HIGH-DIM. PROBABILITY. 29 MAY 2014.

MATRIX CONCENTRATION INEQS. VIA EXCHANGEABLE PAIRS

Joel A. Tropp / CALTECH / COMPUTING + MATH SCI. (CMS)

CLASSICAL RMT: EARLY EXAMPLES

- Hurwitz (19° CEN) — AVG. OVER ORTHOG. GROUP.
- Wishart (1927) — SAMPLE COVAR. NORMAL POP.
- von Neumann – Goldstine (1951) — GAUSS. MODEL FOR ROUNDING ERR
- Wigner (1955) — WIGNER MTX AS MODEL FOR HAMILTONIAN

↳ Examples involve highly symmetric matrix models

By now, we know everything about these rdm matrices

MODERN RMT: NEW APPLICATIONS

- Randomized algorithms — DIM. REDUX, RAND NLA
- Combinatorial optimiz^{n} — RELAX + ROUND METHODS
- Multivariate statistics, ML, SP — DATA MODELS
- Combinatorics, high-dim geometry, quantum IT, ...

↳ Examples have structure but lack symmetries that drive classical theory. But we don't need so much info (LSD)

In many cases, can apply methods from prob. Banach spaces, but these do not yield constants; require expertise.

GOAL: Simple, widely applicable tools for studying a range of rdm. matrix models from applications.

TWO USEFUL PRINCIPLES:

① In applications, common that a rdm. matrix can be written as a sum of indep. rdm. matrices

② There are polynomial, exponential moment ineqs. for spectral norm of a sum of indep. rdm matrices.

TODAY: Study a simple random matrix model using exch. pairs

REFS: MACKEY, Jordan, Chen, Farrell, TROPP, AoP 2014

TROPP 2012 – 2014

- Let $\underline{\underline{A}}_1, \dots, \underline{\underline{A}}_n \in \mathbb{H}_d$ = $d \times d$ Hermitian matrices
- Let $\varepsilon_1, \dots, \varepsilon_n$ be iid Rademacher r.v.s. (UNIF$\{\pm 1\}$)
- Let $\underline{\underline{X}} = \sum_{i=1}^n \varepsilon_i \underline{\underline{A}}_i$

THM (Noncommutative Khintchine ineq.) For $p = 1, 2, 3, \dots$

$$\left(\mathbb{E} \operatorname{tr} \underline{\underline{X}}^{2p}\right)^{1/2p} \le \sqrt{2p-1} \left(\operatorname{tr}\left[\left(\textstyle\sum_{i=1}^n A_i^2\right)^p\right]\right)^{1/2p}$$

$$= \sqrt{2p-1} \left\| \left(\textstyle\sum_{i=1}^n A_i^2\right)^{1/2} \right\|_{S_{2p}}.$$

- ORIG. due to Lust-Piquard, Pisier.
- Optimal const $= \left[(2p-1)!!\right]^{1/2p} \sim \frac{1}{\sqrt{e}} \sqrt{2p-1}$ (Buchholz)
- For $p = \log d$, gives info on spectral norm:

$$\left(\mathbb{E} \|\underline{X}\|^{2p}\right)^{1/2p} \le \sqrt{e} \cdot \sqrt{2p-1} \left\| \left(\textstyle\sum_{i=1}^n A_i^2\right)^{1/2} \right\|$$

- Interpret: Control via noncomm. ℓ_2 norm of coefs.
- Considered a deep result. We'll prove it over next 30 min.

→ matrix extension of Chatterjee 2008.

① Construct exchangeable pair:

$$\underline{\underline{Y}} = \underline{\underline{X}} + (\varepsilon_I' - \varepsilon_I)\underline{\underline{A}}_I \quad \text{where } I \sim \text{UNIF}\{1, 2, \dots, n\}$$

$\varepsilon_1', \dots, \varepsilon_n'$ indep iid Rad.

Key property: Linear regression cond^n

$$\mathbb{E}\left[\underline{\underline{X}} - \underline{\underline{Y}} \mid \varepsilon\right] = \frac{1}{n} \underline{\underline{X}} \qquad \text{(cf. Stein 1972)}$$

↳ Mimics cond^{nal} expect^n property for correlated Gaussians

② Lemma (Method of exchangeable pairs) (cf. Chatterjee 2008)

Let $\underline{F}: \mathbb{H}_d \to \mathbb{H}_d$ (Measurable). Then

$$\mathbb{E}\left[\underline{X} \cdot \underline{F}(\underline{X})\right] = \frac{n}{2} \mathbb{E}\left[(\underline{X} - \underline{Y})(\underline{F}(\underline{X}) - \underline{F}(\underline{Y}))\right]$$

↳ Mimics Gaussian IBP formula.

$(X-Y)^2 \to$ variance. $\frac{F(X) - F(Y)}{X - Y} \to f'$

Proof. $\frac{1}{n}\cdot\mathbb{E}[X F(X)] = \mathbb{E}[\,\mathbb{E}[X-Y \mid X]\cdot F(X)] = \mathbb{E}[(X-Y)F(X)]$

$\frac{1}{n}\cdot\mathbb{E}[XF(X)] \underset{\text{exch.}}{=} \mathbb{E}[(Y-X)F(Y)] = -\mathbb{E}[(X-Y)F(Y)]$

Average.

③ Lemma. (Mean-value trace ineq.)

For (Hermitian matrices) $A, B \in \mathbb{H}_d$, $p \in \mathbb{N}$

$$\operatorname{tr}\left[(A-B)(A^{2p-1}-B^{2p-1})\right] \le \frac{2p-1}{2}\operatorname{tr}\left[(A-B)^2(A^{2p-2}+B^{2p-2})\right]$$

Proof sketch. Scalar case: $p=1$ trivial. $p \ge 2$.

$$(a-b)(a^{2p-1}-b^{2p-1}) = (a-b)^2\cdot\frac{a^{2p-1}-b^{2p-1}}{a-b}$$

$$= (a-b)^2\cdot(2p-1)\cdot\int_0^1 (\tau a+(1-\tau)b)^{2p-2}\,d\tau \qquad \text{(Hermite div. diff.)}$$

$$\le (a-b)^2\cdot(2p-1)\cdot\int_0^1\left[\tau a^{2p-2}+(1-\tau)b^{2p-2}\right]d\tau \qquad \text{Jensen}$$

$$= \tfrac{1}{2}(a-b)^2\cdot(2p-1)\cdot(a^{2p-2}+b^{2p-2}).$$

↳ Lift to matrices via spectral resolution + $\operatorname{tr}[PQ] \ge 0$ for orthog. projectors P, Q.

④ Proof of theorem.

$$\mathbb{E}\operatorname{tr}X^{2p} = \mathbb{E}\operatorname{tr}[X\cdot X^{2p-1}]$$

$$= \frac{n}{2}\mathbb{E}\operatorname{tr}\left[(X-Y)(X^{2p-1}-Y^{2p-1})\right] \qquad \text{(MoEP)}$$

$$\le \frac{n}{4}(2p-1)\,\mathbb{E}\operatorname{tr}\left[(X-Y)^2(X^{2p-2}+Y^{2p-2})\right] \qquad \text{(MVTI)}$$

$$= \frac{n}{2}(2p-1)\,\mathbb{E}\operatorname{tr}\left[(X-Y)^2X^{2p-2}\right] \qquad \text{(exch.)}$$

$$= \frac{n}{2}(2p-1)\,\mathbb{E}\left[(\varepsilon_I-\varepsilon_I')^2\operatorname{tr}(A_I^2X^{2p-2})\right] \qquad \text{(def}^{\underline{n}}\text{)}$$

$$= (2p-1)\cdot\mathbb{E}\left[\operatorname{tr}(nA_I^2\cdot X^{2p-2})\right] \qquad \mathbb{E}'$$

$$= (2p-1)\cdot\mathbb{E}\operatorname{tr}\left[\left(\textstyle\sum_{i=1}^n A_i^2\right)\cdot X^{2p-2}\right] \qquad \mathbb{E}_I$$

$$\le (2p-1)\cdot\left[\operatorname{tr}\left(\textstyle\sum_{i=1}^n A_i^2\right)^p\right]^{1/p}\left[\mathbb{E}\operatorname{tr}X^{2p}\right]^{(p-1)/p} \qquad \text{Hölder}$$

Solve: $\left(\mathbb{E}\operatorname{tr}X^{2p}\right)^{1/2p} \le \sqrt{2p-1}\left[\operatorname{tr}\left(\sum_{i=1}^n A_i^2\right)^p\right]^{1/2p}.$

EXTENSIONS: • Rectangular case....

Cor. $B_1, \dots, B_n \in \mathbb{C}^{d_1 \times d_2}$

$Z = \sum_{i=1}^n \varepsilon_i B_i$.

For $p \in \mathbb{N}$, $\left(\mathbb{E}\, \|Z\|_{S_{2p}}^{2p}\right)^{1/2p} \le \sqrt{2p-1}\ \left\|\left(\sum B_i B_i^*\right)^{1/2}\right\|_{S_{2p}} \vee \left\|\left(\sum B_i^* B_i\right)^{1/2}\right\|_{S_{2p}}$

Pf. Take $A_i = \begin{bmatrix} 0 & B_i \\ B_i^* & 0 \end{bmatrix}$.

• Rosenthal–Pinelis ineq. for matrices

Cor. $\underline{X} = \sum_{i=1}^n \underline{X}_i$ with $\underline{X}_i \in \mathbb{H}$ indep. ~~Bxxxxxx~~

$$\left(\mathbb{E}\, \|\underline{X}\|_{4p}^{4p}\right)^{1/4p} \le \sqrt{4p-1}\ \left\|\left(\sum_i \mathbb{E} X_i^2\right)^{1/2}\right\|_{4p}$$
$$+ (4p-1)\left(\sum_i \mathbb{E}\, \|X_i\|_{4p}^{4p}\right)^{1/4p}.$$

• Moment ineq. associated with matrix Bernstein.

• Others: Chernoff, Bernstein, Bennett, McDiarmid, Efron–Stein, Freedman, via this or related methods ...

EXAMPLE: Rademacher Toeplitz matrix

$$T = \begin{bmatrix} \varepsilon_0 & \varepsilon_1 & \cdots & \\ \varepsilon_1 & \ddots & \ddots & \varepsilon_1 \\ \vdots & \ddots & \varepsilon_1 & \varepsilon_0 \end{bmatrix}_{d \times d} \quad \text{where } \varepsilon_i \text{ iid Rademacher}$$

$$= \varepsilon_0 I + \sum_{i=1}^d \varepsilon_i (S^i) + \sum_{i=1}^d \varepsilon_{-i} (S^*)^i \qquad \text{Sum of indep. rdm matrices}$$

Compute ~~[illegible]~~

Sum of squares: $I^2 + \sum_{i=1}^d (S^i)(S^i)^* + \sum_{i=1}^d (S^*)^i ((S^*)^i)^* = (d-1) I$

$\hookrightarrow\ \mathbb{E}\, \|T\| \le \left(\mathbb{E}\, \|T\|_{S_{2p}}^{2p}\right)^{1/2p} \qquad p = \log d$

$\le \sqrt{(2\log d) - 1}\ \left[\operatorname{tr} (d-1) I\right]^{1/2p}$

$\le e \sqrt{2 \log d} \longrightarrow$ can improve $e \to 1$.

Optimal const (Sen–Virág 2012)

$\geq 0.8288 \sqrt{2 \log d}$