Four Talagrand inequalities under the same umbrella

Michel Ledoux

University of Toulouse, France

Abstract

This note reviews the studies of the last decades emphasizing a common principle based on entropy, logarithmic Sobolev inequality and hypercontractivity, behind four most celebrated inequalities by M. Talagrand: the convex distance inequality, the L^1–L^2 variance inequality, the quadratic transportation cost inequality, and the inequality on the supremum of empirical processes.

1 Four Talagrand inequalities

This section outlines the statements of the four Talagrand inequalities considered in this note, in the original notation of the author. These inequalities were all established (published) between 1991 and 1996, the first three in relatively short articles.

Talagrand's convex distance inequality [81, Theorem 1.1 and (1.3)], [83, Theorem 4.1.1]. Let (Ω_i, μ_i), $i = 1, \ldots, n$, be arbitrary probability spaces and provide their product with the product probability P. Given a set $A \subset \Omega$ and $x \in \Omega$, set

$$U_A(x) \,=\, \big\{ (s_i)_{1 \leq i \leq n} \in \{0,1\}^n ; \exists\, y \in A ; s_i = 0 \Rightarrow x_i = y_i \big\}.$$

Denote by $V_A(x)$ the convex hull of $U_A(x)$ considered as a subset of $\mathbb{R}^n$. The set $V_A(x)$ contains 0 if and only if x belongs to A. Denote then by $d_A(x)$ the Euclidean distance of 0 to $V_A(x)$ (in [83], the notation $f_c(A,x)$ is used instead, the letter c refereeing to "convexity").

For any (measurable) $A \subset \Omega$,

$$\int_\Omega e^{\frac{1}{4} d_A^2} dP \,\leq\, \frac{1}{P(A)} \, . \tag{1}$$

Talagrand's L^1–L^2 variance inequality [82, Theorem 1.5]. Let $\{0,1\}^n$ be the discrete cube equipped with the product probability measure μ_{p} giving weight p to 1 and $1-\mathsf{p}$ to 0, $0<\mathsf{p}<1$. For $1 \leq r \leq \infty$, denote by $\|\cdot\|_r$ the norm in $\mathrm{L}^r(\mu_{\mathsf{p}})$. For every $i=1,\dots,n$ and every $x=(x_1,\dots,x_n)\in\{0,1\}^n$, let $U_i(x)$ be the point of $\{0,1\}^n$ obtained from x by replacing x_i by $1-x_i$ and leaving the other coordinates unchanged. If f is a function on $\{0,1\}^n$, set $\Delta_i f(x) = (1-\mathsf{p})(f(x)-f(U_i(x)))$ if $x_i=1$ and $\Delta_i f(x) = \mathsf{p}(f(x)-f(U_i(x)))$ if $x_i=0$.

There is a numerical constant $K>0$ such that for any function $f:\{0,1\}^n\to\mathbb{R}$ such that $\int_{\{0,1\}^n} f d\mu_{\mathsf{p}} = 0$,

$$\|f\|_2^2 \,\leq\, K \,\log\Big(\frac{2}{\mathsf{p}(1-\mathsf{p})}\Big) \sum_{i=1}^{n} \frac{\|\Delta_i f\|_2^2}{\log\big(e\,\frac{\|\Delta_i f\|_2}{\|\Delta_i f\|_1}\big)}\,. \tag{2}$$

Talagrand's quadratic transportation cost inequality [84, Theorem 1.1]. For two probability measures ν and μ on the Borel sets of $\mathbb{R}^n$, define the transportation cost $T_w(\mu,\nu)$ as the infimum of

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} w(x,y)d\pi(x,y)$$

over all the probability measures π on $\mathbb{R}^n\times\mathbb{R}^n$ such that μ is the first marginal of π and ν is the second, and where $w(x,y)=\sum_{i=1}^n (x_i-y_i)^2$. Let $d\gamma(x) = e^{-|x|^2/2}\frac{dx}{(2\pi)^{n/2}}$ be the standard Gaussian measure on the Borel sets of $\mathbb{R}^n$.

For every probability measure μ absolutely continuous with respect to γ with $f=\frac{d\mu}{d\gamma}$,

$$T_w(\mu,\gamma) \,\leq\, 2\int_{\mathbb{R}^n} f\log f\, d\gamma \,=\, 2\int_{\mathbb{R}^n} \log f\, d\mu. \tag{3}$$

Talagrand's inequality on the supremum of empirical processes [85, Theorem 1.4]. Let $X_1,\dots,X_n$ be independent random variables with values in a measurable space Ω, and let $\mathcal{F}$ be a countable family of measurable functions on Ω. Consider the random variable

$$Z \,=\, \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} f(X_i).$$

Set $U=\sup_{f\in\mathcal{F}}\|f\|_\infty$ and $V=\mathbb{E}\big(\sup_{f\in\mathcal{F}}\sum_{i=1}^n f(X_i)^2\big)$.

For each $t>0$,

$$\mathbb{P}\big(\big|Z-\mathbb{E}(Z)\big|\geq t\big) \,\leq\, K\exp\bigg(-\frac{t}{KU}\log\Big(1+\frac{tU}{V}\Big)\bigg) \tag{4}$$

where $K>0$ is a numerical constant.

2 Introduction

M. Talagrand's mathematical achievements have deeply influenced the scientific developments over the last decades. His work embraces numerous mathematical fields, including measure theory, functional analysis, Banach space geometry, stochastic processes, Boolean analysis, isoperimetric and concentration inequalities, optimal transportation, statistical physics etc.

It is in particular extremely impressive to hear about *four* of his inequalities, all called *Talagrand's inequality*, in rather different areas. The precise statements of these inequalities have been presented in the first section.

– The famous convex distance inequality for product measures, put forward in his celebrated Publication of the Institut des Hautes Études Scientifiques [83], has very fruitful applications in numerous areas of probability theory, statistics or optimization, and combinatorial and discrete mathematics.

– The L^1–L^2 variance inequality on the discrete cube first provided an alternate, sharper, approach to the Kahn-Kalai-Linial theorem on influences in Boolean analysis, and became a central tool in theoretical computer science. As put forward by I. Benjamini, G. Kalai and O. Schramm, it is besides, at this point, the only generic argument towards sub-diffusive and super-concentration phenomena which are ubiquitous to many models of the current research in probability theory (percolation, random matrices, spin glasses etc.).

– Talagrand's quadratic transportation cost has been one founding stone in the interaction between partial differential equations, probability theory and geometry, as first emphasized by F. Otto and C. Villani in connection with the logarithmic Sobolev inequality, leading to the new field of functional inequalities, curvature lower bounds and analysis on metric measure spaces by J. Lott, C. Villani, K. T. Sturm, L. Ambrosio, N. Gigli, G. Savaré.

– The methods developed for concentration inequalities for product measures led to the fundamental Talagrand inequality for the supremum of empirical processes, a major and essential tool in modern infinite-dimensional statistics.

The Talagrand inequalities have both shaped the mathematics of the last decades and stand today as essential and common tools of the current research.

Apart perhaps the L^1–L^2 variance inequality, the Talagrand inequalities were mostly motivated by aspects of the concentration of measure phenomenon and its applications to Banach space geometry and probability in Banach spaces [59, 68, 50, 38, 49, 76, 4]. "The idea of concentration of measure, which was discovered by V. Milman, is arguably one of the great ideas of analysis in our times" (M. Talagrand [86]). The concentration of measure phenomenon has indeed become today a major tool in various areas of mathematics such as asymptotic geometric analysis, probability theory, statistical mechanics, mathematical statistics and learning theory, random matrix theory or quantum information theory, stochastic dynamics, randomized algorithms, complexity... Of isoperimetric origin and flavour, it is suited to the investigation of models involving an infinite number of variables, and emphasizes that, quoting M. Talagrand again [86], "a random variable that depends (in a "smooth" way) on the influence of many independent random variables (but not too much on any of them) is essentially constant". It is indeed a main feature of the four Talagrand inequalities that they are dimension free (constants do not depend on the size of the samples, and the statements extend to infinite-dimensional systems). These inequalities are in particular inspired by, and provide deep and powerful extensions of, the model concentration inequality for Gaussian measures expressing that for any Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}$ with Lipschitz semi-norm $\|F\|_{\mathrm{Lip}}$,

$$\gamma\big(F \geq \int_{\mathbb{R}^n} F d\gamma + r\big) \,\leq\, e^{-r^2/2\|F\|_{\mathrm{Lip}}^2} \tag{5}$$

for all $r \geq 0$ (where $d\gamma(x) = e^{-|x|^2/2} \frac{dx}{(2\pi)^{n/2}}$, the standard Gaussian measure on the Borel sets of $\mathbb{R}^n$).

The Talagrand inequalities look quite different, and have been established by him with different tools and methods (although induction on the dimension towards dimension-free bounds may be detected as a common background). It is the purpose of this note to show, based on the mathematical developments of the last decades, that all these four inequalities may be seen as consequences of a common principle, namely entropy, logarithmic Sobolev inequality and hypercontractivity.

This observation started with the papers [24] and [1] which revived an unpublished letter by I. Herbst to L. Gross in 1975 deducing exponential integrability from the logarithmic Sobolev inequality, just after the fundamental discovery by L. Gross of the latter [39]. The contribution [1] by S. Aida, T. Masuda and I. Shigekawa has been very influential in this regard. The relevance of this observation towards the Gaussian concentration inequality (5) was pointed out next in [45]. Based on this principle, the contribution [47] provided an alternate approach to the inequality on supremum of empirical processes (4), later developed and extended in several steps and contributions summarized in the monograph [20] by S. Boucheron, G. Lugosi and P. Massart. A few years later, these authors also covered the convex distance inequality (1) by this method. (The monograph [20] actually covers in depth various aspects emphasized in this note.) Already present in [82] and earlier [44], the hypercontractivity character of the L^1–L^2 variance inequality (2) is clarified in [12], and led these authors to a proof of sub-linearity of percolation times. It is also mentioned in [12] that Talagrand's variance inequality (2) may be used to recover the associated concentration inequality for percolation time first obtained in [83], linking even more the four inequalities. In the early 2000, F. Otto and C. Villani [65] deduced the Talagrand quadratic transportation cost inequality from the logarithmic Sobolev inequality of L. Gross, bringing to light the deep link between these two families of inequalities.

The purpose of this note is thus to provide a common, self-contained, approach to the four Talagrand inequalities based on entropy, logarithmic Sobolev inequality and hypercontractivity, with for each of them a simple and direct proof. Sub-additivity properties of entropy and logarithmic Sobolev inequality will encapsulate the dimension-free concentration phenomenon, and the Talagrand inequalities.

The next paragraph, Section 3, presents the necessary (elementary) material on the entropy method to this task. It is restricted to the purpose of the subsequent proofs, but wide extensions and applications of the approach have been developed in the literature. The proofs of the Talagrand inequalities are addressed in the following four sections. It should be mentioned that, with respect to the original Talagrand's formulations, the notation and terminology will somewhat be adapted in order to follow the common trends in this regard. This is also motivated by the framework emphasized in Section 3. No attempt is made towards sharp numerical constants (besides the one in (3)), favouring the simplicity of the arguments before technical improvements (with appropriate references for the latter). Each section will be followed by some comments on historical aspects and further developments, with a few (and only a few) general pointers to the relevant literature.

This note should have been written several years ago. Hopefully it is still of some interest.

3 Entropy, logarithmic Sobolev inequality, hypercontractivity

This section describes the basic material towards the entropy approach to the four Talagrand inequalities. As already mentioned, the presentation of the various tools is limited to this specific task, but, as emphasized in the literature, the power and generality of the principle may be extended and applied far outside this given purpose. This material is rather elementary, and most of the statements may be found in standard monographs or lecture notes (see the notes and references at the end of the paragraph) involved with these objects.

(Ω, Σ, μ) will denote a generic probability space. For each $1 \leq p \leq \infty$, $\|\cdot\|_p$ is the norm of the Lebesgue space $\mathrm{L}^p(\mu)$. The various integrability conditions appearing below will be automatically satisfied in all the subsequent illustrations dealing mostly with bounded functions.

Entropy. For a measurable function $f : (\Omega, \Sigma, \mu) \to \mathbb{R}$, non-negative in $\mathrm{L}^1(\mu)$, set

$$\mathrm{Ent}_\mu(f) \,=\, \int_\Omega f \log f \, d\mu - \int_\Omega f \, d\mu \, \log\bigg(\int_\Omega f \, d\mu \bigg) \,\in\, [0, +\infty]$$

(since $u \in [0, \infty) \mapsto u \log u$ is convex bounded from below, with the convention $0 \log 0 = 0$). Observe that $\mathrm{Ent}_\mu(f)$ is homogeneous of order 1. If $d\nu = f d\mu$ for a density f, $\mathrm{Ent}_\mu(f) = \mathrm{H}(\nu \,|\, \mu)$, the relative entropy of ν with respect to μ.

Entropy has a lot of common with variance,

$$\mathrm{Var}_\mu(f) \,=\, \int_\Omega f^2 d\mu - \bigg(\int_\Omega f \, d\mu \bigg)^2$$

(for a function f in $\mathrm{L}^2(\mu)$), in particular the following duality and variational representations.

The duality formula for entropy expresses that

$$\mathrm{Ent}_\mu(f) \,=\, \sup \bigg\{ \int_\Omega f g \, d\mu \, ; \int_\Omega e^g d\mu \leq 1 \bigg\} \tag{6}$$

(where g may be assumed to be bounded from above and below). Indeed, assume by homogeneity that $\int_\Omega f d\mu = 1$. By Young's inequality $uv \leq u \log u - u + e^v$, $u \geq 0$, $v \in \mathbb{R}$, so that, for $\int_\Omega e^g d\mu \leq 1$,

$$\int_\Omega f g \, d\mu \,\leq\, \int_\Omega f \log f \, d\mu - 1 + \int_\Omega e^g d\mu \leq \int_\Omega f \log f \, d\mu \,=\, \mathrm{Ent}_\mu(f).$$

For the converse, set $f_N = \min(\max(f, \frac{1}{N}), N)$, $N \geq 1$, and choose $g = \log\big(\frac{f_N}{\int_\Omega f_N d\mu}\big)$. The claim follows as $N \to \infty$.

For the further purposes, note that this duality formula justifies the well-known entropic inequality, for any $f \geq 0$ with $\int_\Omega f d\mu = 1$ and g measurable such that fg is integrable,

$$\int_\Omega fg\, d\mu \,\leq\, \mathrm{Ent}_\mu(f) + \log\bigg(\int_\Omega e^g d\mu\bigg) \,=\, \int_\Omega f \log f\, d\mu + \log\bigg(\int_\Omega e^g d\mu\bigg). \tag{7}$$

The variational formula on the other hand states that

$$\mathrm{Ent}_\mu(f) \,=\, \inf_{c>0} \int_\Omega \big[f(\log f - \log c) - (f-c)\big] d\mu. \tag{8}$$

Indeed, the infimum of $c \mapsto c - (\log c + 1) \int_\Omega f d\mu$ is attained at $c = \int_\Omega f d\mu$ giving thus rise to $\mathrm{Ent}_\mu(f)$.

Tensorization of entropy. A fundamental feature of entropy (and of variance) is its product or sub-additivity property, main source of the approach to the Talagrand inequalities developed here. Let $(\Omega_i, \Sigma_i, \mu_i)$, $i = 1, \ldots, n$, be probability spaces, and denote by P the product probability measure $\mu_1 \otimes \cdots \otimes \mu_n$ on the product space $X = \Omega_1 \times \cdots \times \Omega_n$ equipped with the product σ-field. A point x in X is denoted $x = (x_1, \ldots, x_n)$, $x_i \in \Omega_i$, $i = 1, \ldots, n$. Given f on the product space, write furthermore f_i, $i = 1, \ldots, n$, for the function on Ω_i defined by

$$f_i(x_i) = f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n),$$

with $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ fixed.

Proposition 1. *For every non-negative function f on the product space X in $\mathrm{L}^1(P)$,*

$$\mathrm{Ent}_P(f) \,\leq\, \sum_{i=1}^n \int_X \mathrm{Ent}_{\mu_i}(f_i) dP.$$

It may be pointed out that on the right-hand side, integration in dP is actually, for each $i = 1, \ldots, n$, over the remaining coordinates $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

Proof. Use the duality formula (6). Given g such that $\int_X e^g dP \leq 1$, set, for every $i = 1 \ldots, n$,

$$g^i(x_i, \ldots, x_n) \,=\, \log\bigg(\frac{\int_X e^{g(x_1,\ldots,x_n)} d\mu_1(x_1) \cdots d\mu_{i-1}(x_{i-1})}{\int_X e^{g(x_1,\ldots,x_n)} d\mu_1(x_1) \cdots d\mu_i(x_i)}\bigg)$$

(well-defined for $\mu_i \otimes \cdots \otimes \mu_n$-almost every $(x_i, \ldots, x_n)$). Then $g \leq \sum_{i=1}^n g^i$ and $\int_{\Omega_i} e^{(g^i)_i} d\mu_i = 1$ for every $i = 1 \ldots, n$. Therefore,

$$\begin{aligned} \int_X fg\, dP &\leq \sum_{i=1}^n \int_X fg^i\, dP \\ &= \sum_{i=1}^n \int_X \bigg(\int_{\Omega_i} f_i (g^i)_i d\mu_i\bigg) dP \\ &\leq \sum_{i=1}^n \int_X \mathrm{Ent}_{\mu_i}(f_i) dP \end{aligned}$$

which is the result. $\square$

Proposition 1 is presented in [47] (in a more general form due to S. Bobkov, the proof presented here being due to S. Kwapien), and deduced from Han's inequality in [20]. It has a classical analogue for the variance, established in the same (even simpler) way, known as the Efron-Stein inequality (cf. [31, 78, 69, 79, 20]) expressing that for every function f on the product space X in $\mathrm{L}^2(P)$,

$$\mathrm{Var}_P(f) \leq \sum_{i=1}^n \int_X \mathrm{Var}_{\mu_i}(f_i) dP. \tag{9}$$

Actually (9) may be deduced from Proposition 1 applied to $f = 1 + \varepsilon g$ with $\varepsilon \to 0$.

The tensorization Proposition 1 admits related formulations which will be of fundamental use in the applications to the Talagrand convex distance inequality and the inequality on the supremum of empirical processes.

First, by Jensen's inequality, for every $i = 1, \ldots, n$ (and $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ fixed),

$$\mathrm{Ent}_{\mu_i}(f_i) \leq \frac{1}{2} \int_{\Omega_i} \int_{\Omega_i} \big[f_i(x_i) - f_i(y_i) \big] \big[\log f_i(x_i) - \log f_i(y_i) \big] d\mu_i(x_i) d\mu_i(y_i)$$

(an analogue of the duplication formula for the variance). Therefore, at a first (mild) level, Proposition 1 yields by symmetry

$$\begin{aligned} &\mathrm{Ent}_P(f) \\ &\leq \sum_{i=1}^n \int_X \int \int_{\{f_i(x_i) \geq f_i(y_i)\}} \big[f_i(x_i) - f_i(y_i) \big] \big[\log f_i(x_i) - \log f_i(y_i) \big] d\mu_i(x_i) d\mu_i(y_i) dP(x) \end{aligned} \tag{10}$$

Proposition 1 may also be combined with the variational representation (8) in the form

$$\mathrm{Ent}_P(f) \leq \sum_{i=1}^n \int_X \left(\inf_{c_i > 0} \int_{\Omega_i} \big[f_i(\log f_i - \log c_i) - (f_i - c_i) \big] \, d\mu_i \right) dP. \tag{11}$$

Of course, for each $i = 1, \ldots, n$, $c_i > 0$ in the infimum may be chosen to depend on the variables $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$.

Logarithmic Sobolev inequality on the two-point space. A logarithmic Sobolev inequality bounds the entropy of a function f by an energy, or Dirichlet form. It is the analogue of the classical Poincaré inequality for the variance. It has been discovered by L. Gross [39] in 1975 together with its equivalence with hypercontractivity (although earlier versions may detected in the literature).

The most basic logarithmic Sobolev inequality takes place on the two-point space $\{-1, +1\}$ with the Bernoulli probability measure $\mu_{\mathsf{p}}(\{+1\}) = \mathsf{p}$, $\mu_{\mathsf{p}}(\{-1\}) = 1 - \mathsf{p} = \mathsf{q}$, $0 < \mathsf{p} < 1$, and states that for any $f : \{-1, +1\} \to \mathbb{R}$,

$$\mathrm{Ent}_{\mu_{\mathsf{p}}}(f^2) \leq \frac{1}{\rho} \mathsf{pq} \big[f(+1) - f(-1) \big]^2 \tag{12}$$

where $\rho = \frac{\mathsf{p} - \mathsf{q}}{\log \mathsf{p} - \log \mathsf{q}}$. The symmetric case $\mathsf{p} = \frac{1}{2}$ is the simplest and most classical one, but as will be seen, general values of p are handled below similarly.

When $\mathsf{p} = \frac{1}{2}$, $\rho = \frac{1}{2}$ so that the constant on the right-hand side of (12) is $\frac{1}{2}$. A proof in this case runs as follows. Setting $f(+1) = \alpha$ and $f(-1) = \beta$, the inequality amounts to

$$\frac{\Phi(\alpha^2) + \Phi(\beta^2)}{2} - \Phi\Big(\frac{\alpha^2 + \beta^2}{2}\Big) \;\leq\; \frac{1}{2}\,(\alpha - \beta)^2$$

where $\Phi(u) = u \log u$, $u \geq 0$. That is, if $r = \frac{1}{2}(\alpha^2 + \beta^2)$ and $s = \frac{1}{2}(\alpha^2 - \beta^2)$

$$\Phi(r + s) + \Phi(r - s) - 2\Phi(r) \;\leq\; (\alpha - \beta)^2.$$

But the left-hand side is

$$\int_0^s \big[\Phi''(r + v) + \Phi''(r - v)\big](b - v)dv$$

and since the function $\Phi'' = \frac{1}{u}$ is convex on the given domain

$$\Phi''(r + v) + \Phi''(r - v) \;\leq\; 2\Phi''(r) \;=\; \frac{2}{r}\,.$$

It remains to observe that $\frac{s^2}{r} \leq (\alpha - \beta)^2$. A direct proof for general p may be found in [3].

Hypercontractivity on the two-point space. The logarithmic Sobolev inequality (12) may be translated equivalently into the famous hypercontractivity property. At this stage, only the symmetric case $\mathsf{p} = \frac{1}{2}$ is described for simplicity. Any function $f : \{-1, +1\} \to \mathbb{R}$ may be represented as $f(x) = a + bx$, $x \in \{-1, +1\}$, $a, b \in \mathbb{R}$. For any $t \geq 0$, define the new function $P_t f(x) = a + e^{-t}bx$, $x \in \{-1, +1\}$. The family $(P_t)_{t \geq 0}$ defines a semigroup of contractions on $\mathrm{L}^p(\mu_{\frac{1}{2}})$ for any $1 \leq p \leq \infty$. It turns out that (12) allows for the strengthening

$$\|P_t f\|_q \;\leq\; \|f\|_p \tag{13}$$

whenever $1 < p < q < \infty$, $e^{2t} \geq \frac{q-1}{p-1}$. The latter (13) may be translated again as a two-point inequality

$$\Big(\frac{1}{2}\,|a + e^{-t}b|^q + \frac{1}{2}\,|a - e^{-t}b|^q\Big)^{1/q} \;\leq\; \Big(\frac{1}{2}\,|a + b|^p + \frac{1}{2}\,|a - b|^p\Big)^{1/p}.$$

This inequality was established directly by A. Bonami [18] and W. Beckner [9] but L. Gross [39] observed that it is actually equivalent to (12) (and as such easier to establish). This connection is developed next, as well as the case $\mathsf{p} \neq \frac{1}{2}$, after the setting is extended to the product model $X = \{-1, +1\}^n$, $n \geq 1$.

Logarithmic Sobolev inequality and hypercontractivity on the discrete cube. On the two-point space $\{-1, +1\}$ equipped with the Bernoulli measure μ_{p}, $0 < \mathsf{p} < 1$, consider the (Markov) operator $\mathrm{L}f = \int_{\{-1,+1\}} f d\mu_{\mathsf{p}} - f$. Note that

$$\int_{\{-1,+1\}} f(-\mathrm{L}f)d\mu_{\mathsf{p}} \;=\; \int_{\{-1,+1\}} (\mathrm{L}f)^2 d\mu_{\mathsf{p}} \;=\; \mathrm{Var}_{\mu_{\mathsf{p}}}(f) \;=\; \mathsf{pq}\big[f(+1) - f(-1)\big]^2. \tag{14}$$

On the product space $X = \{-1, +1\}^n$ with the product measure μ_{p}^n, consider the product operator $\mathrm{L} = \sum_{i=1}^n \mathrm{L}_i$ where L_i is acting on the i-th coordinate of a function $f : X \to \mathbb{R}$ as

$\mathrm{L}_i f = \int_{\{-1,+1\}} f_i(x_i) d\mu_{\mathsf{p}}(x_i) - f$, $i = 1, \ldots, n$ (with the notation of the tensorization paragraph, that is $f_i(x_i) = f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$ with $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ fixed).

The Dirichlet form

$$\mathcal{E}(f,g) \,=\, \int_X f(-\mathrm{L}g) d\mu_{\mathsf{p}}^n$$

for two functions $f, g : X \to \mathbb{R}$ admits various representations of the form

$$\begin{aligned}
&\mathcal{E}(f,g) \\
&= \sum_{i=1}^n \int_X \mathrm{L}_i f \, \mathrm{L}_i g \, d\mu_{\mathsf{p}}^n \\
&= \sum_{i=1}^n \int_X \mathrm{Cor}_{\mu_{\mathsf{p}}}(f_i, g_i) d\mu_{\mathsf{p}}^n \\
&= \frac{1}{2} \sum_{i=1}^n \int_X \int_{\{-1,+1\}} \int_{\{-1,+1\}} \big[f_i(x_i) - f_i(y_i) \big] \big[g_i(x_i) - g_i(y_i) \big] d\mu_{\mathsf{p}}(x_i) d\mu_{\mathsf{p}}(y_i) d\mu_{\mathsf{p}}^n(x).
\end{aligned} \tag{15}$$

In particular, from the Efron-Stein inequality (9), the Poincaré inequality

$$\mathrm{Var}_{\mu_{\mathsf{p}}^n}(f) \,\leq\, \mathcal{E}(f,f) \tag{16}$$

holds true for any $f : X \to \mathbb{R}$.

The operator L generates the Markov semigroup $(P_t)_{t \geq 0}$ defined by

$$P_t \,=\, \sum_{k=0}^\infty \frac{t^k}{k!} \, \mathrm{L}^k \tag{17}$$

($\mathrm{L}^0 = \mathrm{Id}$) in the sense that $\frac{d}{dt} P_t f = \mathrm{L} P_t f = P_t \, \mathrm{L} f$. It is symmetric and invariant with respect to μ_{p}^n, that is for functions $f, g : X \to \mathbb{R}$, $\int_X f P_t g \, d\mu_{\mathsf{p}}^n = \int_X g P_t f \, d\mu_{\mathsf{p}}^n$. It is immediately checked that for $\mu_{\frac{1}{2}}$ on $\{-1,+1\}$, $P_t f(x) = a + e^{-t} bx$ for a function $f = a + bx$, $x \in \{-1,+1\}$.

It is a consequence of the tensorization property, and the representations (14) and (15) of the associated Dirichlet forms, that the logarithmic Sobolev inequality (12) extends to functions f on the product space $X = \{-1,+1\}^n$ equipped with the product measure μ_{p}^n in the form

$$\mathrm{Ent}_{\mu_{\mathsf{p}}^n}(f^2) \,\leq\, \frac{1}{\rho} \, \mathcal{E}(f,f) \tag{18}$$

for any $f : X \to \mathbb{R}$, where it is recalled that $\rho = \frac{\mathsf{p}-\mathsf{q}}{\log \mathsf{p} - \log \mathsf{q}}$ ($= \frac{1}{2}$ if $\mathsf{p} = \frac{1}{2}$). In the same way, the hypercontractivity inequality (13) extends to functions f on $X = \{-1,+1\}^n$, expressing that whenever $1 < p < q < \infty$ and $e^{4\rho t} \geq \frac{q-1}{p-1}$,

$$\|P_t f\|_q \,\leq\, \|f\|_p \,. \tag{19}$$

While the tensorization of hypercontractivity may be achieved independently (cf. e.g. [9, 6]), it is fruitful to deduce it from the logarithmic Sobolev inequality as developed by L. Gross [39].

Given $1 < p < q < \infty$, the key idea of L. Gross is to differentiate in time the quantity $\|P_t f\|_{q(t)}$, $q(t) = (p-1)e^{4\rho t} - 1$, $t \geq 0$. Since $|P_t f| \leq P_t(|f|)$, it is enough to deal with a non-negative function f (actually not identically zero). It holds that

$$\frac{d}{dt} \int_X (P_t f)^{q(t)} d\mu_{\mathsf{p}}^n \;=\; q'(t) \int_X (P_t f)^{q(t)} \log P_t f \, d\mu_{\mathsf{p}}^n + q(t) \int_X (P_t f)^{q(t)-1} \, \mathrm{L} P_t f \, d\mu_{\mathsf{p}}^n$$

and hence

$$\begin{aligned}
&\|P_t f\|_{q(t)}^{q(t)-1} \frac{d}{dt} \|P_t f\|_{q(t)} \\
&\qquad = -\frac{q'(t)}{q(t)^2} \int_X (P_t f)^{q(t)} d\mu_{\mathsf{p}}^n \log \int_X (P_t f)^{q(t)} d\mu_{\mathsf{p}}^n + \frac{1}{q(t)} \frac{d}{dt} \int_X (P_t f)^{q(t)} d\mu_{\mathsf{p}}^n \\
&\qquad = \frac{q'}{q^2} \, \mathrm{Ent}_{\mu_{\mathsf{p}}^n}\big((P_t f)^q\big) + \int_X (P_t f)^{q-1} \, \mathrm{L} P_t f \, d\mu_{\mathsf{p}}^n
\end{aligned}$$

where the short-hand notation $q = q(t)$, $q' = q'(t)$, is used in the last line. Assume now that

$$- \int_X (P_t f)^{q-1} \, \mathrm{L} P_t f \, d\mu_{\mathsf{p}}^n \;=\; \mathcal{E}\big((P_t f)^{q-1}, P_t f\big) \;\geq\; \frac{4(q-1)}{q^2} \, \mathcal{E}\big((P_t f)^{q/2}, (P_t f)^{q/2}\big) \tag{20}$$

so that

$$q^2 \, \|P_t f\|_q^q \, \frac{d}{dt} \|P_t f\|_q \;\leq\; q' \, \mathrm{Ent}_{\mu_{\mathsf{p}}^n}\big((P_t f)^q\big) - 4(q-1) \, \mathcal{E}\big((P_t f)^{q/2}, (P_t f)^{q/2}\big).$$

Applying the logarithmic Sobolev inequality (18) to $(P_t f)^{q(t)/2}$ then indicates that for the choice of $q = q(t) = (p-1)e^{4\rho t} - 1$, the right-hand side of the latter inequality is negative. Therefore the map $t \mapsto \|P_t f\|_{q(t)}$ is decreasing, which amounts to the hypercontractivity inequality (19). The proof shows in the same way that hypercontractivity is actually equivalent to the logarithmic Sobolev inequality.

It remains nevertheless to establish (20) which follows from a convexity argument. Namely, for all $u > v \geq 0$,

$$\begin{aligned}
\left(\frac{u^{q/2} - v^{q/2}}{u - v}\right)^2 &= \left(\frac{q}{2(u-v)} \int_v^u s^{\frac{q}{2}-1} ds\right)^2 \\
&\leq \frac{q^2}{4(u-v)} \int_v^u s^{q-2} ds \\
&= \frac{q^2}{4(q-1)} \, \frac{u^{q-1} - v^{q-1}}{u - v} \, .
\end{aligned}$$

Hence, for any $u, v \in \mathbb{R}$,

$$\big(u^{q-1} - v^{q-1}\big)(u - v) \;\geq\; \frac{4(q-1)}{q^2} \, \big(u^{q/2} - v^{q/2}\big)^2.$$

Recalling (15) then concludes to (20).

Logarithmic Sobolev inequality for the Gaussian measure. If $d\gamma(x) = e^{-|x|^2/2}\frac{dx}{(2\pi)^n}$ is the standard Gaussian (product) measure on the Borel sets of $\mathbb{R}^n$ ($|x|$ being the Euclidean length of $x \in \mathbb{R}^n$), the logarithmic Sobolev inequality for γ states that

$$\mathrm{Ent}_\gamma(f^2) \,\leq\, 2\int_{\mathbb{R}^n} |\nabla f|^2 d\gamma \tag{21}$$

for every smooth (for example locally Lipschitz in $\mathrm{L}^2(\gamma)$) function $f : \mathbb{R}^n \to \mathbb{R}$. It might be worthwhile mentioning that by the tensorization Proposition 1, it is enough to know the inequality in dimension one.

A proof of this inequality, also due to L. Gross [39], may be obtained from the logarithmic Sobolev inequality on the discrete cube via the central limit theorem. Roughly speaking, if $\varphi : \mathbb{R} \to \mathbb{R}$ is smooth with compact support, apply (18) with $\mathsf{p} = \frac{1}{2}$ to $f(x_1, \ldots, x_n) = \varphi\big(\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i\big)$, $(x_1, \ldots, x_n) \in \{-1, +1\}^n$. If $\|\varphi'\|_\infty + \|\varphi''\|_\infty < \infty$, by a Taylor expansion

$$f_i(x_i) - f_i(y_i) \,=\, \frac{x_i - y_i}{\sqrt{n}}\,\varphi'\bigg(\frac{1}{\sqrt{n}}\sum_{j\neq i} x_j\bigg) + O\Big(\frac{1}{n}\Big)$$

uniformly in x_i, y_i, $i = 1, \ldots, n$. Hence,

$$\sum_{i=1}^n \big[f_i(x_i) - f_i(y_i)\big]^2 \,=\, \frac{1}{n}\sum_{i=1}^n (x_i - y_i)^2\,\varphi'\bigg(\frac{1}{\sqrt{n}}\sum_{j\neq i} x_j\bigg)^2 + O\Big(\frac{1}{\sqrt{n}}\Big)$$

and from (15),

$$\mathcal{E}(f,f) \,\leq\, 2\int_X \varphi'\bigg(\frac{1}{\sqrt{n}}\sum_{j=2}^n x_j\bigg)^2 d\mu_{\frac{1}{2}}^n + O\Big(\frac{1}{\sqrt{n}}\Big).$$

By the central limit theorem, $\limsup_{n\to\infty} \mathcal{E}(f,f) \leq 2\int_{\mathbb{R}} {\varphi'}^2 d\gamma$ and similarly $\lim_{n\to\infty} \mathrm{Ent}_{\mu_{\frac{1}{2}}^n}(f^2) = \mathrm{Ent}_\gamma(\varphi^2)$.

In addition to this proof, there are at least 15 different further proofs of the logarithmic Sobolev inequality for the Gaussian measure. The following introduces the analytic (semigroup) proof put forward by D. Bakry and M. Émery [7, 6] (which has been recognized as the simplest one by L. Gross in 2010). It is presented with the Ornstein-Uhlenbeck semigroup (with invariant measure γ); a similar argument may be developed with the standard heat (Brownian) kernel/semigroup (cf. [8]).

Denote by $(P_t)_{t\geq 0}$ the so-called Ornstein-Uhlenbeck semigroup defined by

$$P_t f(x) \,=\, \int_{\mathbb{R}^n} f\big(e^{-t}x + \sqrt{1 - e^{-2t}}\,y\big)d\gamma(y), \quad t \geq 0,\ x \in \mathbb{R}^n \tag{22}$$

(for any $f : \mathbb{R}^n \to \mathbb{R}$ in $\mathrm{L}^1(\gamma)$ for example). The Ornstein-Uhlenbeck semigroup is invariant and symmetric with respect to the standard Gaussian measure γ, with infinitesimal generator $\mathrm{L}f(x) = \Delta f(x) - \langle x, \nabla f(x)\rangle$ for which the integration by parts formula

$$\int_{\mathbb{R}^n} f(-\mathrm{L}g)d\gamma \,=\, \int_{\mathbb{R}^n} \langle \nabla f, \nabla g\rangle\, d\gamma$$

holds true for every smooth functions $f, g : \mathbb{R}^n \to \mathbb{R}$.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a (measurable) function such that $\varepsilon \leq f \leq \frac{1}{\varepsilon}$ for some $\varepsilon > 0$ and $\int_{\mathbb{R}^n} f d\gamma = 1$. For every $t > 0$, $P_t f$ is then a C^∞ function, $\varepsilon \leq P_t f \leq \frac{1}{\varepsilon}$, and, as $t \to \infty$, $P_t f \to \int_{\mathbb{R}^n} f d\gamma = 1$ at every point. Therefore

$$\int_{\mathbb{R}^n} f \log f \, d\gamma \, = \, - \int_0^\infty \bigg(\frac{d}{dt} \int_{\mathbb{R}^n} P_t f \log P_t f \, d\gamma \bigg) dt.$$

By the chain rule

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^n} P_t f \log P_t f \, d\gamma \, &= \int_{\mathbb{R}^n} \mathrm{L} P_t f \log P_t f \, d\gamma + \int_{\mathbb{R}^n} \mathrm{L} P_t f \, d\gamma \\ &= - \int_{\mathbb{R}^n} \frac{|\nabla P_t f|^2}{P_t f} \, d\gamma \end{aligned}$$

where integration by parts is used as well as the fact that $\int_{\mathbb{R}^n} \mathrm{L} g d\gamma = 0$. Next, at any point, $|\nabla P_t f| \leq e^{-t} P_t(|\nabla f|)$ as is clear from the integral representation (22) of P_t, and by the Cauchy-Schwarz inequality along the same representation,

$$\frac{|\nabla P_t f|^2}{P_t f} \, \leq \, e^{-2t} \, \frac{P_t(|\nabla f|)^2}{P_t f} \, \leq \, e^{-2t} P_t \bigg(\frac{|\nabla f|^2}{f} \bigg).$$

Hence, by invariance of P_t with respect to γ,

$$- \frac{d}{dt} \int_{\mathbb{R}^n} P_t f \log P_t f \, d\gamma \, \leq \, e^{-2t} \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} \, d\gamma$$

from which it follows by integration that

$$\int_{\mathbb{R}^n} f \log f \, d\gamma \, \leq \, \frac{1}{2} \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} \, d\gamma. \tag{23}$$

This is a classical alternate formulation of the logarithmic Sobolev inequality with, on the left-hand side the relative entropy of $f d\gamma$ with respect to γ, and on the right-hand side the so-called Fisher information of f.

Let now $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable with gradient in $\mathrm{L}^2(\gamma)$. Apply (23) to, for example,

$$\frac{(P_t f_N)^2 + \varepsilon}{\int_{\mathbb{R}^n} (P_t f_N)^2 d\gamma + \varepsilon}$$

where $\varepsilon, t > 0$ and $f_N = \min(\max(f, -N), N))$. Letting successively $\varepsilon \to 0$ and $t \to 0$ yields that

$$\mathrm{Ent}_\gamma(f_N^2) \, \leq \, 2 \int_{\mathbb{R}^n} |\nabla f|^2 d\gamma$$

uniformly in $N \geq 1$. By means of the entropic inequality (7), for every $\theta \in (0,1)$,

$$(1 - \theta) \int_{\mathbb{R}^n} f_N^2 d\gamma \, \leq \, \int_{\mathbb{R}^n} f_N^2 \mathbb{1}_B \, d\gamma \, \leq \, 2 \int_{\mathbb{R}^n} |\nabla f|^2 d\gamma + \int_{\mathbb{R}^n} f_N^2 d\gamma \, \log\big(1 + (e-1)\gamma(B)\big)$$

where $B = \{f_N^2 \geq \theta \int_{\mathbb{R}^n} f_N^2 d\gamma\}$, from which it easily follows that $\sup_N \int_{\mathbb{R}^n} f_N^2 d\gamma < \infty$, and therefore by monotone convergence that $\int_{\mathbb{R}^n} f^2 d\gamma < \infty$. Then also $\int_{\mathbb{R}^n} f^2 \log f^2 d\gamma < \infty$ which, altogether, concludes to (21).

The constant 2 is optimal in (21) as can be seen from the choice of the functions $f(x) = e^{\langle b, x\rangle}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^n$, which achieve equality.

On the basis of the logarithmic Sobolev inequality (21) for the Gaussian measure, the Gross argument may be developed similarly along the Ornstein-Uhlenbeck semigroup to show that it is hypercontractive, with parameter $\rho = \frac{1}{2}$ in the notation of (19), a property going back to E. Nelson [61, 62] in quantum field theory. (It may be pointed out that in this diffusive setting, the convexity argument (20) is immediate by integration by parts and the chain rule.)

The Herbst argument. The main inspiration connecting a logarithmic Sobolev inequality to a concentration property is the Herbst argument. It was originally observed by I. Herbst towards exponential integrability properties, with, in the notation below, the application of the Gross logarithmic Sobolev inequality for the Gaussian measure to $e^{\lambda F^2}$, $\lambda \in \mathbb{R}$. It is developed here on $e^{\lambda F}$, $\lambda \in \mathbb{R}$, with in this form direct application to concentration inequalities. This observation is at the root of the entropic proof of the Talagrand convex distance inequality and on the supremum of empirical processes, as well actually as the transportation cost inequality. It is presented here on the Gaussian model (see the references for much more).

Let F be a smooth bounded Lipschitz function on $\mathbb{R}^n$ with Lipschitz semi-norm $\|F\|_{\mathrm{Lip}} \leq 1$. Since F is assumed to start with to be regular enough, it can be that $|\nabla F| \leq 1$ at every point. The aim is to apply the logarithmic Sobolev inequality (21) to $f^2 = e^{\lambda F}$ for every $\lambda \in \mathbb{R}$. With the notation $\Lambda(\lambda) = \int_{\mathbb{R}^n} e^{\lambda F} d\gamma$, $\lambda \in \mathbb{R}$, it holds that

$$\mathrm{Ent}_\gamma(f^2) \,=\, \lambda \int_{\mathbb{R}^n} F e^{\lambda F} d\gamma - \Lambda(\lambda)\log\Lambda(\lambda) \,=\, \lambda\Lambda'(\lambda) - \Lambda(\lambda)\log\Lambda(\lambda)$$

while

$$\int_{\mathbb{R}^n} |\nabla f|^2 d\gamma \,=\, \frac{\lambda^2}{4}\int_{\mathbb{R}^n} |\nabla F|^2 e^{\lambda F} d\gamma \,\leq\, \frac{\lambda^2}{4}\,\Lambda(\lambda).$$

Hence from (21),

$$\lambda\Lambda'(\lambda) - \Lambda(\lambda)\log\Lambda(\lambda) \,\leq\, \frac{\lambda^2}{2}\,\Lambda(\lambda).$$

If $H(\lambda) = \frac{1}{\lambda}\log\Lambda(\lambda)$ (with $H(0) = \frac{\Lambda'(0)}{\Lambda(0)} = \int_{\mathbb{R}^n} F d\gamma$), $\lambda \in \mathbb{R}$, then $H'(\lambda) \leq \frac{1}{2}$ for every λ. Therefore, $H(\lambda) - H(0) \leq \frac{\lambda}{2}$ if $\lambda \geq 0$ and $\geq \frac{\lambda}{2}$ if $\lambda \leq 0$ from which

$$\Lambda(\lambda) = \int_{\mathbb{R}^n} e^{\lambda F} d\gamma \,\leq\, e^{\lambda \int_{\mathbb{R}^n} F d\gamma + \frac{\lambda^2}{2}} \tag{24}$$

for every $\lambda \in \mathbb{R}$.

If F is an arbitrary Lipschitz function with $\|F\|_{\mathrm{Lip}} \leq 1$, apply the preceding for example to $P_t(F_N)$, $t > 0$, $N \geq 1$, where P_t is the Ornstein-Uhlenbeck semigroup and $F_N = \min(\max(F, -N), N))$, and let then $t \to 0$ and $N \to \infty$ in (24).

From (24), let $F : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz with Lipschitz semi-norm $\|F\|_{\rm Lip}$. For every $r \geq 0$ and $\lambda \geq 0$, by Markov's inequality,

$$\gamma\big(F \geq \textstyle\int_{\mathbb{R}^n} F d\gamma + r\big) \,\leq\, e^{-\lambda(\int_{\mathbb{R}^n} F d\gamma + r)} \textstyle\int_{\mathbb{R}^n} e^{\lambda F} d\gamma \,\leq\, e^{-\lambda r + \lambda^2/2 \|F\|^2_{\rm Lip}}$$

where (24) is used after homogeneity. Optimizing in λ, the Gaussian concentration inequality (5)

$$\gamma\big(F \geq \textstyle\int_{\mathbb{R}^n} F d\gamma + r\big) \,\leq\, e^{-r^2/2\|F\|^2_{\rm Lip}}, \quad r \geq 0,$$

is recovered in this way. Together with the same inequality for $-F$ and the union bound,

$$\gamma\big(|F - \textstyle\int_{\mathbb{R}^n} F d\gamma| \geq r\big) \,\leq\, 2\, e^{-r^2/2\|F\|^2_{\rm Lip}}, \quad r \geq 0.$$

These inequalities describe the fundamental concentration property of Gaussian measures (cf. e.g. [68, 50, 51, 46, 17, 49]).

Some notes and references. As mentioned at the beginning, this section puts forward the entropy method, and logarithmic Sobolev and hypercontractivity inequalities, only in the context required by the proofs of the Talagrand inequalities emphasized in this note. The framework and methodology may be vastly extended and generalized to various settings and applications. Introductions to logarithmic Sobolev inequalities are for example [28, 40, 71, 48, 3, 41, 8]... The general monographs and courses from the non-exhaustive list [29, 49, 90, 20, 63, 34]... contain further material and suitable pointers to the literature relevant to the topics of this section.

4 The convex distance inequality

In order to address the convex distance inequality via the entropy method displayed in the previous section, it is of interest to first suitably translate the (somewhat obscure) functional d_A of (1).

Recall the framework of a product probability measure $P = \mu_1 \otimes \cdots \otimes \mu_n$ on the product space $X = \Omega_1 \times \cdots \times \Omega_n$. A point x in X has coordinates $x = (x_1, \ldots, x_n)$. There might be some measurability questions in the forthcoming claims and proofs, but as emphasized in [83] these are unessential, and one should "treat all sets and functions as if they were measurable. This is certainly the case if one should assume that the Ω_i's are Polish and the μ_i's are Borel measures, and that one studies only compact sets, which is the only situation that occurs in applications". If necessary, it is therefore even possible to assume all sets finite.

The weighted Hamming distance on X, with weight $a = (a_1, \ldots, a_n) \in [0, \infty)^n$, is defined by

$$d_a(x, y) \,=\, \sum_{i=1}^{n} a_i \, \mathbb{1}_{\{x_i \neq y_i\}}, \quad x, y \in X.$$

For every non-empty (measurable) subset A of X and every $x \in X$, let then

$$F_A(x) \,=\, \sup_{|a|=1} d_a(x, A)$$

where $|a| = \big(\sum_{i=1}^n a_i^2\big)^{1/2}$.

The point is that the functional F_A is actually equal to d_A of (1) [83, Lemma 4.1.2]. Recall that, for $x \in X$,

$$U_A(x) \,=\, \big\{s = (s_i)_{1\leq i\leq n} \in \{0,1\}^n; \exists\, y \in A; s_i = 0 \Rightarrow x_i = y_i\big\}$$

and that $d_A(x)$ is the Euclidean distance of 0 to the convex hull $V_A(x)$ of $U_A(x)$ in $\mathbb{R}^n$. It is easily seen that, in this definition, $U_A(x)$ may be replaced by the collection of the indicator functions $\mathbb{1}_{\{x_i\neq y_i\}}$, $y \in A$. Now, if $d_A(x) < r$ for some $r > 0$, there exists z in $V_A(x)$ with $|z| < r$. Let $a \in [0,\infty)^n$ with $|a| = 1$. Then

$$\inf_{y\in V_A(x)} \langle a, y\rangle \leq \langle a, z\rangle \,\leq\, |z| \,<\, r.$$

Since

$$\inf_{y\in V_A(x)} \langle a, y\rangle \,=\, \inf_{s\in U_A(x)} \langle a, s\rangle \,=\, d_a(x, A), \tag{25}$$

it follows that $F_A(x) < r$. Hence $d_A(x) \geq F_A(x)$. Conversely, assume that $d_A(x) > 0$ (otherwise there will be nothing to prove) and let $\delta > 0$. Take $z \in V_A(x)$ such that $0 < |z|^2 \leq d_A(x)^2 + \delta$. By convexity, for every $\theta \in (0,1)$ and every $y \in V_A(x)$, $\theta y + (1-\theta)z \in V_A(x)$ so that

$$\big|z + \theta(y - z)\big|^2 \,=\, \big|\theta y + (1-\theta)z\big|^2 \,\geq\, d_A(x)^2 \,\geq\, |z|^2 - \delta.$$

Therefore

$$2\theta\langle y - z, z\rangle + \theta^2|y - z|^2 \,\geq\, -\delta.$$

Setting $a = \frac{z}{|z|}$,

$$\langle a, y\rangle \,\geq\, |z| - \frac{\theta|y-z|^2}{2|z|} - \frac{\delta}{2\theta|z|} \,\geq\, |z| - \frac{2\theta n}{|z|} - \frac{\delta}{2\theta|z|}\,.$$

Now, by (25),

$$F_A(x) \,\geq\, d_a(x, A) \,=\, \inf_{y\in V_A(x)} \langle a, y\rangle \,\geq\, |z| - \frac{2\theta n}{|z|} - \frac{\delta}{2\theta|z|} \,\geq\, d_A(x) - \frac{2\theta n}{d_A(x)} - \frac{\delta}{2\theta d_A(x)}\,.$$

Since $\delta > 0$ and $\theta \in (0,1)$ are arbitrary, it follows that $F_A(x) \geq d_A(x)$.

On the basis of this description, the principle will be to apply the tensorization Proposition 1 to the functional F_A together with the Herbst argument. For each $d_a(\cdot, A)$, or any Lipschitz function with respect to the Hamming metric $d_a(x,y)$, a simple Laplace transform argument together with iteration along the coordinates yields Gaussian concentration bounds (cf. [59, 83, 52, 49]). The challenge is to achieve a similar goal uniformly over all a's in the unit sphere of $\mathbb{R}^n$.

To ease the notation, set $F = F_A$ throughout the following steps (assuming that $P(F > 0) > 0$ otherwise there is nothing to prove). Let $\varepsilon > 0$. For each $x = (x_1, \ldots, x_n) \in X$, there exists $a(x) = a = (a_1, \ldots, a_n) \in [0,\infty)^n$ with $|a| = 1$ such that

$$F(x) \,\leq\, d_a(x, A) + \varepsilon.$$

For $1 \leq i \leq n$ and $y_i \in \Omega_i$, set $y = (x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, n)$. Then, with the notation of Proposition 1,

$$F_i(x_i) - F_i(y_i) \,=\, F(x) - F(y) \,\leq\, d_a(x, A) - d_a(y, A) + \varepsilon.$$

By the triangle inequality,

$$F_i(x_i) - F_i(y_i) \,\leq\, d_a(x, y) + \varepsilon \,=\, a_i \, \mathbb{1}_{\{x_i \neq y_i\}} + \varepsilon \,\leq\, a_i + \varepsilon.$$

Apply now Proposition 1 to $f = e^{\lambda F^2}$, $\lambda \geq 0$, in the form of (10). Whenever $F_i(x_i) \geq F_i(y_i)$ (≥ 0) for $i = 1, \ldots, n$, by the mean value inequality and the preceding,

$$\begin{aligned}\big[\lambda F_i(x_i)^2 - \lambda F_i(y_i)^2\big]\big[e^{\lambda F_i(x_i)^2} - e^{\lambda F_i(y_i)^2}\big] \,&\leq\, \lambda^2 \big[F_i(x_i)^2 - F_i(y_i)^2\big]^2 e^{\lambda F_i(x_i)^2} \\ &\leq\, 4\lambda^2 \big(a_i(x) + \varepsilon\big)^2 F_i(x_i)^2 e^{\lambda F_i(x_i)^2}.\end{aligned}$$

As $\sum_{i=1}^n a_i(x)^2 = 1$, it therefore follows from (10) that

$$\mathrm{Ent}_P(e^{\lambda F^2}) \,\leq\, 4\lambda^2 \int_X \big(1 + 2\varepsilon\sqrt{n} + \varepsilon^2 n\big) F^2 e^{\lambda F^2} dP$$

for any $\lambda \geq 0$. As $\varepsilon \to 0$,

$$\mathrm{Ent}_P(e^{\lambda F^2}) \,\leq\, 4\lambda^2 \int_X F^2 e^{\lambda F^2} dP. \tag{26}$$

Set $\Lambda(\lambda) = \int_X e^{\lambda F^2} dP$, $\lambda \in \mathbb{R}$, so that the preceding expresses that, in the range $\lambda \geq 0$,

$$\lambda \Lambda'(\lambda) - \Lambda(\lambda) \log \Lambda(\lambda) \,\leq\, 4\lambda^2 \Lambda'(\lambda).$$

In the spirit of the Herbst argument, it remains to integrate such a differential inequality. If $K(\lambda) = \log \Lambda(\lambda)$ $(> 0$ when $\lambda > 0)$, it reads for $0 < \lambda < \frac{1}{4}$,

$$\lambda(1 - 4\lambda) K'(\lambda) \,\leq\, K(\lambda),$$

that is $(\log K)'(\lambda) \leq \frac{1}{\lambda(1-4\lambda)}$. Hence, for every $0 < \eta \leq \lambda < \frac{1}{4}$,

$$\log K(\lambda) - \log K(\eta) \,\leq\, \int_\eta^\lambda \frac{1}{u(1 - 4u)}\, du \,=\, \log\Big(\frac{\lambda}{1 - 4\lambda} \cdot \frac{1 - 4\eta}{\eta}\Big).$$

As $\eta \to 0$, $K(\eta) = \log \int_X e^{\eta F^2} dP \sim \eta M_2$ where $M_2 = \int_X F^2 dP$. Therefore, in this limit, the preceding inequality yields

$$\log K(\lambda) - \log M_2 \,\leq\, \log\Big(\frac{\lambda}{1 - 4\lambda}\Big),$$

that is

$$K(\lambda) = \log\bigg(\int_X e^{\lambda F^2} dP\bigg) \,\leq\, \frac{\lambda M_2}{1 - 4\lambda}$$

for every $0 \le \lambda < \frac{1}{4}$. For instance with $\lambda = \frac{1}{14}$ (see below for the significance of this choice),

$$\int_X e^{\frac{1}{14}F^2} dP \,\le\, e^{\frac{1}{10}M_2}. \tag{27}$$

As $\lambda \to 0$ in (26), $\mathrm{Var}_P(F^2) \le 8M_2$. Since $F = F_A = 0$ on A it follows that $M_2 \le \frac{8}{P(A)}$. Hence, from (27),

$$\int_X e^{\frac{1}{14}F^2} dP \,\le\, e^{\frac{4}{5P(A)}}. \tag{28}$$

In particular, if $P(A) \ge \frac{1}{2}$, $\int_X e^{\frac{1}{14}F_A^2} dP \le 5$.

These first conclusions are weaker than (1) in terms of numerical constants, but more importantly in terms of the dependence on $P(A)$, especially for sets A with small probability. It is nevertheless already good enough for many of the significant applications of Talagrand's convex distance inequality. It may be pointed out also that, towards (28), the Herbst argument may be developed more simply with $e^{\lambda F}$ rather than $e^{\lambda F^2}$. The presentation here is motivated by homogeneity with the second part which is coming next.

To reach better dependence on $P(A)$ as in (1), it is necessary to develop the previous analysis but for negative values of λ. Consider therefore (10) now applied to $f = e^{-\lambda F^2}$, $\lambda \ge 0$. To this task, it is worthwhile to observe that given $x = (x_1, \dots, x_n)$ and $y_i \in \Omega_i$, $i = 1, \dots, n$,

$$F_i(x_i)^2 - F_i(y_i)^2 \,=\, F(x)^2 - F(y)^2 \,\le\, 1 \tag{29}$$

where as usual $F_i(y_i) = F(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n) = F(y)$. A proof of this claim may be given using the identity $F(x) = F_A(x) = d_A(x)$. Indeed,

$$d_A(x)^2 \,=\, \inf\big\{ |w|^2;\, w \in \mathrm{Conv}\big((\mathbb{1}_{\{x_j \neq z_j\}})_{1 \le j \le n};\, z \in A \big) \big\} \,=\, \inf \Big| \sum \theta_z s(z) \Big|^2$$

where the infimum is over all finite sums $\sum \theta_z s(z)$ with $\theta_z \ge 0$, $\sum \theta_z = 1$, $s(z) = (\mathbb{1}_{\{x_j \neq z_j\}})_{1 \le j \le n}$, $z \in A$. If then $\big| \sum \theta_z s'(z) \big|^2$ witnesses $d_A(y)^2$ (up to some $\delta > 0$),

$$d_A(x)^2 - d_A(y)^2 \,\le\, \Big| \sum \theta_z s(z) \Big|^2 - \Big| \sum \theta_z s'(z) \Big|^2$$

where $s(z)$ differs from $s'(z)$ only on the i-th coordinate which is $\mathbb{1}_{\{x_i \neq z_i\}}$ rather than $\mathbb{1}_{\{y_i \neq z_i\}}$. That is

$$d_A(x)^2 - d_A(y)^2 \,\le\, \Big(\sum \theta_z \mathbb{1}_{\{x_i \neq z_i\}} \Big)^2 - \Big(\sum \theta_z \mathbb{1}_{\{y_i \neq z_i\}} \Big)^2 \,\le\, 1$$

justifying (29).

For $f = e^{-\lambda F^2}$, $\lambda \ge 0$, the reasoning of the case $f = e^{\lambda F^2}$ may then be repeated together with the new information (29) to get that whenever $F_i(x_i) \ge F_i(y_i)$ (≥ 0), $i = 1, \dots, n$,

$$\begin{aligned} \big[-\lambda F_i(x_i)^2 + \lambda F_i(y_i)^2 \big] \big[e^{-\lambda F_i(x_i)^2} - e^{-\lambda F_i(y_i)^2} \big] \,&\le\, \lambda^2 \big[F_i(x_i)^2 - F_i(y_i)^2 \big) \big]^2 e^{-\lambda F_i(y_i)^2} \\ &\le\, \lambda^2 e^{\lambda} \big[F_i(x_i)^2 - F_i(y_i)^2 \big) \big]^2 e^{-\lambda F_i(x_i)^2} \\ &\le\, 4\lambda^2 e^{\lambda} \big(a_i(x) + \varepsilon \big)^2 F_i(x_i)^2 e^{-\lambda F_i(x_i)^2}. \end{aligned}$$

Arguing as before and letting $\varepsilon \to 0$ yields that

$$\mathrm{Ent}_P(e^{-\lambda F^2}) \,\leq\, 4\lambda^2 e^{\lambda} \int_X F^2 e^{-\lambda F^2} dP$$

for every $\lambda \geq 0$.

Set $\Lambda(\lambda) = \int_X e^{-\lambda F^2} dP$, $\lambda \in \mathbb{R}$, so that the preceding expresses that, in the range $0 \leq \lambda \leq \frac{1}{2}$ (for example),

$$\lambda \Lambda'(\lambda) - \Lambda(\lambda) \log \Lambda(\lambda) \,\leq\, -8\lambda^2 \Lambda'(\lambda).$$

This differential inequality is integrated as before, this time with $K(\lambda) = -\log \Lambda(\lambda) \geq 0$, to produce that

$$-K(\lambda) = \log\bigg(\int_X e^{-\lambda F^2} dP\bigg) \,\leq\, -\frac{\lambda M_2}{1 + 8\lambda}$$

for $0 \leq \lambda \leq \frac{1}{2}$.

Since $F = F_A = 0$ on A, it follows that for $\lambda = \frac{1}{2}$, $\frac{1}{10} M_2 \leq \log\big(\frac{1}{P(A)}\big)$. Together with (27),

$$\int_X e^{\frac{1}{14} F_A^2} dP \,\leq\, \frac{1}{P(A)}$$

which, up to the numerical constant, is the announced Talagrand inequality (1).

This thereby concludes the proof of the Talagrand convex hull inequality by the entropy method. It is not clear (see [20]) whether the constant $\frac{1}{4}$ may be reached by this method (nor than $\frac{1}{4}$ is optimal).

Some notes and references. Talagrand's convex distance inequality was first established in [81], following an earlier result on the discrete cube [80, 42] – the note [42] by W. Johnson and G. Schechtman actually motivated M. Talagrand towards the general formulation (observe that $d_A(x) \geq \inf_{y \in \mathrm{Conv}(A)} |x - y|$ on the discrete cube $X = \{0,1\}^n$), with a (rather short) proof going by induction on the dimension together with geometric arguments. The original statement [81, 83] actually involves a stronger family of distances rather than only the quadratic d_A^2. The proof presented here via the entropy method was put forward by S. Boucheron, G. Lugosi and P. Massart in [19] and [20].

Mass transportation proofs of the convex distance inequality have been considered in [54, 55, 26, 25, 73, 27, 37]... (see [20, Chapter 8] for an account).

Talagrand's convex distance inequality was initially motivated by several issues in probability in Banach spaces, in particular [80] the analogue of the Gaussian concentration inequality for norms of series $S = \sum_{i=1}^n \varepsilon_i v_i$ of independent Bernoulli random variables ε_i, $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = 0) = \frac{1}{2}$, with vector-valued coefficients v_i, $i = 1, \ldots, n$,

$$\mathbb{P}\Big(\big|\|S\| - \mathbb{E}\big(\|S\|\big)\big| \geq r\Big) \,\leq\, 4\, e^{-r^2/4\sigma^2}, \quad r \geq 0,$$

where $\sigma^2 = \sup_{\|\xi\| \leq 1} \sum_{i=1}^n \langle \xi, v_i \rangle^2$ (cf. [50]), a significant strengthening of the famous Khintchine-Kahane inequality [43].

Its abstract and powerful potential was then emphasized in the monumental memoir [83] on concentration inequalities for product measures, describing numerous illustrations in discrete and combinatorial probability theory. References to this work give an idea of the impact of the result. It may be refereed in particular to the general reviews and books [83, 79, 52, 49, 60, 30, 87, 88, 20, 2, 89]... for a sample applications and developments.

Among the commonly used forms of the convex distance inequality, one may put forward the following two.

Corollary 2. *Let $P = \mu_1 \otimes \cdots \otimes \mu_n$ be a product probability measure on a product space $X = \Omega_1 \times \cdots \times \Omega_n$. Let $F : X \to \mathbb{R}$ (measurable) be such that for every $x \in X$ there exists $a(x) = a \in [0,\infty)^n$ with $|a| = 1$ such that for every $y \in X$,*

$$F(x) \,\leq\, F(y) + d_a(x,y). \tag{30}$$

Then, if M is a median of F for P, for any $r \geq 0$,

$$P\big(|F - M| \geq r\big) \,\leq\, 4\,e^{-r^2/4}.$$

Replacing F by $-F$, this corollary applies similarly if (30) is changed into $F(y) \leq F(x) + d_a(x,y)$.

Corollary 3. *Let P be any product probability measure supported on $[0,1]^n$. For every convex Lipschitz function F on $\mathbb{R}^n$ with $\|F\|_{\mathrm{Lip}} \leq 1$, and every $r \geq 0$,*

$$P\big(|F - M| \geq r\big) \,\leq\, 4\,e^{-r^2/4}$$

where M is a median of F for P.

At the expense of numerical constants, medians may be replaced by expectations. Integrating in $r \geq 0$ the inequalities of the preceding corollaries yields $|\int_X FdP - M| \leq 4\sqrt{\pi} \leq 8$. Hence

$$P\big(|F - \textstyle\int_X FdP| \geq r\big) \,\leq\, P\big(|F - M| \geq \frac{r}{2}\big) \,\leq\, 4\,e^{-r^2/16}$$

if $r \geq 16$, while $P\big(|F - \int_X FdP| \geq r\big) \leq 1 \leq e^{16}\,e^{-r^2/16}$ if $r \leq 16$. It is also simple to go back from a concentration inequality around the mean to one around a median (cf. [59, 49]).

5 The L^1–L^2 variance inequality

With respect to the original formulation of Talagrand's L^1–L^2 variance inequality in Section 1, the argument is developed here with functions on $\{-1,+1\}^n$ and makes use of the framework presented in Section 3. The end of the proof catches up with the original statement (2). The main point of the proof will be to use hypercontractivity on the expansion of the variance along the semigroup $(P_t)_{t\geq 0}$ defined in (17).

The starting point is therefore the variance representation along $(P_t)_{t\geq 0}$ of a function f on $X = \{-1,+1\}^n$ as

$$\begin{aligned}\int_X f^2 d\mu_{\mathsf{p}}^n - \int_X (P_1 f)^2 d\mu_{\mathsf{p}}^n &= -\int_0^1 \bigg(\frac{d}{dt}\int_X (P_t f)^2 d\mu_{\mathsf{p}}^n\bigg) dt \\ &= -2\int_0^1 \bigg(\int_X P_t f \,\mathrm{L} P_t f d\mu_{\mathsf{p}}^n\bigg) dt \\ &= 2\int_0^1 \sum_{i=1}^n \int_X (\mathrm{L}_i P_t f)^2 d\mu_{\mathsf{p}}^n \, dt.\end{aligned}$$

Assume next that $\int_X f d\mu_{\mathsf{p}}^n = 0$ so that $\int_X P_s f d\mu_{\mathsf{p}}^n = 0$ for every $s \geq 0$ as well. From the Poincaré inequality (16), the derivative of the map $s \mapsto e^{2s}\int_X (P_s f)^2 d\mu_{\mathsf{p}}^n$ is negative, so the map is decreasing and thus

$$\int_X (P_1 f)^2 d\mu_{\mathsf{p}}^n \leq \frac{1}{e^2}\int_X f^2 d\mu_{\mathsf{p}}^n.$$

Therefore

$$\int_X f^2 d\mu_{\mathsf{p}}^n \leq 3\int_0^1 \sum_{i=1}^n \int_X (\mathrm{L}_i P_t f)^2 d\mu_{\mathsf{p}}^n \, dt.$$

Now $\mathrm{L}_i \mathrm{L} f = \mathrm{L}\,\mathrm{L}_i f$ so that $\mathrm{L}_i P_t f = P_t(\mathrm{L}_i f)$ for every $i = 1, \ldots, n$ and $t \geq 0$. It may thus be called on the hypercontractivity property (19) to get that, for every $i = 1, \ldots, n$ and $t > 0$,

$$\int_X (\mathrm{L}_i P_t f)^2 d\mu_{\mathsf{p}}^n = \int_X \big|P_t(\mathrm{L}_i f)\big|^2 d\mu_{\mathsf{p}}^n \leq \bigg(\int_X |\mathrm{L}_i f|^p d\mu_{\mathsf{p}}^n\bigg)^{2/p}$$

where $p = p(t) = 1 + e^{-4\rho t} < 2$. Recall that $\rho = \frac{\mathsf{p}-\mathsf{q}}{\log \mathsf{p} - \log \mathsf{q}}$ ($= \frac{1}{2}$ if $\mathsf{p} = \frac{1}{2}$). After the change of variables $p(t) = v$, it holds that

$$\int_X f^2 d\mu_{\mathsf{p}}^n \leq \frac{1}{\rho}\, e^{4\rho} \sum_{i=1}^n \int_1^2 \bigg(\int_X |\mathrm{L}_i f|^v d\mu_{\mathsf{p}}^n\bigg)^{2/v} dv.$$

This inequality actually basically amounts to the result (and may be used toward an Orlicz space formulation – see the comment below). Indeed, by Hölder's inequality,

$$\bigg(\int_X |\mathrm{L}_i f|^v d\mu_{\mathsf{p}}^n\bigg)^{1/v} = \|\mathrm{L}_i f\|_v \leq \|\mathrm{L}_i f\|_1^\theta \, \|\mathrm{L}_i f\|_2^{2/\theta}$$

where $\theta = \theta(v) \in [0,1]$ is defined by $\frac{1}{v} = \frac{\theta}{1} + \frac{1-\theta}{2}$. Hence

$$\int_1^2 \|\mathrm{L}_i f\|_v^2 \, dv \leq \|\mathrm{L}_i f\|_2^2 \int_1^2 b^{2\theta(v)} dv$$

where $b = \frac{\|\mathrm{L}_i f\|_1}{\|\mathrm{L}_i f\|_2} \leq 1$ ($= 1$ if $\|\mathrm{L}_i f\|_2 = 0$). It remains to evaluate the latter integral with $2\theta(v) = s$,

$$\int_1^2 b^{2\theta(v)} dv \leq \int_0^2 b^s ds \leq \frac{2}{1 + \log(\frac{1}{b})}\,.$$

As a consequence, for any $f : X \to \mathbb{R}$,

$$\mathrm{Var}_{\mu_{\mathsf{p}}^n}(f) \,\leq\, \frac{2}{\rho}\, e^{4\rho} \sum_{i=1}^n \frac{\|\mathrm{L}_i f\|_2^2}{1 + \log \frac{\|\mathrm{L}_i f\|_2}{\|\mathrm{L}_i f\|_1}} \,. \tag{31}$$

It remains to make the comparison with the formulation (2) of Talagrand's inequality. To this task, observe that

$$\mathrm{L}_i f(x) \,=\, (1-\mathsf{p})\big[f(x_1, \ldots, x_{i-1}, -1, x_{i+1}, \ldots, x_n) - f(x)\big]$$

if $x_i = +1$ and

$$\mathrm{L}_i f(x) \,=\, \mathsf{p}\big[f(x_1, \ldots, x_{i-1}, +1, x_{i+1}, \ldots, x_n) - f(x)\big]$$

if $x_i = -1$, that is what is denoted $-\Delta_i f$ in (2) after the change from $\{-1,+1\}$ to $\{0,1\}$. The Talagrand inequality (2) therefore follows having observed that $\frac{1}{\rho}\, e^{4\rho}$ is of the order of $\log \frac{1}{\mathsf{p}}$ as $\mathsf{p} \to 0$. In fact, from the preceding proof, $K = 30$ ($K = 14$ if $\mathsf{p} = \frac{1}{2}$) is a valid numerical constant for (2) (but may be easily improved).

Some notes and references. The original proof in [82] uses some Fourier analysis on the discrete cube (although the author is claiming that it may not be used in the $\mathsf{p} \neq \frac{1}{2}$ case). It does not mention hypercontractivity although it is implicit (Lemma 2.1). A simplified proof, for $\mathsf{p} = \frac{1}{2}$, was proposed in [12] putting forward the hypercontractive argument. The proof presented here is taken from [23], where it is extended to wider settings. In particular, the Talagrand L^1–L^2 inequality also applies to the Gaussian model.

Talagrand's L^1–L^2 variance inequality was motivated by a result of L. Russo [72] on a threshold effect for monotone sets depending little on any given coordinate. The result also provided an alternate proof of the famous result of J. Kahn, G. Kalai and N. Linial [44] about influences on the cube (that already used hypercontractivity). Actually, M. Talagrand's approach is an adaptation of the ideas of [44]. Namely, applying (2) to the (Boolean) function $f = \mathbb{1}_A - \mu(A)$ for some set $A \subset \{0,1\}^n$ with $\mu_{\frac{1}{2}}^n(A) = \alpha \in (0,1)$ (take $\mathsf{p} = \frac{1}{2}$ for simplicity), it follows that

$$\alpha(1-\alpha) \,\leq\, 2K \sum_{i=1}^n \frac{\mathrm{I}_i(A)}{\log\big(\frac{e}{\sqrt{2\mathrm{I}_i(A)}}\big)}$$

where, for each $i = 1, \ldots, n$,

$$\mathrm{I}_i(A) \,=\, \mu_{\frac{1}{2}}^n\big(x \in A; U_i(x) \notin A\big)$$

is the so-called influence of the i-th coordinate on the set A. In particular, there is a coordinate i, $1 \leq i \leq n$, such that

$$\mathrm{I}_i(A) \,\geq\, \frac{\alpha(1-\alpha)}{8K}\, \frac{\log n}{n}$$

which is the main result of [44]. This result remarkably improves by a (optimal) factor $\log n$ what would follow from the Poincaré inequality (16) applied to $f = \mathbb{1}_A$. Since then, the Talagrand

L^1–L^2 variance inequality plays a major role in Boolean analysis and its ramifications with theoretical computer science (cf. e.g. [63]).

The Talagrand inequality (2) indeed represents a sharpening upon the Poincaré inequality (16) (up to numerical constants). It may actually be interpreted as a dual form of the logarithmic Sobolev inequality in the (loose) sense that the latter ensures that if some gradient of a function f is in L^2, then the function belongs to the Orlicz space $L^2 \log L$, while the Talagrand inequality expresses that if the gradient is in $L^2(\log L)^{-1}$, then the function (with zero mean) is in L^2 (cf. Theorem 1.6 in [82]). This dual point of view was emphasized in [16]. A specific feature of the Talagrand inequality (2) is nevertheless that is has a suitable product structure. Alternate forms of the Talagrand inequality as a direct consequence of the logarithmic Sobolev inequality are developed in [33] and [70].

With the work [12] by I. Benjamini, G. Kalai and O. Schramm (see further [BR08]), the Talagrand L^1–L^2 variance inequality has been identified as one of the rare tool towards subdiffusive regimes and super-concentration phenomena, ubiquitous to many models of the current research (percolation, random matrices, spin glasses etc.). The surveys and monographs [11, 22, 34, 5, 77]... give an account on these recent active developments.

6 The quadratic transportation cost inequality

This section presents the entropic proof of Talagrand's transportation inequality (3), deducing it from the logarithmic Sobolev inequality (21).

To start with, recast the transportation metric T_w of (3) as the (quadratic) Kantorovich distance between probability measures μ and ν on the Borel sets of $\mathbb{R}^n$

$$\mathrm{W}_2(\mu,\nu) \,=\, \inf\bigg(\int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2 d\pi(x,y)\bigg)^{1/2}$$

where the infimum is taken over all couplings π on $\mathbb{R}^n\times\mathbb{R}^n$ with respective marginals μ and ν, $|x-y|$ being the Euclidean distance between x and y in $\mathbb{R}^n$.

The classical duality formula (e.g. [90]) for the Kantorovich distance $\mathrm{W}_2(\mu,\nu)$ between two probability measures μ and ν on the Borel sets of $\mathbb{R}^n$ expresses that

$$\frac{1}{2}\,\mathrm{W}_2(\mu,\nu)^2 \,=\, \sup\bigg(\int_{\mathbb{R}^n} Q_1\varphi\, d\mu - \int_{\mathbb{R}^n}\varphi\, d\nu\bigg) \tag{32}$$

where the supremum is taken over all bounded continuous functions $\varphi : \mathbb{R}^n \to \mathbb{R}$ and where

$$Q_s\varphi(x) \,=\, \inf_{y\in\mathbb{R}^n}\Big[\varphi(y) + \frac{1}{2s}\,|x-y|^2\Big], \quad s>0,\ x\in\mathbb{R}^n,$$

is the infimum-convolution Hopf-Lax semigroup. It is standard (cf. e.g. [32, 90]) that $Q_s\varphi(x)$, $s>0$, $x\in\mathbb{R}^n$, solves the Halmiton-Jacobi equation

$$\frac{d}{ds}\,Q_s\varphi \,=\, -\frac{1}{2}\,|\nabla Q_s\varphi|^2 \tag{33}$$

in $(0,\infty)\times\mathbb{R}^n$ with initial condition φ.

On the basis of (32), the entropic inequality (7) ensures that if $d\mu = f d\nu$,

$$\begin{aligned}\int_{\mathbb{R}^n} Q_1\varphi\, d\mu - \int_{\mathbb{R}^n} \varphi\, d\nu \ &= \int_{\mathbb{R}^n} Q_1\varphi\, f\, d\nu - \int_{\mathbb{R}^n} \varphi\, d\nu \\ &\le \int_{\mathbb{R}^n} f\log f\, d\nu + \log \int_{\mathbb{R}^n} e^{Q_1\varphi} d\nu - \int_{\mathbb{R}^n} \varphi\, d\nu.\end{aligned}$$

Assume now that ν is the standard Gaussian measure γ on $\mathbb{R}^n$. The proof will follow the Herbst argument and deduce the Talagrand transportation cost inequality from the logarithmic Sobolev inequality (21). Apply namely the latter $f^2 = e^{s\,Q_s\varphi}$, $s>0$, to get that

$$\int_{\mathbb{R}^n} s\, Q_s\varphi\, e^{s\,Q_s\varphi} d\gamma - \Lambda(s)\log\Lambda(s) \ \le \ \frac{s^2}{2}\int_{\mathbb{R}^n} |\nabla Q_s\varphi|^2 e^{s\,Q_s\varphi} d\gamma$$

where $\Lambda(s) = \int_{\mathbb{R}^n} e^{s\,Q_s\varphi} d\gamma$. But $\partial_s Q_s\varphi = -\frac{1}{2}|\nabla Q_s\varphi|^2$ so that

$$\Lambda'(s) \ = \int_{\mathbb{R}^n} Q_s\varphi\, e^{s\,Q_s\varphi} d\gamma - \frac{s}{2}\int_{\mathbb{R}^n} |\nabla Q_s\varphi|^2 e^{s\,Q_s\varphi} d\gamma,$$

and hence the previous inequality is translated into

$$s\,\Lambda'(s) - \Lambda(s)\log\Lambda(s) \ \le \ 0$$

for every $s>0$. Setting $H(s) = \frac{1}{s}\log\Lambda(s)$, $H(0) = \int_{\mathbb{R}^n}\varphi\, d\gamma$, it therefore holds true that $H'(s)\le 0$, $s>0$. Hence

$$\log\int_{\mathbb{R}^n} e^{Q_1\varphi} d\gamma \ = \ H(1) \ \le \ H(0) \ = \int_{\mathbb{R}^n} \varphi\, d\gamma,$$

and plugging this conclusion into the above entropic inequality yields that

$$\int_{\mathbb{R}^n} Q_1\varphi\, d\mu - \int_{\mathbb{R}^n} \varphi\, d\gamma \ \le \int_{\mathbb{R}^n} f\log f\, d\gamma.$$

Taking the supremum over all bounded continuous $\varphi : \mathbb{R}^n \to \mathbb{R}$, the Kantorovich duality (32) yields

$$\mathrm{W}_2(\mu,\gamma)^2 \ \le \ 2\int_{\mathbb{R}^n} f\log f\, d\gamma$$

which is the Talagrand quadratic transportation cost inequality (3).

As for the logarithmic Sobolev inequality (21), the constant 2 is optimal in (3) as can be seen by the choice for μ of a shift of γ by $b\in\mathbb{R}^n$.

Some notes and references. M. Talagrand's original proof of (3) in [84] uses monotone transport in dimension one together with a tensorization argument. The main result of [84] is actually a corresponding (stronger) inequality for products of the exponential measure, the Gaussian case being presented as a simpler example to deal with first. Motivation comes from the mass transportation approach to the concentration of measure phenomenon put forward by K. Marton [53, 54], and a sharp form of which for these measures.

Mass transportation proofs directly in dimension n have been provided next, in particular by means of the Brenier map [21, 13, 90, 91]... That the logarithmic Sobolev inequality implies the Talagrand quadratic transportation cost inequality is the main achievement of the celebrated paper [65] by F. Otto and C. Villani. The implication holds in a rather general setting. The approach in [65] relies on the formal Otto calculus in Wasserstein space [64], and has been one driving force in the study of functional inequalities and curvature lower bounds in metric measure spaces (cf. [90, 91, 8]). The proof presented here inspired by the Herbst argument is taken from [14, 15]. For more on transportation cost inequalities, see in addition [36].

That mass transportation is at the root of the four Talagrand inequalities is an alternate project of independent interest, already partly understaken in [20, Chapter 8].

7 Supremum of empirical processes

This last section is thus devoted to the proof of the inequality (4), in the following notation according to the choices adopted so far. Let $X_1, \ldots, X_n$ be independent random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in some measurable space $(S, \mathcal{S})$. Let $\mathcal{F}$ be a countable family of real-valued uniformly bounded measurable functions on S, and set

$$Z = \sup_{g \in \mathcal{F}} \sum_{i=1}^{n} g(X_i). \tag{34}$$

By (dominated) convergence and homogeneity, it is enough to consider a finite family $\mathcal{F} = \{g_1, \ldots, g_N\}$ such that $|g_k| \leq 1$, $k = 1, \ldots, N$. (It is also assumed below that Z is not 0 almost surely.)

To make use of the framework of Section 3, denote by $\mu_1, \ldots, \mu_n$ the respective probability distributions of the independent random variables $X_1, \ldots, X_n$, and set $P = \mu_1 \otimes \cdots \otimes \mu_n$ on the product space S^n. In accordance

$$Z = Z(x) = Z(x_1, \ldots, x_n) = \max_{1 \leq k \leq N} \sum_{i=1}^{n} g_k(x_i), \quad x = (x_1, \ldots, x_n) \in S^n$$

(for which nevertheless the probabilistic notation induced by (34) will be used from time to time below). Following the Herbst argument, the task will be to apply the tensorized logarithmic Sobolev inequalities of (10) and (11) to $e^{\lambda Z}$ for every $\lambda \in \mathbb{R}$, in much the same way as for the convex distance inequality in Section 4.

As is classical in the study of exponential inequalities for sums of independent random variables, the inequality (4) entails a Gaussian tail for the small values of r and a Poisson one for the large values.

In the first part of the proof, consider Poisson tails for non-negative random variables. Let thus g_k, $k = 1, \ldots, N$, be such that $0 \leq g_k \leq 1$ and apply (11) to $f = e^{\lambda Z}$, $\lambda \in \mathbb{R}$. For each

$i = 1, \ldots, n$ and $x = (x_1, \ldots, x_n) \in S^n$, choose $c_i = e^{\lambda Z^i(x)}$ where

$$Z^i(x) \, = \, \max_{1 \leq k \leq N} \sum_{j \neq i} g_k(x_j).$$

By definition $Z^i(x)$ only depends on $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ and $0 \leq Z_i(x_i) - Z^i(x) \leq 1$, $i = 1, \ldots, n$. Now

$$\lambda \big[Z_i(x_i) - Z^i(x) \big] e^{\lambda Z_i(x_i)} - \big[e^{\lambda Z_i(x_i)} - e^{\lambda Z^i(x)} \big] \, = \, \phi\big(- \lambda [Z_i(x_i) - Z^i(x)] \big) \, e^{\lambda Z_i(x_i)}$$

where $\phi(u) = e^u - 1 - u$, $u \in \mathbb{R}$. Since ϕ is convex and $\phi(0) = 0$, $\phi(-\lambda u) \leq u \phi(-\lambda)$ for every λ and $0 \leq u \leq 1$, so that

$$\begin{aligned} \sum_{i=1}^n \lambda \big[Z_i(x_i) - Z^i(x) \big] e^{\lambda Z_i(x_i)} - \big[e^{\lambda Z_i(x_i)} - e^{\lambda Z^i(x)} \big] & \\ \leq \, \sum_{i=1}^n \big[Z_i(x_i) - Z^i(x) \big] \phi(-\lambda) \, e^{\lambda Z_i(x_i)} & \\ \leq \, \phi(-\lambda) \, Z(x) \, e^{\lambda Z(x)} & \end{aligned}$$

where it is used that

$$\sum_{i=1}^n \big[Z_i(x_i) - Z^i(x) \big] \, = \, \sum_{i=1}^n \big[Z(x) - Z^i(x) \big] \, \leq \, Z(x).$$

As a consequence therefore of (11), and with probabilistic notation, for any $\lambda \in \mathbb{R}$,

$$\lambda \, \mathbb{E}(Z e^{\lambda Z}) - \mathbb{E}(e^{\lambda Z}) \log \mathbb{E}(e^{\lambda Z}) \, \leq \, \phi(-\lambda) \, \mathbb{E}(Z e^{\lambda Z}).$$

If $\Lambda(\lambda) = \mathbb{E}(e^{\lambda Z})$ and $H(\lambda) = \frac{1}{\lambda} \log \Lambda(\lambda)$, $\lambda \in \mathbb{R}$ ($H(0) = \int_X Z dP$), the preceding reads

$$H'(\lambda) \, \leq \, \frac{\phi(-\lambda)}{\lambda^2} \, \frac{\Lambda'(\lambda)}{\Lambda(\lambda)} \, .$$

Since $H'(\lambda) = -\frac{1}{\lambda} H(\lambda) + \frac{1}{\lambda} \frac{\Lambda'(\lambda)}{\Lambda(\lambda)}$, it follows that for $\lambda > 0$,

$$\frac{H'(\lambda)}{H(\lambda)} \, \leq \, \frac{\phi(-\lambda)}{\lambda(\lambda - \phi(-\lambda))} \, = \, \frac{1}{\lambda - \phi(-\lambda)} - \frac{1}{\lambda} \, . \tag{35}$$

It is easily seen that, again with $\lambda > 0$,

$$\int_0^\lambda \Big[\frac{1}{u - \phi(-u)} - \frac{1}{u} \Big] \, du \, = \, \int_0^\lambda \Big[\frac{1}{1 - e^{-u}} - \frac{1}{u} \Big] \, du \, = \, \lambda - \log \lambda + \log(1 - e^{-\lambda})$$

so that, integrating (35),

$$H(\lambda) \, \leq \, H(0) \frac{1}{\lambda} \, (e^\lambda - 1), \quad \lambda \geq 0.$$

As a conclusion of this analysis, the following statement holds true.

Proposition 4. *If $0 \leq g \leq 1$ for every $g \in \mathcal{F}$,*

$$\mathbb{E}(e^{\lambda Z}) \leq e^{\mathbb{E}(Z)(e^{\lambda}-1)}$$

for every $\lambda \geq 0$. As a consequence, for every $r \geq 0$,

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z) + r\big) \leq \exp\bigg(-\mathbb{E}(Z)\, h\Big(\frac{r}{\mathbb{E}(Z)}\Big)\bigg) \tag{36}$$

where $h(u) = (1+u)\log(1+u) - u$, $u \geq 0$.

The Poisson tail (36) is obtained from Markov's inequality

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z) + r\big) \leq e^{-\lambda(\mathbb{E}(Z)+r)+\mathbb{E}(Z)(e^{\lambda}-1)}$$

and optimization in $\lambda \geq 0$.

For functions g taking values in $[-1,+1]$, a similar scheme may be followed on the basis this time of (10) towards Gaussian tails. Let thus g_k, $k = 1, \ldots, N$, be such that $|g_k| \leq 1$ and apply (10) to $f = e^{\lambda Z}$. If $Z_i(x_i) \geq Z_i(y_i)$, $x_i, y_i \in S$, $i = 1, \ldots, n$, and $\lambda \geq 0$, by the mean value theorem,

$$\lambda\big[Z_i(x_i) - Z_i(y_i)\big]\big[e^{\lambda Z_i(x_i)} - e^{\lambda Z_i(y_i)}\big] \leq \lambda^2\big[Z_i(x_i) - Z_i(y_i)\big]^2 e^{\lambda Z_i(x_i)}.$$

By the definition of $Z = Z(x)$,

$$\sum_{i=1}^{n}\big[Z_i(x_i) - Z_i(y_i)\big]^2 \mathbb{1}_{\{Z_i(x_i)\geq Z_i(y_i)\}} \leq \max_{1\leq k\leq N}\sum_{i=1}^{n}\big[g_k(x_i) - g_k(y_i)\big]^2.$$

Denoting by $\widetilde{W} = \widetilde{W}(x,y)$ the right-hand side of the preceding inequality, (10) indicates that, in probabilistic notation,

$$\lambda\,\mathbb{E}(Ze^{\lambda Z}) - \mathbb{E}(e^{\lambda Z})\log\mathbb{E}(e^{\lambda Z}) \leq \lambda^2\,\mathbb{E}(\widetilde{W}e^{\lambda Z}) \tag{37}$$

for any $\lambda \geq 0$.

Working with $\lambda \leq 0$, it still holds that

$$\lambda\big[Z_i(x_i) - Z_i(y_i)\big]\big[e^{\lambda Z_i(x_i)} - e^{\lambda Z_i(y_i)}\big] \leq \lambda^2 e^{-2\lambda}\big[Z_i(x_i) - Z_i(y_i)\big]^2 e^{\lambda Z_i(x_i)}$$

as $(0 \leq)\, Z_i(x_i) - Z_i(y_i) \leq 2$. Together with (37), it may thus be concluded that, in the range $\lambda \in [-\frac{1}{4}, +\frac{1}{4}]$ (for example),

$$\lambda\,\mathbb{E}(Ze^{\lambda Z}) - \mathbb{E}(e^{\lambda Z})\log\mathbb{E}(e^{\lambda Z}) \leq 2\lambda^2\,\mathbb{E}(\widetilde{W}e^{\lambda Z}).$$

In addition, by independence, $\mathbb{E}(\widetilde{W}e^{\lambda Z}) \leq 2V\,\mathbb{E}(e^{\lambda Z}) + 2\,\mathbb{E}(We^{\lambda Z})$ where $W = W(x) = \max_{1\leq k\leq N}\sum_{i=1}^{n} g_k(x_i)^2$ and, in the notation of (4), $V = \mathbb{E}(W)$. As a consequence, for every $\lambda \in [-\frac{1}{4}, +\frac{1}{4}]$,

$$\lambda\,\mathbb{E}(Ze^{\lambda Z}) - \mathbb{E}(e^{\lambda Z})\log\mathbb{E}(e^{\lambda Z}) \leq 4V\lambda^2\,\mathbb{E}(e^{\lambda Z}) + 4\lambda^2\,\mathbb{E}(We^{\lambda Z}). \tag{38}$$

As in the Herbst argument, the latter inequality (38) is transformed into a differential inequality on Laplace transforms. To start with, observe that for any $\lambda \in \mathbb{R}$,

$$\lambda \, \mathbb{E}(W e^{\lambda Z}) \, \leq \, \lambda \, \mathbb{E}(Z e^{\lambda Z}) - \mathbb{E}(e^{\lambda Z}) \log \mathbb{E}(e^{\lambda Z}) + \mathbb{E}(e^{\lambda Z}) \log \mathbb{E}(e^{\lambda W}) \tag{39}$$

(as another instance of the entropic inequality (7) with $f = \frac{e^{\lambda Z}}{\mathbb{E}(e^{\lambda Z})}$ and $g = \lambda W$).

Start next with the positive values of λ. Setting $\Lambda(\lambda) = \mathbb{E}(e^{\lambda Z})$, $R(\lambda) = \mathbb{E}(e^{\lambda W})$, $\lambda \in \mathbb{R}$, it follows from (38) and (39) that for every $\lambda \in [0, \frac{1}{4}]$,

$$(1 - 4\lambda) \big[\lambda \Lambda'(\lambda) - \Lambda(\lambda) \log \Lambda(\lambda) \big] \, \leq \, 4V \lambda^2 \Lambda(\lambda) + 4\lambda \Lambda(\lambda) \log R(\lambda).$$

Proposition 4 may be applied to W to get that

$$\log R(\lambda) \, \leq \, \mathbb{E}(W)(e^\lambda - 1) \, = \, V(e^\lambda - 1), \quad \lambda \geq 0.$$

Hence

$$(1 - 4\lambda) \big[\lambda \Lambda'(\lambda) - \Lambda(\lambda) \log \Lambda(\lambda) \big] \, \leq \, 4V \big[\lambda^2 + \lambda(e^\lambda - 1) \big] \Lambda(\lambda),$$

and in the usual notation $H(\lambda) = \frac{1}{\lambda} \log \Lambda(\lambda)$, for any $0 \leq \lambda < \frac{1}{4}$,

$$H'(\lambda) \, \leq \, \frac{4V}{1 - 4\lambda} \bigg[1 + \frac{e^\lambda - 1}{\lambda} \bigg].$$

It remains to suitably integrate this differential inequality. Without trying sharp bounds, the right-hand side may simply be upper-bounded by $20V$ in the range $0 \leq \lambda \leq \frac{1}{8}$ so to get that

$$\mathbb{E}(e^{\lambda Z}) \, \leq \, e^{\lambda \mathbb{E}(Z) + 20\lambda^2 V}.$$

Given then $r \geq 0$, use Markov's exponential inequality with $\lambda = \frac{r}{40V}$ if $r \leq 5V$ and $\lambda = \frac{1}{8}$ otherwise to derive that

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z) + r \big) \, \leq \, e^{- \min(\frac{r}{16}, \frac{r^2}{80V})}.$$

Now, the same reasoning may be applied to $-Z$ since (38) holds also for $\lambda \in [-\frac{1}{4}, 0]$. Together with the two parts and the union bound, the following statement follows.

Proposition 5. *In the preceding notation, for every $r \geq 0$*

$$\mathbb{P}\big(\big| Z - \mathbb{E}(Z) \big| \geq r \big) \, \leq \, 2 \, e^{- \min(\frac{r}{16}, \frac{r^2}{80V})}.$$

This proposition is close to the Talagrand inequality (4) with the Gaussian tail for the small values of r ($\leq 5V$), but only an exponential decay for the large values (describing the so-called Bernstein inequality, cf. [20]). To reach the Poisson tail and the full conclusion, it should be combined with Proposition 4.

To this task, let $\tau > 0$ and set

$$Z_\tau^1 \, = \, \max_{1 \leq k \leq N} \sum_{i=1}^n g_k(X_i) \mathbb{1}_{\{|g_k(X_i)| \leq \tau\}}$$

and

$$Z_\tau^2 \;=\; \max_{1\le k\le N} \sum_{i=1}^n \big|g_k(X_i)\big| \mathbb{1}_{\{|g_k(X_i)|>\tau\}}$$

so that $|Z - Z_\tau^1| \le Z_\tau^2$. Then, for $r \ge \mathbb{E}(Z_\tau^2)$,

$$\mathbb{P}\big(\big|Z - \mathbb{E}(Z)\big| \ge 4r\big) \;\le\; \mathbb{P}\big(\big|Z_\tau^1 - \mathbb{E}(Z_\tau^1)\big| \ge r\big) + \mathbb{P}\big(Z_\tau^2 \ge \mathbb{E}(Z_\tau^2) + r\big). \tag{40}$$

By Proposition 5 applied to $\frac{1}{\tau} Z_\tau^1$,

$$\mathbb{P}\big(\big|Z_\tau^1 - \mathbb{E}(Z_\tau^1)\big| \ge r\big) \;\le\; 2\, e^{-\min(\frac{r}{16\tau}, \frac{r^2}{80\tau^2 V})}$$

for every $r \ge 0$, while by Proposition 4 applied to Z_τ^2,

$$\mathbb{P}\big(Z_\tau^2 \ge \mathbb{E}(Z_\tau^2) + r\big) \;\le\; e^{-\frac{r}{2}\log(1+\frac{r}{\mathbb{E}(Z_\tau^2)})}$$

since $h(u) \ge \frac{u}{2}\log(1+u)$, $u \ge 0$.

Choose next $\tau = \sqrt{\frac{4V}{5r}}$ $(r > 0)$. In the range $4r \ge 5V$,

$$\mathbb{E}(Z_\tau^2) \;\le\; \frac{V}{\tau} \;=\; \sqrt{\frac{5}{4}\, rV} \;\le\; r.$$

Hence

$$\frac{r}{2}\log\bigg(1 + \frac{r}{\mathbb{E}(Z_\tau^2)}\bigg) \;\ge\; \frac{r}{2}\log\bigg(1 + \sqrt{\frac{4r}{5V}}\bigg) \;\ge\; \frac{r}{6}\log\bigg(1 + \frac{4r}{V}\bigg).$$

It also holds in this range that

$$\min\bigg(\frac{r}{16\tau}, \frac{r^2}{80\tau^2 V}\bigg) \;\ge\; \frac{r}{75}\log\bigg(1 + \frac{4r}{V}\bigg).$$

Since by the choice of τ, $r \ge \mathbb{E}(Z_\tau^2)$ whenever $4r \ge 5V$, as a consequence of (40),

$$\mathbb{P}\big(\big|Z - \mathbb{E}(Z)\big| \ge 4r\big) \;\le\; 3\, e^{-\frac{r}{75}\log(1+\frac{4r}{V})}.$$

Proposition 5 for the values of $r \le 5V$ shows that

$$\mathbb{P}\big(\big|Z - \mathbb{E}(Z)\big| \ge r\big) \;\le\; 2\, e^{-\frac{r^2}{80V}} \;\le\; 2\, e^{-\frac{r}{80}\log(1+\frac{r}{V})}$$

since $\log(1+u) \le u$, $u \ge 0$. Together with the previous inequality (after the change from $4r$ to r) yields finally that

$$\mathbb{P}\big(\big|Z - \mathbb{E}(Z)\big| \ge r\big) \;\le\; 3\, \exp\bigg(-\frac{r}{300}\log\bigg(1 + \frac{r}{V}\bigg)\bigg)$$

for every $r \ge 0$. This is the Talagrand inequality (4), with $U = 1$ by homogeneity, completing thereby its proof.

Some notes and references. The proof of (4) developed in [85] is rather cumbersome, elaborating on and deepening the investigation [83]. These strengthenings have been clarified since then by means of information-theoretic and infimum-convolution tools in several contributions, including [54, 55, 26, 25, 73, 66, 74, 75, 27, 20, 37]... It is deduced from the convex distance inequality after a symmetrization argument in [67]. The approach relying on the entropy method and logarithmic Sobolev inequality was initiated in [47], and expanded and made precise in [56] and in various subsequent publications (cf. [20]). The truncation argument to suitably combine the Gaussian and Poisson tails is already present in [85]. The article [56] by P. Massart provides a much more careful analysis of the involved numerical constants, of significant relevance in the applications (Proposition 4 is taken from there – it is remarkable that it is already optimal for a class of functions reduced to one element). The reference [20] provides an account on these developments and on the various steps and contributions towards sharper (sometimes optimal) constants in families of inequalities for empirical processes. The Talagrand inequality (4) on the supremum of empirical processes is indeed nowadays a major tool in non-asymptotic statistics, where sensitive numerical constants are of importance. The monographs [57, 58, 20, 35, 92]... and the references therein present numerous illustrations and applications of this most powerful result in modern statistics.

It is worthwhile mentioning that for the applications, standard symmetrization tools allow for the bound

$$V \,=\, \mathbb{E}\bigg(\sup_{f\in\mathcal{F}} \sum_{i=1}^n f(X_i)^2\bigg) \,\leq\, U\,\mathbb{E}(Z) + 8\,\sup_{f\in\mathcal{F}} \sum_{i=1}^n \mathbb{E}\big(f(X_i)^2\big)$$

whenever $\mathbb{E}(f(X_i)) = 0$, $i = 1, \ldots, n$, $f \in \mathcal{F}$ and the class $\mathcal{F}$ is symmetric ($-\mathcal{F} = \mathcal{F}$) (see [50, 20]).

References

[1] S. Aida, T. Masuda, I. Shigekawa. Logarithmic Sobolev inequalities and exponential integrability. *J. Funct. Anal.* 126, 83–101 (1994).

[2] N. Alon, J. Spencer. *The probabilistic method.* Fourth edition. Wiley Series in Discrete Mathematics and Optimization. Wiley (2016).

[3] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, G. Scheffer. Sur les inégalités de Sobolev logarithmiques. *Panoramas et Synthèses* vol. 10. Soc. Math. de France (2000).

[4] S. Artstein-Avidan, A. Giannopoulos, V. Milman. *Asymptotic geometric analysis. Part I.* Mathematical Surveys and Monographs 202. American Mathematical Society (2015).

[5] A. Auffinger, M. Damron, J. Hanson. *50 years of first-passage percolation.* University Lecture Series 68. American Mathematical Society (2018).

[6] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. *École d'Été de Probabilités de Saint-Flour 1992, Lecture Notes in Math.* 1581, 1–114 (1994). Springer.

[7] D. Bakry, M. Émery. Diffusions hypercontractives. *Séminaire de Probabilités XIX, Lecture Notes in Math.* 1123, 177–206 (1985). Springer.

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

[9] W. Beckner. Inequalities in Fourier analysis. *Ann. of Math.* 102, 159–182 (1975).

[10] M. Benaïm, R. Rossignol. Exponential concentration for first passage percolation through modified Poincaré inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* 44, 544–573 (2008).

[11] I. Benjamini, G. Kalai, O. Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.* 90, 5–43 (1999).

[12] I. Benjamini, G. Kalai, O. Schramm. First passage percolation has sublinear distance variance. *Ann. Probab.* 31, 1970–1978 (2003).

[13] G. Blower. The Gaussian isoperimetric inequality and transportation. *Positivity* 7, 203–224 (2003).

[14] S. Bobkov, I. Gentil, M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl.* 80, 669–696 (2001).

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

[16] S. Bobkov, C. Houdré. A converse Gaussian Poincaré-type inequality for convex functions. *Statist. Probab. Lett.* 44, 281–290 (1999).

[17] V. Bogachev. *Gaussian measures.* Mathematical Surveys and Monographs 62. American Mathematical Society (1998).

[18] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier* 20, 335–402 (1971).

[19] S. Boucheron, G. Lugosi, P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.* 31, 1583–1614 (2003).

[20] S. Boucheron, G. Lugosi, P. Massart. *Concentration inequalities. A nonasymptotic theory of independence.* Oxford University Press (2013).

[21] D. Cordero-Erausquin. Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.* 161, 257–269 (2002).

[22] S. Chatterjee. *Superconcentration and related topics.* Springer Monographs in Mathematics. Springer (2014).

[23] D. Cordero-Erausquin, M. Ledoux. Hypercontractive measures, Talagrand's inequality, and influences. *Geometric aspects of functional analysis, Lecture Notes in Math.* 2050, 169–189. Springer (2012).

[24] E. B. Davies, B. Simon. Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. *J. Funct. Anal.* 59, 335–395 (1984).

[25] A. Dembo. Information inequalities and concentration of measure. *Ann. Probab.* 25, 927–939 (1997).

[26] A. Dembo, O. Zeitouni. Transportation approach to some concentration inequalities in product spaces. *Electron. Comm. Probab.* 1, 83–90 (1996).

[27] A. Dembo, O. Zeitouni. *Large deviations techniques and applications.* Corrected reprint of the second (1998) edition. Stochastic Modelling and Applied Probability 38. Springer (2010).

[28] J.-D. Deuschel, D. Stroock. *Large Deviations.* Pure and Applied Mathematics 137. Academic Press (1989).

[29] P. Diaconis, L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* 6, 695–750 (1996).

[30] D. P. Dubhashi, A. Panconesi. *Concentration of measure for the analysis of randomized algorithms.* Cambridge University Press (2009).

[31] B. Efron, C. Stein. The jackknife estimate of variance. *Ann. Statist.* 9, 586–596 (1981).

[32] L. Evans. *Partial differential equations.* Graduate Studies in Mathematics 19. American Mathematical Society (1998).

[33] D. Falik, A. Samorodnitsky. Edge-isoperimetric inequalities and influences. *Combin. Probab. Comput.* 16, 693–712 (2007).

[34] C. Garban, J. Steif. *Noise sensitivity of Boolean functions and percolation.* Institute of Mathematical Statistics Textbooks 5. Cambridge University Press (2015).

[35] E. Giné, R. Nickl. *Mathematical foundations of infinite-dimensional statistical models.* Cambridge Series in Statistical and Probabilistic Mathematics 40. Cambridge University Press (2015).

[36] N. Gozlan, C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields* 16, 635–736 (2010).

[37] N. Gozlan, C. Roberto, P.-M. Samson, P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.* 273, 3327–3405 (2017).

[38] M. Gromov. *Metric Structures for Riemannian and non-Riemannian Spaces.* Birkhäuser (1998).

[39] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.* 97, 1061–1083 (1975).

[40] L. Gross. Logarithmic Sobolev inequalities and contractivity properties of semigroups. *Dirichlet Forms, Lecture Notes in Math.* 1563, 54-88. Springer (1993).

[41] A. Guionnet, B. Zegarlinski. Lectures on logarithmic Sobolev inequalities. *Séminaire de Probabilités XXXVI, Lecture Notes in Math.* 1801, 1–134. Springer (2003).

[42] W. Johnson, G. Schechtman. Remarks on Talagrand's deviation inequality for Rademacher's functions. *Functional Analysis – Longhorn Notes, Lecture Notes in Math.* 1470, 72–77. Springer (1991).

[43] J.-P. Kahane. *Some random series of functions.* Heath Math. Monographs (1968). Second Edition Cambridge University Press (1985).

[44] J. Kahn, G. Kalai, N. Linial. The influence of variables on boolean functions. *29th Symposium on the Foundations of Computer Science*, 68–80 (1988).

[45] M. Ledoux. Remarks on logarithmic Sobolev constants, exponential integrability and bounds on the diameter *J. Math. Kyoto Univ.* 35, 211–220 (1995).

[46] M. Ledoux. Isoperimetry and Gaussian Analysis. *École d'Été de Probabilités de Saint-Flour 1994, Lecture Notes in Math.* 1648, 165–294 (1996). Springer.

[47] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.* 1, 63–87 (1996).

[48] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. *Séminaire de Probabilités XXXIII, Lecture Notes in Math.* 1709, 120–216. Springer (1999).

[49] M. Ledoux. *The concentration of measure phenomenon.* Mathematical Surveys and Monographs 89. American Mathematical Society (2001).

[50] M. Ledoux, M. Talagrand. *Probability in Banach spaces (Isoperimetry and processes).* Ergebnisse der Mathematik und ihrer Grenzgebiete 23. Springer (1991).

[51] M. Lifshits. *Gaussian random functions.* Mathematics and its Applications 322. Kluwer (1995).

[52] C. McDiarmid. Concentration. *Probabilistic Methods for Algorithmic Discrete Mathematics, Algorithms Combin.* 16, 195–248 (1998). Springer.

[53] K. Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory* 32, 445–446 (1986).

[54] K. Marton. Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration. *Ann. Probab.* 24, 857–866 (1996).

[55] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* 6, 556–571 (1996).

[56] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* 28, 863–884 (2000).

[57] P. Massart. Some applications of concentration inequalities to statistics. *Ann. Fac. Sci. Toulouse Math.* 9, 245–303 (2000).

[58] P. Massart. *Concentration inequalities and model selection.* École d'été de Probabilités de Saint-Flour 2003, Lecture Notes in Math. 1896. Springer (2007).

[59] V. D. Milman, G. Schechtman. *Asymptotic theory of finite dimensional normed spaces.* Lecture Notes in Math. 1200. Springer (1986).

[60] M. Molloy, B. Reed. *Graph colouring and the probabilistic method.* Algorithms and Combinatorics 23. Springer (2002).

[61] E. Nelson. A quartic interaction in two dimensions. *Mathematical Theory of Elementary Particles, Dedham 1965. Proc. Conf.*, 69–73. MIT Press, Cambridge (1966).

[62] E. Nelson. The free Markoff field. *J. Funct. Anal.* 12, 211– 227 (1973).

[63] R. O'Donnell. *Analysis of Boolean functions.* Cambridge University Press (2014).

[64] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations* 26, 101–174 (2001).

[65] F. Otto, C. Villani. Generalization of an inequality by Talagrand, and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* 173, 361–400 (2000).

[66] D. Panchenko. A note on Talagrand's concentration inequality. *Electron. Comm. Probab.* 6, 55–65 (2001).

[67] D. Panchenko. Symmetrization approach to concentration inequalities for empirical processes. *Ann. Probab.* 31, 2068–2081 (2003).

[68] G. Pisier. Probabilistic methods in the geometry of Banach spaces. *Probability and Analysis, Lecture Notes in Math.* 1206, 167–241 (1986). Springer.

[69] W. Rhee, M. Talagrand. Martingale inequalities and the jackknife estimate of variance. *Statist. Probab. Lett.* 4, 5–6 (1986).

[70] R. Rossignol. Threshold for monotone symmetric properties through a logarithmic Sobolev inequality. *Ann. Probab.* 34, 1707–1725 (2006).

[71] G. Royer. *An initiation to logarithmic Sobolev inequalities.* SMF/AMS Texts and Monographs 14, American Mathematical Society (2007). (Translated from the 1999 French original).

[72] L. Russo. An approximate zero-one law. *Z. Wahrsch. Verw. Gebiete* 61, 129–139 (1982).

[73] P.-M. Samson. Concentration of measure inequalities for Markov chains and ϕ-mixing processes. *Ann. Probab.* 28, 416–461 (2000).

[74] P.-M. Samson. Concentration inequalities for convex functions on product spaces. *Stochastic inequalities and applications, Progr. Probab.* 56, 33–52. Birkhäuser (2003).

[75] P.-M. Samson. Infimum-convolution description of concentration properties of product probability measures, with applications. *Ann. Inst. H. Poincaré Probab. Statist.* 43, 321–338 (2007).

[76] G. Schechtman. Concentration results and applications. *Handbook of the geometry of Banach spaces* Vol. 2, 1603–1634. North-Holland (2003).

[77] P. Sosoe. Fluctuations in first-passage percolation. *Random growth models*, *Proc. Sympos. Appl. Math.* 75, 69–93. American Mathematical Society (2018)

[78] M. Steele. An Efron-Stein inequality for nonsymmetric statistics. *Ann. Statist.* 14, 753–758 (1986).

[79] M. Steele. *Probability theory and combinatorial optimization.* CBMS-NSF Regional Conference Series in Applied Mathematics 69. SIAM (1997).

[80] M. Talagrand. An isoperimetric theorem on the cube and the Khintchine-Kahane inequalities. *Proc. Amer. Math. Soc.* 104, 905–909 (1988).

[81] M. Talagrand. A new isoperimetric inequality for product measure and the tails of sums of independent random variables. *Geom. Funct. Anal.* 1, 211–223 (1991).

[82] M. Talagrand. On Russo's approximate zero-one law. *Ann. Probab.* 22, 1576–1587 (1994).

[83] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.* 81, 73–205 (1995).

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

[85] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.* 126, 505–563 (1996).

[86] M. Talagrand. A new look at independence. *Ann. Probab.* 24, 1–34 (1996).

[87] T. Tao. *Topics in random matrix theory.* Graduate Studies in Mathematics 132. American Mathematical Society (2012).

[88] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *Compressed sensing*, 210–268. Cambridge University Press (2012).

[89] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*. Cambridge Series in Statistical and Probabilistic Mathematics 47. Cambridge University Press (2018).

[90] C. Villani. *Topics in optimal transportation.* Graduate Studies in Mathematics 58. American Mathematical Society (2003).

[91] C. Villani. *Optimal transport. Old and new.* Grundlehren der mathematischen Wissenschaften 338. Springer (2009).

[92] M. Wainwright. *High-dimensional statistics. A non-asymptotic viewpoint.* Cambridge Series in Statistical and Probabilistic Mathematics 48. Cambridge University Press (2019).

Institut de Mathématiques de Toulouse
Université de Toulouse – Paul-Sabatier, F-31062 Toulouse, France
& Institut Universitaire de France
ledoux@math.univ-toulouse.fr