

# Principal Components Analysis in high dimensions

Ergan Shang, Yan Chen

USTC

October 30, 2022

- 1 Motivation and PCA
  - Dimension reduction
  - Perturbations
- 2 Bounds for generic eigenvectors
  - General result
  - Spiked ensemble
- 3 Sparse PCA
  - General result
  - Spiked model with sparsity

- 1 Motivation and PCA
  - Dimension reduction
  - Perturbations
- 2 Bounds for generic eigenvectors
  - General result
  - Spiked ensemble
- 3 Sparse PCA
  - General result
  - Spiked model with sparsity

# Motivations

We consider  $\Sigma \in S_+^{d \times d}$  which is a positive semidefinite matrix with an ordered eigenvalues  $\gamma_1(\Sigma) \geq \gamma_2(\Sigma) \geq \cdots \gamma_d(\Sigma) \geq 0$  and denote  $S_{d-1}$  as the unit sphere. Assuming random variable  $\mathbf{X}$  with  $\mathbb{E}\mathbf{X} = 0$ , then we obtain the maximal eigenvector as

$$\mathbf{v}^* = \arg \max_{\mathbf{v} \in S_{d-1}} \text{var}(\mathbf{v}^\top \mathbf{X}) = \arg \max_{\mathbf{v} \in S_{d-1}} \mathbf{v}^\top \Sigma \mathbf{v}.$$

More generally, we seek orthonormal matrix  $V \in \mathbb{R}^{d \times r}$  satisfying

$$V = \arg \max_V \mathbb{E} \|V^\top \mathbf{X}\|_2^2 = \arg \max_V \text{tr}(V^\top \Sigma V).$$

By variational representation :

$$\sum_{i=1}^k \gamma_i(\Sigma) = \max \{ \text{tr}(V^\top \Sigma V) : V \in \mathbb{R}^{d \times k}, V^\top V = I \}$$

, we know that  $V = (\mathbf{v}_1, \cdots, \mathbf{v}_r)$  where  $\mathbf{v}_1, \cdots, \mathbf{v}_r$  are first  $r$  eigenvectors of  $\Sigma$  with  $\mathbf{v}_i^\top \mathbf{v}_j = \delta_{ij}$ .

## 1 Low-rank Approximation:

$$Z^* = \arg \min_{r(Z) \leq r} (\|\Sigma - Z\|^2) = \sum_{j=1}^r \gamma_j(\Sigma) v_j v_j^\top,$$

where the matrix norm is invariant under orthonormal transformation. The solution can be derived by spectral decomposition of  $\Sigma = PDP^\top$  and let  $\tilde{Z} = P^\top ZP$ , leading to  $r(Z) = r(\tilde{Z})$ . Under the Frobenius norm, we conclude  $\tilde{Z}$  has to be diagonal to achieve the minimum  $\tilde{Z} = \text{diag}(\gamma_1, \dots, \gamma_r, 0, \dots, 0)$ . So that  $Z^* = P\tilde{Z}P^\top = \sum_{j=1}^r \gamma_j(\Sigma) v_j v_j^\top$ . And our approximated error is

$$\|Z^* - \Sigma\|_F^2 = \sum_{j=r+1}^d \gamma_j^2(\Sigma).$$

## 2 Data Compression:

Given a zero-mean random variable  $\mathbf{X} \in \mathbb{R}^d$ , we consider a projection to a subspace  $\mathbb{V}$  of dimension  $r$ :

$$\mathbb{V}^* = \arg \min_{\mathbb{V}} \mathbb{E} [\|\mathbf{X} - \Pi_{\mathbb{V}}(\mathbf{X})\|_2^2] .$$

We assume that the subspace  $\mathbb{V}^*$  is spanned by orthonormal vectors, i.e., its matrix expression is  $V_r$ , so that  $\Pi_{\mathbb{V}^*}(\mathbf{X}) = V_r V_r^\top \mathbf{X}$ .

$$\begin{aligned} \mathbb{E} [\|\mathbf{X} - V_r V_r^\top \mathbf{X}\|_2^2] &= \mathbb{E} [\mathbf{X}^\top (I - V_r V_r^\top) \mathbf{X}] = \text{tr}((I - V_r V_r^\top) \Sigma) \\ &= \sum_{i=1}^d \gamma_i(\Sigma) - \text{tr}(V_r^\top \Sigma V_r), \end{aligned}$$

so we should maximize  $\text{tr}(V_r^\top \Sigma V_r)$ . By variational representation we know that  $V_r = (v_1, \dots, v_r)$ , whose vectors are top  $r$  eigenvectors of  $\Sigma$ , and  $\mathbb{V}^*$  is spanned by those vectors.

# Approximation and Perturbation

In practice, we do not know the covariance matrix  $\Sigma$  of the population  $\mathbf{X}$ . Instead, we make estimation by  $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top$ . Then a natural question rises: What is the gap between  $\Sigma$  and  $\hat{\Sigma}$ .

Given a symmetric matrix  $R$ , how does its eigenstructure relate to the perturbed matrix  $Q = R + P$ , where  $P$  is another symmetric matrix. In fact

$$\gamma_1(Q) \leq \max_{v \in S_{d-1}} v^\top (R + P)v \leq \max_{v \in S_{d-1}} v^\top Rv + \max_{v \in S_{d-1}} v^\top Pv \leq \gamma_1(R) + \|P\|_2,$$

which means

$$|\gamma_1(Q) - \gamma_1(R)| \leq \|Q - R\|_2,$$

where  $\|\cdot\|_2$  denotes the operator norm of matrix.

# Weyl's Inequality

We claim that

$$\max_{j=1,\dots,d} |\gamma_j(Q) - \gamma_j(R)| \leq \|Q - R\|_2.$$

To prove this, we only need to prove

$$\gamma_j(Q) = \min_{\mathbb{V} \in \mathcal{V}_{j-1}} \max_{u \in \mathbb{V}^\perp \cap \mathcal{S}_{d-1}} u^\top Q u,$$

where  $\mathcal{V}_{j-1}$  means all subspace of dimension  $j-1$ .

For all subspace of dimension  $k-1$   $S_{k-1}$ , let  $S' = \text{span}\{u_1, \dots, u_k\}$ , where eigenvectors of  $Q$  are  $\{u_1, \dots, u_d\}$ . Then  $S' \cap S_{k-1}^\perp \neq 0$ . Thus, there exists  $x = \sum_{i=1}^k \alpha_i u_i \in S_{k-1}^\perp$ ,  $\|x\| = 1$ , satisfying  $x^\top Q x \geq \gamma_k$ , so that  $\max_{u \in \mathbb{V}^\perp \cap \mathcal{S}_{d-1}} u^\top Q u \geq \gamma_k$ . Noting that, the process above is applied to all subspace of dimension  $k-1$ , then we have

$$\min_{\mathbb{V} \in \mathcal{V}_{j-1}} \max_{u \in \mathbb{V}^\perp \cap \mathcal{S}_{d-1}} u^\top Q u \geq \gamma_k.$$

Finally, we take  $S_{k-1} = \text{span}\{u_1, \dots, u_{k-1}\}$  to attain the equality.



- 1 Motivation and PCA
  - Dimension reduction
  - Perturbations
- 2 Bounds for generic eigenvectors
  - General result
  - Spiked ensemble
- 3 Sparse PCA
  - General result
  - Spiked model with sparsity

# Notations

Given  $\Sigma \geq 0$  and  $\gamma_1(\Sigma) \geq \cdots \gamma_d(\Sigma) \geq 0$ , corresponding to its eigenvectors  $\{v_1, \cdots, v_d\}$ , let  $\theta^* \in \mathbb{R}^d$  be its (unique) maximal eigenvector. We have the perturbation as  $\hat{\Sigma} = \Sigma + P$ .

Define eigengap  $\nu = \gamma_1(\Sigma) - \gamma_2(\Sigma)$  assumed to be strictly positive. Define the transformed perturbation matrix

$$\tilde{P} := U^\top P U = \begin{pmatrix} \tilde{p}_{11} & \tilde{p}^\top \\ \tilde{p} & \tilde{P}_{22} \end{pmatrix}$$

where  $\tilde{p}_{11} \in \mathbb{R}$ .

A direct observation is that  $|\tilde{p}_{11}| \leq \|\tilde{P}\|_2$ , because  $|\tilde{p}_{11}| = e_1^\top \tilde{P} e_1 \leq \|\tilde{P}\|_2$ .

# Bound for maximal vector

## Theorem 8.5

Given any  $P \in S^{d \times d}$  such that  $\|P\|_2 < \nu/2$ , the perturbed matrix  $\hat{\Sigma} = \Sigma + P$  has a unique maximal eigenvector  $\hat{\theta}$  satisfying the bound

$$\|\hat{\theta} - \theta^*\|_2 \leq \frac{2\|\tilde{p}\|_2}{\nu - 2\|P\|_2}.$$

Define  $\hat{\Delta} = \hat{\theta} - \theta^*$  and the function

$$\psi(\Delta; P) = \Delta^\top P \Delta + 2\Delta^\top P \theta^*.$$

Moreover, assume that  $\rho = \hat{\theta}^\top \theta^*$ , thus,  $\hat{\theta} = \rho \theta^* + \sqrt{1 - \rho^2} z$  where  $z \in \mathbb{R}^d$  which is orthogonal to  $\theta^*$ .

## Lemma 8.6 (PCA basic inequality)

$$\nu \left( 1 - \left( \hat{\theta}^\top \theta^* \right)^2 \right) \leq |\psi(\hat{\Delta}; P)|. \quad (8.15)$$

Recall  $\tilde{P} = U^\top P U$ , then we have

$$\psi(\Delta; P) = \hat{\Delta}^\top U \tilde{P} U^\top \hat{\Delta} + 2 \hat{\Delta}^\top U \tilde{P} U^\top \theta^*. \quad (8.16)$$

Define  $U = (\theta^*, U_2)$  and  $\tilde{z} = U_2^\top z \in \mathbb{R}^{d-1} \Rightarrow \|\tilde{z}\|_2 = \|z\|_2 \leq 1$ . We can calculate that

$$\psi(\Delta; P) = (\rho^2 - 1) \tilde{p}_{11} + 2\rho \sqrt{1 - \rho^2} \tilde{z}^\top \tilde{p} + (1 - \rho^2) \tilde{z}^\top \tilde{P}_{22} \tilde{z}.$$

Thus,

$$\nu(1 - \rho^2) \stackrel{8.15}{\leq} |\psi(\hat{\Delta}; P)| \leq 2(1 - \rho^2) \|\tilde{P}\|_2 + 2\rho \sqrt{1 - \rho^2} \|\tilde{p}\|_2,$$

which means  $\sqrt{1 - \rho^2} \leq \frac{2\rho\|\tilde{p}\|_2}{\nu - 2\|\tilde{P}\|_2}$ . Recall  $\|\hat{\Delta}\|_2 = \sqrt{2(1 - \rho)}$ , we have

$$\|\hat{\Delta}\|_2 \leq \frac{\sqrt{2}}{\sqrt{1 + \rho}} \sqrt{1 - \rho^2} \leq \frac{\sqrt{2}}{\sqrt{1 + \rho}} \frac{2\rho\|\tilde{p}\|_2}{\nu - 2\|\tilde{P}\|_2} \leq \frac{2\|\tilde{p}\|_2}{\nu - 2\|\tilde{P}\|_2},$$

where the final inequality is because  $2\rho^2 \leq 1 + \rho, \forall \rho \in [0, 1]$ .

Now we turn to the proof of 8.15: by definition we have  $(\theta^*)^\top \hat{\Sigma} \theta^* \leq (\hat{\theta})^\top \hat{\Sigma} \hat{\theta}$ . Under the definition of  $P = \hat{\Sigma} - \Sigma$ , we have

$$\begin{aligned} \text{tr} \left[ \Sigma^\top \left( \theta^* (\theta^*)^\top - \hat{\theta} (\hat{\theta})^\top \right) \right] &= \text{tr} \left[ \left( \Sigma - \hat{\Sigma} \right) \left( \theta^* (\theta^*)^\top - \hat{\theta} (\hat{\theta})^\top \right) \right] \\ &+ \text{tr} \left[ \hat{\Sigma} \left( \theta^* (\theta^*)^\top - \hat{\theta} (\hat{\theta})^\top \right) \right] \leq -\text{tr} \left[ P \left( \theta^* (\theta^*)^\top - \hat{\theta} (\hat{\theta})^\top \right) \right] \\ &= - \left( \hat{\theta}^\top P \hat{\theta} - (\theta^*)^\top P \theta^* \right) = -\psi(\hat{\Delta}; P). \end{aligned} \quad (*)$$

Now we control the LHS in \*, by defining  $\Gamma = \Sigma - \gamma_1 \theta^* (\theta^*)^\top = \sum_{j=2}^d \gamma_j \theta_j \theta_j^\top \Rightarrow \Gamma \theta^* = 0$ . By considering  $x = \sum_{j=1}^d x_j \theta_j$  with  $\theta_1 = \theta^*$ , we have  $x^\top \Gamma x \leq \gamma_2 \Rightarrow \|\Gamma\|_2 \leq \gamma_2$ . Then

$$\begin{aligned} \text{tr} \left[ \Sigma^\top \left( \theta^* (\theta^*)^\top - \hat{\theta} (\hat{\theta})^\top \right) \right] &= \text{tr} [\gamma_1 (1 - \rho^2)] - \text{tr} [\Gamma \hat{\theta} (\hat{\theta})^\top] \\ &= (1 - \rho^2) (\gamma_1 - z^\top \Gamma z) \geq (1 - \rho^2) \nu. \end{aligned}$$

Combining \*, we have

$$(1 - \rho^2) \nu \leq -\psi(\hat{\Delta}; P)$$

which finishes the proof of Lemma.

# Spiked ensemble

A sample  $\mathbf{x}_i \in \mathbb{R}^d$  from the spiked covariance ensemble takes the form

$$\mathbf{x}_i \stackrel{d}{=} \sqrt{\nu} \xi_i \theta^* + w_i,$$

where  $\xi_i \in \mathbb{R}$ ,  $\xi_i \sim (0, 1)$ ,  $w_i \in \mathbb{R}^d$ ,  $w_i \sim (0, I_d)$ ,  $\xi \perp w_i$  and  $\theta^* \in \mathcal{S}_{d-1}$ . It has a form similar to Factor analysis

$$\mathbf{X} - \mu = LF + \epsilon \Rightarrow \Sigma = LL^\top + \psi.$$

Under the spiked ensemble, we have the form of covariance as

$$\Sigma = \nu \theta^* (\theta^*)^\top + I_d.$$

By construction, if we take  $x \in \mathcal{S}_{d-1}$ , we have  $x^\top \Sigma x = \nu(x^\top \theta^*)^2 + 1 \leq \nu + 1$  by CS inequality. We achieve the equality when  $x = \theta^*$ , thus  $\gamma_1(\Sigma) = \nu + 1$ ,  $\gamma_2(\Sigma) = \cdots = \gamma_d(\Sigma) = 1$ . Then  $\gamma_1(\Sigma) - \gamma_2(\Sigma) = \nu$ .

In the following result, we say that  $\mathbf{x}_i \in \mathbb{R}^d$  has sub-Gaussian tails if both  $\xi_i, w_i$  are sub-Gaussian with parameter at most 1.

## Corollary 8.7

Given i.i.d. sample  $\{\mathbf{x}_i\}_{i=1}^n$  from the spiked covariance ensemble with sub-Gaussian tails, suppose that  $n > d$  and  $\sqrt{\frac{\nu+1}{\nu^2}} \sqrt{\frac{d}{n}} \leq \frac{1}{128}$ . Then, with probability at least  $1 - c_1 \exp(-c_2 n \min\{\sqrt{\nu}\delta, \nu\delta^2\})$ , there is a unique maximal eigenvector  $\hat{\theta}$  of the sample covariance matrix  $\hat{\Sigma} = \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top$  such that

$$\|\hat{\theta} - \theta^*\|_2 \leq c_0 \sqrt{\frac{\nu+1}{\nu^2}} \sqrt{\frac{d}{n}} + \delta.$$



# Proof

In order to apply 9, we let  $P = \hat{\Sigma} - \Sigma$ ,  $\tilde{P} = U^\top P U$  and derive upper bound for  $\|P\|_2$  and  $\|\tilde{P}\|_2$ . Define  $\bar{w} = \frac{1}{n} \sum_{i=1}^n \xi_i w_i$ , then  $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\sqrt{\nu} \xi_i \theta^* + w_i)(\sqrt{\nu} \xi_i \theta^* + w_i)^\top$ . We have the decomposition of  $P$  as

$$P = \underbrace{\nu \left( \frac{1}{n} \sum_{i=1}^n \xi_i^2 - 1 \right) \theta^* (\theta^*)^\top}_{P_1} + \underbrace{\sqrt{\nu} (\bar{w} (\theta^*)^\top + \theta^* \bar{w}^\top)}_{P_2} + \underbrace{\left( \frac{1}{n} \sum_{i=1}^n w_i w_i^\top - I_d \right)}_{P_3}$$

Therefore, we have the upper bound as

$$\|P\|_2 \leq \nu \left| \frac{1}{n} \sum_{i=1}^n \xi_i^2 - 1 \right| + 2\sqrt{\nu} \|\bar{w}\|_2 + \left\| \frac{1}{n} \sum_{i=1}^n w_i w_i^\top - I_d \right\|_2. \quad (8.22a)$$

By the notation of  $U = (\theta^*, U_2)$ , we have

$\tilde{p} = \sqrt{\nu} U_2^\top \bar{w} + U_2^\top \left( \frac{1}{n} \sum_{i=1}^n w_i w_i^\top - I \right) \theta^*$ . Noting that  $\|U_2^\top \bar{w}\|_2 \leq \|\bar{w}\|_2$  and also

$$\begin{aligned} \left\| \frac{1}{n} \sum_{i=1}^n U_2^\top w_i \langle w_i, \theta^* \rangle \right\|_2 &\stackrel{CS}{=} \sup_{v \in \mathcal{S}_{d-1}} \left| (U_2 v)^\top \left( \frac{1}{n} \sum_{i=1}^n w_i w_i^\top - I \right) \theta^* \right| \\ &\leq \sup_{v \in \mathcal{S}_{d-1}} \|U_2^\top v\|_2 \left\| \frac{1}{n} \sum_{i=1}^n w_i w_i^\top - I \right\|_2 \leq \left\| \frac{1}{n} \sum_{i=1}^n w_i w_i^\top - I \right\|_2 \end{aligned}$$

where the last inequality is because  $\|U_2^\top v\|_2 \leq \|v\|_2$ . Therefore, we have

$$\|\tilde{p}\|_2 \leq \sqrt{\nu} \|\bar{w}\|_2 + \left\| \frac{1}{n} \sum_{i=1}^n w_i w_i^\top - I \right\|_2. \quad (8.22b)$$

# Concentration Lemma

## Lemma 8.8

Under the conditions of Corollary 8.7, we have

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n \xi_i^2 - 1\right| \geq \delta_1\right) \leq 2\exp(-c_2 n \min\{\delta_1, \delta_1^2\}), \quad (8.23a)$$

$$P\left(\|\bar{\mathbf{w}}\|_2 \geq 2\sqrt{\frac{d}{n}} + \delta_2\right) \leq 2\exp(-c_2 n \min\{\delta_2, \delta_2^2\}), \quad (8.23b)$$

$$P\left(\left\|\frac{1}{n}\sum_{i=1}^n \mathbf{w}_i \mathbf{w}_i^\top - \mathbf{I}\right\|_2 \geq c_3 \sqrt{\frac{d}{n}} + \delta_3\right) \leq 2\exp(-c_2 n \min\{\delta_3, \delta_3^2\}). \quad (8.23c)$$

8.23a is because product of sub-Gaussian is sub-Exponential; 8.23c is the result of Example 6.2 in Page 162.

# Proof

We define

$\phi(\delta_1, \delta_2, \delta_3) = 2e^{-c_2 n \min\{\delta_1, \delta_1^2\}} + 2e^{-c_2 n \min\{\delta_2, \delta_2^2\}} + 2e^{-c_2 n \min\{\delta_3, \delta_3^2\}}$ . We apply Lemma 8.8 with  $\delta_1 = \frac{1}{16}$ ,  $\delta_2 = \frac{\delta}{4\sqrt{\nu}}$ ,  $\delta_3 = \delta/16 \in (0, 1)$ , we have

$$\|P\|_2 \leq \frac{\nu}{16} + 8(\sqrt{\nu} + 1)\sqrt{\frac{d}{n}} + \delta \leq \frac{\nu}{16} + 16(\sqrt{\nu} + 1)\sqrt{\frac{d}{n}} + \delta.$$

As long as  $\sqrt{\frac{\nu+1}{\nu^2}}\sqrt{\frac{d}{n}} \leq \frac{1}{128}$ , we have

$$\|P\|_2 \leq \frac{3}{16}\nu + \delta < \frac{\nu}{4} < \frac{\nu}{2} \quad \forall \delta \in (0, \frac{\nu}{16}).$$

Also, we have

$$\|\tilde{p}\|_2 \leq 2(\sqrt{\nu} + 1)\sqrt{\frac{d}{n}} + \delta \leq 4\sqrt{\nu} + 1\sqrt{\frac{d}{n}} + \delta.$$

Finally, by 9 we finish the proof.

- 1 Motivation and PCA
  - Dimension reduction
  - Perturbations
- 2 Bounds for generic eigenvectors
  - General result
  - Spiked ensemble
- 3 Sparse PCA
  - General result
  - Spiked model with sparsity

- Corollary 8.7 requires that the sample size  $n$  be larger than the dimension  $d$  in order for ordinary PCA to perform well.
- For any fixed signal-to-noise ratio, if the ratio  $d/n$  stays suitably bounded away from zero, then the eigenvectors of the sample covariance in a spiked covariance model become asymptotically orthogonal to their population analogs.
- Via the framework of minimax theory that no method can produce consistent estimators of the population eigenvectors when  $d/n$  stays bounded away from zero.
- So, the simplest such structure is that of sparsity in the eigenvectors, which allows for both effective estimation in high-dimensional settings.

# General result

Consider the constrained problem

$$\hat{\theta} \in \arg \max_{\|\theta\|_2=1} \{\langle \theta, \hat{\Sigma} \theta \rangle\} \quad \text{such that } \|\theta\|_1 \leq R, \quad (1)$$

as well as the penalized variant

$$\hat{\theta} \in \arg \max_{\|\theta\|_2=1} \left\{ \langle \theta, \hat{\Sigma} \theta \rangle - \lambda_n \|\theta\|_1 \right\} \quad \text{such that } \|\theta\|_1 \leq \left( \frac{n}{\log d} \right)^{1/4}. \quad (2)$$

- $R = \|\theta^*\|_1$ .
- The regularization parameter  $\lambda_n$  can be chosen without knowledge of the true eigenvector  $\theta^*$ .

$$\sup_{\substack{\Delta = \theta - \theta^* \\ \|\theta\|_2 = 1}} |\Psi(\Delta; \mathbf{P})| \leq c_0 v \|\Delta\|_2^2 + \varphi_v(n, d) \|\Delta\|_1 + \psi_v^2(n, d) \|\Delta\|_1^2 \quad (3)$$

## Theorem 8.10

Given a matrix  $\Sigma$  with a unique, unit-norm,  $s$ -sparse maximal eigenvector  $\theta^*$  with eigengap  $v$ , let  $\hat{\Sigma}$  be any symmetric matrix satisfying the uniform deviation condition (3) with constant  $c_0 < \frac{1}{6}$ , and  $16s\psi_v^2(n, d) \leq c_0 v$ .

- (a) For any optimal solution  $\hat{\theta}$  to the constrained program (1) with  $R = \|\theta^*\|_1$ ,  $\min \left\{ \left\| \hat{\theta} - \theta^* \right\|_2, \left\| \hat{\theta} + \theta^* \right\|_2 \right\} \leq \frac{8}{v(1-4c_0)} \sqrt{s} \varphi_v(n, d)$ .
- (b) Consider the penalized program (2) with the regularization parameter lower bounded as  $\lambda_n \geq 4 \left( \frac{n}{\log d} \right)^{1/4} \psi_v^2(n, d) + 2\varphi_v(n, d)$ . Then any optimal solution  $\hat{\theta}$  satisfies the bound

$$\min \left\{ \left\| \hat{\theta} - \theta^* \right\|_2, \left\| \hat{\theta} + \theta^* \right\|_2 \right\} \leq \frac{2 \left( \frac{\lambda_n}{\varphi_v(n, d)} + 4 \right)}{v(1-4c_0)} \sqrt{s} \varphi_v(n, d).$$



## Lemma 8.11

Under the conditions of Theorem 8.10, the error vector  $\hat{\Delta} = \hat{\theta} - \theta^*$  satisfies the cone inequality

$$\left\| \hat{\Delta}_{S^c} \right\|_1 \leq 3 \left\| \hat{\Delta}_S \right\|_1 \quad \text{and hence } \left\| \hat{\Delta} \right\|_1 \leq 4\sqrt{s} \left\| \hat{\Delta} \right\|_2.$$

# Proof: Argument for constrained estimator

Note that  $\|\hat{\theta}\|_1 \leq R = \|\theta^*\|_1$  by construction of the estimator, and moreover  $\theta_{S^c}^* = 0$  by assumption. By Lemma 8.11, we have

$$|\Psi(\hat{\Delta}; \mathbf{P})| \leq c_0 v \|\hat{\Delta}\|_2^2 + 4\sqrt{s}\varphi_v(n, d)\|\hat{\Delta}\|_2 + 16s\psi_v^2(n, d)\|\hat{\Delta}\|_2^2.$$

Substituting back into the basic inequality and performing some algebra yields

$$\underbrace{v \left\{ \frac{1}{2} - c_0 - 16\frac{s}{v}\psi_v^2(n, d) \right\}}_{\kappa} \|\hat{\Delta}\|_2^2 \leq 4\sqrt{s}\varphi_v(n, d)\|\hat{\Delta}\|_2.$$

Note that our assumptions imply that  $\kappa > \frac{1}{2}(1 - 4c_0) > 0$ , so that the bound follows.

# Proof: Argument for regularized estimator

With the addition of the regularizer, the basic inequality now takes the slightly modified form

$$\frac{\nu}{2} \|\hat{\Delta}\|_2^2 - |\Psi(\hat{\Delta}; \mathbf{P})| \leq \lambda_n \left\{ \|\theta^*\|_1 - \|\hat{\theta}\|_1 \right\} \leq \lambda_n \left\{ \|\hat{\Delta}_S\|_1 - \|\hat{\Delta}_{S^c}\|_1 \right\},$$

We find that

$$\underbrace{\nu \left\{ \frac{1}{2} - c_0 - \frac{16}{\nu} s \psi_v^2(n, d) \right\}}_{\kappa} \|\hat{\Delta}\|_2^2 \leq \sqrt{s} (\lambda_n + 4\varphi_v(n, d)) \|\hat{\Delta}\|_2.$$

Our assumptions imply that  $\kappa \geq \frac{1}{2} (1 - 4c_0) > 0$ , from which claim (b) follows.

# Proof of Lemma 8.11

Combining the uniform bound with the basic inequality

$$0 \leq \underbrace{\nu\left(\frac{1}{2} - c_0\right)}_{>0} \|\Delta\|_2^2 \leq \varphi_\nu(n, d) \|\Delta\|_1 + \psi_\nu^2(n, d) \|\Delta\|_1^2 + \lambda_n \left\{ \left\| \hat{\Delta}_S \right\|_1 - \left\| \hat{\Delta}_{S^c} \right\|_1 \right\}$$

Introducing the shorthand  $R = \left( \frac{n}{\log d} \right)^{1/4}$ , the feasibility of  $\hat{\theta}$  and  $\theta^*$  implies that  $\|\hat{\Delta}\|_1 \leq 2R$ , and hence

$$\begin{aligned} 0 &\leq \underbrace{\left\{ \varphi_\nu(n, d) + 2R\psi_\nu^2(n, d) \right\}}_{\leq \frac{\lambda n}{2}} \|\hat{\Delta}\|_1 + \lambda_n \left\{ \left\| \hat{\Delta}_S \right\|_1 - \left\| \hat{\Delta}_{S^c} \right\|_1 \right\} \\ &\leq \lambda_n \left\{ \frac{3}{2} \left\| \hat{\Delta}_S \right\|_1 - \frac{1}{2} \left\| \hat{\Delta}_{S^c} \right\|_1 \right\}, \end{aligned}$$

and rearranging yields the claim.

# Spiked model with sparsity

We consider a random vector  $x_i \in \mathbb{R}^d$  generated from the usual spiked ensemble, namely,

$$x_i \stackrel{d}{=} \sqrt{v} \xi_i \theta^* + w_i,$$

where  $\theta^* \in \mathbb{S}^{d-1}$  is an  $s$ -sparse vector, corresponding to the maximal eigenvector of  $\Sigma = \text{cov}(x_i)$ . As before, we assume that both the random variable  $\xi_i$  and the random vector  $w_i \in \mathbb{R}^d$  are independent, each sub-Gaussian with parameter 1, the random vector  $x_i \in \mathbb{R}^d$  has sub-Gaussian tails.

## Corollary 8.12

Consider  $n$  i.i.d. samples  $\{x_i\}_{i=1}^n$  from an  $s$ -sparse spiked covariance matrix with eigengap  $v > 0$  and suppose that  $\frac{s \log d}{n} \leq c \min \left\{ 1, \frac{v^2}{v+1} \right\}$  for a sufficiently small constant  $c > 0$ . Then for any  $\delta \in (0, 1)$ , any optimal solution  $\hat{\theta}$  to the constrained program (1) with  $R = \|\theta^*\|_1$ , or to the penalized program (2) with  $\lambda_n = c_3 \sqrt{v+1} \left\{ \sqrt{\frac{\log d}{n}} + \delta \right\}$ , satisfies the bound

$$\min \left\{ \|\hat{\theta} - \theta^*\|_2, \|\hat{\theta} + \theta^*\|_2 \right\} \leq c_4 \sqrt{\frac{v+1}{v^2}} \left\{ \sqrt{\frac{s \log d}{n}} + \delta \right\},$$

for all  $\delta \in (0, 1)$  with probability at least  $1 - c_1 e^{-c_2(n/s) \min\{\delta^2, v^2, v\}}$ .

We claim that

$$|\Psi(\Delta; \mathbf{P})| \leq \underbrace{\frac{1}{8}}_{c_0} v \|\Delta\|_2^2 + \underbrace{16\sqrt{v+1} \left\{ \sqrt{\frac{\log d}{n}} + \delta \right\}}_{\varphi_v(n,d)} \|\Delta\|_1 + \underbrace{\frac{c'_3}{v} \frac{\log d}{n}}_{\psi_v^2(n,d)} \|\Delta\|_1^2,$$

with probability at least  $1 - c_1 e^{-c_2 n \min\{\delta^2, v^2\}}$ . Here  $(c_1, c_2, c'_3)$  are universal constants.

Check the condition of Theorem 8.10:

$$\frac{9s\psi_v^2(n, d)}{c_0} = \frac{72c'_3}{v} \frac{s \log d}{n} \leq v \left\{ 72c'_3 \frac{v+1}{v^2} \frac{s \log d}{n} \right\} \leq v.$$

$\lambda_n$  satisfies the lower bound requirement in Theorem 8.10. We have

$$\begin{aligned} 4R\psi_v^2(n, d) + 2\varphi_v(n, d) &\leq 4v \sqrt{\frac{n}{\log d}} \frac{c'_3}{v} \frac{\log d}{n} + 24\sqrt{v+1} \left\{ \sqrt{\frac{\log d}{n}} + \delta \right\} \\ &\leq \underbrace{c_3 \sqrt{v+1} \left\{ \sqrt{\frac{\log d}{n}} + \delta \right\}}_{\lambda_n}. \end{aligned}$$



Recall

$$\mathbf{P} = \underbrace{\nu \left( \frac{1}{n} \sum_{i=1}^n \xi_i^2 - 1 \right) \theta^* (\theta^*)^T}_{\mathbf{P}_1} + \underbrace{\sqrt{\nu} \left( \bar{\mathbf{w}} (\theta^*)^T + \theta^* \bar{\mathbf{w}}^T \right)}_{\mathbf{P}_2} + \underbrace{\left( \frac{1}{n} \sum_{i=1}^n \mathbf{w}_i \mathbf{w}_i^T - \mathbf{I}_d \right)}_{\mathbf{P}_3}.$$

Control of first component:

Lemma 8.8 guarantees that  $\left| \frac{1}{n} \sum_{i=1}^n \xi_i^2 - 1 \right| \leq \frac{1}{16}$  with probability at least  $1 - 2e^{-cn}$ . For any vector  $\Delta = \theta - \theta^*$  with  $\theta \in \mathbb{S}^{d-1}$ , we have

$$|\Psi(\Delta; \mathbf{P}_1)| \leq \frac{\nu}{16} \langle \Delta, \theta^* \rangle^2 = \frac{\nu}{16} (1 - \langle \theta^*, \theta \rangle)^2 \leq \frac{\nu}{32} \|\Delta\|_2^2.$$

# Proof:

## Control of second component:

We have

$$\begin{aligned} |\Psi(\Delta; \mathbf{P}_2)| &\leq 2\sqrt{v} \{ \langle \Delta, \bar{w} \rangle \langle \Delta, \theta^* \rangle + \langle \bar{w}, \Delta \rangle + \langle \theta^*, \bar{w} \rangle \langle \Delta, \theta^* \rangle \} \\ &\leq 4\sqrt{v} \|\Delta\|_1 \|\bar{w}\|_\infty + 2\sqrt{v} |\langle \theta^*, \bar{w} \rangle| \frac{\|\Delta\|_2^2}{2}. \end{aligned}$$

### Lemma 8.13

Under the conditions of Corollary 8.12, we have

$$\begin{aligned} \mathbb{P} \left[ \|\bar{w}\|_\infty \geq 2\sqrt{\frac{\log d}{n}} + \delta \right] &\leq c_1 e^{-c_2 n \delta^2} \quad \text{for all } \delta \in (0, 1), \text{ and} \\ \mathbb{P} \left[ |\langle \theta^*, \bar{w} \rangle| \geq \frac{\sqrt{v}}{32} \right] &\leq c_1 e^{-c_2 n v}. \end{aligned}$$

Then

$$|\Psi(\Delta; \mathbf{P}_2)| \leq \frac{v}{32} \|\Delta\|_2^2 + 8\sqrt{v+1} \left\{ \sqrt{\frac{\log d}{n}} + \delta \right\} \|\Delta\|_1.$$

# Proof

Control of third term: Recalling that  $\mathbf{P}_3 = \frac{1}{n} \mathbf{W}^T \mathbf{W} - \mathbf{I}_d$ , we have

$$|\Psi(\Delta; \mathbf{P}_3)| \leq |\langle \Delta, \mathbf{P}_3 \Delta \rangle| + 2 \|\mathbf{P}_3 \theta^*\|_\infty \|\Delta\|_1.$$

Our final lemma controls the two terms in this bound:

## Lemma 8.14

Under the conditions of Corollary 8.12, for all  $\delta \in (0, 1)$ , we have

$$\|\mathbf{P}_3 \theta^*\|_\infty \leq 2 \sqrt{\frac{\log d}{n}} + \delta$$

and

$$\sup_{\Delta \in \mathbb{R}^d} |\langle \Delta, \mathbf{P}_3 \Delta \rangle| \leq \frac{\nu}{16} \|\Delta\|_2^2 + \frac{c'_3}{\nu} \frac{\log d}{n} \|\Delta\|_1^2,$$

where both inequalities hold with probability greater than  $1 - c_1 e^{-c_2 n \min(y, \nu^2, \delta^2)}$ .

Combining this lemma, yields the bound

$$|\Psi(\Delta; \mathbf{P}_3)| \leq \frac{\nu}{16} \|\Delta\|_2^2 + 8 \left\{ \sqrt{\frac{\log d}{n}} + \delta \right\} \|\Delta\|_1 + \frac{c'_3}{\nu} \frac{\log d}{n} \|\Delta\|_1^2.$$

Thank you !