X|\,\|1\|_\psi}\Big) \ge \psi\Big(\frac{\mathbb{E}|X - \operatorname{med} X|}{\mathbb{E}|X - \operatorname{med} X|\,\|1\|_\psi}\Big) = \psi\big(\frac{1}{\|1\|_\psi}\big) = 1\,,$$

it implies $\mathbb{E}|X - \operatorname{med} X|\,\|1\|_\psi \le \|X - \operatorname{med} X\|_\psi$.

Next, assume that $\operatorname{med} X \ge \mathbb{E}\,X$ (or else replace X by $-X$). Then,

$$\begin{aligned}\frac{1}{2} &\le \mathbb{P}\{X \ge \operatorname{med} X\} \le \mathbb{P}\{|X - \mathbb{E}\,X| \ge \operatorname{med} X - \mathbb{E}\,X\} \\ &\le \frac{1}{\psi((\operatorname{med} X - \mathbb{E}\,X)/\|X - \mathbb{E}\,X\|_\psi)}\,,\end{aligned}$$

so that

$$|\operatorname{med} X - \mathbb{E}\,X| \le \psi^{-1}(2)\,\|X - \mathbb{E}\,X\|_\psi\,.$$

Therefore,

$$\|X - \operatorname{med} X\|_\psi \le \|X - \mathbb{E}\,X\|_\psi + \|\mathbb{E}\,X - \operatorname{med} X\|_\psi \le \big(1 + \|1\|_\psi\,\psi^{-1}(2)\big)\,\|X - \mathbb{E}\,X\|_\psi\,.$$

Note, however, that $\|1\|_\psi = 1/\psi^{-1}(1)$. Since $\psi(\psi^{-1}(2)/2) \le 1$ by convexity (and the property $\psi(0) = 0$), it implies $\psi^{-1}(2) \le 2\psi^{-1}(1)$, and we obtain the result. $\square$

1.2.2 The Herbst argument

In this section, we specialize to the case where $(\mathcal{X}, \mathsf{d})$ is the Euclidean space $\mathbb{R}^d$.

To put it succinctly, the idea of the Herbst argument is to apply functional inequalities, such as the Poincaré inequality or the log-Sobolev inequality, to the moment-generating function of a 1-Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$ in order to deduce a concentration inequality for f. We illustrate this with the log-Sobolev inequality, which implies, for any $\lambda \in \mathbb{R}$,

$$\begin{aligned} \operatorname{ent}_\pi \exp(\lambda f) &\le 2C_{\mathsf{LSI}}\, \mathbb{E}_\pi\Big[\Big\|\frac{\lambda \exp(\lambda f/2)}{2}\, \nabla f\Big\|^2\Big] = \frac{C_{\mathsf{LSI}}\,\lambda^2}{2}\, \mathbb{E}_\pi[\exp(\lambda f)\, \|\nabla f\|^2] \\ &\le \frac{C_{\mathsf{LSI}}\,\lambda^2}{2}\, \mathbb{E}_\pi \exp(\lambda f)\,. \end{aligned} \tag{1.1}$$

The next lemma shows how to apply this inequality.

Lemma 1.2 (Herbst argument)**.** Suppose that a random variable X satisfies

$$\operatorname{ent} \exp(\lambda X) \le \frac{\lambda^2 \sigma^2}{2}\, \mathbb{E} \exp(\lambda X) \qquad \text{for all } \lambda \ge 0\,.$$

Then, it holds that

$$\mathbb{E} \exp\{\lambda\,(X - \mathbb{E}\, X)\} \le \exp \frac{\lambda^2 \sigma^2}{2} \qquad \text{for all } \lambda \ge 0\,.$$

In particular, via a standard Chernoff inequality,

$$\mathbb{P}\{X \ge \mathbb{E}\, X + t\} \le \exp\big(-\frac{t^2}{2\sigma^2}\big) \qquad \text{for all } t \ge 0\,.$$

Proof. Let $\tau(\lambda) := \lambda^{-1} \ln \mathbb{E} \exp\{\lambda\,(X - \mathbb{E}\, X)\}$. We leave it to the reader to check the calculus identity

$$\tau'(\lambda) = \frac{1}{\lambda^2}\, \frac{\operatorname{ent} \exp(\lambda X)}{\mathbb{E} \exp(\lambda X)}\,. \tag{1.2}$$

Since $\tau(\lambda) \to 0$ as $\lambda \searrow 0$, the assumption of the lemma yields $\tau(\lambda) \le \lambda\sigma^2/2$. □

The calculation in (1.1) shows that the assumption of the Herbst argument is satisfied for *all* 1-Lipschitz functions f, with $\sigma^2 = C_{\mathsf{LSI}}$. Hence, we deduce a concentration inequality for Lipschitz functions, which we formally state in the next theorem together with the corresponding result under a Poincaré inequality.

Theorem 1.2. Let $\pi \in \mathcal{P}(\mathbb{R}^d)$, and let $f : \mathbb{R}^d \to \mathbb{R}$ be a 1-Lipschitz function.

1. If π satisfies a Poincaré inequality with constant C_{PI}, then for all $t \geq 0$,
$$\pi\{f - \mathbb{E}_\pi f \geq t\} \leq 3 \exp\Bigl(-\frac{t}{\sqrt{C_{\mathsf{PI}}}}\Bigr).$$

2. If π satisfies a log-Sobolev inequality with constant C_{LSI}, then for all $t \geq 0$,
$$\pi\{f - \mathbb{E}_\pi f \geq t\} \leq \exp\Bigl(-\frac{t^2}{2C_{\mathsf{LSI}}}\Bigr).$$

The Poincaré case is left as Exercise 1.1.

1.2.3 Transport inequalities and concentration

Next, we show that a T_1 transport inequality is equivalent to sub-Gaussian concentration of Lipschitz functions, which was proven by Bobkov and Götze. The proof shows that in a sense, the two statements are dual to each other.

Theorem 1.3 (Bobkov–Götze [BG99])**.** Let $\pi \in \mathcal{P}_1(\mathcal{X})$. The following are equivalent.

1. The function f is σ^2-sub-Gaussian with respect to π, in the sense that
$$\mathbb{E}_\pi \exp\{\lambda\,(f - \mathbb{E}_\pi f)\} \leq \exp\frac{\lambda^2\sigma^2}{2} \qquad \text{for all } \lambda \in \mathbb{R},$$
for every 1-Lipschitz function $f : \mathcal{X} \to \mathbb{R}$.

2. The measure π satisfies $\mathsf{T}_1(\sigma^2)$.

Proof. Let $\mathrm{Lip}_1(\mathcal{X})$ denote the space of 1-Lipschitz and mean-zero functions on $\mathcal{X}$. Lipschitz concentration can be stated as
$$\sup_{\lambda\in\mathbb{R}} \sup_{f\in\mathrm{Lip}_1(\mathcal{X})} \Bigl\{\ln \int \exp(\lambda f)\,\mathrm{d}\pi - \frac{\lambda^2\sigma^2}{2}\Bigr\} \leq 0\,.$$

By Donsker–Varadhan duality, this is equivalent to
$$\sup_{\lambda\in\mathbb{R}} \sup_{f\in\mathrm{Lip}_1(\mathcal{X})} \sup_{\nu\in\mathcal{P}(\mathcal{X})} \Bigl\{\lambda\,\Bigl(\int f\,\mathrm{d}\nu - \int f\,\mathrm{d}\pi\Bigr) - \mathsf{KL}(\nu \,\|\, \pi) - \frac{\lambda^2\sigma^2}{2}\Bigr\} \leq 0\,,$$

where we recall that $\int f \, \mathrm{d}\pi = 0$ for $f \in \mathrm{Lip}_1(\mathcal{X})$. If we first evaluate the supremum over $\lambda \in \mathbb{R}$, then we obtain the statement

$$\sup_{f \in \mathrm{Lip}_1(\mathcal{X})} \sup_{\nu \in \mathcal{P}(\mathcal{X})} \Bigl\{\frac{1}{2\sigma^2} \Bigl(\int f \, \mathrm{d}\nu - \int f \, \mathrm{d}\pi\Bigr)^2 - \mathsf{KL}(\nu \,\|\, \pi)\Bigr\} \le 0\,,$$

If we next evaluate the supremum over functions $f \in \mathrm{Lip}_1(\mathcal{X})$ using the Kantorovich duality formula for W_1, we obtain

$$\sup_{\nu \in \mathcal{P}(\mathcal{X})} \Bigl\{\frac{W_1^2(\nu, \pi)}{2\sigma^2} - \mathsf{KL}(\nu \,\|\, \pi)\Bigr\} \le 0\,,$$

which is the T_1 inequality. □

Using the fact that the W_1 distance for the trivial metric $\mathrm{d}(x, y) = \mathbb{1}\{x \neq y\}$ coincides with the TV distance[2], the Bobkov–Götze theorem implies that two classical inequalities in probability theory, Hoeffding's inequality and Pinsker's inequality, are in fact equivalent to each other (see Exercise 1.2).

Although the T_1 inequality implies sub-Gaussian concentration for *all* Lipschitz functions, it is in fact equivalent to sub-Gaussian concentration of a single function, the distance function $\mathrm{d}(\cdot, x_0)$ for some $x_0 \in \mathcal{X}$. The next theorem is not used often because the quantitative dependence of the equivalence can be crude, but it is worth knowing. A proof can be found in, e.g., [BV05].

Theorem 1.4. Let $\pi \in \mathcal{P}_1(\mathcal{X})$ and $x_0 \in \mathcal{X}$. The following are equivalent:

1. π satisfies a T_1 inequality.

2. There exists $c > 0$ such that $\mathbb{E}_\pi \exp(c\, \mathrm{d}(\cdot, x_0)^2) < \infty$.

Transport inequalities offer a flexible and powerful method for characterizing and proving concentration inequalities, as we will see in the next section. Before doing so, however, we wish to also demonstrate how concentration of measure, formulated via blow-up of sets, can be deduced directly from a T_1 inequality.

Suppose that $\mathsf{T}_1(\sigma^2)$ holds, i.e.,

$$W_1^2(\mu, \pi) \le 2\sigma^2 \, \mathsf{KL}(\mu \,\|\, \pi) \qquad \text{for all } \mu \in \mathcal{P}_1(\mathcal{X})\,,\ \mu \ll \pi\,.$$

[2] One has to be slightly careful since for the trivial metric, $(\mathcal{X}, \mathrm{d})$ is usually not separable.

For any disjoint sets A, B, with $\pi(A)\,\pi(B) > 0$, if we let $\pi(\cdot \mid A)$ (resp. $\pi(\cdot \mid B)$) denote the distribution π conditioned on A (resp. B), then

$$\begin{aligned}\mathsf{d}(A,B) &\le W_1\big(\pi(\cdot \mid A), \pi(\cdot \mid B)\big) \le W_1\big(\pi(\cdot \mid A), \pi\big) + W_1\big(\pi(\cdot \mid B), \pi\big) \\ &\le \sqrt{2\sigma^2\,\mathsf{KL}\big(\pi(\cdot \mid A) \,\big\|\, \pi\big)} + \sqrt{2\sigma^2\,\mathsf{KL}\big(\pi(\cdot \mid B) \,\big\|\, \pi\big)}\,.\end{aligned}$$

However,

$$\mathsf{KL}\big(\pi(\cdot \mid A) \,\big\|\, \pi\big) = \int_A \frac{\pi(\mathrm{d}x)}{\pi(A)} \ln \frac{1}{\pi(A)} = \ln \frac{1}{\pi(A)}\,,$$

so that

$$\mathsf{d}(A,B) \le \sqrt{2\sigma^2 \ln \frac{1}{\pi(A)}} + \sqrt{2\sigma^2 \ln \frac{1}{\pi(B)}}\,.$$

In particular, if we take $B = (A^\varepsilon)^{\mathsf{c}}$ where $\pi(A) \ge \frac{1}{2}$, then $\mathsf{d}(A,B) \ge \varepsilon$. Hence, for all $\varepsilon \ge 2\sqrt{2\sigma^2 \ln 2}$, it holds that $\frac{\varepsilon}{2} \le \sqrt{2\sigma^2 \ln \frac{1}{\pi(B)}}$, or

$$\pi\big((A^\varepsilon)^{\mathsf{c}}\big) \le \exp\big(-\frac{\varepsilon^2}{8\sigma^2}\big) \qquad \text{for all } \varepsilon \ge \sqrt{8 \ln 2}\,\sigma\,. \tag{1.3}$$

1.3 Tensorization and Gozlan's theorem

Our goal is now to investigate the relationship between concentration and tensorization. Although results like the Bobkov–Götze theorem (Theorem 1.3) provide us with powerful tools to establish concentration results, so far there is nothing inherently *high-dimensional* about these phenomena.

Indeed, to discuss dimensionality, we should move to the product space $\mathcal{X}^N$ and ask when concentration results can hold *independently* of N. If such a statement holds, then the concentration inequality typically becomes stronger[3] as N becomes larger.

For instance, when $\mathcal{X} = \mathbb{R}$, then we know that the Poincaré and log-Sobolev inequalities both tensorize: if they hold for $\pi \in \mathcal{P}(\mathbb{R})$ with a constant C, then they also hold for $\pi^{\otimes N} \in \mathcal{P}(\mathbb{R}^N)$ with the *same* constant C. Since these inequalities imply powerful concentration results (Theorem 1.2), they yield examples of genuinely high-dimensional concentration.

For transport inequalities, the tensorization for the T_1 inequality is unsatisfactory in the sense that once we equip $\mathcal{X}^N$ with the product metric $\mathsf{d}(x_{1:N}, x'_{1:N})^2 := \sum_{i=1}^N \mathsf{d}(x_i, x'_i)^2$,

[3] Here, the word "stronger" is not precisely defined but it means something akin to "more useful" or "produces more surprising consequences".

the validity of $\mathsf{T}_1(C)$ for $\pi \in \mathcal{P}(\mathcal{X})$ does *not* imply the validity of $\mathsf{T}_1(C)$ for $\pi^{\otimes N} \in \mathcal{P}(\mathcal{X}^N)$ with the same constant C. In fact, from Example 1.1, we expect that the T_1 constant for $\pi^{\otimes N}$ can grow as $\sqrt{N}$. On the other hand, Marton's tensorization (Theorem 1.1) shows that the T_2 inequality tensorizes. Since the T_2 inequality on $\mathcal{X}^N$ implies the T_1 inequality on $\mathcal{X}^N$ (trivially), it in turn implies high-dimensional concentration via the Bobkov–Götze equivalence (Theorem 1.3).

In this section, we prove the surprising fact that high-dimensional concentration is actually *equivalent* to the T_2 inequality, in a sense that we shall make precise shortly.

First, we need a few preliminary results, which we shall not prove. The first one is a straightforward technical lemma (see Exercise 1.4).

Lemma 1.3. Let $\pi \in \mathcal{P}_2(\mathcal{X})$.

1. The mapping $(x_1, \dots, x_N) \mapsto W_2(N^{-1} \sum_{i=1}^N \delta_{x_i}, \pi)$ is $N^{-1/2}$-Lipschitz.
2. (Wasserstein law of large numbers) Suppose that $(X_i)_{i=1}^\infty \overset{\text{i.i.d.}}{\sim} \pi$, and that for some $x_0 \in \mathcal{X}$ and some $\varepsilon > 0$, it holds that $\mathbb{E}[\mathsf{d}(x_0, X_1)^{2+\varepsilon}] < \infty$. Then,

$$\mathbb{E}\, W_2\Big(\frac{1}{N}\sum_{i=1}^N \delta_{X_i}, \pi\Big) \to 0 \qquad \text{as } N \to \infty\,.$$

The second result, Sanov's theorem, is a foundational theorem from large deviations. Although Sanov's theorem is of fundamental importance in its own right, it would take us too far afield to develop large deviations theory here, so we invoke it as a black box.

Theorem 1.5 (Sanov's theorem)**.** Let $(X_i)_{i=1}^\infty \overset{\text{i.i.d.}}{\sim} \pi$ and let $\pi_N := N^{-1}\sum_{i=1}^N \delta_{X_i}$ denote the empirical measure. Then, for any Borel set $A \subseteq \mathcal{P}(\mathcal{X})$, it holds that

$$\begin{aligned} -\inf_{\operatorname{int} A} \mathsf{KL}(\cdot \,\|\, \pi) &\le \liminf_{N\to\infty} \frac{1}{N} \ln \mathbb{P}\{\pi_N \in A\} \\ &\le \limsup_{N\to\infty} \frac{1}{N} \ln \mathbb{P}\{\pi_N \in A\} \le -\inf_{\bar{A}} \mathsf{KL}(\cdot \,\|\, \pi)\,. \end{aligned}$$

We are now ready to establish the equivalence.

Theorem 1.6 (Gozlan)**.** The measure $\pi \in \mathcal{P}_2(\mathcal{X})$ satisfies $\mathsf{T}_2(\sigma^2)$ if and only if for all $N \in \mathbb{N}^+$ and all 1-Lipschitz $f : \mathcal{X}^N \to \mathbb{R}$, the centered function $f - \mathbb{E}_{\pi^{\otimes N}} f$ is σ^2-sub-Gaussian under $\pi^{\otimes N}$.

Proof. It remains to prove the converse implication. Fix $t > 0$ and apply the assumption statement to the $N^{-1/2}$-Lipschitz function $(x_1, \ldots, x_N) \mapsto W_2(N^{-1} \sum_{i=1}^N \delta_{x_i}, \pi)$. It implies

$$\mathbb{P}\{W_2(\pi_N, \pi) > t\} \le \exp\Bigl(-\frac{N\,\{t - \mathbb{E}\,W_2(\pi_N, \pi)\}^2}{2\sigma^2}\Bigr),$$

where $\pi_N := N^{-1} \sum_{i=1}^N \delta_{X_i}$, with $(X_i)_{i\in\mathbb{N}^+} \overset{\text{i.i.d.}}{\sim} \pi$. On the other hand, the lower semicontinuity of W_2 implies that $\{\nu \in \mathcal{P}(\mathcal{X}) \mid W_2(\mu, \nu) > t\}$ is open. By Sanov's theorem (Theorem 1.5), we obtain

$$\begin{aligned} -\inf\{\mathsf{KL}(\nu \,\|\, \pi) \mid W_2(\nu, \pi) > t\} &\le \liminf_{N\to\infty} \frac{1}{N} \ln \mathbb{P}\{W_2(\pi_N, \pi) > t\} \\ &\le -\limsup_{N\to\infty} \frac{\{t - \mathbb{E}\,W_2(\pi_N, \pi)\}^2}{2\sigma^2} = -\frac{t^2}{2\sigma^2}\,, \end{aligned}$$

where the last inequality comes from the Wasserstein law of large numbers (our assumption implies that π has sub-Gaussian tails, which in particular means $\mathbb{E}[\mathsf{d}(x, X_1)^p] < \infty$ for any $x \in \mathcal{X}$ and any $p \ge 1$).

We have proven that $W_2(\nu, \pi) > t$ implies $\mathsf{KL}(\nu \,\|\, \pi) \ge t^2/(2\sigma^2)$, which is seen to be equivalent to the T_2 inequality. □

Observe in particular that this theorem implies the Otto–Villani theorem: due to tensorization and the Herbst argument (Lemma 1.2), a log-Sobolev inequality implies high-dimensional sub-Gaussian concentration of Lipschitz functions, which by Gözlan's theorem is equivalent to a T_2 inequality.

1.4 Exercises

Exercise 1.1 (Herbst argument)**.** Consider the Herbst argument from §1.2.2.

1. Verify the calculus identity (1.2) in the Herbst argument.

2. Suppose that X is a real-valued random variable satisfying the following condition: for all $\lambda \ge 0$, it holds that

$$\operatorname{var}\exp\frac{\lambda X}{2} \le \frac{\lambda^2\sigma^2}{4}\,\mathbb{E}\exp(\lambda X)\,.$$

Let $\eta(\lambda) := \mathbb{E}\exp(\lambda X)$ and deduce an inequality for $\eta(\lambda)$ in terms of $\eta(\lambda/2)$. Solve this recursion to prove that for $\lambda < 2/\sigma$,

$$\mathbb{E}\exp\{\lambda\,(X - \mathbb{E}\,X)\} \le \frac{2 + \lambda\sigma}{2 - \lambda\sigma}\,.$$

3. Prove the Poincaré case of Theorem 1.2.

Exercise 1.2 (Hoeffding's lemma and Pinsker's inequality)**.** This exercise establishes the equivalence of Pinsker's inequality with a statement about sub-Gaussian concentration.

1. **Hoeffding's lemma** states that for any mean-zero random variable X with values in $[a, b]$ a.s., it holds that X is $(b-a)^2/4$-sub-Gaussian. Prove this lemma as follows. For $\lambda \in \mathbb{R}$, let $\psi(\lambda) := \ln \mathbb{E} \exp(\lambda X)$. Differentiate ψ twice and show that $\psi''(\lambda)$ can be interpreted as the variance of a random variable under a change of measure and hence $\psi''(\lambda) \le (b-a)^2/4$.

2. **Pinsker's inequality** states that for any two probability measures μ and ν on the same space, $\|\mu - \nu\|_{\mathrm{TV}}^2 \le \frac{1}{2} \mathsf{KL}(\mu \,\|\, \nu)$. Prove this inequality as follows. First, by the data-processing inequality, for any event A,
$$\mathsf{KL}(\mu \,\|\, \nu) \ge \mathsf{KL}\big((\mathbb{1}_A)_\# \mu \bigm\| (\mathbb{1}_A)_\# \nu\big) = \mathsf{KL}\big(\mathsf{Bernoulli}(\mu(A)) \bigm\| \mathsf{Bernoulli}(\nu(A))\big)\,.$$
Next, for any $q \in (0,1)$, differentiate $p \mapsto k_q(p) := \mathsf{KL}(\mathsf{Bernoulli}(p) \,\|\, \mathsf{Bernoulli}(q))$ twice to show that k_q is 4-strongly convex, and deduce that $k_q(p) \ge 2\,|p-q|^2$. Finally, take the supremum over events A.

3. Apply the Bobkov–Götze theorem (Theorem 1.3) to show that Hoeffding's lemma and Pinsker's inequality are equivalent to each other.

Exercise 1.3 (bounded differences inequality)**.** This exercise establishes a broadly useful concentration inequality.

1. Prove the **Azuma–Hoeffding inequality**: let $(\mathscr{F}_i)_{i=0}^n$ be a filtration, let $(\Delta_i)_{i=1}^n$ be a martingale difference sequence (that is, Δ_i is $\mathscr{F}_i$-measurable and $\mathbb{E}[\Delta_i \mid \mathscr{F}_{i-1}] = 0$), and assume that for each i there exist $\mathscr{F}_{i-1}$-measurable random variables A_i and B_i such that $A_i \le \Delta_i \le B_i$ a.s. Then, $\sum_{i=1}^n \Delta_i$ is $\sum_{i=1}^n \|B_i - A_i\|_{L^\infty(\mathbb{P})}^2/4$-sub-Gaussian.

   Hint: Apply Hoeffding's lemma from Exercise 1.2 conditionally.

2. Use this to prove the **bounded differences inequality**: if $X_1, \dots, X_n$ are independent, then $f(X_1, \dots, X_n) - \mathbb{E} f(X_1, \dots, X_n)$ is $\sum_{i=1}^n \|D_i f\|_{\sup}^2/4$-sub-Gaussian.

   Hint: Recall the proof of the Efron–Stein inequality.

3. Next, apply Marton's tensorization (Theorem 1.1) to Pinsker's inequality from Exercise 1.2 (see Example 1.1) to obtain a transport inequality for the product space $\mathcal{X}^N$. Using the Bobkov–Götze equivalence (Theorem 1.3), give a second proof of the bounded differences inequality.

Exercise 1.4 (a loose end in Gozlan's theorem)**.** Prove the first statement of Lemma 1.3.

Exercise 1.5 (inequivalence between PI and T_1)**.** We show that the Poincaré inequality and the T_1 inequality are incomparable, i.e., one does not necessarily imply the other.

1. Use Theorem 1.4 to provide an example of a measure $\pi \in \mathcal{P}_1(\mathbb{R}^d)$ which satisfies a T_1 inequality but which does not satisfy a Poincaré inequality.

   Hint: Explain why a Poincaré inequality necessarily requires the support of the measure to be connected.

2. For the converse direction, let μ be the exponential distribution on $\mathbb{R}$, so that the density is $\mu(x) = \exp(-x)\,\mathbb{1}\{x > 0\}$. Let $f : \mathbb{R}_+ \to \mathbb{R}$; we may assume that $f(0) = 0$. Now apply the identity $f(x)^2 = 2\int_0^x f(s)\,f'(s)\,\mathrm{d}s$ to the integral $\int f^2\,\mathrm{d}\mu$ and prove that μ satisfies $\mathsf{PI}(4)$. Explain why μ cannot satisfy a T_1 inequality.

References

[BG99] S. G. Bobkov and F. Götze. "Exponential integrability and transportation cost related to logarithmic Sobolev inequalities". In: *J. Funct. Anal.* 163.1 (1999), pp. 1–28.

[BV05] F. Bolley and C. Villani. "Weighted Csiszár–Kullback–Pinsker inequalities and applications to transportation inequalities". In: *Ann. Fac. Sci. Toulouse Math. (6)* 14.3 (2005), pp. 331–352.

[Che24] S. Chewi. *Log-concave sampling*. Forthcoming, 2024.

[Mar96] K. Marton. "Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration". In: *Ann. Probab.* 24.2 (1996), pp. 857–866.

[Mil09] E. Milman. "On the role of convexity in isoperimetry, spectral gap and concentration". In: *Invent. Math.* 177.1 (2009), pp. 1–43.