

# Fluctuations of eigenvalues and second order Poincaré inequalities

Sourav Chatterjee (UC Berkeley)

# Linear statistics of random eigenvalues

- ▶ Let  $A$  be a random matrix with eigenvalues  $\lambda_1, \dots, \lambda_n$ .
- ▶ Let  $f$  be any function on  $\mathbb{R}$  and let  $W = \sum_{i=1}^n f(\lambda_i)$ . This is called a “linear statistic of the eigenvalues of  $A$ ”.
- ▶ Important in theory and applications.
- ▶ Theorem (Sinai & Soshnikov '98): If  $A$  is symmetric and the entries of  $A$  are i.i.d. with mean 0 and variance  $1/n$ , then under some further assumptions,
  - ▶  $\text{Var}(W)$  converges to a positive limit  $\sigma^2$  as  $n \rightarrow \infty$ .
  - ▶ Moreover,  $W - \mathbb{E}(W)$  converges in law to  $N(0, \sigma^2)$ .
- ▶ Proved using the method of moments.

- ▶ Similar results hold for Wishart matrices (Jonsson '82, Bai & Silverstein '04), random unitary matrices (Diaconis & Evans '01), beta ensembles (Johansson '98), Ginibre ensemble (Rider & Virág '06), Hankel matrices (Basor & Chen '05), band matrices (Anderson & Zeitouni '06), etc.
- ▶ The proofs are rather difficult and each case needs its own proof.
- ▶ Is it possible to devise a **soft** and **unified** method of proof? The present work is an attempt in that direction.

# 'Second order Poincaré inequality'

- ▶ Suppose  $\mathbf{X} = (X_1, \dots, X_n)$  is a vector of independent random variables, and  $g(\mathbf{X})$  is a function of  $\mathbf{X}$ .
- ▶ The Poincaré inequality tells us that if  $X_1, \dots, X_n$  are i.i.d. standard gaussians, then

$$\text{Var}(g(\mathbf{X})) \leq \sum_{i=1}^n \mathbb{E} \left( \frac{\partial g}{\partial x_i} \right)^2.$$

So, smallness of  $\nabla g \implies$  small variance of  $g(\mathbf{X})$ .

- ▶ Is it possible that if the **Hessian**

$$\nabla^2 g := \left( \frac{\partial^2 g}{\partial x_i \partial x_j} \right)_{1 \leq i, j \leq n}$$

is 'small' in some sense, then  $g(\mathbf{X})$  is approximately gaussian?

- ▶ This is what we introduce as the notion of a **second order Poincaré inequality**.

## Some evidence

- ▶ Suppose  $B$  is an  $n \times n$  real symmetric matrix, and

$$g(\mathbf{x}) = \mathbf{x}^t B \mathbf{x}.$$

- ▶ Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a vector of i.i.d. standard gaussian r.v.
- ▶ When is  $g(\mathbf{X})$  approximately gaussian?
- ▶ If  $\lambda_1, \lambda_2, \dots, \lambda_n$  are the eigenvalues of  $B$  with eigenvectors  $\mathbf{u}_1, \dots, \mathbf{u}_n$ , then

$$g(\mathbf{X}) = \sum_{i=1}^n \lambda_i Y_i^2,$$

where  $Y_i = \mathbf{u}_i^t \mathbf{X}$ .

- ▶  $Y_1, \dots, Y_n$  are again i.i.d. standard gaussian.
- ▶ This seems to suggest that  $g(\mathbf{X})$  is approximately gaussian if and only if 'no eigenvalue dominates in the sum'.
- ▶ Indeed,  $g(\mathbf{X})$  is approximately gaussian if and only if  $\max_i \lambda_i^2 \ll \sum_{i=1}^n \lambda_i^2$ .

## Evidence contd...

- ▶ Now  $\nabla^2 g(\mathbf{x}) \equiv 2B$ , where  $\nabla^2 g$  denotes the Hessian matrix of  $g$ .
- ▶ Thus,  $\|\nabla^2 g(\mathbf{X})\| = 2 \max_i |\lambda_i|$ .
- ▶ Again,

$$\text{Var}(g(\mathbf{X})) = 2 \sum_{i=1}^n \lambda_i^2.$$

- ▶ For general  $g$ , is it possible that gaussianity of  $g(\mathbf{X})$  holds whenever

$$\text{typical value of } \|\nabla^2 g(\mathbf{X})\|^2 \ll \text{Var}(g(\mathbf{X}))?$$

- ▶ What we have in mind...

vector  $\mathbf{X} \rightarrow$  matrix  $A(\mathbf{X}) \rightarrow$  linear statistic  $\sum_i f(\lambda_i) =: g(\mathbf{X})$ .

# Towards a general theorem: Stein's method

- ▶ If  $Z \sim N(0, 1)$ , then  $\mathbb{E}(\varphi(Z)Z) = \mathbb{E}(\varphi'(Z))$  for all  $\varphi$ .
- ▶ Stein's idea: If  $\mathbb{E}(\varphi(W)W) \approx \mathbb{E}(\varphi'(W))$  for many  $\varphi$ 's, then  $W$  is approximately  $N(0, 1)$ .
- ▶ Many variants, e.g.
  - ▶ Exchangeable pairs (Stein)
  - ▶ Zero bias couplings (Goldstein & Reinert)
  - ▶ Size bias couplings (Goldstein & Rinott)
  - ▶ Generator approach (Barbour)
  - ▶ Dependency graphs (Baldi & Rinott; Arratia, Goldstein, & Gordon)
- ▶ Common complaint: Hard to apply to problems that are not tailor-made for Stein's method.

# Duality between normal approximation and concentration

- ▶ Define

$$S_p(W) := \sup\{|\mathbb{E}(\varphi(W)W - \varphi'(W))| : \|\varphi'(W)\|_p \leq 1\}.$$

- ▶ From Stein's lemma:  $d_{TV}(W, Z) \leq 2S_p(W)$  for every  $p > 1$ .  
(Recall:  $d_{TV}(X, Y) = \sup_A |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|$ .)

## Theorem

If  $W$  has mean zero, unit variance, and density  $\rho$ , then

$$S_p(W) = \|h(W) - \mathbb{E}h(W)\|_q,$$

where

$$h(x) = \frac{\int_x^\infty y\rho(y)dy}{\rho(x)},$$

and  $\frac{1}{p} + \frac{1}{q} = 1$ .

- ▶ This shows that approximate normality of  $W$  = concentration of  $h(W)$ .
- ▶ Proof is based on  $L^p$ - $L^q$  duality from functional analysis.



- ▶ Concentration problems, unlike normal approximation problems, are *transferable* via conditional expectation. That is, if we can write

$$h(W) = \mathbb{E}(T \mid W),$$

where  $T$  is a an explicit object arising from the given problem, then

$$\|h(W) - \mathbb{E}h(W)\|_q \leq \|T - \mathbb{E}T\|_q.$$

# The key lemma

## Lemma

Suppose  $W = f(X_1, \dots, X_n)$ , where  $X_i$ 's are i.i.d.  $N(0, 1)$ , and  $f$  is smooth. Assume  $\mathbb{E}(W) = 0$  and  $\mathbb{E}(W^2) = 1$ . Let  $\mathbf{Z}$  be a vector of  $n$  i.i.d. standard gaussians and define  $T : \mathbb{R}^n \rightarrow \mathbb{R}$  as

$$T(\mathbf{x}) := \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{x}) \int_0^1 \frac{1}{2\sqrt{t}} \mathbb{E} \left( \frac{\partial f}{\partial x_i}(\sqrt{t}\mathbf{x} + \sqrt{1-t}\mathbf{Z}) \right) dt.$$

Then  $h(W) = \mathbb{E}(T(X_1, \dots, X_n) | W)$ .

(Recall: This implies that  $d_{TV}(W, N(0, 1)) \leq 2\sqrt{\text{Var}(h(W))}$ .)

# Sketch of proof

- ▶ For  $0 \leq t \leq 1$ , let  $\mathbf{X}^t = \sqrt{t}\mathbf{X} + \sqrt{1-t}\mathbf{Z}$ . Then

$$\begin{aligned}\mathbb{E}(\varphi(W)W) &= \mathbb{E}(\varphi(W)(f(\mathbf{X}^1) - f(\mathbf{X}^0))) \\ &= \mathbb{E}\left(\varphi(W) \int_0^1 \frac{d}{dt} f(\mathbf{X}^t) dt\right).\end{aligned}$$

- ▶ Note that

$$\frac{d}{dt} f(\mathbf{X}^t) = \sum_{i=1}^n \left( \frac{X_i}{2\sqrt{t}} - \frac{Z_i}{2\sqrt{1-t}} \right) \frac{\partial}{\partial x_i} f(\mathbf{X}^t).$$

- ▶ Applying integration-by-parts, we can now get

$$\begin{aligned}\mathbb{E}(\varphi(W)W) &= \mathbb{E}\left(\varphi'(W) \sum_{i=1}^n \int_0^1 \frac{1}{2\sqrt{t}} \frac{\partial f}{\partial x_i}(\mathbf{X}) \frac{\partial f}{\partial x_i}(\mathbf{X}^t) dt\right) \\ &= \mathbb{E}(\varphi'(W)T(\mathbf{X})).\end{aligned}$$

- ▶ Since  $h$  is characterized by  $\mathbb{E}(\varphi(W)W) = \mathbb{E}(\varphi'(W)h(W))$ , this shows that  $h(W) = \mathbb{E}(T|W)$ .

# Simple examples

$f(X_1, \dots, X_n)$	$T(X_1, \dots, X_n)$
$\frac{\sum_{i=1}^n X_i}{\sqrt{n}}$	1
$\frac{\sum_{i=1}^n X_i^2 - n}{\sqrt{2n}}$	$\frac{\sum_{i=1}^n X_i^2}{n}$
$\frac{\sum_{i=1}^n X_i X_{i+1}}{\sqrt{n}}$	$\frac{1}{2n} \sum_{i=2}^n (X_{i-1} + X_{i+1})^2 + \frac{X_1^2 + X_{n+1}^2}{2n}$

# How does one bound $\text{Var}(T)$ in general?

- ▶ **Answer:** By the gaussian **Poincaré inequality**. If  $\mathbf{X} = (X_1, \dots, X_n)$  is a vector of i.i.d.  $N(0, 1)$  r.v., and  $T : \mathbb{R}^n \rightarrow \mathbb{R}$  is absolutely continuous, then

$$\text{Var}(T(\mathbf{X})) \leq \mathbb{E} \|\nabla T(\mathbf{X})\|^2.$$

(Recall:  $\nabla T = (\partial T / \partial x_1, \dots, \partial T / \partial x_n)$  is the gradient of  $T$ .)

# A second order Poincaré inequality

## Theorem

Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a vector of i.i.d. standard gaussian random variables. Take any  $g \in C^2(\mathbb{R}^n)$  and let  $\nabla g$  and  $\nabla^2 g$  denote the *gradient* and *Hessian* of  $g$ . Let

$$\kappa_1 = (\mathbb{E}\|\nabla g(\mathbf{X})\|^4)^{1/4}, \text{ and}$$

$$\kappa_2 = (\mathbb{E}\|\nabla^2 g(\mathbf{X})\|^4)^{1/4}.$$

Suppose  $W = g(\mathbf{X})$  has a finite fourth moment and let  $\sigma^2 = \text{Var}(W)$ . Let  $Z$  be a normal random variable having the same mean and variance as  $W$ . Then

$$d_{TV}(W, Z) \leq \frac{2\sqrt{5}\kappa_1\kappa_2}{\sigma^2}.$$

**Remark:** Think of  $\kappa_1$  and  $\kappa_2$  as the *typical sizes* of  $\|\nabla g\|$  and  $\|\nabla^2 g\|$ .

# An example

- ▶ Consider the function

$$g(\mathbf{x}) = \sum_{i=1}^{n-2} x_i x_{i+1} x_{i+2},$$

and let  $W = g(\mathbf{X})$ .

- ▶ Easily, we have

$$\kappa_1 \sim \text{typical size of } \|\nabla g\| = O(\sqrt{n}),$$

and

$$\kappa_2 \sim \text{typical size of } \|\nabla^2 g\| = O(1).$$

- ▶ Again,  $\sigma^2 = \text{Var}(W) \geq Cn$  for some positive constant  $C$ .
- ▶ Thus, if  $Z \sim N(0, \sigma^2)$ , then

$$d_{TV}(W, Z) \leq \frac{2\sqrt{5}\kappa_1\kappa_2}{\sigma^2} \leq \frac{\text{const.}}{\sqrt{n}}.$$

# Application to random matrices

- ▶ General plan:

vector  $\mathbf{X} \rightarrow$  matrix  $A(\mathbf{X}) \rightarrow$  linear statistic  $\sum_i f(\lambda_i) =: g(\mathbf{X})$ .

Use previous theorem to prove CLT for  $g(\mathbf{X})$ .

- ▶ I have worked out the details for:
  1. Wigner matrices.
  2. Wishart matrices.
  3. Double Wishart matrices.
  4. Gaussian matrices with arbitrary correlation structure.
  5. Gaussian Toeplitz matrices.
- ▶ Convergence rates are also obtained. For example, for Wigner matrices, the TV rate of convergence is  $n^{-1}$  for gaussian and  $n^{-1/2}$  for non-gaussian.
- ▶ The result for Toeplitz matrices is new. However, there are no formulas for the limiting variance.
- ▶ Details of computations are not presentable in a seminar.



# A special class of distributions

- ▶ Let  $\mathcal{L}(c_1, c_2)$  be the class of probability measures on  $\mathbb{R}$  that arise as laws random variables like  $u(Z)$ , where  $Z \sim N(0, 1)$  and  $u \in C^2(\mathbb{R})$  satisfies

$$|u'(x)| \leq c_1 \text{ and } |u''(x)| \leq c_2.$$

# A general 'tool' for linear statistics

- ▶ Let  $n$  be a fixed positive integer and  $\mathcal{J}$  be a finite indexing set.
- ▶ For each  $i, j$ , we have a  $C^2$  map  $a_{ij} : \mathbb{R}^{\mathcal{J}} \rightarrow \mathbb{C}$ .
- ▶ For each  $\mathbf{x} \in \mathbb{R}^{\mathcal{J}}$ , let

$$A(\mathbf{x}) = (a_{ij}(\mathbf{x}))_{1 \leq i, j \leq n}.$$

- ▶ Let  $f$  be an analytic function on the real line.
- ▶ Let  $g(\mathbf{x}) = \operatorname{Re} \operatorname{Tr}[f(A(\mathbf{x}))]$ .

# Measuring the size of the derivatives of $g$

- ▶ First, let

$$\mathcal{R} = \{\alpha \in \mathbb{C}^{\mathcal{J}} : \sum_{u \in \mathcal{J}} |\alpha_u|^2 = 1\} \text{ and}$$
$$\mathcal{S} = \{\beta \in \mathbb{C}^{n \times n} : \sum_{i,j=1}^n |\beta_{ij}|^2 = 1\}.$$

- ▶ Next, define three functions

$$\gamma_0(\mathbf{x}) := \sup_{u \in \mathcal{J}, \|B\|=1} \left| \text{Tr} \left( B \frac{\partial A}{\partial x_u} \right) \right|,$$

$$\gamma_1(\mathbf{x}) := \sup_{\alpha \in \mathcal{R}, \beta \in \mathcal{S}} \left| \sum_{u \in \mathcal{J}} \sum_{i,j=1}^n \alpha_u \beta_{ij} \frac{\partial a_{ij}}{\partial x_u} \right|, \text{ and}$$

$$\gamma_2(\mathbf{x}) := \sup_{\alpha, \alpha' \in \mathcal{R}, \beta \in \mathcal{S}} \left| \sum_{u,v \in \mathcal{J}} \sum_{i,j=1}^n \alpha_u \alpha'_v \beta_{ij} \frac{\partial^2 a_{ij}}{\partial x_u \partial x_v} \right|.$$

## Measuring derivatives contd...

- ▶ Let  $f(z) = \sum_{m=0}^{\infty} b_m z^m$  be an entire function.
- ▶ Let

$$f_1(z) = \sum_{m=1}^{\infty} m |b_m| z^{m-1} \quad \text{and} \quad f_2(z) = \sum_{m=2}^{\infty} m(m-1) |b_m| z^{m-2}.$$

- ▶ Let  $a(\mathbf{x}) = f_1(\|A(\mathbf{x})\|)$  and  $b(\mathbf{x}) = f_2(\|A(\mathbf{x})\|)$ .
- ▶ Define three more functions

$$\eta_0(\mathbf{x}) = \gamma_0(\mathbf{x}) a(\mathbf{x}),$$

$$\eta_1(\mathbf{x}) = \gamma_1(\mathbf{x}) a(\mathbf{x}) \sqrt{n}, \quad \text{and}$$

$$\eta_2(\mathbf{x}) = \gamma_2(\mathbf{x}) a(\mathbf{x}) \sqrt{n} + \gamma_1(\mathbf{x})^2 b(\mathbf{x}).$$

# The end product

- ▶ Let  $\mathbf{X} = (X_u)_{u \in \mathcal{J}}$  be independent r.v. in  $\mathcal{L}(c_1, c_2)$ .
- ▶ Let

$$\kappa_0 = (\mathbb{E}(\eta_0(\mathbf{X})^2 \eta_1(\mathbf{X})^2))^{1/2},$$

$$\kappa_1 = (\mathbb{E} \eta_1(\mathbf{X})^4)^{1/4}, \quad \text{and}$$

$$\kappa_2 = (\mathbb{E} \eta_2(\mathbf{X})^4)^{1/4}.$$

## Theorem

Let  $W = g(\mathbf{X}) = \text{Re Tr}[f(A(\mathbf{X}))]$ . Suppose  $W$  has finite fourth moment and let  $\sigma^2 = \text{Var}(W)$ . Let  $Z$  be a gaussian r.v. with the same mean and variance as  $W$ . Then

$$d_{TV}(W, Z) \leq \frac{2\sqrt{5}(c_1 c_2 \kappa_0 + c_1^3 \kappa_1 \kappa_2)}{\sigma^2}.$$

# Summarizing...

- ▶ The general theorem gives a way of proving CLTs for linear statistics of random matrices that are expressible as functions of independent random variables.
- ▶ Works by bounding first and second order derivatives.
- ▶ Gives total variation **error bounds**.
- ▶ Main weakness: Need *a priori* lower bound for the variance of the linear statistic.
- ▶ Other problems: (i) Hard to apply to matrices that are not easily expressible as functions of independent random variables, e.g. random orthogonal and unitary matrices. (ii) Restrictions on the distributions of the entries.
- ▶ Paper available on arxiv at the URL <http://arxiv.org/abs/0705.1224>  
(To appear in PTRF.)