Ramon van Handel's Remarks on the Discrete Cube

Gregory Rosenthal

October 13, 2020

Abstract

We transcribe a series of four lectures by Ramon van Handel titled "Remarks on the Discrete Cube" [8], the content of which is summarized below.

Lecture 1

The "discrete cube" is the set $\{-1,1\}^n$. We will consider the following classes of functions on the cube, in increasing order of generality:

$$\begin{aligned} &f:\{-1,1\}^n \to \{0,1\} && \text{(boolean functions)},\\ &f:\{-1,1\}^n \to \mathbb{R} && \text{(real-valued functions)},\\ &f:\{-1,1\}^n \to (X,\|\cdot\|_X) && \text{(vector-valued functions)}, \end{aligned}$$

where X is an arbitrary Banach space with norm $\|\cdot\|_X$.

A fundamental fact about real-valued functions on the cube is the Poincaré inequality, which will be stated shortly. In these lectures we will do the following:

1. Prove L^p analogues of the Poincaré inequality for vector-valued functions on the cube. This result is due to Ivanisvili, van Handel and Volberg [10], as is the proof given here.

2. Prove a certain strengthening of the Poincaré inequality for boolean functions on the cube. This result is due to Eldan and Gross [4], and generalizes previous results of Kahn, Kalai and Linial [11] and Talagrand [17]. The proof given by Eldan and Gross uses stochastic calculus, but here we present a new simplification of their proof which uses techniques of Ivanisvili, van Handel and Volberg in place of stochastic calculus.

Real-valued functions on the cube are commonly analyzed using (discrete) Fourier analysis [14, 7], and the Poincaré inequality is easy to prove in this way. In contrast, except for a single application of hypercontractivity near the end, in these lectures we will use only elementary probability and calculus, and in particular no Fourier analysis.

For $f : \{-1,1\}^n \to (X, \|\cdot\|_X)$ let $\mathbf{E}f = 2^{-n}\sum_{\varepsilon\in\{-1,1\}^n} f(\varepsilon)$ denote the expectation of f under the uniform distribution, and if f is real-valued then let $\operatorname{Var} f = \mathbf{E}f^2 - (\mathbf{E}f)^2$ denote the variance of f under the uniform distribution. (We also use $\mathbf{E}$ to denote expected value more generally.) For $1 \le i \le n$ define the i'th "discrete derivative" of a function $f : \{-1,1\}^n \to (X, \|\cdot\|_X)$ as follows: for all $\varepsilon = (\varepsilon_1,\dots,\varepsilon_n) \in \{-1,1\}^n$,

$$D_i f(\varepsilon) = \frac{f(\varepsilon) - f(\varepsilon_1,\dots,\varepsilon_{i-1},-\varepsilon_i,\varepsilon_{i+1},\dots,\varepsilon_n)}{2}.$$

Theorem 1 (Poincaré inequality)**.** *Let* $f : \{-1,1\}^n \to \mathbb{R}$*; then* $\operatorname{Var} f \le \mathbf{E}\sum_{i=1}^n (D_i f)^2$.

Let $Df = (D_1 f, \dots, D_n f)$ and let $\|\cdot\|$ denote the Euclidean norm. Then we may also write the Poincaré inequality as $\operatorname{Var} f \le \mathbf{E}\|Df\|^2$, so one interpretation of the Poincaré inequality is that "Lipschitz" functions have constant variance.

If f takes values in $\{-1,1\}$ then $(D_i f(\varepsilon))^2 = \mathbb{1}_{f(\varepsilon)\neq f(\varepsilon_1,\dots,-\varepsilon_i,\dots,\varepsilon_n)}$. Therefore another interpretation of the Poincaré inequality is that if f represents a voting rule in a two-candidate election, and if votes are independent and uniform random, then on average there are at least $\operatorname{Var} f$ voters i such that flipping only the i'th vote would change the outcome of the election. If both candidates have probability $1/2$ of winning the election then $\operatorname{Var} f = 1$, in which case at least one voter has probability at least $1/n$ of casting a decisive vote. The previously mentioned result of Kahn, Kalai and Linial [11] improves this $1/n$ lower bound to $\Omega\left(\frac{\log n}{n}\right)$.

We begin by proving the Poincaré inequality, in a manner which is much less efficient than the Fourier-analytic proof but which introduces machinery used to prove the main results of these lectures. Suppose we have a smooth function $\varphi : [0,\infty) \to \mathbb{R}$ such that $\varphi(0) = \mathbf{E}f^2$ and $\varphi(\infty) \coloneqq \lim_{t\to\infty}\varphi(t) = (\mathbf{E}f)^2$. Then,

$$\operatorname{Var} f = \mathbf{E}f^2 - (\mathbf{E}f)^2 = \varphi(0) - \varphi(\infty) = -\int_0^\infty \frac{d\varphi(t)}{dt}dt,$$

so it suffices to bound $d\varphi(t)/dt$.

For $t \ge 0$ let $\xi(t) = (\xi_1(t),\dots,\xi_n(t)) \in \{-1,1\}^n$ be a random variable where each $\xi_i(t)$ is independently 1 with probability $\frac{1+e^{-t}}{2}$ and -1 with probability $\frac{1-e^{-t}}{2}$, i.e. $\mathbf{E}_\xi \xi_i(t) = e^{-t}$. For $\varepsilon \in \{-1,1\}^n$ let $P_t f(\varepsilon) = \mathbf{E}_\xi f(\varepsilon\xi(t))$ where $\varepsilon\xi(t) \coloneqq (\varepsilon_1\xi_1(t),\dots,\varepsilon_n\xi_n(t))$. Then $P_0 f = f$ and $P_\infty f = \mathbf{E}f$, so we may define $\varphi(t) \coloneqq \mathbf{E}[(P_t f)^2]$, implying that

$$\operatorname{Var} f = -\int_0^\infty \frac{d}{dt}\mathbf{E}[(P_t f)^2]dt = -2\int_0^\infty \mathbf{E}\left[P_t f \frac{d}{dt} P_t f\right] dt.$$

Remark. For intuition's sake, we now give an equivalent definition of $P_t f$ in terms of the following continuous-time random walk $Y(t) = (Y_1(t),\dots,Y_n(t))$ on the cube, where $t \ge 0$ represents time. To each coordinate from 1 to n, assign a "clock" which "ticks" at times determined by a rate-1 Poisson process,[1] independently of the other $n-1$ clocks. Whenever the i'th clock ticks, resample Y_i uniformly at random. If $Y(0)=\varepsilon$ then $Y(t)$ is distributed identically to $\varepsilon\xi(t)$, because the i'th clock ticks before time t with probability $1-e^{-t}$, and because $Y_i(t)$ equals ε_i before the i'th clock's first tick and is uniform random after the i'th clock's first tick. Therefore $P_tf(\varepsilon)=\mathbf{E}[f(Y(t)) \mid Y(0)=\varepsilon]$.

It is easy to verify that $D_i^2=D_i$ and $D_iD_j=D_jD_i$. Let $\Delta=-\sum_{i=1}^n D_i$. In the next lecture we will prove the following:

Lemma 2. *For all* $f:\{-1,1\}^n \to (X, \|\cdot\|_X)$*,*

0. $\mathbf{E}P_tf=\mathbf{E}f$*,*

1. $\frac{d}{dt}P_tf=\Delta P_tf$*,*

2. $D_iP_tf=P_tD_if$*,*

and for all $f,g:\{-1,1\}^n\to\mathbb{R}$*,*

3. $\mathbf{E}[f\Delta g]=-\sum_{i=1}^n \mathbf{E}[D_if\cdot D_ig]$*,*

4. $(D_iP_tf)^2\le e^{-2t}P_t(D_if)^2$ *pointwise.*

Remark. The case of Items 0 to 2 where $X=\mathbb{R}$ is sufficient for our proof of the Poincaré inequality, and can be proved perhaps more easily using Fourier analysis,[2] but we will use the generalization to arbitrary Banach spaces later in these lectures.

Item 1 is called the heat equation. The transformation Δ is called the Laplacian because it equals $-\sum_{i=1}^n D_i^2$, analogous to the standard calculus definition of the Laplacian. Item 3 is analogous to integration by parts, since $\Delta=-\sum_{i=1}^n D_i^2$.

Proof of the Poincaré inequality. By Lemma 2,

$$\begin{aligned}
\operatorname{Var} f &= -2\int_0^\infty \mathbf{E}\left[P_tf\frac{d}{dt}P_tf\right]dt && \text{(proved above)}\\
&= -2\int_0^\infty \mathbf{E}[P_tf\Delta P_tf]dt && \text{(Item 1)}\\
&= 2\int_0^\infty \sum_i \mathbf{E}[(D_iP_tf)^2]dt && \text{(Item 3)}\\
&\le 2\int_0^\infty \sum_i \mathbf{E}[e^{-2t}P_t(D_if)^2]dt && \text{(Item 4)}
\end{aligned}$$

[1] I.e. the times between ticks of any given clock are independent rate-1 exponential random variables.

[2] For $S\subseteq\{1,\dots,n\}$ let $\chi_S(\varepsilon)=\prod_{i\in S}\varepsilon_i$. Then $D_i\chi_S=\mathbb{1}_{i\in S}\chi_S$, and $\Delta\chi_S=-|S|\chi_S$, and $P_t\chi_S=e^{-t|S|}\chi_S$, from which Items 0 to 3 follow easily using the Fourier expansion of f.

$$= \int_0^\infty 2e^{-2t} \sum_i \mathbf{E}[(D_i f)^2] dt \qquad \text{(Item 0)}$$

$$= \sum_i \mathbf{E}[(D_i f)^2]. \qquad \square$$

Lecture 2

We now prove Lemma 2:

Proof of Item 0. For any fixed $\xi \in \{-1,1\}^n$, if ε is uniform random on $\{-1,1\}^n$ then so is $\varepsilon\xi$, so $\mathbf{E}P_t f = \mathbf{E}_{\xi,\varepsilon} f(\varepsilon\xi(t)) = \mathbf{E}_\xi \mathbf{E} f = \mathbf{E} f$. $\square$

Proof of Item 1. By applying the definition of $P_t f$ and again substituting $\varepsilon\xi$ for ξ,

$$P_t f(\varepsilon) = \sum_{\xi \in \{-1,1\}^n} \prod_{j=1}^n \frac{1 + \xi_j e^{-t}}{2} f(\varepsilon\xi) = \sum_{\xi \in \{-1,1\}^n} \prod_{j=1}^n \frac{1 + \varepsilon_j \xi_j e^{-t}}{2} f(\xi),$$

so by the product rule,

$$\frac{d}{dt} P_t f(\varepsilon) = -\sum_{i=1}^n \sum_{\xi \in \{-1,1\}^n} \frac{\varepsilon_i \xi_i e^{-t}}{2} \prod_{j \neq i} \frac{1 + \varepsilon_j \xi_j e^{-t}}{2} f(\xi) = -\sum_{i=1}^n D_i P_t f(\varepsilon). \qquad \square$$

Proof of Item 2. Let $e_i \in \{-1,1\}^n$ have a -1 in position i and 1s elsewhere, i.e. $D_i f(\varepsilon) = \frac{f(\varepsilon) - f(\varepsilon e_i)}{2}$. Then,

$$D_i P_t f(\varepsilon) = \frac{P_t f(\varepsilon) - P_t f(\varepsilon e_i)}{2} = \mathbf{E}_\xi \frac{f(\varepsilon\xi(t)) - f(\varepsilon e_i \xi(t))}{2} = \mathbf{E}_\xi D_i f(\varepsilon\xi(t)) = P_t D_i f(\varepsilon). \quad \square$$

Remark. When f is real-valued, the following is an alternate proof of Item 2. Interpret P_t and Δ as $2^n \times 2^n$ real matrices, acting on the space of functions from $\{-1,1\}^n$ to $\mathbb{R}$. We just proved that $\frac{d}{dt} P_t = \Delta P_t$, and since P_0 is the identity it follows that $P_t = e^{t\Delta}$. Since $D_1, \ldots, D_n$ commute it then follows that $P_t = \prod_{i=1}^n e^{-tD_i}$, so P_t and D_i commute.

Proof of Item 3. Define e_i as in the proof of Item 2. If ε is uniform random on $\{-1,1\}^n$ then so is εe_i, and clearly $D_i g(\varepsilon)$ is antisymmetric in ε_i, so

$$\mathbf{E}[f \Delta g] = -\sum_{i=1}^n \mathbf{E}_\varepsilon[f(\varepsilon e_i) D_i g(\varepsilon e_i)] = \sum_{i=1}^n \mathbf{E}_\varepsilon[f(\varepsilon e_i) D_i g(\varepsilon)].$$

Therefore,

$$\mathbf{E}[f \Delta g] = \frac{\mathbf{E}[f \Delta g] + \mathbf{E}[f \Delta g]}{2} = \sum_{i=1}^n \mathbf{E}_\varepsilon \left[\frac{f(\varepsilon e_i) - f(\varepsilon)}{2} D_i g(\varepsilon) \right] = -\sum_{i=1}^n \mathbf{E}_\varepsilon[D_i f(\varepsilon) D_i g(\varepsilon)].$$

$\square$

Proof of Item 4. The value $\varepsilon_i D_i f(\varepsilon)$ does not depend on ε_i, because

$$\varepsilon_i D_i f(\varepsilon) = \frac{f(\varepsilon_1, \ldots, \varepsilon_{i-1}, 1, \varepsilon_{i+1}, \ldots, \varepsilon_n) - f(\varepsilon_1, \ldots, \varepsilon_{i-1}, -1, \varepsilon_{i+1}, \ldots, \varepsilon_n)}{2}.$$

Therefore, for all $\varepsilon \in \{-1,1\}^n$,

$$\begin{aligned} D_i P_t f(\varepsilon) = P_t D_i f(\varepsilon) &= \mathbf{E}_\xi D_i f(\varepsilon\xi(t)) = \mathbf{E}_\xi[\varepsilon_i \xi_i(t) \cdot \varepsilon_i \xi_i(t) D_i f(\varepsilon\xi(t))] \\ &= \mathbf{E}_\xi[\varepsilon_i \xi_i(t)] \cdot \mathbf{E}_\xi[\varepsilon_i \xi_i(t) D_i f(\varepsilon\xi(t))] = e^{-t}\mathbf{E}_\xi[\xi_i(t) D_i f(\varepsilon\xi(t))], \end{aligned}$$

so by Jensen's inequality,

$$(D_i P_t f(\varepsilon))^2 \le e^{-2t}\mathbf{E}_\xi[(D_i f(\varepsilon\xi(t)))^2] = P_t(D_i f)^2(\varepsilon). \qquad \square$$

Finally, we remark that the Poincaré inequality is sharp for linear functions: if $a_1, \ldots, a_n \in \mathbb{R}$ and $f(\varepsilon) = \sum_{i=1}^n a_i \varepsilon_i$ then $\operatorname{Var} f = \sum_{i=1}^n a_i^2 = \sum_{i=1}^n (D_i f)^2$.

We now consider analogues of the Poincaré inequality for vector-valued functions on the cube, i.e. for $f : \{-1,1\}^n \to (X, \|\cdot\|_X)$. A natural hypothesis is that

$$\mathbf{E}\|f - \mathbf{E}f\|_X^2 \overset{?}{\le} C \sum_{i=1}^n \mathbf{E}\|D_i f\|_X^2, \tag{1}$$

where each occurrence of C throughout these lectures represents a distinct positive universal constant. For example, the Poincaré inequality says that Eq. (1) holds when $(X, \|\cdot\|_X) = (\mathbb{R}, |\cdot|)$. When specialized to linear functions, Eq. (1) would imply that $\mathbf{E}_\varepsilon \|\sum_{i=1}^n \varepsilon_i x_i\|_X^2 \le C \sum_{i=1}^n \|x_i\|_X^2$ for all $x_1, \ldots, x_n \in X$. However, if $(X, \|\cdot\|_X) = (\mathbb{R}^n, \|\cdot\|_1)$ and x_i is the i'th standard basis vector, then $\|\sum_{i=1}^n \varepsilon_i x_i\|_1^2 = \|\varepsilon\|_1^2 = n^2$ and $\sum_{i=1}^n \|x_i\|_1^2 = n$, so Eq. (1) is not universally true.

This raises the following question: can we prove an analogue of the Poincaré inequality for arbitrary functions from $\{-1,1\}^n$ to $(X, \|\cdot\|_X)$ in terms of the behavior of linear functions from $\{-1,1\}^n$ to $(X, \|\cdot\|_X)$? More concretely, for $p \ge 1$ let $(X, \|\cdot\|_X)$ have *type* p if $\mathbf{E}_\varepsilon \|\sum_{i=1}^n \varepsilon_i x_i\|_X^p \le C \sum_{i=1}^n \|x_i\|_X^p$ for all $x_1, \ldots, x_n \in X$. For example, every space has type 1 by the triangle inequality, Hilbert spaces have type 2, and no space has type greater than 2 due to the case where $x_1 = \cdots = x_n \ne 0$. One may also verify that every space with type q has type p for $p \le q$,[3] and that if $p \le 2$ then $(\mathbb{R}^n, \|\cdot\|_p)$ has type p.[4]

Theorem 3 (Conjectured by Enflo [5], proved by Ivanisvili, van Handel and Volberg [10])**.** *If $(X, \|\cdot\|_X)$ has type p, then $\mathbf{E}\|f - \mathbf{E}f\|_X^p \le C \sum_{i=1}^n \mathbf{E}\|D_i f\|_X^p$ for all $f : \{-1,1\}^n \to X$.*

[3] Take $1/q$'th powers on both sides of the definition of type q, and apply the monotonicity in p of L^p and ℓ^p norms.

[4] Reduce to the one-dimensional case, and apply the Khintchine inequality. To see that $(\mathbb{R}^n, \|\cdot\|_p)$ does not have type greater than p when n is large, let x_i be the i'th standard basis vector.

For example, Enflo's conjecture trivially holds for linear functions, and implies that Eq. (1) holds for spaces of type 2.

When trying to prove a conjecture about functions on the cube, one approach is to first prove a similar statement for functions with n independent standard Gaussian inputs, and then modify the proof to hold for functions on the cube. For a function f on $\mathbb{R}^n$ let $\mathbf{E}f$ denote the expectation of f under this input distribution, and let $\partial_i f$ denote the i'th partial derivative of f.

Theorem 4 (Pisier [15, Theorem 2.2])**.** *Let $f : \mathbb{R}^n \to (X, \|\cdot\|_X)$ be a "sufficiently smooth" function such that $\mathbf{E}f$ exists. Let $g = (g_1, \ldots, g_n), g' = (g'_1, \ldots, g'_n)$ where $g_1, \ldots, g_n, g'_1, \ldots, g'_n$ are independent standard Gaussians. Then for all $p \geq 1$,*

$$\mathbf{E}\,\|f - \mathbf{E}f\|_X^p \leq \left(\frac{\pi}{2}\right)^p \mathbf{E} \left\| \sum_{i=1}^n g'_i \partial_i f(g) \right\|_X^p .$$

Note that Pisier's inequality does not require $(X, \|\cdot\|_X)$ to have type p. If $(X, \|\cdot\|_X)$ has "Gaussian type p", i.e. if $\mathbf{E}\|\sum_{i=1}^n g_i x_i\|_X^p \leq C \sum_{i=1}^n \|x_i\|_X^p$ for all $x_1, \ldots, x_n \in X$, then conditioning on g in Pisier's inequality gives $\mathbf{E}\|f - \mathbf{E}f\|_X^p \leq C\left(\frac{\pi}{2}\right)^p \sum_{i=1}^n \mathbf{E}\|\partial_i f\|_X^p$, which is a Gaussian analogue of Enflo's conjecture. Pisier's inequality is tight for linear functions up to a factor of $\left(\frac{\pi}{2}\right)^p$, and even the factor $\left(\frac{\pi}{2}\right)^p$ is sharp in certain cases as well.[5] Before proving Pisier's inequality we apply it to two more examples:

Example. Let $(X, \|\cdot\|_X) = (\mathbb{R}, |\cdot|)$. If g is fixed then $\sum_{i=1}^n g'_i \partial_i f(g)$ is Gaussian with mean 0 and variance $\sum_{i=1}^n (\partial_i f(g))^2 = \|\nabla f(g)\|^2$, or equivalently $\|\nabla f(g)\|$ times a standard Gaussian, so $\mathbf{E}|f - \mathbf{E}f|^p \leq \left(\frac{\pi}{2}\right)^p \mathbf{E}|Z|^p \cdot \mathbf{E}\|\nabla f\|^p$ where Z denotes a standard Gaussian. For example, since $\mathbf{E}|Z| = \sqrt{2/\pi}$, taking $p = 1$ gives $\mathbf{E}|f - \mathbf{E}f| \leq \sqrt{\pi/2} \cdot \mathbf{E}\|\nabla f\|$.

Example. The noncommutative Khintchine inequality [9] states that $\mathbf{E}\,\|\sum_{i=1}^n g_i A_i\|_{op} \leq O\left(\sqrt{\log d}\right) \|\sum_{i=1}^n A_i^2\|_{op}^{1/2}$ for $A_1, \ldots, A_n \in \mathbb{R}^{d\times d}_{\text{sym}}$, where "op" and "sym" are short for "operator norm" and "symmetric" respectively. Therefore, for all "nice" $f : \mathbb{R}^n \to \mathbb{R}^{d\times d}_{\text{sym}}$,

$$\mathbf{E}\|f - \mathbf{E}f\|_{op} \leq \frac{\pi}{2}\mathbf{E} \left\| \sum_{i=1}^n g'_i \partial_i f(g) \right\|_{op} \leq O\left(\sqrt{\log d}\right) \cdot \mathbf{E} \left\| \sum_{i=1}^n (\partial_i f(g))^2 \right\|_{op}^{1/2} .$$

Proof of Pisier's inequality. Let $g(\theta) = g\cos\theta + g'\sin\theta$ and $g'(\theta) = dg(\theta)/d\theta = -g\sin\theta + g'\cos\theta$. By the fundamental theorem of calculus and the chain rule,

$$f(g') - f(g) = \int_0^{\pi/2} \frac{d}{d\theta} f(g(\theta))d\theta = \int_0^{\pi/2} \sum_{i=1}^n \partial_i f(g(\theta)) g'_i(\theta) d\theta,$$

[5] E.g. let $(X, \|\cdot\|_X) = (\mathbb{R}, |\cdot|), p = 1, n = 1, f(x) = \max(\min(Kx, 1), 0)$ and let $K \to \infty$.

so by the triangle inequality and Jensen's inequality,

$$\|f(g') - f(g)\|_X^p \le \left(\frac{2}{\pi} \int_0^{\pi/2} \frac{\pi}{2} \left\| \sum_{i=1}^n \partial_i f(g(\theta)) g_i'(\theta) \right\|_X d\theta \right)^p$$
$$\le \frac{2}{\pi} \int_0^{\pi/2} \left(\frac{\pi}{2} \left\| \sum_{i=1}^n \partial_i f(g(\theta)) g_i'(\theta) \right\|_X \right)^p d\theta.$$

The pairs $(g(\theta), g'(\theta))$ and (g, g') are identically distributed, because

$$\begin{pmatrix} g(\theta) \\ g'(\theta) \end{pmatrix} = \begin{pmatrix} \cos\theta \cdot I & \sin\theta \cdot I \\ -\sin\theta \cdot I & \cos\theta \cdot I \end{pmatrix} \begin{pmatrix} g \\ g' \end{pmatrix}$$

and the standard multivariate Gaussian distribution is invariant under orthogonal transformations. Therefore,

$$\mathbf{E}\|f(g') - f(g)\|_X^p \le \left(\frac{\pi}{2}\right)^p \mathbf{E} \left\| \sum_{i=1}^n \partial_i f(g) g_i' \right\|_X^p.$$

Finally, by the triangle inequality and Jensen's inequality,

$$\mathbf{E}_{g'} \left\| f(g') - \mathbf{E}_g f(g) \right\|_X^p \le \mathbf{E}_{g'} \left(\mathbf{E}_g \left\| f(g') - f(g) \right\|_X \right)^p \le \mathbf{E}_{g'} \mathbf{E}_g \left\| f(g') - f(g) \right\|_X^p. \qquad \square$$

Lecture 3

Recall that we want an analogue of Pisier's inequality for functions $f : \{-1,1\}^n \to (X, \|\cdot\|_X)$. Fix some $p \ge 1$. A natural hypothesis is that if $\varepsilon, \delta \in \{-1,1\}^n$ are independent and uniform random then

$$\mathbf{E}\|f - \mathbf{E}f\|_X^p \overset{?}{\le} C\mathbf{E} \left\| \sum_{i=1}^n \delta_i D_i f(\varepsilon) \right\|_X^p, \tag{2}$$

from which Enflo's conjecture (Theorem 3) would follow by conditioning on ε and applying the definition of type p. An obstacle to mimicking the proof of Pisier's inequality is that the cube $\{-1,1\}^n$ is not rotationally invariant in continuous space. Encouragingly, Pisier [15] proved that Eq. (2) holds for some $C \le O(\log^p n)$, despite this obstacle. However, Talagrand [17, Section 6] proved that when $(X, \|\cdot\|_X) = (\mathbb{R}^{2^n}, \|\cdot\|_\infty)$ there exists f such that Eq. (2) holds *only* when $C \ge \Omega(\log^p n)$.[6]

[6] See the original lecture [8, Lecture 3, 9:00–10:30 and 40:45–42:20] for a description of Banach spaces for which Eq. (2) holds with $C = \Theta(1)$.

Remark. Efraim and Lust-Piquard [3] used ideas from quantum information to adopt Pisier's proof to real-valued functions on the cube, by associating each $\{-1,1\}$-valued coordinate of the cube with a measurement of a σ_x or σ_z observable, and rotating continuously between these noncommuting observables. However, this approach does not seem to generalize to vector-valued functions.

Therefore we formulate a different analogue of Pisier's inequality for functions on the cube. Recall from Lecture 1 that $\xi_1(t), \ldots, \xi_n(t) \in \{-1,1\}$ are i.i.d. random variables with distribution $\mathbb{P}(\xi_i(t) = \pm 1) = \frac{1 \pm e^{-t}}{2}$, and that $P_t f(\varepsilon) = \mathbf{E}_\xi f(\varepsilon \xi(t))$. Also let

$$\delta_i(t) = \frac{\xi_i(t) - \mathbf{E}\xi_i(t)}{\operatorname{Var}^{1/2} \xi_i(t)} = \frac{\xi_i(t) - e^{-t}}{\sqrt{1 - e^{-2t}}}.$$

Let $\mu(dt) = \frac{2}{\pi} \frac{e^{-t}}{\sqrt{1-e^{-2t}}} dt$, and note that μ is a probability measure on $[0, \infty)$ because

$$\int_0^\infty \sqrt{\frac{e^{-2t}}{1 - e^{-2t}}} \cdot dt = \int_1^0 \sqrt{\frac{u}{1-u}} \cdot \frac{du}{-2u} = \int_0^1 \frac{du}{2\sqrt{u(1-u)}} = -\arcsin(\sqrt{1-u})\big|_0^1 = \frac{\pi}{2}.$$

Theorem 5 (Ivanisvili, van Handel and Volberg [10])**.** *For all $f : \{-1,1\}^n \to (X, \|\cdot\|_X)$ and $p \geq 1$,*

$$\left(\mathbf{E}\|f - \mathbf{E}f\|_X^p\right)^{1/p} \leq \frac{\pi}{2} \int \left(\mathbf{E} \left\| \sum_{i=1}^n \delta_i(t) D_i f(\varepsilon) \right\|_X^p \right)^{1/p} \mu(dt),$$

where $\varepsilon \in \{-1,1\}^n$ is uniform random and independent of $\delta(t)$.

Taking p'th powers and applying Jensen's inequality yields a statement identical to Eq. (2), except that the distribution of δ is different. Enflo's conjecture then follows from a routine symmetrization argument [10, Section 3], which is not presented here. It is also an easy exercise to derive Pisier's inequality from Theorem 5 using the central limit theorem.

Proof. By observations from Lecture 1 and the beginning of Lecture 2,

$$f(\varepsilon) - \mathbf{E}f = P_0 f(\varepsilon) - P_\infty f(\varepsilon) = -\int_0^\infty \frac{d}{dt} P_t f(\varepsilon) dt = \int_0^\infty \sum_{i=1}^n D_i P_t f(\varepsilon) dt,$$

and

$$D_i P_t f(\varepsilon) = D_i^2 P_t f(\varepsilon) = D_i P_t D_i f(\varepsilon) = D_i \left(\sum_{\xi \in \{-1,1\}^n} \prod_{j=1}^n \frac{1 + \varepsilon_j \xi_j e^{-t}}{2} D_i f(\xi) \right)$$
$$= \sum_{\xi \in \{-1,1\}^n} \prod_{j=1}^n \frac{1 + \varepsilon_j \xi_j e^{-t}}{2} \cdot \frac{\varepsilon_i \xi_i e^{-t}}{1 + \varepsilon_i \xi_i e^{-t}} D_i f(\xi)$$

$$= \mathbf{E}_\xi \left[D_i f(\varepsilon\xi(t)) \cdot \frac{\xi_i(t)e^{-t}}{1+\xi_i(t)e^{-t}} \right],$$

and

$$\frac{\xi_i(t)e^{-t}}{1+\xi_i(t)e^{-t}} = \frac{\xi_i(t)e^{-t}(1-\xi_i(t)e^{-t})}{(1+\xi_i(t)e^{-t})(1-\xi_i(t)e^{-t})} = \frac{e^{-t}(\xi_i(t)-e^{-t})}{1-e^{-2t}} = \frac{e^{-t}}{\sqrt{1-e^{-2t}}}\delta_i(t),$$

so

$$\begin{aligned} f(\varepsilon) - \mathbf{E}f &= \int_0^\infty \mathbf{E}_\xi \left[\sum_{i=1}^n D_i f(\varepsilon\xi(t)) \cdot \delta_i(t) \right] \frac{e^{-t}}{\sqrt{1-e^{-2t}}} dt \\ &= \frac{\pi}{2} \int \mathbf{E}_\xi \left[\sum_{i=1}^n D_i f(\varepsilon\xi(t)) \cdot \delta_i(t) \right] \mu(dt). \end{aligned}$$

By the triangle inequality,

$$\|f(\varepsilon) - \mathbf{E}f\|_X \le \frac{\pi}{2} \int \mathbf{E}_\xi \left\| \sum_{i=1}^n D_i f(\varepsilon\xi(t)) \cdot \delta_i(t) \right\|_X \mu(dt),$$

so by Minkowski's inequality and then Jensen's inequality,

$$\begin{aligned} \left(\mathbf{E}_\varepsilon \|f(\varepsilon) - \mathbf{E}f\|_X^p\right)^{1/p} &\le \frac{\pi}{2} \int \left(\mathbf{E}_\varepsilon \left(\mathbf{E}_\xi \left\| \sum_{i=1}^n D_i f(\varepsilon\xi(t)) \cdot \delta_i(t) \right\|_X \right)^p \right)^{1/p} \mu(dt) \\ &\le \frac{\pi}{2} \int \left(\mathbf{E}_{\varepsilon,\xi} \left\| \sum_{i=1}^n D_i f(\varepsilon\xi(t)) \cdot \delta_i(t) \right\|_X^p \right)^{1/p} \mu(dt). \end{aligned}$$

Finally, the result follows because $\varepsilon\xi(t)$ is uniform random conditioned on $\xi(t)$. □

Remark. Implicit above is that

$$D_i P_t f(\varepsilon) = \mathbf{E}_\xi \left[f(\varepsilon\xi(t)) \cdot \frac{\xi_i(t)e^{-t}}{1+\xi_i(t)e^{-t}} \right]$$

for all $f : \{-1,1\}^n \to (X, \|\cdot\|_X)$. (We proved this with $D_i f$ in place of f on the right, but the argument generalizes easily.) This is analogous to the well-known formula

$$\partial_i(\varphi * f)(x) = \partial_i \int_{\mathbb{R}^n} \varphi(x-y) f(y) dy_1 \cdots dy_n = \int_{\mathbb{R}^n} \frac{y_i - x_i}{a} \varphi(x-y) f(y) dy_1 \cdots dy_n$$

for all "nice" functions $f : \mathbb{R}^n \to \mathbb{R}$, where $\varphi(x) = (2\pi a)^{-n/2} e^{-\|x\|^2/2a}$ is the density at x of the multivariate Gaussian distribution with mean 0 and covariance matrix aI, and $*$

denotes convolution. In particular, using the triangle inequality, both formulas allow us to bound the "smoothness" (i.e. some norm of the derivatives) of $P_t f$ or $\varphi * f$ in terms of the values of f itself rather than f's derivatives. Additionally, if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable then $\partial_i(\varphi * f) = \varphi * \partial_i f$ (since convolution is commutative), analogous to the fact that D_i and P_t commute.

Recall that we used Pisier's inequality to prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is "nice" and takes n independent standard Gaussian inputs then $\mathbf{E}|f - \mathbf{E}f| \leq \sqrt{\pi/2} \cdot \mathbf{E}\|\nabla f\|$. By Theorem 5 an analogous inequality holds for functions $f : \{-1,1\}^n \to \mathbb{R}$:

$$\mathbf{E}|f - \mathbf{E}f| \leq \frac{\pi}{2} \int \mathbf{E}\left|\sum_{i=1}^{n} \delta_i(t) D_i f(\varepsilon)\right| \mu(dt) \leq \frac{\pi}{2} \int \mathbf{E}_\varepsilon \sqrt{\mathbf{E}_\delta \left|\sum_{i=1}^{n} \delta_i(t) D_i f(\varepsilon)\right|^2} \cdot \mu(dt)$$

$$= \frac{\pi}{2} \int \mathbf{E}_\varepsilon \|Df(\varepsilon)\| \mu(dt) = \frac{\pi}{2} \mathbf{E}\|Df\|. \tag{3}$$

Eq. (3) was first proved by Efraim and Lust-Piquard [3] (using noncommutative probability, as previously discussed). The constant $\pi/2$ can be slightly improved using a tighter bound than Jensen's inequality for the second inequality above, but it is unknown whether the constant can be improved all the way to $\sqrt{\pi/2}$ like in the Gaussian inequality.

We now turn our attention to the second main topic of these lectures, namely a strengthening of the Poincaré inequality for boolean functions on the cube. For $f : \{-1,1\}^n \to \{0,1\}$, the value $\|Df(\varepsilon)\|^2$ equals $1/4$ times the number of coordinates i such that $f(\varepsilon_1, \ldots, \varepsilon_{i-1}, 1, \varepsilon_{i+1}, \ldots, \varepsilon_n) \neq f(\varepsilon_1, \ldots, \varepsilon_{i-1}, -1, \varepsilon_{i+1}, \ldots, \varepsilon_n)$. Therefore the Poincaré inequality ($\operatorname{Var} f \leq \mathbf{E}\|Df\|^2$) may be far from tight for f, because $\operatorname{Var} f \leq 1/4$ whereas $\mathbf{E}\|Df\|^2$ may be arbitrarily large as n goes to infinity. For example, if f is the majority function ($f(\varepsilon) = \mathbb{1}_{\sum_{i=1}^n \varepsilon_i > 0}$) then $\|Df(\varepsilon)\|^2 = \Theta(n) \cdot \mathbb{1}_{\sum_{i=1}^n \varepsilon_i \approx 0}$, so $\mathbf{E}\|Df\|^2 = \Theta\left(2^{-n}\binom{n}{n/2} n\right) = \Theta(\sqrt{n})$ by Stirling's approximation.

To obtain a tighter bound for arbitrary $f : \{-1,1\}^n \to \{0,1\}$, note that $\operatorname{Var} f = \mathbf{E}f(1 - \mathbf{E}f) = \frac{1}{2}\mathbf{E}|f - \mathbf{E}f|$, so it follows from Eq. (3) that $\operatorname{Var} f \leq \frac{\pi}{4}\mathbf{E}\|Df\|$. This inequality is tight up to a constant factor for the majority function, and was first proved by Talagrand [17] with $\sqrt{2}$ in place of $\pi/4$.

Lecture 4

The following inequality also improves on the Poincaré inequality in certain cases:

Theorem 6 (Falik and Samorodnitsky [6][7])**.** *For all $f : \{-1,1\}^n \to \mathbb{R}$,*

$$\operatorname{Var} f \cdot \log\left(\frac{\operatorname{Var} f}{\sum_{i=1}^n \mathbf{E}[|D_i f|]^2}\right) \leq 2\mathbf{E}\|Df\|^2.$$

[7] Defining $\mathcal{E}(f,f)$ and d_i as in [6, Section 2], it is easy to verify that $\mathcal{E}(f,f) = 4\mathbf{E}\|Df\|^2$ and $\mathbf{E}|d_i| \leq \mathbf{E}|D_i f|$. Theorem 6 then follows from [6, Theorem 2.2] with the constant $C = 2$ from [6, Section 3.1].

Similar bounds were previously obtained by Kahn, Kalai and Linial [11], Talagrand [19], Benjamini, Kalai and Schramm [1], and Rossignol [16], in this order chronologically. Theorem 6 is tight up to a constant factor for the tribes function, i.e. the function $f : \{-1,1\}^{w\times s} \to \{0,1\}, f(\varepsilon) = \bigvee_{i=1}^{s}\bigwedge_{j=1}^{w}\varepsilon_{ij}$ where $s = \Theta(2^w)$ is such that $\mathbf{E}f \approx 1/2$. However, Theorem 6 is far from tight for the majority function. In contrast, the bound $\operatorname{Var} f \le C\mathbf{E}\|Df\|$ for boolean functions f (proved above) is tight for majority but not for tribes.

In this lecture we will prove the following inequality, which (for boolean functions f, and up to constants) implies both the bound $\operatorname{Var} f \le C\mathbf{E}\|Df\|$ and the $\operatorname{Var} f = \Theta(1)$ case of Theorem 6, and hence is tight for both majority and tribes:

Theorem 7 (Conjectured by Talagrand [18], proved by Eldan and Gross [4])**.** *For all $f : \{-1,1\}^n \to \{0,1\}$,*

$$\operatorname{Var} f \cdot \sqrt{\log\left(1 + C + \frac{C}{\sum_{i=1}^{n}\mathbf{E}[|D_i f|]^2}\right)} \le C\mathbf{E}\|Df\|.$$

We break the proof down into components as follows:

Claim 8. *For all $f : \{-1,1\}^n \to \{0,1\}$ and $t \ge 0$,*

1. $\operatorname{Var} f = \frac{1}{2}\mathbf{E}|f - P_t f| + \operatorname{Var} P_{t/2} f$,
2. $\mathbf{E}|f - P_t f| \le C\sqrt{t}\cdot\mathbf{E}\|Df\|$,
3. $\operatorname{Var} P_t f \le \operatorname{Var} f \cdot \left(4\sum_{i=1}^{n}\mathbf{E}[|D_i f|]^2\right)^{\theta(t)/2}$ *for* $\theta(t) := \frac{1-e^{-2t}}{1+e^{-2t}}$.

Bibliographic notes. Item 1 is well known, e.g. [13, Eq. 7]. Pisier [15] proved an analogue of Item 2 in the Gaussian case, using Theorem 4. Item 3 strengthens a result of Eldan and Gross [4], which itself strengthens a result of Keller and Kindler [12].

Remark. Item 1 generalizes the observation that $\operatorname{Var} f = \frac{1}{2}\mathbf{E}|f - \mathbf{E}f|$, which we have already seen. Our proof of Item 2 holds even if f is real-valued rather than boolean. Item 3 may be of independent interest. The function θ equals tanh, but we use the θ notation for brevity.

Remark. For a different bound on $\operatorname{Var} f - \operatorname{Var} P_t f$, note that since

$$\frac{d}{dt}\operatorname{Var} P_t f = \frac{d}{dt}\left(\mathbf{E}(P_t f)^2 - (\mathbf{E}P_t f)^2\right) = \frac{d}{dt}\left(\mathbf{E}(P_t f)^2 - (\mathbf{E}f)^2\right) = \frac{d}{dt}\mathbf{E}(P_t f)^2,$$

evaluating the integral from Lecture 1 up to t rather than up to infinity reveals that

$$\operatorname{Var} f - \operatorname{Var} P_t f \le 2t\mathbf{E}\|Df\|^2.$$

Thus, if $\operatorname{Var} P_t f$ is small for some $t \le o(1)$, then we obtain the bound $\operatorname{Var} f \lesssim 2t\mathbf{E}\|Df\|^2$ which improves on the Poincaré inequality.

Proof of Theorem 7 assuming Claim 8. Let $K = \sum_{i=1}^n \mathbf{E}[|D_i f|]^2$. If $K > 0.01$ then the result follows because $\operatorname{Var} f \leq C\mathbf{E}\|Df\|$, so we may assume that $K \leq 0.01$. First we give an informal argument using the small t approximation $\theta(t) \approx t$ (unfortunately, the inequality $\theta(t) \leq t$ is in the wrong direction for this to be rigorous), and then we give a rigorous proof. By Claim 8,

$$\operatorname{Var} f \lesssim C\sqrt{t} \cdot \mathbf{E}\|Df\| + \operatorname{Var} f \cdot (4K)^{t/2},$$

so plugging in $t = \log(4)/\log(1/4K)$ gives

$$\operatorname{Var} f \lesssim C\sqrt{1/\log(C/K)} \cdot \mathbf{E}\|Df\| + \frac{1}{2}\operatorname{Var} f,$$

and the result follows by subtracting $\frac{1}{2}\operatorname{Var} f$ from both sides.

To make this rigorous it suffices to prove that if $t = C/\log(1/4K)$ (for an appropriate constant C) then $(4K)^{\theta(t)/2} \leq 1/2$. Let $c_1 > 0$ be a universal constant such that $0.04^{\theta(c_1)/2} \leq 1/2$. Let $c_2 > 0$ be a universal constant such that $\theta(t) \geq c_2 t$ for all $0 \leq t \leq c_1$; to see that such a constant exists, note that $\theta(t) \geq \frac{1-e^{-2t}}{2}$, and that $e^{-2t} \leq 1 - Ct$ for all $0 \leq t \leq c_1$ and an appropriate constant C. Now let $t = \frac{2\log(2)/c_2}{\log(1/4K)}$: if $t \geq c_1$ then $(4K)^{\theta(t)/2} \leq 0.04^{\theta(c_1)/2} \leq 1/2$, and if $t \leq c_1$ then $(4K)^{\theta(t)/2} \leq (4K)^{c_2 t/2} = 1/2$. □

Proof of Item 1. Since $P_t f(\varepsilon)$ is a convex combination of values of f, we have $0 \leq P_t f \leq 1$. Therefore, if $f = 1$ then $|f - P_t f| = 1 - P_t f$, and if $f = 0$ then $|f - P_t f| = P_t f$, so

$$|f - P_t f| = f(1 - P_t f) + (1-f)P_t f = f + P_t f - 2fP_t f.$$

We now use the fact that $\mathbf{E}[f \cdot P_t f] = \mathbf{E}(P_{t/2} f)^2$. (Here is one way to see this: first recall from a remark early in Section 2 that $P_t = e^{t\Delta}$, so $P_t = P_{t/2}^2$. Next observe that $\mathbf{E}[f \cdot P_{t/2} g] = \mathbf{E}[P_{t/2} f \cdot g]$ for all $g : \{-1,1\}^n \to \mathbb{R}$.) Recalling that $\mathbf{E}P_t f = \mathbf{E}f$, we obtain

$$\frac{1}{2}\mathbf{E}|f - P_t f| = \mathbf{E}f - \mathbf{E}(P_{t/2} f)^2 = \mathbf{E}f - \mathbf{E}(P_{t/2} f)^2 - (\mathbf{E}f)^2 + (\mathbf{E}P_{t/2} f)^2 = \operatorname{Var} f - \operatorname{Var} P_{t/2} f.$$

□

Proof of Item 2. By the same reasoning as in Lecture 3,

$$f(\varepsilon) - P_t f(\varepsilon) = -\int_0^t \frac{d}{ds} P_s f(\varepsilon) ds = \int_0^t \frac{e^{-s}}{\sqrt{1-e^{-2s}}} \cdot \mathbf{E}_\xi \left[\sum_{i=1}^n \delta_i(s) D_i f(\varepsilon \xi(s))\right] ds,$$

so

$$\mathbf{E}|f - P_t f| \leq \int_0^t \frac{e^{-s}}{\sqrt{1-e^{-2s}}} \cdot \mathbf{E}\left|\sum_{i=1}^n \delta_i(s) D_i f(\varepsilon)\right| ds \leq \mathbf{E}\|Df\| \cdot \int_0^t \frac{e^{-s}}{\sqrt{1-e^{-2s}}} ds.$$

Finally,

$$\int_0^t \frac{e^{-s}}{\sqrt{1-e^{-2s}}} ds = \int_0^t \frac{ds}{\sqrt{e^{2s}-1}} \leq \int_0^t \frac{ds}{\sqrt{2s}} = C\sqrt{t}.$$

□

All that remains is to prove Item 3. We use without proof the following case of hypercontractivity (see e.g. [14, Chapters 9-10]), where $\|f\|_p$ denotes $(\mathbf{E}|f|^p)^{1/p}$:

Fact 9 (Hypercontractivity)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $t \geq 0$; then $\|P_t f\|_2 \leq \|f\|_{1+e^{-2t}}$.*

Remark. The intuition behind our use of hypercontractivity is as follows. Results from Lecture 1 imply that $\frac{d}{dt} \operatorname{Var} P_t f = -2\mathbf{E}\|DP_t f\|^2 \leq -2 \operatorname{Var} P_t f$ (where the inequality is the Poincaré inequality[8]), so $\operatorname{Var} P_t f \leq e^{-2t} \operatorname{Var} f$. This can be rewritten as

$$\mathbf{E}(P_t f)^2 \leq e^{-2t}\mathbf{E}[f^2] + (1 - e^{-2t})\mathbf{E}[f]^2,$$

i.e. the quantity $\mathbf{E}(P_t f)^2$ interpolates between $\mathbf{E}[f^2]$ and $\mathbf{E}[f]^2$ according to an arithmetic mean. We use hypercontractivity to improve this to a *geometric* mean when $f \geq 0$. This is a major improvement, because it shows that when $\mathbf{E}[f]^2 \leq o(\mathbf{E}[f^2])$, the quantity $\mathbf{E}(P_t f)^2$ halves in time $o(1)$.

Corollary 10 (AM-GM principle)**.** *For all $f : \{-1,1\}^n \to \mathbb{R}$ and $t \geq 0$,*

$$\mathbf{E}[(P_t f)^2] \leq \mathbf{E}[f^2]^{1-\theta(t)}\mathbf{E}[|f|]^{2\theta(t)}.$$

Proof. Note that $1 + e^{-2t} = 2(1-s) + s$ for $s = 1 - e^{-2t}$. By Hölder's inequality,

$$\mathbf{E}[|f|^{1+e^{-2t}}] = \mathbf{E}[|f|^{2(1-s)}|f|^s] \leq \mathbf{E}[f^2]^{1-s}\mathbf{E}[|f|]^s,$$

so by Fact 9,

$$\|P_t f\|_2^2 \leq \|f\|_{1+e^{-2t}}^2 \leq \mathbf{E}[f^2]^{1-\theta(t)}\mathbf{E}[|f|]^{2\theta(t)}.$$

□

Falik and Samorodnitsky [6] and Rossignol [16] observed that such principles can be tensorized. These authors did this at the level of the log-Sobolev inequality, but here we do it directly:

Lemma 11 (Essentially Falik–Samorodnitsky/Rossignol)**.** *For all $f : \{-1,1\}^n \to \mathbb{R}$,*

$$\operatorname{Var} P_t f \leq (\operatorname{Var} f)^{1-\theta(t)} \left(\sum_{i=1}^n \mathbf{E}[|D_i f|]^2 \right)^{\theta(t)}.$$

Proof. We use a standard argument involving the Doob martingale, e.g. like in the proof of the Efron-Stein inequality [2]. Let $\mathbf{E}_i f(\varepsilon) = \mathbf{E}_{\delta_{i+1},\ldots,\delta_n} f(\varepsilon_1, \ldots, \varepsilon_i, \delta_{i+1}, \ldots, \delta_n)$ (where the δ_j are independent and uniform random) and $\Gamma_i f = \mathbf{E}_i f - \mathbf{E}_{i-1} f$, and note that $f - \mathbf{E}f = \sum_{i=1}^n \Gamma_i f$. Observe that $\operatorname{Var} f = \sum_{i=1}^n \mathbf{E}(\Gamma_i f)^2$, because for all $i < j$,

$$\mathbf{E}[\Gamma_i f \cdot \Gamma_j f] = \mathbf{E}_{\varepsilon_1,\ldots,\varepsilon_i}[\Gamma_i f(\varepsilon)\mathbf{E}_{\varepsilon_{i+1},\ldots,\varepsilon_j}\Gamma_j f(\varepsilon)] = 0.$$

[8] After first proving Item 3 without assuming Theorem 6, we will then show what happens if we use Theorem 6 in place of the Poincaré inequality here.

Furthermore, an easy generalization of the proof that $\mathbf{E}P_t f = \mathbf{E}f$ implies that $\mathbf{E}_i P_t f = P_t \mathbf{E}_i f$, so $\Gamma_i P_t f = P_t \Gamma_i f$. Therefore, by Corollary 10 and Hölder,

$$\operatorname{Var} P_t f = \sum_{i=1}^n \mathbf{E}[(P_t \Gamma_i f)^2] \le \sum_{i=1}^n \mathbf{E}[(\Gamma_i f)^2]^{1-\theta(t)} \mathbf{E}[|\Gamma_i f|]^{2\theta(t)} \le (\operatorname{Var} f)^{1-\theta(t)} \left(\sum_{i=1}^n \mathbf{E}[|\Gamma_i f|]^2 \right)^{\theta(t)}.$$

Finally, $\mathbf{E}|\Gamma_i f| \le \mathbf{E}|D_i f|$ because

$$\begin{aligned}|\Gamma_i f(\varepsilon)| &= |\mathbf{E}_{\delta_{i+1},\dots,\delta_n} D_i f(\varepsilon_1, \dots, \varepsilon_i, \delta_{i+1}, \dots, \delta_n)| \\ &\le \mathbf{E}_{\delta_{i+1},\dots,\delta_n} |D_i f(\varepsilon_1, \dots, \varepsilon_i, \delta_{i+1}, \dots, \delta_n)|.\end{aligned}$$

□

Now we invoke the fact that f is boolean to complete the proof of Item 3:

Proof. Let $K = \sum_{i=1}^n \mathbf{E}[|D_i f|]^2$. If $\operatorname{Var} f \ge \frac{1}{2}\sqrt{K}$ then by Lemma 11,

$$\operatorname{Var} P_t f \le \operatorname{Var} f \cdot \left(\frac{K}{\operatorname{Var} f} \right)^{\theta(t)} \le \operatorname{Var} f \cdot (4K)^{\theta(t)/2}.$$

Alternatively, if $\operatorname{Var} f \le \frac{1}{2}\sqrt{K}$ then by Corollary 10 applied to $f - \mathbf{E}f$,

$$\operatorname{Var} P_t f \le (\operatorname{Var} f)^{1-\theta(t)} (2 \operatorname{Var} f)^{2\theta(t)} \le \operatorname{Var} f \cdot (4K)^{\theta(t)/2},$$

where we used $\mathbf{E}|f - \mathbf{E}f| = 2 \operatorname{Var} f$. □

We conclude with an alternate proof of (something similar to) Lemma 11:

Proof. Recall from the discussion following Fact 9 that $\frac{d}{dt} \operatorname{Var} P_t f = -2\mathbf{E}\|DP_t f\|^2$. It follows from applying Theorem 6 to $P_t f$ that

$$\frac{d}{dt} \operatorname{Var} P_t f \le -\operatorname{Var} P_t f \cdot \log \left(\frac{\operatorname{Var} P_t f}{\sum_{i=1}^n \mathbf{E}[|D_i P_t f|]^2} \right).$$

Furthermore,

$$\mathbf{E}_\varepsilon |P_t D_i f(\varepsilon)| = \mathbf{E}_\varepsilon |\mathbf{E}_\xi D_i f(\varepsilon \xi(t))| \le \mathbf{E}_{\varepsilon,\xi} |D_i f(\varepsilon \xi(t))| = \mathbf{E}|D_i f|,$$

so

$$\frac{d}{dt} \operatorname{Var} P_t f \le -\operatorname{Var} P_t f \cdot \log \left(\frac{\operatorname{Var} P_t f}{K} \right)$$

where $K = \sum_{i=1}^n \mathbf{E}[|D_i f|]^2$. The solution to the above differential inequality (with initial condition $\operatorname{Var} P_0 f = \operatorname{Var} f$) is

$$\operatorname{Var} P_t f \le K \left(\frac{\operatorname{Var} f}{K} \right)^{e^{-t}} = (\operatorname{Var} f)^{e^{-t}} K^{1-e^{-t}},$$

which essentially matches Lemma 11 when t is small. □

Acknowledgments

Thanks to Ramon van Handel for very helpful discussions and feedback regarding these notes. In particular, the proof of Item 3 of Claim 8 given here differs somewhat from that in the original lecture, and was provided by Ramon for the purpose of these notes. Thanks to Xinyu Wu and Henry Yuen for helpful feedback as well.

References

[1] Itai Benjamini, Gil Kalai, and Oded Schramm. "First passage percolation has sublinear distance variance". In: *Ann. Probab.* 31.4 (2003), pp. 1970–1978. DOI: 10.1214/aop/1068646373 (p. 11).

[2] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press, 2013. Chap. 3. DOI: 10.1093/acprof:oso/9780199535255.001.0001 (p. 13).

[3] Limor Ben Efraim and Françoise Lust-Piquard. "Poincaré type inequalities on the discrete cube and in the CAR algebra". In: *Probab. Theory Related Fields* 141.3-4 (2008), pp. 569–602. DOI: 10.1007/s00440-007-0094-x. arXiv: math/0702233 (pp. 8, 10).

[4] Ronen Eldan and Renan Gross. "Concentration on the Boolean hypercube via pathwise stochastic analysis". In: STOC 2020. 2020. arXiv: 1909.12067 (pp. 1, 11).

[5] Per Enflo. "On infinite-dimensional topological groups". In: *Séminaire sur la Géométrie des Espaces de Banach (1977–1978).* École Polytech., Palaiseau, 1978, Exp. No. 10–11, 11 (p. 5).

[6] Dvir Falik and Alex Samorodnitsky. "Edge-isoperimetric inequalities and influences". In: *Combin. Probab. Comput.* 16.5 (2007), pp. 693–712. DOI: 10.1017/S0963548306008340. arXiv: math/0512636 (pp. 10, 13).

[7] Christophe Garban and Jeffrey E. Steif. *Noise sensitivity of Boolean functions and percolation.* Cambridge University Press, New York, 2015. DOI: 10.1017/CBO9781139924160 (p. 1).

[8] Ramon van Handel. *Remarks on the Discrete Cube.* Probability, Geometry, and Computation in High Dimensions Boot Camp. Simons Institute for the Theory of Computing. 2020. URL: https://simons.berkeley.edu/workshops/hd-2020-bc (pp. 1, 7).

[9] Ramon van Handel. "Structured random matrices". In: *Convexity and concentration.* Vol. 161. IMA Vol. Math. Appl. Springer, New York, 2017, pp. 107–156. DOI: 10.1007/978-1-4939-7005-6_4. arXiv: 1610.05200 (p. 6).

[10] Paata Ivanisvili, Ramon van Handel, and Alexander Volberg. "Rademacher type and Enflo type coincide". In: *Ann. of Math. (2)* 192.2 (2020), pp. 665–678. DOI: 10.4007/annals.2020.192.2.8. arXiv: 2003.06345 (pp. 1, 5, 8).

[11] Jeff Kahn, Gil Kalai, and Nati Linial. "The influence of variables on Boolean functions". In: FOCS 1988. 1988, pp. 68–80 (pp. 1, 2, 11).

[12] Nathan Keller and Guy Kindler. "Quantitative relation between noise sensitivity and influences". In: *Combinatorica* 33.1 (2013), pp. 45–71. DOI: 10.1007/s00493-013-2719-2 (p. 11).

[13] Michel Ledoux. "A simple analytic proof of an inequality by P. Buser". In: *Proc. Amer. Math. Soc.* 121.3 (1994), pp. 951–959. DOI: 10.2307/2160298 (p. 11).

[14] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, New York, 2014. DOI: 10.1017/CBO9781139814782 (pp. 1, 13).

[15] Gilles Pisier. "Probabilistic methods in the geometry of Banach spaces". In: *Probability and analysis (Varenna, 1985)*. Vol. 1206. Lecture Notes in Math. Springer, Berlin, 1986, pp. 167–241. DOI: 10.1007/BFb0076302 (pp. 6, 7, 11).

[16] Raphaël Rossignol. "Threshold for monotone symmetric properties through a logarithmic Sobolev inequality". In: *Ann. Probab.* 34.5 (2006), pp. 1707–1725. DOI: 10.1214/009117906000000287. arXiv: math/0511607 (pp. 11, 13).

[17] Michel Talagrand. "Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis' graph connectivity theorem". In: *Geom. Funct. Anal.* 3.3 (1993), pp. 295–314. DOI: 10.1007/BF01895691 (pp. 1, 7, 10).

[18] Michel Talagrand. "On boundaries and influences". In: *Combinatorica* 17.2 (1997), pp. 275–285. DOI: 10.1007/BF01200910 (p. 11).

[19] Michel Talagrand. "On Russo's approximate zero-one law". In: *Ann. Probab.* 22.3 (1994), pp. 1576–1587. URL: https://www.jstor.org/stable/2245033 (p. 11).