Matrix Eigenvector Perturbation

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

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

where $\lambda_i \geq \ldots \geq \lambda_p$ and $\tilde{\lambda}_i \geq \ldots \geq \tilde{\lambda}_p$. Let eigengap $\gamma := \min(\lambda_i - \lambda_{i+1} : 1 \leq i \leq r)$.

Eigenvector Perturbation (Davis and Kahan 70)

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

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

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

Question: What about $\ell^\infty$ bound? Any sharper result than $\|v_i\|_\infty \leq \|v_i\|_2$?

Our goal: For incoherent low-rank matrices, a sharp $\ell^\infty$ 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. logarthmic in $p$.
- Let $\gamma$ be the smallest gap in $\{\lambda_1, \ldots, \lambda_s, 0\}$.
- Matrix infinity norm: $\|E\|_\infty = \sup_{\|x\|_\infty \leq 1} \|E x\|_\infty = \max_{i} \max_{j} |E_{ij}|$.

Matrix coherence: (Candes and Recht 09)

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

$$\mu(V) = \frac{p}{r} \max_{i,j} \|V_{ij}\|_2.$$

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

$\ell^\infty$ Eigenvector Perturbation

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

$$C(\mu, r) = O(\mu^{1.5/r^{0.5}}),$$

$$\max_{i \leq s \leq r} \|v_i - \tilde{v}_i\|_\infty \leq C(\mu, r)\|E\|_\infty \gamma^{0.5/r^{0.5}}.$$

Remark: A similar bound for SVD.

Simulations

- $p$ runs from 200 to 2000 with increment 200.
- $A = \sum_{i=1}^{p} (1 - E) v_i v_i^T$; $v_i$ is an eigenvector of an i.i.d.
- normal random matrix.

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

Generating $E$: (a) random number in $[0,1]$ by randomly selecting $s$ entries each row: (b) $E_i = U|U|^T$.

- Repeat the largest error over 100 runs.

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: $y_i = B f_i + u_i$, where $y_i, u_i \in \mathbb{R}^d$, $f_i \in \mathbb{R}^d$ and $B \in \mathbb{R}^{p \times d}$.

Unobserved: $f_i$ is i.i.d. and observed: $\{f_i\}_{i=1}^{n}$.

Assuming $\text{Cov}(u, u) = 0$ and $\Sigma_f = I_d$.

$\Sigma = BB^T + \Sigma_u$, where $r$ is a constant and $\Sigma_u$ is sparse.

Challenge: Estimate $\Sigma$ with heavy-tailed data.

- Taming heavy-tailedness: Huber’s $M$-estimator with a diverging parameter. For any $i$ and $Z_i$ with $\mu^* = E Z_i$. Let $\tilde{\mu}(x) = 2a|x| - a^2$ when $|x| \geq \alpha$ and $x^2$ when $|x| \leq \alpha$.

$$\tilde{\mu} = \arg \min_{\mu} \sum_{i=1}^{n} |Z_i - \mu|.$$

- Concentration bound: (Fan, Li and Wang 14) Suppose $\epsilon \in (0, 1)$ and $n \geq 8 \log(\epsilon^{-1})$. Choose $\alpha = \sqrt{\log p} \log(\epsilon^{-1})/n$. Then, $E|\tilde{\mu} - \mu^*| \leq 4n^{0.5}\log(\epsilon^{-1})/n \geq 1 - 2\epsilon$.

- Elementary-wise estimation: $\Rightarrow$ initial robust $\Sigma$.

Similar results: Catoni 12.

Estimation Procedure and Theory

- Decomposition:

$$\tilde{\Sigma} = B \Sigma_f B^T + \tilde{\Sigma}_u = \Sigma_u + (\tilde{\Sigma} - \Sigma).$$

- Question: How to denoise? How to disentangle?

- Blessing of pervasive assumption: the top $r$ eigenvalues of $\Sigma$ grow linearly with $p$, and the elements of $B$ are uniformly bounded.

- Step 1: rank $r$ eigendecomposition: extract low rank $\tilde{U} \tilde{\Lambda} \tilde{U}^T$ from $\tilde{\Sigma}$.

- Step 2:

$$\tilde{\Sigma}_u = \tilde{\Sigma} - B \Sigma_f B^T.$$

- Step 3:

$$\Sigma = \tilde{\Sigma} + \tilde{\Sigma}_u = \tilde{\Sigma} + (\tilde{\Sigma} - \Sigma) = \tilde{U} \tilde{\Lambda} \tilde{U}^T.$$

- Why it works?: Apply $\ell^\infty$ perturbation bound (see above).

Assumption: Let $m_n = \max_{i \leq s \leq r} \|\Sigma_{\delta i}^{-\frac{1}{2}}\|_2$ for some $q \in [0, 1]$. Assume a bound on $\|\Sigma_{\delta i}^{-\frac{1}{2}}\|_2$ above and below from

$$\|v_i - \tilde{v}_i\|_\infty \leq O(\sqrt{\frac{\log p/n + \sqrt{p} \|\Sigma_u\|_2}{p\sqrt{p}}}).$$

which is sharp for analysis.

$\Sigma_{\delta i}^{-\frac{1}{2}}$.

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

$$\|\Sigma - \tilde{\Sigma}\|_\infty = O_p(w_n),$$

$$\|\Sigma - \tilde{\Sigma}\|_2 = O_p(\sqrt{\frac{\log p}{n} + m_n w_n^{-q}}),$$

$$\|\Sigma - \tilde{\Sigma}\|_2 = O_p(m_n w_n^{-q}).$$

$\|\Sigma\|_2 = \|P\|^{1/2}/\|\Sigma^{1/2} A^{1/2}\|^{1/2}$ is the relative Frobenius norm.