Random Matrices from the Classical Compact Groups: a Panorama

Part IV: Geometric Applications of Measure Concentration

Mark Meckes

Case Western Reserve University

Oxford, 15 February 2021

Today’s first slogan

Not-too-low-rank projections act almost like isometries.

Concentration of a norm

Let $P_k \in M_{n,k}(\mathbb{R})$ be the first k columns of a random $U \in \mathbb{O}(n)$.

Recall from last time:

$$\mathbb{P}\left[\left|\|P_k x\|_2 - \sqrt{\frac{k}{n}}\|x\|_2\right| \geq \left(\sqrt{\frac{k}{n}}\|x\|_2\right)\varepsilon\right] \leq 2e^{-ck\varepsilon^2}.$$

Concentration of many norms

If $x_1, \ldots, x_m \in S^{n-1}$, then with probability at least

$$1 - 2me^{-ck\varepsilon^2}$$

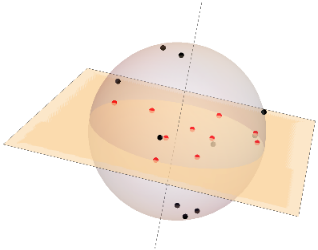
we have

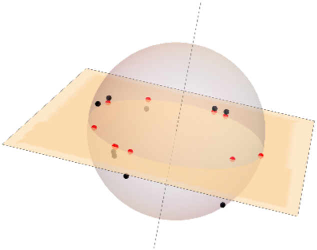
$$1 - \varepsilon \leq \frac{\|P_k x_i\|_2}{\sqrt{\frac{k}{n}}\,\|x_i\|_2} \leq 1 + \varepsilon$$

for every i.

High-dimensional intuition

This phenomenon is surprising to our two/three-dimensional brains:





but makes more sense from a properly high-dimensional perspective.

The Johnson–Lindenstrauss lemma

Applying the argument to the $\binom{m}{2}$ points $x_i - x_j$, we get:

Theorem ($\sim$ Johnson–Lindenstrauss)

If $k \geq \frac{C}{\varepsilon^2} \log m$*, then with probability at least*

$$1 - 2e^{-ck\varepsilon^2}$$

we have

$$1 - \varepsilon \leq \frac{\left\|P_k(x_i - x_j)\right\|_2}{\sqrt{\frac{k}{n}}\left\|(x_i - x_j)\right\|_2} \leq 1 + \varepsilon$$

for every i and j.

Dimension reduction

The punchline:

Projecting $\{x_i\}_{i=1}^m \subseteq \mathbb{R}^n$ onto a $\approx \log m$-dimensional subspace barely changes the distances between the points. (Probably.)

Why you should care:

Algorithms that depend only on distances between n-dimensional data points can be run on the $\approx \log m$-dimensional projections instead, lifting the curse of dimensionality!

Restricted Isometry Property

Combining the same ideas with a discretization argument yields:

Theorem (∼ Candès–Tao)

If $k \geq Cs\log\left(\frac{cn}{s}\right)$, *then with probability at least*

$$1 - 2e^{-ck}$$

we have

$$0.9 \leq \frac{\|P_k x\|_2}{\sqrt{\frac{k}{n}}\,\|x\|_2} \leq 1.1$$

for every $x \in \mathbb{R}^n$ *with* $\leq s$ *nonzero components.*

Sparse signal recovery

Corollary

If $k \geq Cs \log\left(\frac{cn}{s}\right)$*, then with probability at least* $1 - 2e^{-ck}$ *the following holds:*
If $x \in \mathbb{R}^n$ *has* $\leq s$ *nonzero components, then*

$$x = \underset{x' : P_k x' = P_k x}{\operatorname{argmin}} \left\| x' \right\|_1 .$$

Why you should care:

Under the assumption that x is sparse, it can (probably!) be recovered from $P_k x$ via a convex program.

Today’s second slogan

Not-too-high-dimensional sections/projections are almost all alike.

The Dvoretzky–Milman theorem

Let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^n$.

For normalization, assume $\|\cdot\|_2 \leq \|\cdot\|$.

$M := \mathbb{E}\,\|\theta\|$ for $\theta \sim \mathrm{unif}(S^{n-1})$.

Theorem (V. Milman, Gordon)

Suppose $k \leq c\varepsilon^2 M^2 n$ and let $E \in G_{n,k}$ be random. Then with probability at least $1 - 2e^{-ck}$,

$$1 - \varepsilon \leq \frac{\|x\|}{M\,\|x\|_2} \leq 1 + \varepsilon$$

for every $x \in E$.

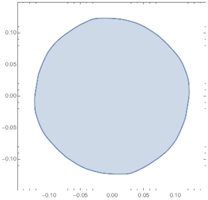
The Dvoretzky–Milman theorem

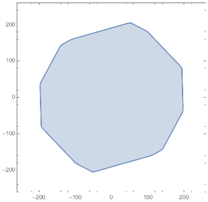
The punchline:

On a random $\approx \varepsilon^2 M^2 n$-dimensional subspace, $\|\cdot\|$ is basically the same as $\|\cdot\|_2$.

Or:

A random $\approx \varepsilon^2 M^2 n$-dimensional section of a symmetric convex body is basically a Euclidean ball.





Random 2-dimensional subspaces of ℓ_1^{100} and $\ell_\infty^{1000000}$.

A version for projections follows by duality.

Dvoretzky's theorem

The Dvoretzky–Rogers lemma roughly says that we can arrange to have $M \leq c\sqrt{\frac{\log n}{n}}$.

Theorem (Dvoretzky)

If B is an infinite-dimensional Banach space, then for every k, B has k-dimensional subspaces which are arbitrarily close to being Hilbert spaces.

Sketch of proof of Dvoretzky–Milman

Fix $F \in G_{n,k}$. Then $E \sim U(F)$.

$\mathbb{P}\left[\left| \|Ux\| - \|Uy\| \right| \geq t \right] \leq 2e^{-cnt^2/\|x-y\|_2^2}$

Thus $\{\|Ux\| - M\}$ is a subgaussian random process indexed by $x \in S^{n-1} \cap F$ with $d(x,y) = n^{-1/2}\|x-y\|_2$.

Dudley's entropy bound $\Rightarrow$

$$\mathbb{E} \sup_{x \in S^{n-1} \cap F} \left| \|Ux\| - M \right| \leq C\sqrt{\frac{k}{n}}.$$

So for a typical E, $\|y\| \approx M$ for every $y \in S^{n-1} \cap E$.

$\square$

Projections of measures

Observation (Sudakov, Diaconis–Freedman, ...):

If you project a high-dimensional probability measure / data set onto one or two dimensions, the result nearly always looks Gaussian.

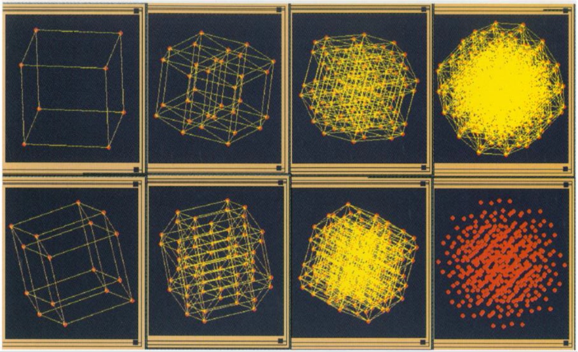


Figure from Buja, Cook, and Swayne “Interactive High-dimensional Data Visualization”, 1996.

Measure-theoretic Dvoretzky theorem

The bounded Lipschitz distance between X and Y is

$$d_{BL}(X, Y) = \sup_{\|\psi\|_L, \|\psi\|_\infty \leq 1} |\mathbb{E}\psi(X) - \mathbb{E}\psi(Y)|.$$

Theorem (E. Meckes)

Suppose that $X \in \mathbb{R}^n$ satisfies

$$\mathbb{E}X = 0, \qquad \mathbb{E}X_i X_j = \delta_{ij}, \qquad \mathbb{E}\left| \|X\|_2^2 - n \right| \leq C \frac{n}{(\log n)^{1/3}},$$

and that $k \leq (2 - \varepsilon)\frac{\log n}{\log \log n}$.

Then for almost all $E \in G_{n,k}$, $d_{BL}(\pi_E(X), Z_E)$ is small, where Z_E is a standard Gaussian vector in E.

Measure-theoretic Dvoretzky theorem

If $k \geq (2+\varepsilon)\frac{\log n}{\log \log n}$, there is an X such that $d_{BL}(\pi_E(X), Z_E) \geq c$ for every $E \in G_{n,k}$.

Theorem (Klartag, ...)

If $X \in \mathbb{R}^n$ satisfies $\mathbb{E}X = 0$, $EX_iX_j = \delta_{ij}$, and is log-concave, and $k \leq cn^\alpha$, then $d_{TV}(\pi_E(X), Z_E)$ is small for almost all $E \in G_{n,k}$.

Outline of proof of measure-theoretic Dvoretzky

First step — annealed version:

Let μ_E be the distribution of $\pi_E(X) \in \mathbb{R}^k$.
Then $d_{BL}(\mathbb{E}\mu_E, N(0, I_k))$ is small ($\sim$ Poincaré limit).

Second step — average distance to the average:

$\mathbb{E} d_{BL}(\mu_E, \mathbb{E}\mu_E) = \mathbb{E} \sup_{\psi} |\psi(\pi_E(X)) - \mathbb{E}\psi(\pi_E(X))|$ is the expected supremum of a centered subgaussian random process (concentration on $\mathbb{SO}(n)$: $E = U(F)$).
We can bound it using Dudley's entropy bound.

Third step — from annealed to quenched:
$d_{BL}(\mu_E, \mu)$ is a Lipschitz function of U, and hence is usually not much bigger than its mean.

□

Additional reference

- Roman Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge, 2018.