On the Discrepancy of Random Matrices with Many Columns

Cole Franks and Michael Saks

March 23, 2019

Discrepancy

Discrepancy

- discrepancy of a matrix: extent to which the rows can be simultaneously split into two equal parts.
- Formally, let $\|\cdot\|_*$ be a norm, and let

$$\mathrm{disc}_*(M) = \min_{v \in \{+1,-1\}^n} \|Mv\|_*$$

  (M is an $m \times n$ matrix).

Goal: prove $\mathrm{disc}_*(M)$ is small in certain situations, and find the good assignments v efficiently.

Examples and Applications

- $$\mathrm{disc}_\infty \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} = 1$$

- Extractors: the best extractor for two independent n-bit sources with min-entropy k has error rate $\mathrm{disc}_\infty(M)$ where M is a
  1. $\binom{2^n}{2^k}^2 \times 2^{2n}$ matrix
  2. with one row for each rectangle $A \times B \subset \{0,1\}^n \times \{0,1\}^n$ with $|A| = |B| = 2^k$,
  3. each row is a $2^n \times 2^n$ matrix with (x, y) entry equal to $\frac{1}{2^{2k}} 1_A(x) 1_B(y)$.

  number of rows is $\gg$ number of columns, random coloring optimal but useless!

Upper bounds

Definition

herdisc(M): maximum discrepancy of any subset of columns of M.

Beck-Fiala Theorem: $M_{ij} \in [-1, 1]$ and $\leq t$ nonzero entries per column,

$$\text{herdisc}(M) \leq 2t - 1.$$

Beck-Fiala Conjecture: If M as above,

$$\text{herdisc}(M) = O(\sqrt{t})$$

Komlos Conjecture: M with unit vector columns,

$$\text{herdisc}(M) = O(1)$$

Banaczszyk's Theorem: If M as above,

$$\text{herdisc}(M) = O(\sqrt{\log m})$$

Discrepancy of random matrices

Let M be a random t-sparse matrix

$$\begin{array}{cc} m & \underbrace{\begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}}_{n} \end{array}$$

Theorem (Ezra, Lovett 2015)

Few columns: If $n = O(m)$, then with probability $1 - \exp(-\Omega(t))$.

$$\text{herdisc}(M) = O(\sqrt{t \log t}).$$

Many columns: If $n = \Omega\left(\binom{m}{t} \log \binom{m}{t}\right)$ then with pr. $1 - \binom{m}{t}^{-\Omega(1)}$,

$$\text{disc}(M) \leq 2$$

Why not herdisc for many columns?

General setup

- $\mathcal{L} \subset \mathbb{R}^m$ is a nondegenerate lattice,
- X is a finitely supported r.v. on $\mathcal{L}$ such that $\operatorname{span}_{\mathbb{Z}} X = \mathcal{L}$.
- n columns of M are drawn i.i.d from X.

Question

How does $\operatorname{disc}_*(M)$ *behave for various ranges of* n*?*

This talk: $n \gg m$

For $n \gg m$ the problem becomes a closest vector problem on $\mathcal{L}$.

Definition

$\rho_*(\mathcal{L})$ is the covering radius of $\mathcal{L}$ in the norm $\|\cdot\|_*$.

Fact

$\mathrm{disc}_*(M) \leq 2\rho_*(\mathcal{L})$ *with high probability as* $n \to \infty$.

Proof: For every subset $S \subset \operatorname{supp} X$, pick E_S an *even integer* combination of the elements of $\operatorname{supp} X$ that is $2\rho_*(\mathcal{L})$ away from $\sum S$. Let B be a bound on all these coefficients. Each element of $\operatorname{supp} X$ appears $B+1$ times with high probability in n; remove one of each column that appeared an odd number of times and set the labels on the remaining columns so that they sum to $-E_S$. Naïvely, n has to be huge. not tight!

Question

For a given random variable X, how large must n be before $\mathrm{disc}_*(M) \leq 2\rho_*(\mathcal{L})$ *with high probability?*

t-sparse vectors, ℓ_∞

- $\mathcal{L}$ is $\{\boldsymbol{x} \in \mathbb{Z}^m : \sum x_i \equiv 0 \bmod t\}$
- $\rho_\infty(\mathcal{L}) = 1$

By fact, $\mathrm{disc}_\infty(M) \leq 2$ eventually.
*EL*15 showed this happens for $n \geq \Omega(\binom{m}{t} \log \binom{m}{t})$. exponential dependence on t!
This work: $n = \Omega(m^3 \log^2 m)$

Our results

Our Results

Random t-sparse matrices:

Theorem (FS18)

Let M be a random t-sparse matrix. If $n = \Omega(m^3 \log^2 m)$, then

$$\mathrm{disc}_\infty(M) \leq 2$$

with probability at least $1 - O\left(\sqrt{\frac{m \log n}{n}}\right)$.

Actually usually $\mathrm{disc}_\infty(M) = 1$.

Related work: Hoberg and Rothvoss '18 obtained $\Omega(m^2 \log m)$ for M with i.i.d $\{0, 1\}$ entries.

More generally

$\mathcal{L}, M, X$ as before, and define

1. $L = \max_{v \in \operatorname{supp} X} \|v\|_2$

   e.g. $\sqrt{t}$ for t-sparse

2. *distortion* $R_* = \max_{\|v\|_2=1} \|v\|_*$.

   e.g. $\sqrt{m}$ for $* = \infty$

3. *spanningness*: $s(X)$ "how far X is from proper sublattice."

   will be $\leq 1/m$ for t-sparse

Theorem (FS18)

Suppose $\mathbb{E}XX^\dagger = I_m$. *Then* $\operatorname{disc}_*(M) \leq 2\rho_*(\mathcal{L})$ *with probability* $1 - O\left(L\sqrt{\frac{\log n}{n}}\right)$ *for*

$$n \geq N = \operatorname{poly}(m, s(X)^{-1}, R_*, \rho_*(\mathcal{L}), \log\det\mathcal{L}).$$

To apply the theorem to non-isotropic X,
consider the transformed r.v. $\Sigma^{-1/2}X$, where $\Sigma = \mathbb{E}XX^{\dagger}$.

Proof outline

Need to show: for most *fixed* M, the r.v. $M\boldsymbol{y}$, $\boldsymbol{y} \in_R \{\pm 1\}^n$, gets within $2\rho_*(\mathcal{L})$ of the origin with positive probability.

Use local central limit theorem:

1. Intuitively the $M\boldsymbol{y}$ (sampled at same time) approaches lattice Gaussian:
$$\Pr[M\boldsymbol{y} = \boldsymbol{\lambda}] \propto\approx e^{-\frac{1}{2}\boldsymbol{\lambda}^\dagger \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}}$$
for $\boldsymbol{\lambda} \in M\mathbf{1} + 2\mathcal{L}$
2. For *most* M, $M\boldsymbol{y}$ also behaves like this!
3. Then done: $\boldsymbol{\lambda} \in M\mathbf{1} + 2\mathcal{L}$ contains, near origin, elements of $*$-norm $2\rho_*(\mathcal{L})$.

Local central limit theorem

We propose an LCLT that takes a matrix parameter M, and show it holds for most M.

- Proof of LCLT $\approx$ proof of LCLT in [Kuperberg, Lovett, Peled, '12].
- Differences:
  - theirs was for FIXED *very wide* matrices.
  - Ours holds for MOST *less wide* matrices.

Motivation for our LCLT

Obstruction to LCLTs:

If X lies on a proper sublattice $\mathcal{L}' \subsetneq \mathcal{L}$, in trouble.

Need an *approximate* version of the assumption that this doesn't happen.

Spanningness

Definition

Dual lattice: $\mathcal{L}^* := \{\boldsymbol{\theta} : \forall \boldsymbol{\lambda} \in \mathcal{L}, \langle \boldsymbol{\lambda}, \boldsymbol{\theta} \rangle \in \mathbb{Z}\}$.

Definition

$f_X(\boldsymbol{\theta}) := \sqrt{\mathbb{E}[|\langle X, \boldsymbol{\theta} \rangle \bmod 1|^2]}$, where mod1 $\to [-1/2, 1/2)$

$f_X(\boldsymbol{\theta}) = 0 \implies \boldsymbol{\theta} \in \mathcal{L}^*$.

$f_X(\boldsymbol{\theta}) \approx 0 \implies \langle X, \boldsymbol{\theta} \rangle \approx\in \mathbb{Z}$.

Thus, obstruction is $\boldsymbol{\theta}$ far from $\mathcal{L}^*$ with $f_X(\boldsymbol{\theta})$ small.

Spanningness: recall $f_X(\boldsymbol{\theta}) := \sqrt{\mathbb{E}[|\langle X, \boldsymbol{\theta}\rangle \bmod 1|^2]}$

Say θ is pseudodual if

$$f_X(\boldsymbol{\theta}) \leq \frac{1}{2} d(\theta, \mathcal{L}^*).$$

(Why pseudodual? Near $\mathcal{L}^*$, $f_X(\boldsymbol{\theta}) \approx d(\theta, \mathcal{L}^*)$.)

Spanningness:

$$s(X) := \inf_{\mathcal{L}^* \not\ni \theta \text{ pseudodual}} f_X(\theta).$$

CLT

For a matrix M, define the multidimensional Gaussian density

$$G_M(\lambda) = \frac{2^{m/2}\det(\mathcal{L})}{\pi^{m/2}\sqrt{\det(MM^\dagger)}} e^{-2\lambda^\dagger (MM^\dagger)^{-1}\lambda}$$

on $\mathbb{R}^m$ (Gaussian with covariance $\frac{1}{2}MM^\dagger$).

Theorem ([FS18])

With probability $1 - O\left(L\sqrt{\frac{\log n}{n}}\right)$ *over the choice of* M,

1. $\frac{1}{2}nI_m \preceq MM^\dagger \preceq 2nI_m$
2. $$\left|\Pr_{y_i \in \{\pm 1/2\}}[M\boldsymbol{y} = \boldsymbol{\lambda}] - G_M(\lambda)\right| = G_M(0)\cdot O\left(\frac{m^2L^2}{n}\right)$$

   for all $\boldsymbol{\lambda} \in \frac{1}{2}M + \mathcal{L}$.

prvided $n \geq N_0 = \mathrm{poly}(m, s(X)^{-1}, L, \log\det\mathcal{L})$.

Proof of local limit theorem

Definition (Fourier transform!)

If Y is a random variable on $\mathbb{R}^m$, $\widehat{Y} : \mathbb{R}^m \to \mathbb{C}$ is

$$\widehat{Y}(\boldsymbol{\theta}) = \mathbb{E}[e^{2\pi i \langle Y, \boldsymbol{\theta} \rangle}].$$

Fact (Fourier inversion:)

if Y takes values on $\mathcal{L}$, then

$$\Pr(Y = \boldsymbol{\lambda}) = \det(\mathcal{L}) \int_D \widehat{Y}(\boldsymbol{\theta}) e^{-2\pi i \langle \boldsymbol{\lambda}, \boldsymbol{\theta} \rangle} d\boldsymbol{\theta}$$

Here D is any fundamental domain of the dual lattice $\mathcal{L}^*$.

Neat/obvious: true even if Y takes values on an affine shift $v + \mathcal{L}$.

Take Fourier transform

For fixed M, Fourier transform of $M\boldsymbol{y}$ for $\boldsymbol{y} \in_R \{\pm 1/2\}$?
Say i^{th} column is $\boldsymbol{x}_i$.

$$\begin{aligned}
\widehat{M\boldsymbol{y}}(\boldsymbol{\theta}) &= \mathbb{E}_{\boldsymbol{y}}\left[e^{2\pi i \langle \sum_{j=1}^{n} y_j \boldsymbol{x}_j, \boldsymbol{\theta}\rangle}\right] \\
&= \prod_{j=1}^{n} \mathbb{E}_{y_j}[e^{2\pi i y_j \langle \boldsymbol{x}_j, \theta\rangle}] \\
&= \prod_{j=1}^{n} \cos(\pi \langle \boldsymbol{x}_j, \theta\rangle).
\end{aligned}$$

Use Fourier inversion

Let $\varepsilon > 0$, to be picked with hindsight (think $n^{-1/4}$)

$$\begin{aligned}
\left|\frac{1}{\det \mathcal{L}} \Pr(My = \boldsymbol{\lambda}) - G_M(\boldsymbol{\lambda})\right| &= \left|\int_D e^{-2\pi i \langle \boldsymbol{\lambda}, \theta \rangle} (\widehat{My}(\boldsymbol{\theta}) - \widehat{G_M}(\boldsymbol{\theta})) d\boldsymbol{\theta}\right| \\
&\leq \int_{B(\varepsilon)} |\widehat{My}(\boldsymbol{\theta}) - \widehat{G_M}(\boldsymbol{\theta})| d\boldsymbol{\theta} && (I_1) \\
&+ \int_{\mathbb{R}^m \setminus B(\varepsilon)} |\widehat{G_M}(\boldsymbol{\theta})| d\boldsymbol{\theta} && (I_2) \\
&+ \int_{D \setminus B(\varepsilon)} |\widehat{My}(\boldsymbol{\theta})| d\boldsymbol{\theta} && (I_3)
\end{aligned}$$

If $D \subset B(\varepsilon)$. D is the Voronoi cell in $\mathcal{L}^*$.

rest of the proof is to show these are small!

- First two easy from the eigenvalue property.
- $\mathbb{E}_M[I_3] \leq e^{-\varepsilon^2 n}$ if $\varepsilon \leq s(X)$.

Applying the main theorem

Random t-sparse matrices

From now on we just want to bound the spanningness. We'll do it for t-sparse vectors - the framework is that of [KLP12].

Lemma

Let X be a random t-sparse vector. Then $s(X) = \Omega(\frac{1}{m})$.

Framework from [KLP12] for bounding spanningness

Recall what $s(X) \geq \frac{1}{m}$ means. We need to show that if $\boldsymbol{\theta}$ is pseudodual, i.e., $f_X(\boldsymbol{\theta}) \leq \|\boldsymbol{\theta}\|/2$ but *not dual*, then $f_X(\boldsymbol{\theta}) \geq \alpha/m$.

Proof outline: (recall $f_X(\boldsymbol{\theta}) := \sqrt{\mathbb{E}[|\langle X, \boldsymbol{\theta}\rangle \bmod 1|^2]}$)

- if all $|\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle \bmod 1| \leq 1/4$ for all $x \in \operatorname{supp} X$, then $f_X(\boldsymbol{\theta}) \geq d(\boldsymbol{\theta}, \mathcal{L}^*)$, so $\boldsymbol{\theta}$ not pseudodual unless dual.
- X is $\frac{1}{2m}$-spreading: for all θ,

$$f_X(\boldsymbol{\theta}) \geq \frac{1}{2m} \sup_{x \in \operatorname{supp} X} |\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle \bmod 1|$$

Together, if $\boldsymbol{\theta}$ is pseudodual, then $f_X(\boldsymbol{\theta}) \geq \frac{1}{8m}$. $\square$

Showing X is spreading

1. The argument in [KLP12] shows that X is $\frac{1}{(m \log m)^{3/2}}$-spreading, but is much more general.
2. A direct proof yields the $\frac{1}{m}$.

Random unit vectors

A result for a non-lattice distribution:

Theorem (FS18)

Let M be a matrix with i.i.d random unit vector columns. Then

$$\operatorname{disc} M = O(e^{-\sqrt{\frac{n}{m^3}}})$$

with probability at least $1 - O\left(L\sqrt{\frac{\log n}{n}}\right)$ *provided* $n = \Omega(m^3 \log^2 m)$,

Open problems

- Can the colorings guaranteed by our theorems be produced efficiently? The probability a random coloring is good decreases with n as $\sqrt{n}^{-m}$, which is not good enough.
- As a function of m, how many columns are required such that $\mathrm{disc}(M) \leq 2$ for t-sparse vectors with high probability?

Thank you!