

## A Proofs of Theorem 2.1 and Theorem 2.2

The key idea in the proof of Theorem 2.1 is to find an “envelope”  $m_1 \leq k \leq m_2$  in the spectrum of  $\mathbf{A}$  surrounding  $k$ , such that the eigenvalues within the envelope are relatively close. Define

$$\begin{aligned} m_1 &= \operatorname{argmax}_{0 \leq j \leq k} \{\sigma_j(\mathbf{A}) \geq (1 + 2\epsilon)\sigma_{k+1}(\mathbf{A})\}; \\ m_2 &= \operatorname{argmax}_{k \leq j \leq n} \{\sigma_j(\mathbf{A}) \geq \sigma_k(\mathbf{A}) - 2\epsilon\sigma_{k+1}(\mathbf{A})\}, \end{aligned}$$

where we let  $\sigma_0(\mathbf{A}) = \infty$  for convenience. Let  $\mathcal{U}_m, \widehat{\mathcal{U}}_m$  be basis of the top  $m$ -dimensional linear subspaces of  $\mathbf{A}$  and  $\widehat{\mathbf{A}}$ , respectively. Also denote  $\mathcal{U}_{n-m}$  and  $\widehat{\mathcal{U}}_{n-m}$  as basis of the orthogonal complement of  $\mathcal{U}_m$  and  $\widehat{\mathcal{U}}_m$ .

**Lemma A.1.** *If  $\|\widehat{\mathbf{A}} - \mathbf{A}\|_2 \leq \epsilon^2 \sigma_{k+1}(\mathbf{A})$  for  $\epsilon \in (0, 1)$  then  $\|\widehat{\mathbf{U}}_{n-k}^\top \mathbf{U}_{m_1}\|_2, \|\widehat{\mathbf{U}}_k^\top \mathbf{U}_{n-m_2}\|_2 \leq \epsilon$ .*

*Proof.* We apply an asymmetric version of Davis-Kahan inequality (Lemma C.1), with  $\mathbf{X} = \mathbf{A}$ ,  $\mathbf{Y} = \widehat{\mathbf{A}}$ ,  $i = m_1$  and  $j = k$ . By Weyl’s inequality, we know that  $\sigma_{k+1}(\widehat{\mathbf{A}}) \leq \sigma_{k+1}(\mathbf{A}) + \|\widehat{\mathbf{A}} - \mathbf{A}\|_2 \leq (1 + \epsilon^2)\sigma_{k+1}(\mathbf{A}) \leq (1 + \epsilon)\sigma_{k+1}(\mathbf{A})$ . Subsequently,  $\|\widehat{\mathbf{U}}_{n-k}^\top \mathbf{U}_{m_1}\|_2 \leq \frac{\epsilon^2 \sigma_{k+1}(\mathbf{A})}{\sigma_{m_1}(\mathbf{A}) - (1 + \epsilon)\sigma_{k+1}(\mathbf{A})} \leq \epsilon$ . Similarly, applying Lemma C.1 with  $\mathbf{X} = \widehat{\mathbf{A}}$ ,  $\mathbf{Y} = \mathbf{A}$ ,  $i = k$  and  $j = m_2$  we have that  $\|\widehat{\mathbf{U}}_k^\top \mathbf{U}_{n-m_2}\|_2 \leq \epsilon$ .  $\square$

Let  $\mathcal{U}_{m_1:m_2}$  be the linear subspace of  $\mathbf{A}$  associated with eigenvalues  $\sigma_{m_1+1}(\mathbf{A}), \dots, \sigma_{m_2}(\mathbf{A})$ . Intuitively, we choose a  $(k - m_1)$ -dimensional linear subspace in  $\mathcal{U}_{m_1:m_2}$  that is “most aligned” with the top- $k$  subspace  $\widehat{\mathcal{U}}_k$  of  $\widehat{\mathbf{A}}$ . Formally, define

$$\mathcal{W} = \operatorname{argmax}_{\dim(\mathcal{W})=k-m_1, \mathcal{W} \in \mathcal{U}_{m_1:m_2}} \sigma_{k-m_1}(\mathbf{W}^\top \widehat{\mathbf{U}}_k).$$

$\mathbf{W}$  is then a  $d \times (k - m_1)$  matrix with orthonormal columns that corresponds to a basis of  $\mathcal{W}$ .  $\mathcal{W}$  is carefully constructed so that it is closely aligned with  $\widehat{\mathcal{U}}_k$ , yet still lies in  $\mathcal{U}_k$ . In particular, Lemma 3.2 shows that  $\sin \angle(\mathcal{W}, \widehat{\mathcal{U}}_k) = \|\widehat{\mathbf{U}}_{n-k}^\top \mathbf{W}\|_2$  is upper bounded by  $\epsilon$ .

**Lemma A.2.** *If  $\|\widehat{\mathbf{A}} - \mathbf{A}\|_2 \leq \epsilon^2 \sigma_{k+1}(\mathbf{A})$  for  $\epsilon \in (0, 1)$  then  $\|\widehat{\mathbf{U}}_{n-k}^\top \mathbf{W}\|_2 \leq \epsilon$ .*

*Proof.* First note that  $\|\widehat{\mathbf{U}}_{n-k}^\top \mathbf{W}\|_2 \leq \sqrt{1 - \sigma_{k-m_1}(\widehat{\mathbf{U}}_k^\top \mathbf{W})^2}$  because

$$\begin{aligned} \|\widehat{\mathbf{U}}_{n-k}^\top \mathbf{W}\|_2^2 &= \sup_{\|\mathbf{x}\|_2=1} \|\widehat{\mathbf{U}}_{n-k}^\top \mathbf{W} \mathbf{x}\|_2^2 = \sup_{\|\mathbf{x}\|_2=1} \left\{ \|\mathbf{W} \mathbf{x}\|_2^2 - \|\widehat{\mathbf{U}}_k^\top \mathbf{W} \mathbf{x}\|_2^2 \right\} \\ &\leq \sup_{\|\mathbf{x}\|_2=1} \|\mathbf{W} \mathbf{x}\|_2^2 - \inf_{\|\mathbf{x}\|_2=1} \|\widehat{\mathbf{U}}_k^\top \mathbf{W} \mathbf{x}\|_2^2 = 1 - \sigma_{k-m_1}(\widehat{\mathbf{U}}_k^\top \mathbf{W})^2. \end{aligned}$$

Subsequently, it suffices to prove that  $\sigma_{k-m_1}(\widehat{\mathbf{U}}_k^\top \mathbf{W}) \geq \sqrt{1 - \epsilon^2}$ . By Weyl’s monotonicity theorem (Lemma C.4), we have that

$$\sigma_k(\widehat{\mathbf{U}}_k^\top \mathbf{U}_{m_2}) \leq \sigma_{m_1+1}(\widehat{\mathbf{U}}_k^\top \mathbf{U}_{m_1}) + \sigma_{k-m_1}(\widehat{\mathbf{U}}_k^\top \mathbf{U}_{m_1:m_2}).$$

In addition,  $\sigma_{m_1+1}(\widehat{\mathbf{U}}_k^\top \mathbf{U}_{m_1}) = 0$  because  $\operatorname{rank}(\widehat{\mathbf{U}}_k^\top \mathbf{U}_{m_1}) \leq m_1$  and  $\sigma_{k-m_1}(\widehat{\mathbf{U}}_k^\top \mathbf{U}_{m_1:m_2}) = \sigma_{k-m_1}(\widehat{\mathbf{U}}_k^\top \mathbf{W})$  because of the definition of  $\mathbf{W}$ . Subsequently,

$$\begin{aligned} \sigma_{k-m_1}(\widehat{\mathbf{U}}_k^\top \mathbf{W})^2 &\geq \sigma_k(\widehat{\mathbf{U}}_k^\top \mathbf{U}_{m_2})^2 = \inf_{\|\mathbf{x}\|_2=1} \|\mathbf{U}_{m_2}^\top \widehat{\mathbf{U}}_k \mathbf{x}\|_2^2 = \inf_{\|\mathbf{x}\|_2=1} \left\{ \|\widehat{\mathbf{U}}_k \mathbf{x}\|_2^2 - \|\mathbf{U}_{n-m_2}^\top \widehat{\mathbf{U}}_k \mathbf{x}\|_2^2 \right\} \\ &\geq \inf_{\|\mathbf{x}\|_2=1} \left\{ \|\widehat{\mathbf{U}}_k \mathbf{x}\|_2^2 \right\} - \sup_{\|\mathbf{x}\|_2=1} \left\{ \|\mathbf{U}_{n-m_2}^\top \widehat{\mathbf{U}}_k \mathbf{x}\|_2^2 \right\} \geq 1 - \epsilon^2. \end{aligned}$$

Here in the last inequality we invoke Lemma 3.1. The proof is then complete.  $\square$

Define

$$\widetilde{\mathbf{A}} = \mathbf{A}_{m_1} + \mathbf{W} \mathbf{W}^\top \mathbf{A} \mathbf{W} \mathbf{W}^\top.$$

The following lemma lists some of the properties of  $\widetilde{\mathbf{A}}$ .

**Lemma A.3.** *It holds that*

1.  $\dim(\text{Range}(\tilde{\mathbf{A}})) = k$  and  $\dim(\text{Range}(\mathbf{W})) = k - m_1$ ;
2.  $\mathcal{U}_{m_1} \subseteq \text{Range}(\tilde{\mathbf{A}}) \subseteq \mathcal{U}_{m_2}$  and  $\text{Range}(\tilde{\mathbf{A}} - \mathbf{A}_{m_1}) \subseteq \mathcal{U}_{m_1:m_2}$ , where  $\mathcal{U}_{m_2} = \mathcal{U}_{m_1} \oplus \mathcal{U}_{m_1:m_2}$ .
3.  $\|\hat{\mathbf{U}}_k^\top \tilde{\mathbf{U}}_\perp\|_2, \|\tilde{\mathbf{U}}^\top \hat{\mathbf{U}}_{n-k}\|_2 \leq 2\epsilon$ , where  $\tilde{\mathbf{U}}$  and  $\tilde{\mathbf{U}}_\perp$  are orthonormal basis of  $\text{Range}(\tilde{\mathbf{A}})$  and  $\text{Null}(\tilde{\mathbf{A}})$ , respectively.

*Proof.* Properties 1 and 2 are obviously true by the definition of  $\mathcal{W}$  and  $\tilde{\mathbf{A}}$ . For property 3, note that both  $\|\hat{\mathbf{U}}_k^\top \tilde{\mathbf{U}}_\perp\|_2$  and  $\|\tilde{\mathbf{U}}^\top \hat{\mathbf{U}}_{n-k}\|_2$  are equal to  $\sin \angle(\tilde{\mathcal{U}}, \hat{\mathcal{U}}_k)$ . Hence it suffices to show that  $\|\hat{\mathbf{U}}_{n-k}^\top \tilde{\mathbf{U}}\|_2 \leq 2\epsilon$ . Invoking Lemmas 3.1 and 3.2 we have that  $\|\hat{\mathbf{U}}_{n-k}^\top \tilde{\mathbf{U}}\|_2 \leq \|\hat{\mathbf{U}}_{n-k}^\top \mathbf{U}_{m_1}\|_2 + \|\hat{\mathbf{U}}_{n-k}^\top \mathbf{W}\|_2 \leq \epsilon + \epsilon = 2\epsilon$ .  $\square$

Decompose  $\|\hat{\mathbf{A}}_k - \mathbf{A}\|_F$  as

$$\|\hat{\mathbf{A}}_k - \mathbf{A}\|_F \leq \|\mathbf{A} - \tilde{\mathbf{A}}\|_F + \|\hat{\mathbf{A}}_k - \tilde{\mathbf{A}}\|_F \leq \|\mathbf{A} - \tilde{\mathbf{A}}\|_F + \sqrt{2k} \|\hat{\mathbf{A}}_k - \tilde{\mathbf{A}}\|_2. \quad (12)$$

Here the last inequality holds because both  $\hat{\mathbf{A}}_k$  and  $\tilde{\mathbf{A}}$  have rank at most  $k$ . Lemmas 3.3 and 3.4 give separate upper bounds for  $\|\mathbf{A} - \tilde{\mathbf{A}}\|_F$  and  $\|\hat{\mathbf{A}}_k - \tilde{\mathbf{A}}\|_2$ .

**Lemma A.4.** *If  $\|\hat{\mathbf{A}} - \mathbf{A}\|_2 \leq \epsilon^2 \sigma_{k+1}(\mathbf{A})^2$  for  $\epsilon \in (0, 1/4]$  then  $\|\mathbf{A} - \tilde{\mathbf{A}}\|_F \leq (1 + 32\epsilon) \|\mathbf{A} - \mathbf{A}_k\|_F$ .*

*Proof.* Let  $\mathcal{U}_{m_1:m_2}$  be the  $(m_2 - m_1)$ -dimensional linear subspace such that  $\mathcal{U}_{m_2} = \mathcal{U}_{m_1} \oplus \mathcal{U}_{m_1:m_2}$ . Define  $\mathbf{A}_{m_1:m_2} = \mathbf{U}_{m_1:m_2} \Sigma_{m_1:m_2} \mathbf{U}_{m_1:m_2}^\top$ , where  $\Sigma_{m_1:m_2} = \text{diag}(\sigma_{m_1+1}(\mathbf{A}), \dots, \sigma_{m_2}(\mathbf{A}))$  and  $\mathbf{U}_{m_1:m_2}$  is an orthonormal basis associated with  $\mathcal{U}_{m_1:m_2}$ . We then have

$$\begin{aligned} \|\mathbf{A} - \tilde{\mathbf{A}}\|_F^2 &= \|\mathbf{A}_{n-m_1} - \mathbf{W}\mathbf{W}^\top \mathbf{A}\mathbf{W}\mathbf{W}^\top\|_F^2 \\ &\stackrel{(a)}{=} \|\mathbf{A}_{n-m_2}\|_F^2 + \|\mathbf{A}_{m_1:m_2} - \mathbf{W}\mathbf{W}^\top \mathbf{A}\mathbf{W}\mathbf{W}^\top\|_F^2 \\ &\stackrel{(b)}{=} \|\mathbf{A} - \mathbf{A}_{m_2}\|_F^2 + \|\mathbf{A}_{m_1:m_2} - \mathbf{W}\mathbf{W}^\top \mathbf{A}_{m_1:m_2} \mathbf{W}\mathbf{W}^\top\|_F^2 \\ &\stackrel{(c)}{=} \|\mathbf{A} - \mathbf{A}_{m_2}\|_F^2 + \|\mathbf{A}_{m_1:m_2}\|_F^2 - \|\mathbf{W}\mathbf{W}^\top \mathbf{A}_{m_1:m_2} \mathbf{W}\mathbf{W}^\top\|_F^2. \end{aligned}$$

Here in (a) we apply  $\text{Range}(\tilde{\mathbf{A}} - \mathbf{A}_{m_1}) \subseteq \mathcal{U}_{m_1:m_2}$  and the Pythagorean theorem (Lemma C.2) with  $\mathbf{P} = \mathbf{U}_{m_1:m_2}$ , in (b) we apply  $\mathcal{W} \subseteq \mathcal{U}_{m_1:m_2}$ , and in (c) we apply the Pythagorean theorem again with  $\mathbf{P} = \mathbf{W}$ . Note that  $\|\mathbf{W}\mathbf{W}^\top \mathbf{A}_{m_1:m_2} \mathbf{W}\mathbf{W}^\top\|_F^2 = \|\mathbf{W}^\top \mathbf{A}_{m_1:m_2} \mathbf{W}\|_F^2$ . Applying Poincaré separation theorem (Lemma C.3) where  $\mathbf{X} = \Sigma_{m_1:m_2}$  and  $\mathbf{P} = \mathbf{U}_{m_1:m_2}^\top \mathbf{W}$ , we have  $\|\mathbf{W}^\top \mathbf{A}_{m_1:m_2} \mathbf{W}\|_F^2 \geq \sum_{j=m_2-k+1}^{m_2-m_1} \sigma_j(\mathbf{A}_{m_1:m_2})^2 = \sum_{j=m_1+m_2-k+1}^{m_2} \sigma_j(\mathbf{A})^2$ . Subsequently,

$$\begin{aligned} \|\mathbf{A} - \tilde{\mathbf{A}}\|_F^2 &\leq \|\mathbf{A} - \mathbf{A}_{m_2}\|_F^2 + \sum_{j=m_1+1}^{m_1+m_2-k} \sigma_j(\mathbf{A})^2 \leq \|\mathbf{A} - \mathbf{A}_{m_2}\|_F^2 + (m_2 - k) \sigma_{m_1+1}(\mathbf{A})^2 \\ &\stackrel{(a')}{\leq} \|\mathbf{A} - \mathbf{A}_{m_2}\|_F^2 + (m_2 - k)(1 + 2\epsilon)^2 \sigma_{k+1}(\mathbf{A})^2 \\ &\stackrel{(b')}{\leq} \|\mathbf{A} - \mathbf{A}_{m_2}\|_F^2 + (m_2 - k) \left( \frac{1 + 2\epsilon}{1 - 2\epsilon} \right)^2 \sigma_{m_2}(\mathbf{A})^2 \\ &\stackrel{(c')}{\leq} \|\mathbf{A} - \mathbf{A}_{m_2}\|_F^2 + (m_2 - k) \sigma_{m_2}(\mathbf{A})^2 + 32(m_2 - k) \epsilon \sigma_{m_2}(\mathbf{A})^2 \\ &\stackrel{(d')}{\leq} (1 + 32\epsilon) \|\mathbf{A} - \mathbf{A}_k\|_F^2. \end{aligned}$$

Here in (a') we apply the definition of  $m_1$  that  $\sigma_{m_1+1} \leq (1 + 2\epsilon) \sigma_{k+1}(\mathbf{A})$ , in (b') we apply the definition of  $m_2$  that  $\sigma_{m_2}(\mathbf{A}) \geq \sigma_k(\mathbf{A}) - 2\epsilon \sigma_{k+1}(\mathbf{A}) \geq (1 - 2\epsilon) \sigma_{k+1}(\mathbf{A})$ , and (c') is due to the fact that  $\left( \frac{1+2\epsilon}{1-2\epsilon} \right)^2 \leq 1 + 32\epsilon$  for all  $\epsilon \in (0, 1/4]$ . Finally, (d') holds because  $(m_2 - k) \sigma_{m_2}(\mathbf{A})^2 \leq \sum_{j=k+1}^{m_2} \sigma_j(\mathbf{A})^2$  and  $\|\mathbf{A} - \mathbf{A}_k\|_F^2 = \|\mathbf{A} - \mathbf{A}_{m_2}\|_F^2 + \sum_{j=k+1}^{m_2} \sigma_j(\mathbf{A})^2$ .  $\square$

**Lemma A.5.** *If  $\|\widehat{\mathbf{A}} - \mathbf{A}\|_2 \leq \epsilon^2 \sigma_{k+1}(\mathbf{A})$  for  $\epsilon \in (0, 1/4]$  then  $\|\widehat{\mathbf{A}}_k - \widetilde{\mathbf{A}}\|_2 \leq 102\epsilon^2 \|\mathbf{A} - \mathbf{A}_k\|_2$ .*

*Proof.* Recall the definition that  $\widetilde{\mathcal{U}} = \text{Range}(\widetilde{\mathbf{A}})$  and  $\widetilde{\mathcal{U}}_\perp = \text{Null}(\widetilde{\mathbf{A}})$ . Consider  $\|\mathbf{v}\|_2 = 1$  such that  $\mathbf{v}^\top (\widehat{\mathbf{A}}_k - \widetilde{\mathbf{A}}) \mathbf{v} = \|\widehat{\mathbf{A}}_k - \widetilde{\mathbf{A}}\|_2$ . Because  $\mathbf{v}$  maximizes  $\mathbf{v}^\top (\widehat{\mathbf{A}}_k - \widetilde{\mathbf{A}}) \mathbf{v}$  over all unit-length vectors, it must lie in the range of  $(\widehat{\mathbf{A}}_k - \widetilde{\mathbf{A}})$  because otherwise the component outside the range will not contribute. Therefore, we can choose  $\mathbf{v}$  that  $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$  where  $\mathbf{v}_1 \in \text{Range}(\widehat{\mathbf{A}}_k) = \widehat{\mathcal{U}}_k$  and  $\mathbf{v}_2 \in \text{Range}(\widetilde{\mathbf{A}}) = \widetilde{\mathcal{U}}$ . Subsequently, we have that

$$\mathbf{v} = \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^\top \mathbf{v} + \widetilde{\mathbf{U}} \widetilde{\mathbf{U}}^\top \widehat{\mathbf{U}}_{n-k} \widehat{\mathbf{U}}_{n-k}^\top \mathbf{v} \quad (13)$$

$$= \widetilde{\mathbf{U}} \widetilde{\mathbf{U}}^\top \mathbf{v} + \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^\top \widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top \mathbf{v}. \quad (14)$$

Consider the following decomposition:

$$\left| \mathbf{v}^\top (\widehat{\mathbf{A}}_k - \widetilde{\mathbf{A}}) \mathbf{v} \right| \leq \left| \mathbf{v}^\top (\widehat{\mathbf{A}} - \mathbf{A}) \mathbf{v} \right| + \left| \mathbf{v}^\top (\widehat{\mathbf{A}}_k - \widehat{\mathbf{A}}) \mathbf{v} \right| + \left| \mathbf{v}^\top (\mathbf{A} - \widetilde{\mathbf{A}}) \mathbf{v} \right|.$$

The first term  $|\mathbf{v}^\top (\widehat{\mathbf{A}} - \mathbf{A}) \mathbf{v}|$  is trivially upper bounded by  $\|\widehat{\mathbf{A}} - \mathbf{A}\|_2 \leq \epsilon^2 \sigma_{k+1}(\mathbf{A})$ . For the second term, we have

$$\begin{aligned} \left| \mathbf{v}^\top (\widehat{\mathbf{A}}_k - \widetilde{\mathbf{A}}) \mathbf{v} \right| &= \left| \mathbf{v}^\top \widehat{\mathbf{U}}_{n-k} \widehat{\Sigma}_{n-k} \widehat{\mathbf{U}}_{n-k}^\top \mathbf{v} \right| \\ &\stackrel{(a)}{=} \left| \mathbf{v}^\top \widehat{\mathbf{U}}_{n-k} \widehat{\mathbf{U}}_{n-k}^\top \widetilde{\mathbf{U}} \widetilde{\mathbf{U}}^\top \widehat{\mathbf{U}}_{n-k} \widehat{\Sigma}_{n-k} \widehat{\mathbf{U}}_{n-k}^\top \widetilde{\mathbf{U}} \widetilde{\mathbf{U}}^\top \widehat{\mathbf{U}}_{n-k} \widehat{\mathbf{U}}_{n-k}^\top \mathbf{v} \right| \\ &\leq \left\| \widehat{\mathbf{U}}_{n-k}^\top \widetilde{\mathbf{U}} \right\|_2^4 \left\| \widehat{\mathbf{U}}_{n-k} \right\|_2 \stackrel{(b)}{\leq} 16\epsilon^4 \sigma_{k+1}(\widehat{\mathbf{A}}) \stackrel{(c)}{\leq} 16\epsilon^4 (1 + \epsilon^2) \sigma_{k+1}(\mathbf{A}). \end{aligned}$$

Here in (a) we apply Eq. (10); in (b) we apply Property 3 of Lemma A.3, and (c) is due to Weyl's inequality (Lemma C.4) that  $\sigma_{k+1}(\widehat{\mathbf{A}}) \leq \sigma_{k+1}(\mathbf{A}) + \|\widehat{\mathbf{A}} - \mathbf{A}\|_2 \leq (1 + \epsilon^2) \sigma_{k+1}(\mathbf{A})$ .

For the third term, note that  $\widetilde{\mathbf{A}} = \widetilde{\mathbf{U}} \widetilde{\mathbf{U}}^\top \mathbf{A} \widetilde{\mathbf{U}} \widetilde{\mathbf{U}}^\top$  because  $\text{Range}(\widetilde{\mathbf{A}}) \subseteq \mathcal{U}_{m_2} \subseteq \text{Range}(\mathbf{A})$  by Lemma A.3. Subsequently,

$$\mathbf{A} - \widetilde{\mathbf{A}} = \underbrace{\widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top \mathbf{A} \widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top}_{\mathbf{B}_1} + \underbrace{\widetilde{\mathbf{U}} \widetilde{\mathbf{U}}^\top \mathbf{A} \widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top}_{\mathbf{B}_2} + \underbrace{\widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top \mathbf{A} \widetilde{\mathbf{U}} \widetilde{\mathbf{U}}^\top}_{\mathbf{B}_2^\top}.$$

It then suffices to upper bound  $|\mathbf{v}^\top \mathbf{B}_1 \mathbf{v}|$  and  $|\mathbf{v}^\top \mathbf{B}_2 \mathbf{v}|$  separately. For  $\mathbf{B}_1$  we have

$$\begin{aligned} \left| \mathbf{v}^\top \mathbf{B}_1 \mathbf{v} \right| &\stackrel{(a')}{=} \left| \mathbf{v}^\top \widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^\top \widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top \mathbf{A} \widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^\top \widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top \mathbf{v} \right| \\ &\leq \left\| \widetilde{\mathbf{U}}_\perp^\top \widehat{\mathbf{U}}_k \right\|_2^4 \left\| \widetilde{\mathbf{U}}_\perp^\top \mathbf{A} \widetilde{\mathbf{U}}_\perp \right\|_2 \\ &\stackrel{(b')}{\leq} 16\epsilon^4 \left\| \widetilde{\mathbf{U}}_\perp^\top \mathbf{A} \widetilde{\mathbf{U}}_\perp \right\|_2 \stackrel{(c')}{\leq} 16\epsilon^4 \sigma_{m_1+1}(\mathbf{A}) \stackrel{(d')}{\leq} 16\epsilon^4 (1 + 2\epsilon) \sigma_{k+1}(\mathbf{A}). \end{aligned}$$

Here in (a') we apply Eq. (11); in (b') we apply Property 3 of Lemma A.3; (c') follows the property that  $\widetilde{\mathcal{U}}_\perp \in \mathcal{U}_{n-m_1}$ , and finally (d') follows from the definition of  $m_1$  that  $\sigma_{m_1+1}(\mathbf{A}) \leq (1 + 2\epsilon) \sigma_{k+1}(\mathbf{A})$ .

For  $\mathbf{B}_2$ , we have that

$$\begin{aligned} \left| \mathbf{v}^\top \mathbf{B}_2 \mathbf{v} \right| &= \left| \mathbf{v}^\top \widetilde{\mathbf{U}} \widetilde{\mathbf{U}}^\top \mathbf{A} \widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^\top \widetilde{\mathbf{U}}_\perp \widetilde{\mathbf{U}}_\perp^\top \mathbf{v} \right| \\ &\leq \left\| \mathbf{A} \widetilde{\mathbf{U}}_\perp \right\|_2 \left\| \widetilde{\mathbf{U}}_\perp^\top \widehat{\mathbf{U}}_k \right\|_2^2 \leq \epsilon^2 (1 + 8\epsilon) \sigma_{k+1}(\mathbf{A}). \end{aligned}$$

Combining all inequalities and noting that  $\epsilon \in (0, 1/4]$ , we obtain

$$\begin{aligned} \|\widehat{\mathbf{A}}_k - \widetilde{\mathbf{A}}\|_2 &\leq \epsilon^2 \sigma_{k+1}(\mathbf{A}) + 16\epsilon^4 (1 + 2\epsilon + \epsilon^2) \sigma_{k+1}(\mathbf{A}) + 32\epsilon^2 (1 + 8\epsilon) \sigma_{k+1}(\mathbf{A}) \\ &\leq 102\epsilon^2 \sigma_{k+1}(\mathbf{A}). \end{aligned}$$

□

*Proof. of Theorem 2.2* The proof of Theorem 2.2 is similar and even simpler than that of Theorem 2.1. First observing that with the large spectral gap,  $\tilde{\mathbf{A}} = \mathbf{A}_k$ . Next we replace by replacing the assumption  $\|\hat{\mathbf{A}} - \mathbf{A}\|_2 \leq \epsilon^2 \sigma_{k+1}(\mathbf{A})$  in Lemma 3.4 with  $\|\hat{\mathbf{A}} - \mathbf{A}\|_2 \leq \epsilon (\sigma_k(\mathbf{A}) - \sigma_{k+1}(\mathbf{A}))$  using the exactly the same arguments we have

$$\|\hat{\mathbf{A}}_k - \mathbf{A}_k\|_2 \leq 102\epsilon (\sigma_k(\mathbf{A}) - \sigma_{k+1}(\mathbf{A})).$$

Therefore, we have

$$\|\hat{\mathbf{A}}_k - \mathbf{A}_k\|_F \leq 102\sqrt{2k}\epsilon (\sigma_k(\mathbf{A}) - \sigma_{k+1}(\mathbf{A})).$$

Lastly, apply triangle inequality:

$$\begin{aligned} \|\hat{\mathbf{A}}_k - \mathbf{A}\|_F &\leq \|\mathbf{A} - \mathbf{A}_k\|_F + \|\hat{\mathbf{A}}_k - \mathbf{A}_k\|_F \\ &\leq \|\mathbf{A} - \mathbf{A}_k\|_F + 102\sqrt{2k}\epsilon (\sigma_k(\mathbf{A}) - \sigma_{k+1}(\mathbf{A})). \end{aligned}$$

□

## B Proof of corollaries

*Proof. of Corollary 2.1.* We first verify the condition that  $\delta \leq \epsilon^2 \sigma_{k+1}(\mathbf{A})$  for  $\epsilon = 1/4$  and the particular choice of  $k$ . Because  $k \leq \lfloor C_1 \delta^{-1/\beta} \rfloor - 1$ , we have that  $\sigma_{k+1}(\mathbf{A}) \geq (C_1 \delta^{-1/\beta})^{-\beta}$ . By carefully chosen  $C_1$  (depending on  $\beta$ ) the inequality  $\sigma_{k+1}(\mathbf{A}) \geq \delta/16$  holds.

If  $k = n - 1$  then by Theorem 2.1,  $\|\hat{\mathbf{A}}_k - \mathbf{A}\|_F \leq O(\sqrt{n} \cdot n^{-\beta}) = O(n^{-\frac{2\beta-1}{2}})$ . In the rest of the proof we assume  $k = \lfloor C_1 \delta^{-1/\beta} \rfloor - 1$ . We then have

$$\|\mathbf{A} - \mathbf{A}_k\|_F = \sqrt{\sum_{j=k+1}^n \sigma_j(\mathbf{A})^2} = \sqrt{\sum_{j=k+1}^n j^{-2\beta}} \leq \sqrt{\int_k^\infty x^{-2\beta} dx} = \sqrt{\frac{k^{-(2\beta-1)}}{2\beta-1}} \leq C(\beta) \delta^{\frac{2\beta-1}{2\beta}}.$$

Here  $C(\beta) > 0$  is a constant that only depends on  $\beta$ . In addition,

$$\sqrt{k} \|\mathbf{A} - \mathbf{A}_k\|_2 \leq \sqrt{k} \cdot k^{-\beta} = k^{-(\beta-1/2)} \leq \tilde{C}(\beta) \delta^{\frac{2\beta}{2\beta-1}}.$$

Applying Theorem 2.1 we complete the proof of Corollary 2.1. □

*Proof. of Corollary 2.2* We first verify the condition that  $\delta \leq \epsilon^2 \sigma_{k+1}(\mathbf{A})$  for  $\epsilon = 1/4$  and the particular choice of  $k$ . Because  $k \leq \lfloor c^{-1} \log(1/\delta) - c^{-1} \log \log(1/\delta) \rfloor - 1$ , we have that  $\sigma_{k+1}(\mathbf{A}) \geq \delta \log(1/\delta)$ . Hence, for  $\delta \in (0, e^{-16})$  it holds that  $\sigma_{k+1}(\mathbf{A}) \geq \delta/16$ .

If  $k = n - 1$  then by Theorem 2.1,  $\|\hat{\mathbf{A}}_k - \mathbf{A}\|_F \leq O(\sqrt{n} \cdot \exp\{-cn\})$ . In the rest of the proof we assume  $k = \lfloor C_2 \log(1/\delta) \rfloor - 1$ . We then have

$$\|\mathbf{A} - \mathbf{A}_k\|_F = \sqrt{\sum_{j=k+1}^n \sigma_j(\mathbf{A})^2} = \sqrt{\sum_{j=k+1}^n \exp\{-2cj\}} \leq \sqrt{\frac{\exp\{-2ck\}}{1 - e^{-2c}}} \leq C(c) \delta \log(1/\delta),$$

where  $C(c) > 0$  is a constant that only depends on  $c$ . In addition,

$$\sqrt{k} \|\mathbf{A} - \mathbf{A}_k\|_2 \leq \sqrt{k} \cdot \exp\{-ck\} \leq \delta \log(1/\delta) \cdot \sqrt{c^{-1} \log(1/\delta)} \leq \tilde{C}(c) \delta \sqrt{\log(1/\delta)^3}.$$

Applying Theorem 2.1 we complete the proof of Corollary 2.2. □

## C Technical lemmas

**Lemma C.1** (Asymmetric Davis-Kahan inequality). *Fix  $i \leq j \leq n$  and suppose  $\mathbf{X}, \mathbf{Y}$  are symmetric  $n \times n$  matrices, with eigen-decomposition  $\mathbf{X} = \mathbf{P}_i \mathbf{\Lambda}_i \mathbf{P}_i^\top + \mathbf{P}_{n-i} \mathbf{\Lambda}_{n-i} \mathbf{P}_{n-i}^\top$  and  $\mathbf{Y} = \mathbf{Q}_j \mathbf{\Xi}_j \mathbf{Q}_j^\top + \mathbf{Q}_{n-j} \mathbf{\Xi}_{n-j} \mathbf{Q}_{n-j}^\top$ . If  $\sigma_i(\mathbf{X}) > \sigma_{j+1}(\mathbf{Y})$  then*

$$\|\mathbf{Q}_{n-j}^\top \mathbf{P}_i\|_2 \leq \frac{\|\mathbf{X} - \mathbf{Y}\|_2}{\sigma_i(\mathbf{X}) - \sigma_{j+1}(\mathbf{Y})}.$$

*Proof.* Consider

$$\|\mathbf{Q}_{n-j}^\top(\mathbf{X} - \mathbf{Y})\mathbf{P}_i\|_2 = \|\mathbf{Q}_{n-j}^\top\mathbf{P}_i\mathbf{\Lambda}_i - \mathbf{\Xi}_{n-j}\mathbf{Q}_{n-j}^\top\mathbf{P}_i\|_2 \geq \|\mathbf{Q}_{n-j}^\top\mathbf{P}_i\|_2 (\sigma_i(\mathbf{X}) - \sigma_{j+1}(\mathbf{Y})).$$

Because  $\sigma_i(\mathbf{X}) > \sigma_{j+1}(\mathbf{Y})$ , we have that

$$\|\mathbf{Q}_{n-j}^\top\mathbf{P}_i\|_2 \leq \frac{\|\mathbf{Q}_{n-j}^\top(\mathbf{X} - \mathbf{Y})\mathbf{P}_i\|_2}{\sigma_i(\mathbf{X}) - \sigma_{j+1}(\mathbf{Y})} \leq \frac{\|\mathbf{X} - \mathbf{Y}\|_2}{\sigma_i(\mathbf{X}) - \sigma_{j+1}(\mathbf{Y})}.$$

□

**Lemma C.2** (Pythagorean theorem). *Fix  $n \geq m$ . Suppose  $\mathbf{X}$  is a symmetric  $n \times n$  matrix and  $\mathbf{P}$  is an  $n \times m$  matrix satisfying  $\mathbf{P}^\top\mathbf{P} = \mathbf{I}$ . Then  $\|\mathbf{X}\|_F^2 = \|\mathbf{X} - \mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top\|_F^2 + \|\mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top\|_F^2$ .*

*Proof.* Expanding  $\|\mathbf{X}\|_F^2$  we have that

$$\begin{aligned} \|\mathbf{X}\|_F^2 &= \|(\mathbf{X} - \mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top) + \mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top\|_F^2 \\ &= \|\mathbf{X} - \mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top\|_F^2 + \|\mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top\|_F^2 + 2\text{tr}[(\mathbf{X} - \mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top)\mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top]. \end{aligned}$$

It suffices to prove that the trace term is zero:

$$\begin{aligned} \text{tr}[(\mathbf{X} - \mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top)\mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top] &= \text{tr}(\mathbf{X}\mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top) - \text{tr}(\mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top\mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top) \\ &\stackrel{(*)}{=} \text{tr}(\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}) - \text{tr}(\mathbf{P}^\top\mathbf{X}\mathbf{P}\mathbf{P}^\top\mathbf{X}\mathbf{P}) \\ &= 0. \end{aligned}$$

Here (\*) is due to  $\mathbf{P}^\top\mathbf{P} = \mathbf{I}$ . □

**Lemma C.3** (Poincaré separation theorem). *Fix  $n \geq m$ . Suppose  $\mathbf{X}$  is a symmetric  $n \times n$  matrix,  $\mathbf{P}$  is an  $n \times m$  matrix that satisfies  $\mathbf{P}^\top\mathbf{P} = \mathbf{I}$ , and  $\mathbf{Y} = \mathbf{P}^\top\mathbf{X}\mathbf{P}$ . Let  $\sigma_1(\mathbf{X}) \geq \dots \geq \sigma_n(\mathbf{X})$  and  $\sigma_1(\mathbf{Y}) \geq \dots \geq \sigma_m(\mathbf{Y})$  be the eigenvalues of  $\mathbf{X}$  and  $\mathbf{Y}$  in descending order. Then*

$$\sigma_i(\mathbf{X}) \geq \sigma_i(\mathbf{Y}) \geq \sigma_{n-m+i}(\mathbf{X}), \quad i = 1, \dots, m.$$

**Lemma C.4** (Weyl's monotonicity theorem). *Suppose  $\mathbf{X}, \mathbf{Y}$  are  $n \times n$  symmetric matrices, and let  $\sigma_1(\mathbf{X}) \geq \dots \geq \sigma_n(\mathbf{X})$ ,  $\sigma_1(\mathbf{Y}) \geq \dots \geq \sigma_n(\mathbf{Y})$  and  $\sigma_1(\mathbf{X} + \mathbf{Y}) \geq \dots \geq \sigma_n(\mathbf{X} + \mathbf{Y})$  denote the eigenvalues of  $\mathbf{X}, \mathbf{Y}$  and  $\mathbf{X} + \mathbf{Y}$  in descending order. Then*

$$\sigma_{i+j-1}(\mathbf{X} + \mathbf{Y}) \leq \sigma_i(\mathbf{X}) + \sigma_j(\mathbf{Y}), \quad 1 \leq i, j \leq n, i + j - 1 \leq n.$$

*In particular, setting  $i = 1$  one obtains the commonly used Weyl's inequality:  $|\sigma_j(\mathbf{X} + \mathbf{Y}) - \sigma_j(\mathbf{X})| \leq \|\mathbf{Y}\|_2$ .*