Poincaré Inequalities

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: 2026-02-02.

1 Motivation

In the previous lecture, we developed a general strategy for variance bounds of the form

$$Z = f(X_1, \dots, X_n), \qquad (X_i) \text{ independent.}$$

The core result was *tensorization of variance*:

$$\mathrm{Var}(Z) \leqslant \sum_{i=1}^{n} \mathbb{E}\big[\mathrm{Var}(Z \mid X^{(i)})\big] = \sum_{i=1}^{n} \mathbb{E}\big[(Z - \mathbb{E}^{(i)} Z)^2\big], \tag{1}$$

where

$$X^{(i)} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n), \qquad \mathbb{E}^{(i)}[\cdot] = \mathbb{E}[\cdot \mid X^{(i)}].$$

Last time, we turned (1) into usable bounds by comparing Z to resampled versions (Efron–Stein–Steele) and by bounding coordinatewise ranges (bounded differences).

In this lecture, we take a different step: we show how to upper bound the conditional variances in (1) using *derivatives* (or “gradients”) of f. This leads to *Poincaré inequalities*, which are variance bounds of the schematic form

$$\mathrm{Var}\big(f(X)\big) \leqslant C\, \mathbb{E}\|\nabla f(X)\|_2^2,$$

where the constant C depends on the distribution of X and is often dimension-free.

2 Convex Poincaré inequality

We begin with a variant of tensorization that will be useful for bounding $\mathrm{Var}(Z \mid X^{(i)})$ by comparing Z to an explicitly chosen proxy depending only on $X^{(i)}$.

2.1 A useful variant of tensorization

Recall the variational identity (from the previous lecture): for any square-integrable random variable W,

$$\mathrm{Var}(W) = \inf_{u \in \mathbb{R}} \mathbb{E}[(W - u)^2],$$

with unique minimizer $u^\star = \mathbb{E}W$. Conditioning on $X^{(i)}$ yields the conditional analogue:

$$\mathrm{Var}(Z \mid X^{(i)}) = \inf_{g:\, g=g(X^{(i)})} \mathbb{E}^{(i)}\big[(Z-g)^2\big],$$

where the infimum is taken over all (measurable) functions of $X^{(i)}$.

In particular, *any* choice of a proxy Z_i that depends only on $X^{(i)}$ yields an upper bound:

$$\mathrm{Var}(Z \mid X^{(i)}) \leqslant \mathbb{E}^{(i)}\big[(Z-Z_i)^2\big] \qquad \text{whenever } Z_i = f_i(X^{(i)}).$$

Plugging this inequality into (1) gives the "guess functions" bound

$$\mathrm{Var}(Z) \leqslant \sum_{i=1}^{n} \mathbb{E}\big[(Z-Z_i)^2\big], \qquad \text{for any choices } Z_i = f_i(X^{(i)}). \tag{2}$$

The point of (2) is that it replaces $\mathbb{E}^{(i)}Z$ by an *explicit* proxy Z_i. The art is to choose Z_i so that

- Z_i depends only on $X^{(i)}$ (so it is an admissible "guess" for Z given $X^{(i)}$), and
- $Z - Z_i$ can be controlled in terms of derivatives or simple geometric bounds.

2.2 Convex Poincaré inequality

For bounded independent coordinates $X_i \in [0,1]$, variance bounds in terms of *discrete* coordinate changes (bounded differences) are always available. To obtain a bound in terms of *derivatives*, we impose an additional structural assumption: separate convexity.

Setting. Let $X = (X_1, \ldots, X_n)$ have independent coordinates with values in $[0,1]$. Let $f : [0,1]^n \to \mathbb{R}$ be *separately convex*, meaning: for each i, the map

$$t \mapsto f(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n)$$

is convex on $[0,1]$ for every fixed choice of the other coordinates. Assume for now that the partial derivatives $\partial_i f$ exist. Write

$$\nabla f(x) = \big(\partial_1 f(x), \ldots, \partial_n f(x)\big), \qquad \|\nabla f(x)\|_2^2 = \sum_{i=1}^{n} (\partial_i f(x))^2.$$

Theorem 2.1 (Convex Poincaré inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables taking values in $[0,1]$. Let $f : [0,1]^n \to \mathbb{R}$ be separately convex, and assume $\partial_i f$ exist. Then*

$$\mathrm{Var}\big(f(X)\big) \leqslant \mathbb{E}\|\nabla f(X)\|_2^2. \tag{3}$$

Proof. We apply the guess-functions bound (2). Fix i, and condition on $X^{(i)}$. Consider the one-dimensional function

$$g_i(t) := f(X_1, \ldots, X_{i-1}, t, X_{i+1}, \ldots, X_n), \qquad t \in [0,1].$$

By separate convexity, g_i is convex on $[0,1]$, hence it attains its minimum on $[0,1]$. Let $X_i^\star = X_i^\star(X^{(i)}) \in [0,1]$ be a minimizer and define the proxy

$$Z_i := g_i(X_i^\star) = \inf_{t\in[0,1]} g_i(t).$$

By construction, Z_i depends only on $X^{(i)}$, so it is admissible in (2). Moreover, $Z = g_i(X_i) \geqslant Z_i$, so $Z - Z_i \geqslant 0$.

Since g_i is convex and differentiable, it lies above its tangent line at $t = X_i$:

$$g_i(s) \ \geqslant \ g_i(X_i) + g_i'(X_i)\,(s - X_i) \qquad \text{for all } s \in [0,1].$$

Apply this inequality with $s = X_i^\star$ and rearrange:

$$Z - Z_i = g_i(X_i) - g_i(X_i^\star) \ \leqslant \ g_i'(X_i)\,(X_i - X_i^\star).$$

Using $Z - Z_i \geqslant 0$ and taking absolute values on the right-hand side,

$$Z - Z_i \ \leqslant \ |g_i'(X_i)|\,|X_i - X_i^\star|.$$

Noting that $g_i'(X_i) = \partial_i f(X)$ and that $|X_i - X_i^\star| \leqslant 1$, we obtain

$$(Z - Z_i)^2 \ \leqslant \ (\partial_i f(X))^2\,(X_i - X_i^\star)^2 \ \leqslant \ (\partial_i f(X))^2.$$

Now sum over i and apply (2):

$$\operatorname{Var}(Z) \ \leqslant \ \sum_{i=1}^n \mathbb{E}[(Z - Z_i)^2] \ \leqslant \ \sum_{i=1}^n \mathbb{E}[(\partial_i f(X))^2] = \mathbb{E}\|\nabla f(X)\|_2^2,$$

which is (3). □

Scaled version. If $X_i \in [a,b]$ almost surely (instead of $[0,1]$), the same argument yields

$$\operatorname{Var}\big(f(X)\big) \ \leqslant \ (b-a)^2\,\mathbb{E}\|\nabla f(X)\|_2^2,$$

because $|X_i - X_i^\star| \leqslant b - a$.

Removing differentiability. The differentiability assumption can be removed by a standard approximation argument: mollify (convolve) f with a smooth kernel to obtain smooth, separately convex functions, apply the inequality to the mollified functions, and pass to the limit.

2.3 Convex Lipschitz functions

A common way to use (3) is through a Lipschitz bound. Assume $f : [0,1]^n \to \mathbb{R}$ is L-Lipschitz with respect to $\|\cdot\|_2$, i.e.

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

Whenever f is differentiable at x, this implies $\|\nabla f(x)\|_2 \leqslant L$. Therefore, for separately convex Lipschitz f,

$$\operatorname{Var}\big(f(X)\big) \ \leqslant \ \mathbb{E}\|\nabla f(X)\|_2^2 \ \leqslant \ L^2.$$

More generally, for $X_i \in [a,b]$, the bound becomes $\operatorname{Var}(f(X)) \leqslant L^2(b-a)^2$.

2.4 Example: the spectral norm of a random matrix

Let $A \in \mathbb{R}^{m\times n}$ have independent entries $A_{ij} \in [0,1]$. Consider the spectral (operator) norm

$$Z = \|A\| = \sup_{\|u\|_2=1} \|Au\|_2.$$

View A as a vector in $\mathbb{R}^{mn}$ equipped with the Euclidean norm $\|\cdot\|_2$, which corresponds to the Frobenius norm $\|\cdot\|_F$ on matrices. Then:

- The map $A \mapsto \|A\|$ is convex because it is a norm.
- It is 1-Lipschitz with respect to $\|\cdot\|_F$:

$$\big|\|A\| - \|B\|\big| \;\leqslant\; \|A-B\| \;\leqslant\; \|A-B\|_F.$$

Therefore the convex Lipschitz corollary applies (with $L=1$ and $b-a=1$), giving the dimension-free bound

$$\mathrm{Var}(\|A\|) \;\leqslant\; 1.$$

This illustrates how a Poincaré-type estimate can be far sharper than a crude range estimate (e.g. $\|A\| \leqslant \|A\|_F \leqslant \sqrt{mn}$ deterministically).

3 Gaussian Poincaré inequality

The convex Poincaré inequality above applies broadly to bounded independent variables, but it requires convexity to replace discrete coordinate changes by derivatives. For Gaussian input, we get a stronger statement: *no convexity is needed.*

Theorem 3.1 (Gaussian Poincaré inequality)**.** *Let $X \sim \mathcal{N}(0, I_n)$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Assume $\mathbb{E}\|\nabla f(X)\|_2^2 < \infty$. Then*

$$\mathrm{Var}(f(X)) \;\leqslant\; \mathbb{E}\|\nabla f(X)\|_2^2. \tag{4}$$

3.1 Gaussian Lipschitz functions

If f is L-Lipschitz on $\mathbb{R}^n$, then $\|\nabla f(x)\|_2 \leqslant L$ for almost every x (Rademacher's theorem). Approximating f by smooth convolutions and applying (4) yields

$$\mathrm{Var}(f(X)) \;\leqslant\; L^2.$$

This is tight: for a linear function $f(x) = a^\top x$,

$$\mathrm{Var}(a^\top X) = \|a\|_2^2 \qquad \text{and} \qquad \|\nabla f(x)\|_2^2 = \|a\|_2^2,$$

so equality holds in (4).

3.2 Proof of Theorem 3.1

The proof uses tensorization twice: first to reduce to dimension one, and then (in dimension one) to compare a Gaussian to a normalized Rademacher sum.

Step 1: reduce to the one-dimensional inequality. Assume the following 1D statement:

(1D Gaussian Poincaré) If $G \sim \mathcal{N}(0,1)$ and $h : \mathbb{R} \to \mathbb{R}$ is continuously differentiable with $\mathbb{E}[h'(G)^2] < \infty$, then

$$\mathrm{Var}(h(G)) \leqslant \mathbb{E}[h'(G)^2]. \tag{5}$$

Now let $X = (X_1, \ldots, X_n) \sim \mathcal{N}(0, I_n)$ and write $Z = f(X)$. Apply tensorization (1):

$$\mathrm{Var}(Z) \leqslant \sum_{i=1}^{n} \mathbb{E}\big[\mathrm{Var}(Z \mid X^{(i)})\big].$$

Fix i and condition on $X^{(i)}$. For fixed $X^{(i)}$, define the one-dimensional function

$$h_i(t) := f(X_1, \ldots, X_{i-1}, t, X_{i+1}, \ldots, X_n).$$

Conditional on $X^{(i)}$, the variable $X_i \sim \mathcal{N}(0,1)$ is independent of $X^{(i)}$, so we may apply (5) to $h_i(X_i)$:

$$\mathrm{Var}(Z \mid X^{(i)}) = \mathrm{Var}(h_i(X_i) \mid X^{(i)}) \leqslant \mathbb{E}\big[h_i'(X_i)^2 \mid X^{(i)}\big].$$

But $h_i'(X_i) = \partial_i f(X)$, hence

$$\mathrm{Var}(Z \mid X^{(i)}) \leqslant \mathbb{E}\big[(\partial_i f(X))^2 \mid X^{(i)}\big].$$

Taking expectations and summing over i yields

$$\mathrm{Var}(f(X)) \leqslant \sum_{i=1}^{n} \mathbb{E}[(\partial_i f(X))^2] = \mathbb{E}\|\nabla f(X)\|_2^2,$$

which is (4). Thus it remains to prove the one-dimensional inequality (5).

Step 2: prove the 1D inequality via a Rademacher approximation. Let $h : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable with compact support, and set

$$K := \sup_{x \in \mathbb{R}} |h''(x)| < \infty.$$

Let $\varepsilon_1, \ldots, \varepsilon_m$ be i.i.d. Rademacher random variables (equiprobable on $\{-1, +1\}$) and define

$$S_m := \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \varepsilon_j.$$

By the central limit theorem, $S_m \Rightarrow G$ where $G \sim \mathcal{N}(0,1)$.

Apply tensorization of variance to $h(S_m)$ as a function of the independent inputs $(\varepsilon_j)_{j=1}^m$:

$$\mathrm{Var}(h(S_m)) \leqslant \sum_{j=1}^{m} \mathbb{E}\big[\mathrm{Var}(h(S_m) \mid \varepsilon^{(j)})\big],$$

where $\varepsilon^{(j)} = (\varepsilon_1, \ldots, \varepsilon_{j-1}, \varepsilon_{j+1}, \ldots, \varepsilon_m)$. Condition on $\varepsilon^{(j)}$ and write

$$S_m = T + \frac{1}{\sqrt{m}} \varepsilon_j, \qquad T = \frac{1}{\sqrt{m}} \sum_{k \neq j} \varepsilon_k.$$

Under this conditioning, ε_j is equiprobable on $\{-1, +1\}$, so

$$\mathrm{Var}(h(S_m) \mid \varepsilon^{(j)}) = \frac{1}{4}\Big(h(T + \tfrac{1}{\sqrt{m}}) - h(T - \tfrac{1}{\sqrt{m}})\Big)^2.$$

Since S_m is one of the points $T \pm \frac{1}{\sqrt{m}}$, the other point differs from S_m by $\frac{2}{\sqrt{m}}$. Taylor's theorem with remainder yields

$$\Big|h(T + \tfrac{1}{\sqrt{m}}) - h(T - \tfrac{1}{\sqrt{m}})\Big| \leqslant \frac{2}{\sqrt{m}}\,|h'(S_m)| + \frac{2K}{m}.$$

Therefore,

$$\mathrm{Var}(h(S_m) \mid \varepsilon^{(j)}) \leqslant \frac{1}{4}\Big(\frac{2}{\sqrt{m}}\,|h'(S_m)| + \frac{2K}{m}\Big)^2 = \frac{1}{m}\Big(|h'(S_m)| + \frac{K}{\sqrt{m}}\Big)^2.$$

Summing over $j = 1, \ldots, m$ yields

$$\mathrm{Var}(h(S_m)) \leqslant \mathbb{E}\Big[\Big(|h'(S_m)| + \frac{K}{\sqrt{m}}\Big)^2\Big]. \tag{6}$$

Now let $m \to \infty$. Because h and h' are continuous and bounded (compact support), the CLT implies

$$\mathbb{E}[h(S_m)] \to \mathbb{E}[h(G)], \quad \mathbb{E}[h(S_m)^2] \to \mathbb{E}[h(G)^2], \quad \mathbb{E}[h'(S_m)^2] \to \mathbb{E}[h'(G)^2],$$

and similarly $\mathbb{E}|h'(S_m)| \to \mathbb{E}|h'(G)|$. Thus $\mathrm{Var}(h(S_m)) \to \mathrm{Var}(h(G))$ and the right-hand side of (6) converges to $\mathbb{E}[h'(G)^2]$. Taking limits in (6) gives

$$\mathrm{Var}(h(G)) \leqslant \mathbb{E}[h'(G)^2],$$

which proves (5) for smooth, compactly supported h.

Finally, one extends (5) to general continuously differentiable h with $\mathbb{E}[h'(G)^2] < \infty$ by a standard approximation scheme: truncate h outside $[-M, M]$ and smooth the truncations by convolution, then pass $M \to \infty$ using dominated convergence.

This completes the proof of Theorem 3.1.

4 Look ahead

In the last lecture, we controlled variance using *discrete* coordinate perturbations (resampling and bounded differences). In this lecture, we saw that in certain structured settings—bounded inputs with convexity, and Gaussian inputs without any convexity—the same variance-control philosophy can be expressed in terms of *ordinary derivatives*:

$$\mathrm{Var}(f(X)) \lesssim \mathbb{E}\|\nabla f(X)\|_2^2.$$

This connects concentration to functional inequalities and geometric regularity of f.

In upcoming lectures, we will move beyond variance bounds and develop sharper results that control not only the *scale* of fluctuations but also their *tails* (often sub-Gaussian or sub-exponential), using more refined arguments based on exponential moments.

Source material

Parts of this lecture are based on references: Boucheron et al. (2013); Raginsky and Sason (2013); Tropp (2023); van Handel (2016), 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.

Raginsky, M. and Sason, I. (2013). Concentration of measure inequalities in information theory, communications, and coding. *Foundations and Trends in Communications and Information Theory*, 10(1-2):1–247.

Tropp, J. A. (2023). Probability in high dimensions. Caltech CMS Lecture Notes 2021-01.

van Handel, R. (2016). Probability in high dimension. Lecture Notes (Princeton University). APC 550.