ℓ^∞ Eigenvector Perturbation and Robust Covariance Estimation

Jianqing Fan, Weichen Wang and Yiqiao Zhong

Princeton University

Matrix Eigenvector Perturbation

Symmetric Matrices $A, \widetilde{A}, E$ with size p and $\widetilde{A} = A + E$. Eigendecomposition:

$$A = \sum_{i=1}^{p} \lambda_i v_i v_i^T, \quad \widetilde{A} = \sum_{i=1}^{p} \widetilde{\lambda}_i \widetilde{v}_i \widetilde{v}_i^T,$$

where $\lambda_1 \geq \ldots \geq \lambda_p$ and $\widetilde{\lambda}_1 \geq \ldots \geq \widetilde{\lambda}_p$. Let eigengap $\gamma := \min\{\lambda_i - \lambda_{i+1} : 1 \leq i \leq r\}$.

Eigenvector Perturbation: (Davis and Kahan 70')

$$\|v_i - \eta_i \widetilde{v}\|_2 \leq \frac{2\sqrt{2}\,\|E\|_2}{\gamma}, \quad i = 1, \ldots, r$$

where $\eta_i \in \{\pm 1\}$ are suitable signs.

Remarks: Originally for eigenspaces; extension for SVD: Wedin 72'.

Question: What about ℓ^∞ bound? Any sharper result than $\|\cdot\|_\infty \leq \|\cdot\|_2$?

Our goal: For incoherent low-rank matrices, a sharp ℓ^∞ bound.

Notations and assumptions

- Suppose the rank $r := \text{rank}(A)$ is small, e.g. bounded by a constant.
- Let p by r matrix $V = [v_1, \ldots, v_r]$. The coherence $\mu(V)$ is small, e.g. logarithmic in p.
- Let γ to be the smallest gap in $\{\lambda_1, \ldots, \lambda_r, 0\}$.
- Matrix infinity norm:

$$\|E\|_\infty = \sup_{\|x\|_\infty \leq 1} \|Ex\|_\infty = \max_i \sum_{j=1}^{p} |E_{ij}|.$$

Matrix coherence: (Candes and Recht 09')

Let $V = [v_1, \ldots, v_r]$ be r columns of orthonormal vectors in $\mathbf{R}^p$. The *coherence* of V is defined as

$$\mu(V) = \frac{p}{r} \max_i \sum_{j=1}^{r} V_{ij}^2.$$

- Small coherence means v_i not aligned with any coordinate.
- Complementary structure to sparsity.

ℓ^∞ Eigenvector Perturbation

If $\gamma > C(\mu, r)\|E\|_\infty$, then for suitable $\eta_i \in \{\pm 1\}$, and $C(\mu, r) = O(\mu^{1.5} r^{3.5})$.

$$\max_{1 \leq k \leq r} \|v_k - \eta_k \widetilde{v}_k\|_\infty \leq C(\mu, r) \frac{\|E\|_\infty}{\gamma\sqrt{p}}.$$

- **Remark**: A similar bound for SVD.
- **Simulations:**
  - p runs from 200 to 2000 with increment 200.
  - $A = \sum_{k=1}^{3}(4-k)\gamma v_k v_k^T$; v_k is an eigenvector of an iid normal random matrix.
  - Generating E: (a) random number in $[0, L]$ by randomly selecting s entries each row; (b) $E_{ij} = L'\rho^{|i-j|}$.
  - Report the largest error over 100 runs.

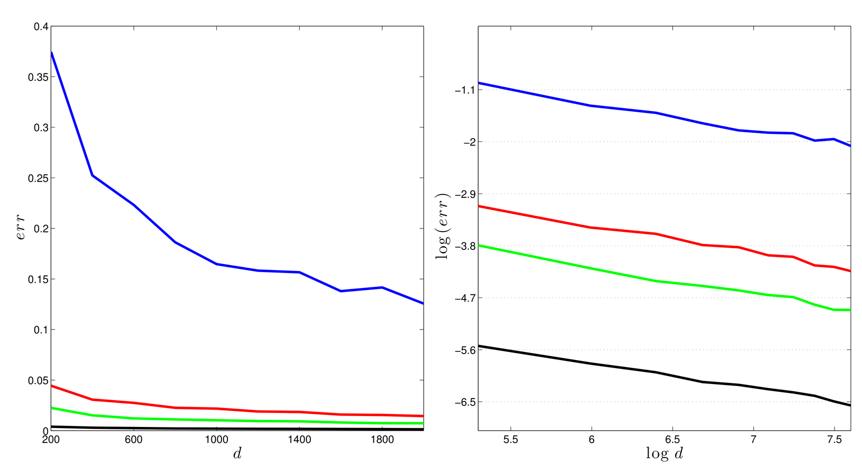


Figure 1: *The slope is around* -0.5. *Blue:* $\gamma = 10$; *red:* $\gamma = 50$; *green:* $\gamma = 100$; *and black:* $\gamma = 500$.

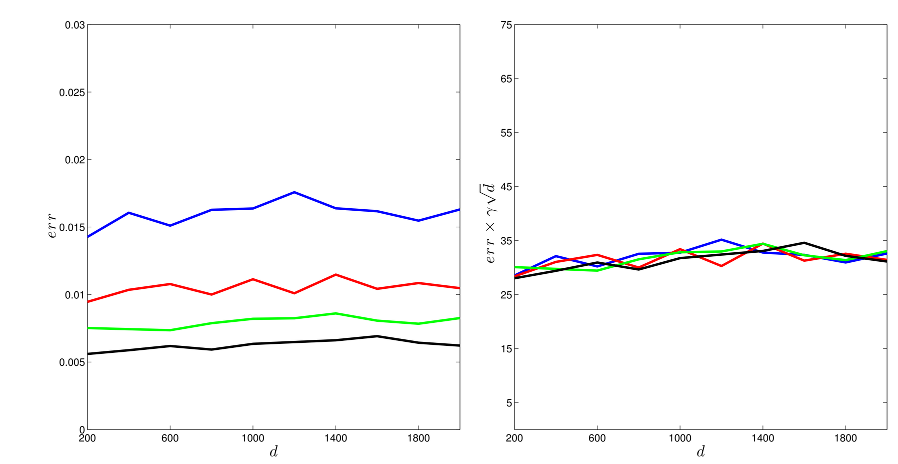


Figure 2: $\gamma\sqrt{p}$ *is fixed for each line. The right plot shows the error multiplied by* $\gamma\sqrt{p}$ *against* p. $\gamma\sqrt{p}$ *is* 2000 *for blue;* 3000 *for red;* 4000 *for green;* 5000 *for black.*

Application: Robust Covariance Matrix Estimation

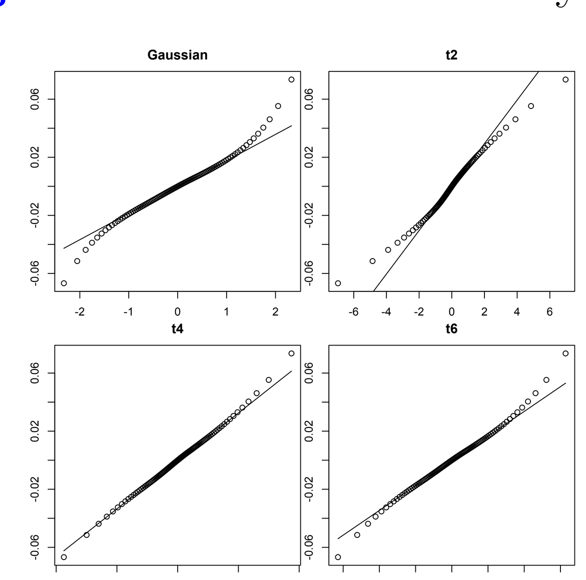
- **Factor Model**:

$$y_i = Bf_i + u_i$$

where $y_i, u_i \in \mathbf{R}^p$, $f_i \in \mathbf{R}^r$ and $B \in \mathbf{R}^{p \times r}$.
- $\{y_i\}_{i=1}^n$ are i.i.d. and observed; $\{f_i\}_{i=1}^n$ unobserved.
- Assuming $\text{Cov}(f_i, u_i) = 0$ and $\Sigma_f = I_r$,

$$\Sigma = BB^T + \Sigma_u,$$

where r is a constant and Σ_u is sparse.
- **Challenge**: Estimate Σ with heavy-tailed data.



- **Taming heavy-tailedness**: Huber's M-estimator with a diverging parameter.
- For any i.i.d. $Z_1, \ldots, Z_n$ with $\mu^\star = \mathbb{E}Z_i$. Let $l_\alpha(x) = 2\alpha|x| - \alpha^2$ when $|x| \geq \alpha$ and x^2 when $|x| \leq \alpha$.

$$\hat{\mu} = \text{argmin}_\mu \sum_{t=1}^{n} l_\alpha(Z_t - \mu).$$

- **Concentration bound**: (Fan, Li and Wang 14') Suppose $\epsilon \in (0, 1)$ and $n \geq 8\log(\epsilon^{-1})$. Choose $\alpha = \sqrt{(nv^2)/\log(\epsilon^{-1})}$, where v^2 is an upper bound of $\text{cov}(Z_t)$. Then,

$$P\left(|\hat{\mu} - \mu^\star| \leq 4v\sqrt{\frac{\log(\epsilon^{-1})}{n}}\right) \geq 1 - 2\epsilon.$$

- Elementary-wise estimation $\Rightarrow$ initial robust $\hat{\Sigma}$.
- Similar results: Catoni 12'.

Estimation Procedure and Theory

- **Decomposition**:

$$\hat{\Sigma} = \underbrace{B\Sigma_f B^T}_{\text{Low rank}} + \underbrace{\Sigma_u}_{\text{Sparse}} + \underbrace{(\hat{\Sigma} - \Sigma)}_{\text{Noise}}$$

- **Question**: How to denoise? How to disentangle?
- **Blessing of pervarsive assumption**: the top r eigenvalues of Σ grows linearly with p; and the elements of B are uniformly bounded.
- **Step 1**: rank r eigendecomposition: extract low rank $\widehat{U}\widehat{\Lambda}\widehat{U}$ from $\hat{\Sigma}$.
- **Step 2**:

$$\hat{\Sigma} - \underbrace{B\Sigma_f B^T}_{\approx \widehat{U}\widehat{\Lambda}\widehat{U}} = \underbrace{\Sigma_u + (\hat{\Sigma} - \Sigma)}_{\mathcal{T}(\hat{\Sigma}_u)\text{:thresholding}}$$

- **Step 3**:

$$\hat{\Sigma}^{\mathcal{T}} = \mathcal{T}(\hat{\Sigma}_u) + \widehat{U}\widehat{\Lambda}\widehat{U}^T.$$

- **Why it works?**: Apply ℓ^∞ perturbation bound to low rank $A := BB^T$ and $E := \Sigma_u + (\hat{\Sigma} - \Sigma)$.

$$\|\hat{v}_i - v_i\|_\infty = O_P\left(\frac{p\sqrt{\log p/n} + \sqrt{p}\|\Sigma_u\|}{p\sqrt{p}}\right),$$

which is sharp for analysis.
- **Assumption**: Let $m_q = \max_{i \leq p} \sum_{j \leq p} (\Sigma_u)_{ij}^q$ for some $q \in [0, 1]$. Assume pervasiveness, bounded fourth moments, r being a constant, and $\|\Sigma_u\|$ bounded above and below from 0.

Let $w_n = \sqrt{\log p/n} + 1/\sqrt{p}$. If $m_q w_n^{1-q} = o(1)$, we have

$$\|\hat{\Sigma}^{\top} - \Sigma\|_{\max} = O_P\Big(w_n\Big),$$

$$\|\hat{\Sigma}^{\top} - \Sigma\|_{\Sigma} = O_P\left(\frac{\sqrt{p}\log p}{n} + m_q w_n^{1-q}\right),$$

$$\|(\hat{\Sigma}^{\top})^{-1} - \Sigma^{-1}\|_2 = O_P\Big(m_q w_n^{1-q}\Big).$$

- $\|A\|_\Sigma = p^{-1/2}\|\Sigma^{-1/2} A \Sigma^{-1/2}\|_F$ is the relative Frobenius norm.