Another observation about operator compressions

Elizabeth Meckes
joint work with Mark Meckes

Case Western Reserve University

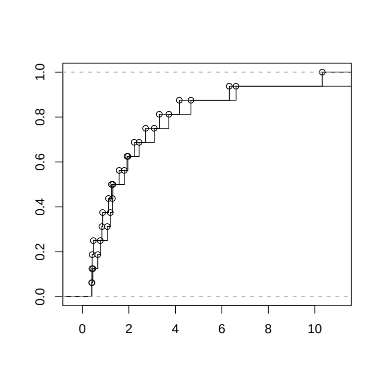
An observation about submatrices by Chatterjee and Ledoux

An observation about submatrices by Chatterjee and Ledoux

Let M be an $n \times n$ Hermitian matrix, and let $1 \ll k \leq n$. Then the empirical spectral distributions of most $k \times k$ principal submatrices of M are about the same.

An observation about submatrices by Chatterjee and Ledoux

Let M be an $n \times n$ Hermitian matrix, and let $1 \ll k \leq n$. Then the empirical spectral distributions of most $k \times k$ principal submatrices of M are about the same.



More formally:

More formally:

Theorem (Chatterjee-Ledoux)

More formally:

Theorem (Chatterjee-Ledoux)

For M given, let A be chosen uniformly at random from all $k \times k$ principal submatrices. Let F_A denote the empirical distribution function of A; that is,

$$F_A(x) = \frac{1}{k}\left|\{j : \lambda_j(A) \leq x\}\right|.$$

More formally:

Theorem (Chatterjee-Ledoux)

For M given, let A be chosen uniformly at random from all $k \times k$ principal submatrices. Let F_A denote the empirical distribution function of A; that is,

$$F_A(x) = \frac{1}{k}\left|\{j : \lambda_j(A) \leq x\}\right|.$$

Let $F(x) := \mathbb{E}F_A(x)$.

More formally:

Theorem (Chatterjee-Ledoux)

For M given, let A be chosen uniformly at random from all $k \times k$ principal submatrices. Let F_A denote the empirical distribution function of A; that is,

$$F_A(x) = \frac{1}{k}\left|\{j : \lambda_j(A) \leq x\}\right|.$$

Let $F(x) := \mathbb{E}F_A(x)$. Then for $r > 0$,

$$\mathbb{P}[\|F_A - F\|_\infty \geq k^{-1/2} + r] \leq 12\sqrt{k}e^{-r\sqrt{k/8}}$$

More formally:

Theorem (Chatterjee-Ledoux)

For M given, let A be chosen uniformly at random from all $k \times k$ principal submatrices. Let F_A denote the empirical distribution function of A; that is,

$$F_A(x) = \frac{1}{k}\big|\{j : \lambda_j(A) \leq x\}\big|.$$

Let $F(x) := \mathbb{E}F_A(x)$. Then for $r > 0$,

$$\mathbb{P}[\|F_A - F\|_\infty \geq k^{-1/2} + r] \leq 12\sqrt{k}e^{-r\sqrt{k/8}}$$

and

$$\mathbb{E}\|F_A - F\|_\infty \leq \frac{13 + \sqrt{8}\log(k)}{\sqrt{k}}.$$

Outline of the proof

Outline of the proof

1. Functions of finite state space Markov chains are concentrated at their means with respect to the stationary distribution of the chain, with bounds in terms of the spectral gap of the chain.

Outline of the proof

1. Functions of finite state space Markov chains are concentrated at their means with respect to the stationary distribution of the chain, with bounds in terms of the spectral gap of the chain.
2. The transposition random walk on S_n has uniform measure as stationary distribution.

Outline of the proof

1. Functions of finite state space Markov chains are concentrated at their means with respect to the stationary distribution of the chain, with bounds in terms of the spectral gap of the chain.
2. The transposition random walk on S_n has uniform measure as stationary distribution.
3. Good bounds (due to Diaconis and Shahshahani) on the spectral gap of the transposition random walk are available.

Outline of the proof

1. Functions of finite state space Markov chains are concentrated at their means with respect to the stationary distribution of the chain, with bounds in terms of the spectral gap of the chain.
2. The transposition random walk on S_n has uniform measure as stationary distribution.
3. Good bounds (due to Diaconis and Shahshahani) on the spectral gap of the transposition random walk are available.
4. It's not too hard to get from concentration of $F_A(x)$ near $F(x)$ to concentration of $\|F_A - F\|_\infty$.

Away from matrices: the coordinate-free viewpoint

Away from matrices: the coordinate-free viewpoint

Let T be a self-adjoint operator on an n-dimensional Hilbert space $\mathcal{H}$. Let E be a subspace of $\mathcal{H}$, and let $\pi_E : \mathcal{H} \to E$ denote orthogonal projection.

Away from matrices: the coordinate-free viewpoint

Let T be a self-adjoint operator on an n-dimensional Hilbert space $\mathcal{H}$. Let E be a subspace of $\mathcal{H}$, and let $\pi_E : \mathcal{H} \to E$ denote orthogonal projection.

The *compression* of T to E is the operator

$$T_E := \pi_E T|_E = \pi_E T \pi_E^*.$$

The spectral distribution of T_E is defined to be the measure

$$\mu_E := \frac{1}{k} \sum_{j=1}^{k} \delta_{\lambda_j(T_A)},$$

where $\lambda_1(T_A) \geq \cdots \geq \lambda_k(T_A)$ are the eigenvalues of T_A.

Our observation

Our observation

For T a given self-adjoint operator on an n-dimensional Hilbert space $\mathcal{H}$ and $1 \leq k \leq n$, most compressions of T to k-dimensional subspaces have spectral distributions which are about the same.

More formally:

More formally:

Theorem (E. M. -M. M.)

More formally:

Theorem (E. M. -M. M.)

If E is a k-dimensional random subspace of $\mathcal{H}$ distributed according to the rotation-invariant probability measure on the Grassmannian, μ_E is the spectral measure of T_E and $\mu := \mathbb{E}\mu_E$, then

More formally:

Theorem (E. M. -M. M.)

If E is a k-dimensional random subspace of $\mathcal{H}$ distributed according to the rotation-invariant probability measure on the Grassmannian, μ_E is the spectral measure of T_E and $\mu := \mathbb{E}\mu_E$, then

$$\mathbb{E}d_1(\mu_E, \mu) \leq A\frac{\sigma_k(T)^{4/7}\rho(T)^{3/7}}{(kn)^{2/7}},$$

More formally:

Theorem (E. M. -M. M.)

If E is a k-dimensional random subspace of $\mathcal{H}$ distributed according to the rotation-invariant probability measure on the Grassmannian, μ_E is the spectral measure of T_E and $\mu := \mathbb{E}\mu_E$, then

$$\mathbb{E}d_1(\mu_E, \mu) \leq A\frac{\sigma_k(T)^{4/7}\rho(T)^{3/7}}{(kn)^{2/7}},$$

and

$$\mathbb{P}\left[d_1(\mu_E, \mu) \geq A\frac{\sigma_k(T)^{4/7}\rho(T)^{3/7}}{(kn)^{2/7}} + t\right] \leq B\exp\left[-C\frac{knt^2}{\sigma_k^2(T)}\right].$$

More formally:

Theorem (E. M. -M. M.)

If E is a k-dimensional random subspace of $\mathcal{H}$ distributed according to the rotation-invariant probability measure on the Grassmannian, μ_E is the spectral measure of T_E and $\mu := \mathbb{E}\mu_E$, then

$$\mathbb{E}d_1(\mu_E, \mu) \leq A\frac{\sigma_k(T)^{4/7}\rho(T)^{3/7}}{(kn)^{2/7}},$$

and

$$\mathbb{P}\left[d_1(\mu_E, \mu) \geq A\frac{\sigma_k(T)^{4/7}\rho(T)^{3/7}}{(kn)^{2/7}} + t\right] \leq B\exp\left[-C\frac{knt^2}{\sigma_k^2(T)}\right].$$

Here,

$$\rho(T) := \frac{1}{2}\left[\lambda_1(T) - \lambda_n(T)\right] \qquad \sigma_k(T) := \inf_\lambda \sqrt{\sum_{i=1}^{k} s_i^2(T - \lambda I)}.$$

A note on distance

A note on distance

We use the Kantorovich-Rubenstein distance

$$\begin{aligned} d_1(\mu,\nu) &= \inf_{\pi} \int_{\mathbb{R}\times\mathbb{R}} |x-y| d\pi(x,y) \\ &= \sup_{f} \left| \int f d\mu - \int f d\nu \right| \\ &= \|F_\mu - F_\nu\|_{L_1(\mathbb{R})}, \end{aligned}$$

where π varies over probability measures on $\mathbb{R} \times \mathbb{R}$ with margins μ and ν, and f varies over functions on $\mathbb{R}$ with $\|f'\|_\infty \leq 1$.

A note on distance

We use the Kantorovich-Rubenstein distance

$$\begin{aligned} d_1(\mu,\nu) &= \inf_{\pi} \int_{\mathbb{R}\times\mathbb{R}} |x-y| d\pi(x,y) \\ &= \sup_{f} \left| \int f d\mu - \int f d\nu \right| \\ &= \|F_\mu - F_\nu\|_{L_1(\mathbb{R})}, \end{aligned}$$

where π varies over probability measures on $\mathbb{R} \times \mathbb{R}$ with margins μ and ν, and f varies over functions on $\mathbb{R}$ with $\|f'\|_\infty \leq 1$.

This distance is not directly comparable to the Kolmogorov distance $\|F_\mu - F_\nu\|_\infty$ in general, although some comparison can be made here due to the finite support of the measures in question.

Outline of Proof

Outline of Proof

We also use measure concentration as a key ingredient, but the measure in question is the rotation-invariant probability measure on the Grassmannian $\mathfrak{G}_k(\mathcal{H})$ of k-dimensional subspaces of $\mathcal{H}$.

Outline of Proof

We also use measure concentration as a key ingredient, but the measure in question is the rotation-invariant probability measure on the Grassmannian $\mathfrak{G}_k(\mathcal{H})$ of k-dimensional subspaces of $\mathcal{H}$.

Recall that one can define a metric $d(E, F)$ on $\mathfrak{G}_k(\mathcal{H})$ by

$$d(E, F) := \inf \sqrt{\sum_{i=1}^{k} \|e_i - f_i\|^2},$$

where the infimum is over orthonormal bases $\{e_i\}_{i=1}^k$ and $\{f_i\}_{i=1}^k$ of E and F, respectively.

Outline of Proof

We also use measure concentration as a key ingredient, but the measure in question is the rotation-invariant probability measure on the Grassmannian $\mathfrak{G}_k(\mathcal{H})$ of k-dimensional subspaces of $\mathcal{H}$.

Recall that one can define a metric $d(E, F)$ on $\mathfrak{G}_k(\mathcal{H})$ by

$$d(E, F) := \inf \sqrt{\sum_{i=1}^{k} \|e_i - f_i\|^2},$$

where the infimum is over orthonormal bases $\{e_i\}_{i=1}^{k}$ and $\{f_i\}_{i=1}^{k}$ of E and F, respectively.

One has concentration about a fixed value for functions on $\mathfrak{G}_k(\mathcal{H})$ which are Lipschitz with respect to the distance $d(\cdot, \cdot)$.

Specifically:

Specifically:

Theorem (Gromov-Milman)

Let $f : \mathfrak{G}_k(\mathcal{H}) \to \mathbb{R}$ be 1-Lipschitz with respect to $d(\cdot,\cdot)$, and let E be distributed according to the rotation-invariant probability measure on $\mathfrak{G}_k(\mathcal{H})$. Then there is are absolute constants C, c such that

$$\mathbb{P}\left[\left|f(E) - \mathbb{E}f(E)\right| \geq t\right] \leq Ce^{-cnt^2}.$$

Specifically:

Theorem (Gromov-Milman)

Let $f : \mathfrak{G}_k(\mathcal{H}) \to \mathbb{R}$ be 1-Lipschitz with respect to $d(\cdot,\cdot)$, and let E be distributed according to the rotation-invariant probability measure on $\mathfrak{G}_k(\mathcal{H})$. Then there is are absolute constants C, c such that

$$\mathbb{P}\left[\left|f(E) - \mathbb{E}f(E)\right| \geq t\right] \leq Ce^{-cnt^2}.$$

We want to apply this theorem to the function $f(E) := d_1(\mu_E, \mu)$, where d_1 is the Kantorovich-Rubenstein distance, μ_E is the spectral distribution of the compression of T to E, and $\mu = \mathbb{E}\mu_E$.

The Lipschitz constant of $f(E) = d_1(\mu_E, \mu)$ can be bounded using the coupling $\pi = \frac{1}{k}\sum_{i=1}^{k} \delta_{(\lambda_i(T_E), \lambda_i(T_F))}$.

The Lipschitz constant of $f(E) = d_1(\mu_E, \mu)$ can be bounded using the coupling $\pi = \frac{1}{k} \sum_{i=1}^{k} \delta_{(\lambda_i(T_E), \lambda_i(T_F))}$.
A little matrix analysis shows that

$$d_1(\mu_E, \mu_F) \leq \frac{2\sigma_k}{\sqrt{k}} d(E, F).$$

The Lipschitz constant of $f(E) = d_1(\mu_E, \mu)$ can be bounded using the coupling $\pi = \frac{1}{k}\sum_{i=1}^{k} \delta_{(\lambda_i(T_E),\lambda_i(T_F))}$.
A little matrix analysis shows that

$$d_1(\mu_E, \mu_F) \leq \frac{2\sigma_k}{\sqrt{k}} d(E, F).$$

Together with the Gromov-Milman measure concentration result, this yields

The Lipschitz constant of $f(E) = d_1(\mu_E, \mu)$ can be bounded using the coupling $\pi = \frac{1}{k}\sum_{i=1}^{k} \delta_{(\lambda_i(T_E),\lambda_i(T_F))}$.
A little matrix analysis shows that

$$d_1(\mu_E, \mu_F) \leq \frac{2\sigma_k}{\sqrt{k}} d(E, F).$$

Together with the Gromov-Milman measure concentration result, this yields

$$\mathbb{P}\left[\left|d_1(\mu_E, \mu) - \mathbb{E}d_1(\mu_E, \mu)\right| > t\right] \leq C \exp\left[-c\frac{nkt^2}{\sigma_k^2}\right].$$

$$\mathbb{P}\left[\left|d_1(\mu_E,\mu)-\mathbb{E}d_1(\mu_E,\mu)\right|>t\right]\leq C\exp\left[-c\frac{nkt^2}{\sigma_k^2}\right].$$

$$\mathbb{P}\left[\left|d_1(\mu_E,\mu) - \mathbb{E}d_1(\mu_E,\mu)\right| > t\right] \le C\exp\left[-c\frac{nkt^2}{\sigma_k^2}\right].$$

That is, $d_1(\mu_E,\mu)$ is concentrated at its mean, and the challenge is to bound that mean.

$$\mathbb{P}\left[\left|d_1(\mu_E,\mu) - \mathbb{E}d_1(\mu_E,\mu)\right| > t\right] \le C\exp\left[-c\frac{nkt^2}{\sigma_k^2}\right].$$

That is, $d_1(\mu_E,\mu)$ is concentrated at its mean, and the challenge is to bound that mean.

Recall that $d_1(\mu_E,\mu) = \sup_f \left|\int f d\mu_E - \int f d\mu\right|$; we need to bound the expected maximum of a stochastic process indexed by $\{f : \|f'\|_\infty \le 1\}$.

Dudley's entropy bound

Dudley's entropy bound

Theorem (Dudley)

Let $\{X_t\}_{t\in T}$ be a stochastic process indexed by a metric space T with distance d. Suppose that there is a constant c such that X_t satisfies the increment condition

$$\forall u, \quad \mathbb{P}\left[|X_t - X_s| \geq u\right] \leq c \exp\left(-\frac{u^2}{2d(s,t)^2}\right).$$

Dudley's entropy bound

Theorem (Dudley)

Let $\{X_t\}_{t \in T}$ be a stochastic process indexed by a metric space T with distance d. Suppose that there is a constant c such that X_t satisfies the increment condition

$$\forall u, \quad \mathbb{P}\left[|X_t - X_s| \geq u\right] \leq c \exp\left(-\frac{u^2}{2d(s,t)^2}\right).$$

Then there is a constant C such that

$$\mathbb{E} \sup_{t \in T} X_t \leq C \int_0^\infty \sqrt{\log N(T, d, \epsilon)}\, d\epsilon,$$

where $N(T, d, \epsilon)$ is the ϵ-covering number of T with respect to the distance d.

The process $X_f := \int f d\mu_E - \int f d\mu$.

The process $X_f := \int f d\mu_E - \int f d\mu$.

- As a function of E, X_f is Lipschitz with Lipschitz constant $\frac{2\sigma_k}{\sqrt{k}}$ whenever f 1-Lipschitz.

The process $X_f := \int f d\mu_E - \int f d\mu$.

- As a function of E, X_f is Lipschitz with Lipschitz constant $\frac{2\sigma_k}{\sqrt{k}}$ whenever f 1-Lipschitz.
- Concentration of measure on $\mathfrak{G}_k(\mathcal{H})$ thus gives that

$$\mathbb{P}[|X_f - X_g| > u] \leq \mathbb{P}[|X_{f-g}| \geq u] \leq C \exp\left[-c\frac{nku^2}{\sigma_k^2 |f-g|_L}\right],$$

where $|f - g|$ is the Lipschitz constant of $f - g$.

The process $X_f := \int f d\mu_E - \int f d\mu$.

- As a function of E, X_f is Lipschitz with Lipschitz constant $\frac{2\sigma_k}{\sqrt{k}}$ whenever f 1-Lipschitz.
- Concentration of measure on $\mathfrak{G}_k(\mathcal{H})$ thus gives that

$$\mathbb{P}[|X_f - X_g| > u] \leq \mathbb{P}[|X_{f-g}| \geq u] \leq C \exp\left[-c \frac{nku^2}{\sigma_k^2 |f-g|_L}\right],$$

where $|f - g|$ is the Lipschitz constant of $f - g$.

$\Longrightarrow$ The process satisfies the sub-Gaussian increment condition with respect to $\|\cdot\|' := \frac{\sigma_k}{\sqrt{kn}}\|\cdot\|_{C^1}$.

The process $X_f := \int f d\mu_E - \int f d\mu$.

- As a function of E, X_f is Lipschitz with Lipschitz constant $\frac{2\sigma_k}{\sqrt{k}}$ whenever f 1-Lipschitz.
- Concentration of measure on $\mathfrak{G}_k(\mathcal{H})$ thus gives that

$$\mathbb{P}[|X_f - X_g| > u] \leq \mathbb{P}[|X_{f-g}| \geq u] \leq C \exp\left[-c\frac{nku^2}{\sigma_k^2 |f-g|_L}\right],$$

where $|f-g|$ is the Lipschitz constant of $f-g$.

$\Longrightarrow$ The process satisfies the sub-Gaussian increment condition with respect to $\|\cdot\|' := \frac{\sigma_k}{\sqrt{kn}}\|\cdot\|_{C^1}$.

Bad News: The covering number of $\{f : \|f'\|_\infty \leq 1\}$ with respect to $\|\cdot\|'$ is infinite. In fact, it suffices to consider $\{f : \|f\|_{C^1} \leq 1 + 2\rho\}$, but that still has infinite covering number.

Getting around it: Approximation

Getting around it: Approximation

The index set $\{f : \|f\|_{C^2} \leq 1\}$ has finite covering number with respect to $\|\cdot\|'$, and estimates for it are available.

Getting around it: Approximation

The index set $\{f : \|f\|_{C^2} \leq 1\}$ has finite covering number with respect to $\|\cdot\|'$, and estimates for it are available.

Those estimates yield:

$$\mathbb{E}\sup\{X_f : \|f\|_{C^2} \leq 1\} \lesssim \frac{\sigma_k\sqrt{\rho+1}}{\sqrt{kn}}.$$

Getting around it: Approximation

The index set $\{f : \|f\|_{C^2} \leq 1\}$ has finite covering number with respect to $\|\cdot\|'$, and estimates for it are available.

Those estimates yield:

$$\mathbb{E}\sup\{X_f : \|f\|_{C^2} \leq 1\} \lesssim \frac{\sigma_k\sqrt{\rho+1}}{\sqrt{kn}}.$$

Smoothing functions in $\{f : \|f\|_{C^1} \leq 1 + 2\rho\}$ and optimizing over the various parameters yields

$$\mathbb{E}d_1(\mu_E, \mu) \lesssim \frac{\sigma_k^{4/7}\rho^{3/7}}{(kn)^{2/7}}.$$

Comparisons with Chatterjee-Ledoux

Comparisons with Chatterjee-Ledoux

- The random subspace onto which the operator is projected is distributed differently:

Comparisons with Chatterjee-Ledoux

- The random subspace onto which the operator is projected is distributed differently: in C-L, E is uniformly chosen from the coordinate-subspaces of dimension k, whereas for us, E is uniformly chosen from *all* subspaces of dimension k. As such, quantitative comparison is necessarily rough.

Comparisons with Chatterjee-Ledoux

- The random subspace onto which the operator is projected is distributed differently: in C-L, E is uniformly chosen from the coordinate-subspaces of dimension k, whereas for us, E is uniformly chosen from *all* subspaces of dimension k. As such, quantitative comparison is necessarily rough.
- The C-L result is only interesting for $k \gg 1$, whereas our result has content for any k as long as $n \gg 1$.

Comparisons with Chatterjee-Ledoux

- The random subspace onto which the operator is projected is distributed differently: in C-L, E is uniformly chosen from the coordinate-subspaces of dimension k, whereas for us, E is uniformly chosen from *all* subspaces of dimension k. As such, quantitative comparison is necessarily rough.
- The C-L result is only interesting for $k \gg 1$, whereas our result has content for any k as long as $n \gg 1$.
- Our result is sensitive, via the appearance of $\sigma_k(T)$ and $\rho(T)$, to the proximity of T to the space of scalar matrices

Thank you.