Exponential Concentration III

SDS 391P.6, Spring 2026

Pratik Patil

These notes are a work in progress and are provided as-is for instructional purposes only. They are not (yet) at the level of a scholarly document. In particular, the notes draw from various sources and do not (yet) have sufficient references to the original sources. Additionally, almost surely the notes have errors and they are only probably approximately correct. The notes will be updated regularly as the course progresses. Last updated: 2026-02-20.

1 Motivation

In the last few lectures, our concentration toolbox has been largely *analytic*: we controlled exponential moments (log-MGFs) using entropy, tensorization, and functional inequalities such as modified log-Sobolev inequalities (MLSIs). This viewpoint is powerful because it is flexible and computational: once we get an entropy bound for $e^{\lambda Z}$, Herbst's argument turns it into a quadratic bound on the centered log-MGF, and Chernoff turns that into Gaussian tails.

There is a second, more *geometric* explanation for why concentration happens, often called the *isoperimetric viewpoint*. Here the primitive object is not an exponential moment, but the growth of sets under small metric enlargements. The guiding principle is the following: if every set of probability at least 1/2 becomes overwhelmingly likely after a small metric blow-up, then every Lipschitz function must concentrate.

This lecture develops this principle in two classical settings:

- Gaussian space: We use the Gaussian isoperimetric inequality (half-spaces minimize blow-ups) to derive the familiar Gaussian concentration for *all* Lipschitz functions $f(g)$ with $g \sim \mathcal{N}(0, I_n)$.
- Product spaces under convexity: We use Talagrand's celebrated convex distance inequality to derive Gaussian-type concentration for convex Lipschitz (more generally quasi-convex Lipschitz) functions of independent bounded coordinates.

A recurring theme is that many of the inequalities we already proved by entropy/MLSI have isoperimetric counterparts. For example, Gaussian MLSI/log-Sobolev and Gaussian isoperimetry both lead to the same dimension-free concentration phenomenon, but from very different perspectives.

2 Isoperimetry as blow-up control

2.1 Metric blow-ups and the concentration function

Let $(\mathcal{X}, d)$ be a metric space and let μ be a probability measure on $\mathcal{X}$. For a measurable set $A \subset \mathcal{X}$ and $t \geqslant 0$, define its t-blow-up (neighborhood)

$$A_t \ := \ \{x \in \mathcal{X} : \ d(x, A) \leqslant t\}, \qquad d(x, A) := \inf_{y \in A} d(x, y).$$

The *concentration function* of the metric probability space is

$$\alpha(t) \;:=\; \sup_{A\subset\mathcal{X}:\ \mu(A)\geqslant 1/2} \mu(A_t^c).$$

Thus $\alpha(t)$ measures the worst-case probability mass that can remain *far* from a set of probability at least 1/2.

Intuition: if $\alpha(t)$ decays rapidly (say $\alpha(t) \lesssim e^{-ct^2}$), then the space has a strong "blow-up" phenomenon: every not-too-small set expands to cover almost all the probability mass after enlarging by distance t.

2.2 From blow-ups to concentration

A function $f : \mathcal{X} \to \mathbb{R}$ is *L-Lipschitz* if

$$|f(x) - f(y)| \leqslant L\, d(x, y) \qquad \text{for all } x, y \in \mathcal{X}.$$

A *median* of $f(X)$ (where $X \sim \mu$) is any number m_f such that $\mu\{f \leqslant m_f\} \geqslant 1/2$ and $\mu\{f \geqslant m_f\} \geqslant 1/2$.

Theorem 2.1 (Lévy's inequality)**.** *Let $X \sim \mu$ and let $f : \mathcal{X} \to \mathbb{R}$ be L-Lipschitz. Then for all $t \geqslant 0$,*

$$\mathbb{P}\{f(X) \geqslant m_f + t\} \leqslant \alpha(t/L), \qquad \mathbb{P}\{f(X) \leqslant m_f - t\} \leqslant \alpha(t/L).$$

Proof. Let $A = \{x : f(x) \leqslant m_f\}$, so $\mu(A) \geqslant 1/2$ by definition of the median. If $x \in A_t$, then there exists $y \in A$ with $d(x, y) \leqslant t$, hence by Lipschitzness

$$f(x) \leqslant f(y) + L\, d(x, y) \leqslant m_f + Lt.$$

Therefore $A_t \subset \{f \leqslant m_f + Lt\}$, so

$$\mathbb{P}\{f(X) > m_f + Lt\} \leqslant \mathbb{P}\{X \notin A_t\} \leqslant \alpha(t).$$

Replace Lt by t to obtain the first inequality. The second inequality follows by applying the first to $-f$. □

Why the median. Isoperimetry naturally talks about sets of measure at least 1/2. Sublevel sets at the median have exactly that property, so medians are the cleanest reference point for isoperimetric concentration. (We have already seen the relation between the median and in mean Homework 1. For sub-Gaussian tails they differ by $O(L)$; see (3) to recall the proof.)

2.3 Converse: concentration determines isoperimetry

The concentration function is not just a convenient bound; it is essentially equivalent to Lipschitz concentration.

Theorem 2.2 (Converse)**.** *If $\beta : \mathbb{R}_+ \to [0, 1]$ is such that for every 1-Lipschitz f we have*

$$\mathbb{P}\{f(X) \geqslant m_f + t\} \leqslant \beta(t) \qquad (t \geqslant 0),$$

then $\alpha(t) \leqslant \beta(t)$ for all $t \geqslant 0$.

Proof. For any set $A \subset \mathcal{X}$, the function $f_A(x) = d(x, A)$ is 1-Lipschitz. If $\mu(A) \geqslant 1/2$, then 0 is a median of $f_A(X)$, hence

$$\mu(A_t^c) = \mathbb{P}\{d(X, A) \geqslant t\} \leqslant \beta(t).$$

Taking the supremum over A with $\mu(A) \geqslant 1/2$ yields $\alpha(t) \leqslant \beta(t)$. □

So, to prove concentration for Lipschitz functions, it suffices to control blow-ups of sets; and to prove an isoperimetric inequality, it suffices to control Lipschitz concentration.

3 Gaussian isoperimetry and Gaussian concentration

3.1 Gaussian isoperimetry

Let γ_n denote the standard Gaussian measure on $\mathbb{R}^n$: if $g \sim \mathcal{N}(0, I_n)$ then $\gamma_n(A) = \mathbb{P}\{g \in A\}$.

The Gaussian isoperimetric theorem (Borell, Sudakov–Tsirelson) identifies the extremal sets for blow-ups under γ_n: half-spaces minimize the Gaussian measure of A_t among sets of fixed measure.

Theorem 3.1 (Gaussian isoperimetric inequality)**.** *For any measurable $A \subset \mathbb{R}^n$ and any $t \geqslant 0$,*

$$\gamma_n(A_t) \;\geqslant\; \Phi\big(\Phi^{-1}(\gamma_n(A)) + t\big),$$

where Φ is the standard normal CDF. Equality holds for half-spaces.

A particularly important special case is $\gamma_n(A) \geqslant 1/2$, where $\Phi^{-1}(\gamma_n(A)) \geqslant 0$. Then

$$\gamma_n(A_t^c) \;\leqslant\; 1 - \Phi(t).$$

In other words, the Gaussian concentration function satisfies

$$\alpha_{\gamma_n}(t) = 1 - \Phi(t). \tag{1}$$

3.2 Concentration of Lipschitz functions of Gaussians

Combine (1) with Lévy's inequality.

Theorem 3.2 (Gaussian concentration around the median)**.** *Let $g \sim \mathcal{N}(0, I_n)$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz with respect to $\|\cdot\|_2$. Let m_f be a median of $f(g)$. Then for all $t \geqslant 0$,*

$$\mathbb{P}\{f(g) \geqslant m_f + t\} \leqslant 1 - \Phi(t/L), \qquad \mathbb{P}\{f(g) \leqslant m_f - t\} \leqslant 1 - \Phi(t/L),$$

and hence

$$\mathbb{P}\{|f(g) - m_f| \geqslant t\} \leqslant 2\big(1 - \Phi(t/L)\big) \leqslant 2\exp\Big(-\frac{t^2}{2L^2}\Big). \tag{2}$$

Proof. By Theorem 2.1 and (1),

$$\mathbb{P}\{f(g) \geqslant m_f + t\} \leqslant \alpha_{\gamma_n}(t/L) = 1 - \Phi(t/L).$$

Apply the same to $-f$ for the lower tail. Finally, use the standard bound $1 - \Phi(u) \leqslant e^{-u^2/2}$ for $u \geqslant 0$. □

Median versus mean. The isoperimetric route naturally yields concentration around the median. Since (2) gives a Gaussian tail, the mean and median are automatically close. For example, integrating the one-sided bound yields

$$\mathbb{E}[(f(g)-m_f)_+] = \int_0^\infty \mathbb{P}\{f(g)-m_f \geqslant t\}\,dt \leqslant \int_0^\infty \big(1-\Phi(t/L)\big)\,dt = \frac{L}{\sqrt{2\pi}},$$

and similarly $\mathbb{E}[(m_f - f(g))_+] \leqslant L/\sqrt{2\pi}$. Thus

$$|\mathbb{E}f(g) - m_f| \leqslant \mathbb{E}|f(g)-m_f| \leqslant \sqrt{\frac{2}{\pi}}\,L. \tag{3}$$

In particular, the median and mean differ by at most a universal constant times L.

Combining (2) and (3), one easily converts concentration around m_f into concentration around $\mathbb{E}f(g)$ (at the cost of changing constants). We will freely use the following standard corollary.

Corollary 3.3 (Gaussian concentration around the mean)**.** *Under the assumptions of Theorem 3.2, there exist absolute constants $c, C > 0$ such that for all $t \geqslant 0$,*

$$\mathbb{P}\{|f(g)-\mathbb{E}f(g)| \geqslant t\} \leqslant 2\exp\Big(-c\,\frac{t^2}{L^2}\Big), \qquad \|f(g)-\mathbb{E}f(g)\|_{\psi_2} \leqslant CL.$$

Connection to the MLSI/log-Sobolev route. In the previous lecture we derived (via Gaussian MLSI + Herbst) the sharper, normalized sub-Gaussian parameter:

$$\log \mathbb{E}e^{\lambda(f(g)-\mathbb{E}f(g))} \leqslant \frac{\lambda^2 L^2}{2} \qquad (\lambda \in \mathbb{R}),$$

which implies (2) (with mean in place of median) with the optimal exponent $t^2/(2L^2)$. The isoperimetric viewpoint explains why such a dimension-free phenomenon is even plausible: in Gaussian space, half-spaces expand as slowly as possible, and everything else expands faster.

4 Convex distance and convex Lipschitz concentration

Gaussian measure has strong concentration for *all* Lipschitz functions. For product measures on $[0,1]^n$, the situation is subtler. From bounded differences (McDiarmid) we know that if changing coordinate i changes f by at most c_i, then

$$\mathbb{P}\{|f(X)-\mathbb{E}f(X)| \geqslant t\} \;\leqslant\; 2\exp\Big(-\frac{t^2}{2\sum_i c_i^2}\Big),$$

which typically scales like $\exp(-t^2/n)$ if $c_i \asymp 1$.

Talagrand discovered that *convexity* can dramatically improve this picture: for convex (or quasi-convex) functions that are Lipschitz with respect to the *Euclidean* metric, one can get Gaussian tails *without* the $1/n$ penalty. The key tool is a refined isoperimetric inequality in product spaces known as the *convex distance inequality*.

4.1 Weighted Hamming distances and Talagrand's convex distance

Let $x, y \in \mathcal{X}^n$. For a weight vector $\alpha \in [0,\infty)^n$, define the weighted Hamming distance

$$d_\alpha(x,y) := \sum_{i=1}^{n} \alpha_i \, \mathbf{1}_{\{x_i \neq y_i\}}, \qquad d_\alpha(x,A) := \inf_{y \in A} d_\alpha(x,y).$$

Talagrand's *convex distance* from x to A is

$$d_T(x,A) \ := \sup_{\alpha \in [0,\infty)^n : \ \|\alpha\|_2 = 1} d_\alpha(x,A).$$

This is a way to measure distance to a set while allowing the coordinate weights to be chosen *adaptively* to the point x (but in a convex/Euclidean manner via the constraint $\|\alpha\|_2 = 1$).

Theorem 4.1 (Talagrand convex distance inequality)**.** *Let $X = (X_1, \ldots, X_n)$ have independent coordinates taking values in a measurable set $\mathcal{X}$. Then for every measurable $A \subset \mathcal{X}^n$ and every $t \geqslant 0$,*

$$\mathbb{P}\{X \in A\} \cdot \mathbb{P}\{d_T(X,A) \geqslant t\} \ \leqslant \ \exp\Big(-\frac{t^2}{4}\Big).$$

Talagrand's inequality can be proved by entropy methods as well (via self-bounding functions), but the statement is fundamentally *isoperimetric*: it controls the measure of blow-ups of sets (in a refined metric) in a product space.

4.2 Convex sets: Euclidean distance is controlled by convex distance

We now specialize to $[0,1]^n$ and relate Euclidean distance to d_T for convex sets.

For $A \subset [0,1]^n$ and $x \in [0,1]^n$, let

$$D(x,A) := \inf_{y \in A} \|x - y\|_2$$

be the Euclidean distance.

Lemma 4.2. *If $A \subset [0,1]^n$ is convex, then for every $x \in [0,1]^n$,*

$$D(x,A) \ \leqslant \ d_T(x,A).$$

Proof sketch. Because A is convex, the Euclidean projection argument implies we may restrict to points in A via barycenters: $D(x,A) = \inf_{\nu \in \mathcal{M}(A)} \|x - \mathbb{E}_\nu Y\|_2$, where ν ranges over probability measures on A and $Y \sim \nu$. Since $x_i, Y_i \in [0,1]$, we have $|x_i - \mathbb{E}_\nu Y_i| \leqslant \mathbb{E}_\nu \mathbf{1}_{\{x_i \neq Y_i\}}$, and therefore by Cauchy–Schwarz,

$$\|x - \mathbb{E}_\nu Y\|_2 \leqslant \sqrt{\sum_{i=1}^{n} \left(\mathbb{E}_\nu \mathbf{1}_{\{x_i \neq Y_i\}}\right)^2} = \sup_{\alpha : \ \|\alpha\|_2 \leqslant 1} \sum_{i=1}^{n} \alpha_i \, \mathbb{E}_\nu \mathbf{1}_{\{x_i \neq Y_i\}}.$$

Taking $\inf_\nu$ on both sides yields $D(x,A) \leqslant d_T(x,A)$. ∎

4.3 Convex (quasi-convex) Lipschitz functions

A function $f : [0,1]^n \to \mathbb{R}$ is *quasi-convex* if all of its sublevel sets are convex: $\{x : f(x) \leqslant s\}$ is convex for every $s \in \mathbb{R}$. (Convex functions are quasi-convex, but quasi-convex is the exact condition we need.)

Theorem 4.3 (Convex Lipschitz concentration via Talagrand)**.** *Let $X = (X_1, \ldots, X_n)$ have independent coordinates taking values in $[0,1]$. Let $f : [0,1]^n \to \mathbb{R}$ be quasi-convex and L-Lipschitz with respect to $\|\cdot\|_2$:*

$$|f(x) - f(y)| \leqslant L\|x - y\|_2 \qquad (x, y \in [0,1]^n).$$

Let m_f be a median of $f(X)$. Then for all $t \geqslant 0$,

$$\mathbb{P}\{f(X) \geqslant m_f + t\} \leqslant 2\exp\Big(-\frac{t^2}{4L^2}\Big), \qquad \mathbb{P}\{f(X) \leqslant m_f - t\} \leqslant 2\exp\Big(-\frac{t^2}{4L^2}\Big).$$

Equivalently, $\|f(X) - m_f\|_{\psi_2} \lesssim L$.

Proof. Fix $s \in \mathbb{R}$ and define the sublevel set $A_s := \{x : f(x) \leqslant s\}$, which is convex by quasi-convexity. For any x, pick $y \in A_s$ with $\|x - y\|_2 \leqslant D(x, A_s) + \varepsilon$. By Lipschitzness,

$$f(x) \leqslant f(y) + L\|x - y\|_2 \leqslant s + L\,D(x, A_s) + L\varepsilon.$$

Letting $\varepsilon \downarrow 0$ and using Lemma 4.2,

$$f(x) \leqslant s + L\,d_T(x, A_s).$$

Therefore, for $t \geqslant 0$,

$$\{f(X) \geqslant s + t\} \subseteq \Big\{d_T(X, A_s) \geqslant \frac{t}{L}\Big\}.$$

Apply Theorem 4.1:

$$\mathbb{P}\{f(X) \leqslant s\} \cdot \mathbb{P}\{f(X) \geqslant s + t\} \leqslant \exp\Big(-\frac{t^2}{4L^2}\Big).$$

Now choose $s = m_f$. Then $\mathbb{P}\{f(X) \leqslant m_f\} \geqslant 1/2$, so

$$\mathbb{P}\{f(X) \geqslant m_f + t\} \leqslant 2\exp\Big(-\frac{t^2}{4L^2}\Big).$$

For the lower tail, apply the same argument to $-f$ (note that $-f$ is quasi-concave, but we can equivalently apply the upper tail bound to f using $s = m_f - t$), yielding the symmetric inequality. □

Comparison with bounded differences (McDiarmid). If f is 1-Lipschitz in Euclidean norm, Theorem 4.3 gives $\exp(-ct^2)$ tails around the median, independent of n. In contrast, McDiarmid with $c_i \asymp 1$ gives $\exp(-ct^2/n)$. The improvement is not magic: it is purchased by *geometry* (Euclidean Lipschitz) and *shape* (quasi-convexity).

Connection to our MLSI-based convex concentration. Earlier we derived a *convex MLSI* for bounded variables, which yields Gaussian-type *upper tail* bounds for convex Lipschitz functions. Talagrand's inequality yields *two-sided* control (around the median) and can be strictly stronger in this convex-Lipschitz regime. At a high level, both are ways of saying that convex functions behave like averages: they do not allow the adversarial oscillations that prevent dimension-free concentration for general Lipschitz functions on product spaces.

5 Summary and look ahead

This lecture presented an isoperimetric route to concentration.

- In a metric probability space, the concentration function

$$\alpha(t) = \sup_{\mu(A) \geqslant 1/2} \mu(A_t^c)$$

encodes worst-case blow-up behavior of sets. Lévy's inequality turns a bound on $\alpha(t)$ into tail bounds for Lipschitz functions (around their median), and a converse shows this relationship is essentially tight.

- For Gaussian measure, the Gaussian isoperimetric theorem identifies half-spaces as minimizers of blow-ups, yielding $\alpha(t) = 1 - \Phi(t)$. Consequently, every L-Lipschitz function of a standard Gaussian vector has Gaussian tails with variance proxy L^2 (dimension-free).

- For product measures on $[0, 1]^n$, Talagrand's convex distance inequality gives Gaussian-type blow-up control in a refined metric. This implies dimension-free Gaussian concentration for quasi-convex Lipschitz functions.

There is a broader picture connecting these ideas: isoperimetry, log-Sobolev/MLSI, and transportation inequalities (e.g. Talagrand T_2) are different "languages" for concentration. We will return to the transportation method later (time permitting), especially since it provides a clean bridge between geometry and exponential-moment control.

Source material

Parts of this lecture are based on references: Boucheron et al. (2013); Vershynin (2018), in addition to the author's accumulated experience working on related topics.

References

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities - A Nonasymptotic Theory of Independence*. Oxford University Press.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.