Communication-efficient distributed eigenspace estimation with arbitrary node failures

Supplementary material

Vasileios Charisopoulos
Operations Research & Information Engineering
Cornell University
Ithaca, NY 14853
vc333@cornell.edu

Anil Damle
Computer Science
Cornell University
Ithaca, NY 14853
damle@cornell.edu

Contents

A Auxiliary results

In this section, we present a few supporting results. The first result is a path independence lemma for perturbations of eigenvectors. It first appeared in [7]; the eigengap condition in the statement of the Lemma is justified in [1, Lemma 5].

Lemma A.1 (Path independence)**.** *Let $A \in \mathbb{R}^{d \times d}$ be a fixed* symmetric *matrix and let $\hat{A} := A + E$, where E is a symmetric perturbation. Suppose that we can write*

$$\hat{A} - A = E_0 + E_1 = F_0 + F_1,$$

where E_0, E_1, F_0, F_1 are symmetric matrices, and define the intermediate *matrices*

$$\hat{A}_1 := A + E_0,\ \hat{A}_2 = \hat{A}_1 + E_1, \quad \tilde{A}_1 := A + F_0,\ \tilde{A}_2 = \tilde{A}_1 + F_1.$$

Fix any $V \in O(d, r)$ whose columns span the principal r-dimensional invariant subspace of A and construct the leading eigenvector matrices $\hat{V}_1$, $\hat{V}_2 \in O(d, r)$ of $\hat{A}_1$ and $\hat{A}_2$ such that

$$\min_{U \in \mathbb{O}_r} \left\| \hat{V}_1 U - V \right\|_F = \left\| \hat{V}_1 - V \right\|_F, \quad \min_{U \in \mathbb{O}_r} \left\| \hat{V}_2 U - \hat{V}_1 \right\|_F = \left\| \hat{V}_2 - \hat{V}_1 \right\|_F.$$

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Further, let $\tilde{V}_1$ and $\tilde{V}_2$ be the leading eigenvector matrices of $\tilde{A}_1$ and $\tilde{A}_2$, constructed in a similar fashion. Then, $\hat{V}_2$ and $\tilde{V}_2$ both span principal invariant subspaces of $\hat{A} = A + E$. Moreover, they satisfy

$$\hat{V}_2 = \tilde{V}_2 + T, \quad \|T\|_2 \lesssim \frac{\varepsilon^2}{\delta^2}, \quad \varepsilon := \max\{\|E_0\|_2, \|E_1\|_2, \|F_0\|_2, \|F_1\|_2\},$$

as long as A satisfies $\delta_r(A) \geq 4\varepsilon$.

Lemma A.2. *Suppose that $U \in \mathbb{O}_{d,r}$ satisfies $\operatorname{dist}(U, V) \leq \varepsilon < 1/2$, where V is the principal eigenvector matrix of a symmetric matrix A with eigengap $\delta_r(A) := \lambda_r(A) - \lambda_{r+1}(A) > 0$. Then there exists a symmetric matrix B such that the following hold:*

1. *$\|A - B\|_2 \leq 8\|A\|_2 \varepsilon$ and $\delta_r(B) = \delta_r(A)$.*
2. *U is the principal eigenvector matrix of B.*

Proof. We prove Item 1 first. To that end, we can write $A = A_1 + A_2$, where $A_1 := V\Sigma_1 V^\mathsf{T}$ and $A_2 := V_\perp \Sigma_2 V_\perp^\mathsf{T}$. We consider the following matrix B:

$$B = U\Sigma_1 U^\mathsf{T} + U_\perp \Sigma_2 U_\perp^\mathsf{T}, \tag{1}$$

where $U_\perp^\mathsf{T} U = 0$ and $U_\perp \in \mathbb{O}_{d,d-r}$. From (1) and the condition $\Sigma_1 \succ \Sigma_2$, it follows that U is a principal eigenvector matrix for B. Moreover, the gap condition on A immediately translates to the claimed gap condition for B.

It remains to bound the distance between A and B. We write

$$\|A - B\|_2 \leq \left\|U\Sigma_1 U^\mathsf{T} - A_1\right\|_2 + \left\|U_\perp \Sigma_2 U_\perp^\mathsf{T} - A_2\right\|_2.$$

To upper bound the first term on the right-hand side above, we use the spectral projectors $P_U := UU^\mathsf{T}$ and $P_{U_\perp} = I - P_U$ to decompose it into

$$\begin{aligned}
\left\|U\Sigma_1 U^\mathsf{T} - A_1\right\|_2 &\leq \left\|U\Sigma_1 U^\mathsf{T} - P_U A_1\right\|_2 + \|P_{U_\perp} A_1\|_2 \\
&\leq \left\|U\Sigma_1 U^\mathsf{T} - P_U A_1 P_U\right\|_2 + \|P_U A_1 P_{U_\perp}\|_2 + \|P_{U_\perp} A_1\|_2 \\
&\leq \left\|\Sigma_1 - U^\mathsf{T} V\Sigma_1 V^\mathsf{T} U\right\|_2 + 2\|\Sigma_1\|_2 \|P_{U_\perp} V\|_2 \\
&\leq \left\|\Sigma_1 (I - V^\mathsf{T} U)\right\|_2 + \left\|(I - V^\mathsf{T} U)\Sigma_1 V^\mathsf{T} U\right\|_2 + 2\|\Sigma_1\|_2 \|P_{U_\perp} V\|_2 \\
&\leq 2\|\Sigma_1\|_2 \varepsilon + 4\|\Sigma_1\|_2 \varepsilon^2,
\end{aligned}$$

where the last inequality follows from the inequality $\|P_{U_\perp} V\|_2 = \operatorname{dist}(U, V) = \varepsilon$ and Lemma A.3. A similar argument shows that $\|U_\perp \Sigma_2 U_\perp^\mathsf{T} - A_2\|_2 \leq 2\|\Sigma_2\|_2 \varepsilon + 4\|\Sigma_2\|_2 \varepsilon^2$. Taking into account the bound $\varepsilon < 1/2$ completes the proof. ☐

Lemma A.3 (Modified $\sin\theta$ distance)**.** *Let $U, V \in O(d, r)$ satisfy $\operatorname{dist}(U, V) = \alpha < 1$. Then the following holds:*

$$\left\|I - U^\mathsf{T} V\right\|_2 \leq 2\alpha^2.$$

Proof. Let $P\Sigma Q^\mathsf{T}$ be the singular value decomposition of $U^\mathsf{T} V$. Recall that [2, Eq. (2.5)]:

$$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r); \quad \sigma_i = \cos(\theta_i),$$

where $\theta_i \in [0, \pi/2]$ and $\|\sin\Theta\|_2 = \alpha$, following [2, Lemma 2.5]. From our assumptions, it follows that

$$\begin{aligned}
\left\|I - U^\mathsf{T} V\right\|_2 &= \left\|P(I - \Sigma)Q^\mathsf{T}\right\|_2 \\
&= \|I - \Sigma\|_2 \\
&= \max_{i \in [r]} \{1 - \cos(\theta_i)\} \\
&= \max_{i \in [r]} \{2\sin^2(\theta_i/2)\} \\
&\leq 2 \max_{i \in [r]} \sin^2(\theta_i) \\
&= 2\|\sin\Theta\|_2^2,
\end{aligned}$$

with the last inequality following from $0 \leq \sin(\theta/2) \leq \sin(\theta)$ for any $\theta \in [0, \pi/2]$. ☐

B Additional proofs

This section includes proofs that were omitted from the main text.

B.1 Proof of Proposition 1

Proof. Define $\mathcal{C} := \{i \in [m] \mid \mathrm{dist}(Y_i, V) \leq \varepsilon\}$ with $|\mathcal{C}| > \frac{m}{2}$. Now, we consider any pair (i, j) with $i, j \in \mathcal{C}$. By the triangle inequality,

$$\mathrm{dist}(Y_i, Y_j) \leq \mathrm{dist}(Y_i, V) + \mathrm{dist}(Y_j, V) \leq 2\varepsilon, \quad \text{for all } i, j \in \mathcal{C}. \tag{2}$$

Now, fix $i_\star$ to be any index for which $\mathrm{dist}(Y_i, Y_j) \leq 2\varepsilon$ for at least $m/2$ other indices $j \neq i_\star$ (such an index always exists because $|\mathcal{C}| \geq \frac{m}{2} + 1$). For any such $i_\star$, there must be another index j satisfying $\mathrm{dist}(Y_j, V) \leq \varepsilon$ and $\mathrm{dist}(Y_j, Y_{i_\star}) \leq 2\varepsilon$. Therefore,

$$\mathrm{dist}(Y_{i_\star}, V) \leq \mathrm{dist}(Y_{i_\star}, Y_j) + \mathrm{dist}(Y_j, V) \leq 3\varepsilon.$$

□

B.2 Proof of Lemma 1

Proof. Recall that V is an eigenvector matrix of A that satisfies

$$\min_{Z \in \mathbb{O}_r} \|V - V_{\mathsf{ref}} Z\|_{\mathrm{F}} = \|V - V_{\mathsf{ref}}\|_{\mathrm{F}}.$$

From Lemma A.2, it follows that the columns of V_{ref} span the principal eigenspace of a matrix B with nontrivial eigengap that satisfies $\|A - B\|_2 \lesssim \|A\|_2 \varepsilon$. We now relate V_i^{ideal} to V_i^{corr} using the aforementioned path independence result.

To that end, note that V_i^{ideal} is the leading eigenvector matrix of

$$A_i := A + (A_i - A) + 0,$$

that has been maximally aligned with V (in the sense of Frobenius distance). On the other hand, the Procrustes estimates V_i^{corr} are given by the leading eigenvector matrices of

$$A_i := A + (B - A) + (A_i - B),$$

since V_{ref} is the leading eigenvector of B nearest to V and V_i^{corr} is formed as the leading eigenvector of A_i nearest to V_{ref}. Applying Lemma A.1 with $E_0 := A_i - A$, $E_1 = \mathbf{0}$, $F_0 := B - A$ and $F_1 := A_i - B$, we obtain

$$V_i^{\mathsf{corr}} = V_i^{\mathsf{ideal}} + \mathcal{O}\left(\frac{1}{\delta^2} \max\left\{\|A_i - A\|_2^2, \|B - A\|_2^2, \|A_i - B\|_2^2\right\}\right).$$

Finally, we note the following upper bound

$$\|A_i - B\|_2^2 \lesssim \|A_i - A\|_2^2 + \|A - B\|_2^2 \lesssim \max\left(\|A_i - A\|_2^2, \|A\|_2^2 \varepsilon^2\right),$$

which concludes the proof. □

B.3 Proof of Proposition 3

Proof. Let j_{crit} be the smallest index for which $2^j \geq \max\{\lambda_{\mathsf{lb}}, \|\Sigma_{G_0}\|_2\}$. For a fixed corruption fraction α and failure probability p, define the events

$$\mathcal{E}_j := \left\{\left\|\theta_{2^j} - \frac{1}{|G_0|}\sum_{i \in G_0} X_i\right\|_2 \leq f(2^j; p, \alpha)\right\}, \quad \mathcal{E} := \bigcap_{j = j_{\mathsf{crit}}}^{j_{\mathsf{hi}}} \mathcal{E}_j.$$

From Theorem 3 in the main text and a union bound, it follows that

$$\begin{aligned}
\mathbb{P}(\mathcal{E}) &\geq 1 - \sum_{j \in \{j_{\mathsf{crit}}, \dots, j_{\mathsf{hi}}\}} \mathbb{P}\left(\left\|\theta_{2^j} - \frac{1}{|G_0|}\sum_{i \in G_0} X_i\right\|_2 \geq f(2^j; p, \alpha)\right) \\
&\geq 1 - (j_{\mathsf{hi}} - j_{\mathsf{crit}}) \cdot p
\end{aligned}$$

$$\geq 1 - 2\log_2\left(\frac{\lambda_{\mathsf{ub}}}{\lambda_{\mathsf{lb}}}\right)p.$$

Let us write $\theta_* := \frac{1}{|G_0|}\sum_{i\in G_0} X_i$. Conditioned on the event $\mathcal{E}$, for any $j, j' \geq j_{\mathsf{crit}}$ we have

$$\|\theta_{2^j} - \theta_{2^{j'}}\|_2 \leq \|\theta_{2^j} - \theta_*\|_2 + \|\theta_{2^{j'}} - \theta_*\|_2 \leq f(2^j; p, \alpha) + f(2^{j'}; p, \alpha).$$

Consequently, it follows that $2^{j_{\mathsf{crit}}}$ satisfies the condition of the estimator, and therefore

$$\hat{\lambda} \leq 2^{j_{\mathsf{crit}}} \leq 2\max\left\{\lambda_{\mathsf{lb}}, \|\Sigma_{G_0}\|_2\right\}.$$

Finally, the desired claim follows since

$$\begin{aligned}
\|\theta_{\hat{\lambda}} - \theta_*\|_2 &\leq \|\theta_{\hat{\lambda}} - \theta_{2^{j_{\mathsf{crit}}}}\|_2 + \|\theta_{2^{j_{\mathsf{crit}}}} - \theta_*\|_2 \\
&\leq f(\hat{\lambda}; p, \alpha) + 2f(2^{j_{\mathsf{crit}}}; p, \alpha) \\
&\leq 3f(2^{j_{\mathsf{crit}}}; p, \alpha) \\
&\leq 171\sqrt{\|\Sigma_{G_0}\|_2}\left(\alpha + \frac{4\log(1/p)}{m}\right)^{1/2}.
\end{aligned}$$

□

The next Lemma provides an upper bound on the operator norm of the empirical covariance $\Sigma_{\mathcal{I}_{\mathsf{good}}}$.

Lemma B.1. *Suppose that $\widehat{V}_{\mathsf{ref}}$ satisfies $\delta_r(A) \geq 8\,\mathrm{dist}(\widehat{V}_{\mathsf{ref}}, V)$. Then we have*

$$\|\Sigma_{\mathcal{I}_{\mathsf{good}}}\|_2 \leq \left\|\frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} V_i V_i^{\mathsf{T}} - VV^{\mathsf{T}}\right\|_2 + 2\left\|\frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} \widetilde{V}_i - V\right\|_2. \tag{3}$$

Proof. Let μ denote the empirical mean over $\mathcal{I}_{\mathsf{good}}$. We have

$$\begin{aligned}
\mu &= \frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} \widetilde{V}_i, \\
\Sigma_{\mathcal{I}_{\mathsf{good}}} &= \frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} (\widetilde{V}_i - \mu)(\widetilde{V}_i - \mu)^{\mathsf{T}} \\
&= \frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} \widetilde{V}_i\widetilde{V}_i^{\mathsf{T}} - \mu\mu^{\mathsf{T}} \\
&= \frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} \widetilde{V}_i\widetilde{V}_i^{\mathsf{T}} - VV^{\mathsf{T}} + VV^{\mathsf{T}} - \mu\mu^{\mathsf{T}} \\
&= \frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} \widetilde{V}_i\widetilde{V}_i^{\mathsf{T}} - VV^{\mathsf{T}} + (V - \mu)(V + \mu)^{\mathsf{T}},
\end{aligned}$$

where $V \in \mathbb{O}_{d,r}$ spans the principal eigenspace of A and satisfies

$$\min_{Z\in\mathbb{O}_r}\left\|VZ - \widehat{V}_{\mathsf{ref}}\right\|_{\mathrm{F}} = \left\|V - \widehat{V}_{\mathsf{ref}}\right\|_{\mathrm{F}}.$$

We now bound the spectral norm of $\Sigma_{\mathcal{I}_{\mathsf{good}}}$. Indeed, we have

$$\begin{aligned}
\|\Sigma_{\mathcal{I}_{\mathsf{good}}}\|_2 &\leq \left\|\frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} \widetilde{V}_i\widetilde{V}_i^{\mathsf{T}} - VV^{\mathsf{T}}\right\|_2 + \|V + \mu\|_2\left\|\frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} \widetilde{V}_i - V\right\|_2 \\
&\leq \left\|\frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} V_iV_i^{\mathsf{T}} - VV^{\mathsf{T}}\right\|_2 + 2\left\|\frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i\in\mathcal{I}_{\mathsf{good}}} \widetilde{V}_i - V\right\|_2,
\end{aligned}$$

using the fact that $\widetilde{V}_i\widetilde{V}_i^{\mathsf{T}} = V_iV_i^{\mathsf{T}}$ for all $i \in \mathcal{I}_{\mathsf{good}}$. □

B.4 Proof of Theorem 3

In this section, we modify the proof of [6, Theorem 4] to derive guarantees for robust mean estimation with matrix-valued inputs. We recall some notation used therein: given the set of "good" samples G_0 and the initial sample $S_0 = \{1, \dots, m\}$, we denote

$$\begin{aligned} S_k &= \{\text{points remaining after } k \text{ recursive calls to } \texttt{Filter}\}, \\ G_k &= S_k \cap G_0, \\ B_k &= S_k \setminus G_0, \\ \alpha &= \frac{m - |G_0|}{m}. \end{aligned} \tag{4}$$

Moreover, given any set $S \subset [m]$, we write

$$\Sigma_S := \frac{1}{|S|} \sum_{i \in S} (X_i - \mu_S)(X_i - \mu_S)^{\mathsf{T}}, \quad \text{where} \quad \mu_S := \frac{1}{|S|} \sum_{i \in S} X_i. \tag{5}$$

In our proofs, we frequently employ the total variation distance d_{TV}. For discrete distributions P_1, P_2 on a common sample space Ω, d_{TV} is given by

$$d_{\mathrm{TV}}(P_1, P_2) = \frac{1}{2} \|P_1 - P_2\|_1 = \frac{1}{2} \sum_{x \in \Omega} |P_1(x) - P_2(x)|. \tag{6}$$

Finally, we define the events $\mathcal{E}_k$, where $k \in \mathbb{N}$, as below:

$$\mathcal{E}_k := \left\{ \sum_{i \in G_k} \tau_i \geq \frac{1}{\gamma} \sum_{j \in S_k} \tau_j \right\}, \quad k = 0, 1, \dots \tag{7}$$

Our proof essentially traces the proof of [6, Theorem 4] but for the case of matrix-valued inputs to the `Filter` algorithm. The first result has already been shown in [6], as its proof is independent of the shape of the inputs.

Lemma B.2 (See [6, Lemma 6]). *Let $T := \inf \{k \in \mathbb{N} \mid \mathcal{E}_k \text{ is true}\}$. Then we have:*

$$\mathbb{P}\left(T \geq 3(m - |G_0|) + 18 \log(1/p)\right) \leq p. \tag{8}$$

The remainder of the proof is devoted to showing that, as soon as some $\mathcal{E}_k$ is true, `Filter` will terminate with a good estimate. Throughout, we condition on the event

$$\mathcal{E} := \{T \leq T_p\}, \quad \text{where} \quad T_p := 3(m - |G_0|) + 18 \log(1/p), \tag{9}$$

which holds with probability at least $1 - p$.

Theorem B.1. *Suppose that α, p and N satisfy*

$$3\alpha + \frac{18 \log(1/p)}{m} \leq \frac{1}{4}. \tag{10}$$

Then the following hold simultaneously with probability at least $1 - p$:

1. $\texttt{Filter}(S_0, \|\Sigma_{G_0}\|_2)$ *terminates after at most T_p iterations;*
2. *The output of* $\texttt{Filter}(S_0, \|\Sigma_{G_0}\|_2)$, $\theta_{\|\Sigma_{G_0}\|_2}$, *satisfies*

$$\left\| \theta_{\|\Sigma_{G_0}\|_2} - \frac{1}{|G_0|} \sum_{i \in G_0} X_i \right\|_2 \leq 18 \sqrt{5 \|\Sigma_{G_0}\|_2} \left(\alpha + \frac{4 \log(1/p)}{m} \right)^{1/2}. \tag{11}$$

Remark 1. *While we prove the Theorem for the case $\lambda_{\mathsf{ub}} = \|\Sigma_{G_0}\|_2$, a straightforward modification of the proof shows that when $\lambda_{\mathsf{ub}} \geq \|\Sigma_{G_0}\|_2$, we have*

$$\left\| \theta_{\lambda_{\mathsf{ub}}} - \frac{1}{|G_0|} \sum_{i \in G_0} X_i \right\|_2 \leq 18 \sqrt{5 \lambda_{\mathsf{ub}}} \left(\alpha + \frac{4 \log(1/p)}{m} \right)^{1/2}.$$

Proof of Theorem B.1. We condition on the event $\mathcal{E}$ from (9), which holds with probability at least $1-p$. This implies that there is some index $k \leq T_p$ such that

$$\sum_{i \in G_k} \tau_i \geq \frac{1}{\gamma} \sum_{j \in S_k} \tau_j.$$

From Lemma B.7, we obtain that the empirical covariance satisfies $\|\Sigma_{S_k}\|_2 \leq 18 \|\Sigma_{G_0}\|_2$, and thus the algorithm terminates after at most k steps. We have the following cases:

1. The termination condition was first triggered at the k^{th} step. In that case, Lemma B.7 directly implies the desired inequality.
2. The algorithm terminated at some index $\ell < k$. Then it follows from Lemma B.8 that

$$\eta := d_{\mathrm{TV}}(\mathrm{Unif}(S_\ell), \mathrm{Unif}(G_0)) \leq 5\alpha + \frac{20\log(1/p)}{N}. \tag{12}$$

   At the same time, Lemma B.3 implies that

$$\left\|\theta_{\|\Sigma_{G_0}\|_2} - \frac{1}{|G_0|}\sum_{i \in G_0} X_i\right\|_2 \leq \frac{\sqrt{\eta}}{1-\sqrt{\eta}} \cdot \left(\|\Sigma_{S_\ell}\|_2^{1/2} + \|\Sigma_{G_0}\|_2^{1/2}\right). \tag{13}$$

   From the termination condition, we obtain that

$$\|\Sigma_{S_\ell}\|_2 \leq 18 \|\Sigma_{G_0}\|_2 . \tag{14}$$

   Combining Eqs. (12) to (14) yields the desired bound.

□

The next few Lemmas are supporting statements used in the proof of Theorem B.1.

Lemma B.3. *Let $S = \{X_1, \ldots, X_m\}$ where $X_i \in \mathbb{R}^{d \times r}$ and suppose that P_1, P_2 are discrete distributions supported over $[m]$ with $d_{\mathrm{TV}}(P_1, P_2) = \eta$. Then the following holds:*

$$\|\mathbb{E}_{P_1}[X_i] - \mathbb{E}_{P_2}[X_i]\|_2 \leq \frac{\sqrt{\eta}}{1-\sqrt{\eta}} \cdot \left(\|\Sigma_{P_1}\|_2^{1/2} + \|\Sigma_{P_2}\|_2^{1/2}\right), \tag{15}$$

where the matrices Σ_{P_i} are defined as:

$$\Sigma_{P_i} = \mathbb{E}_{X \sim P_i}\left[(X - \mathbb{E}_{P_i}[X])(X - \mathbb{E}_{P_i}[X])^{\mathsf{T}}\right].$$

Proof. Following the proof of [5, Lemma 2.1], we consider a coupling between P_1 and P_2 such that $\mathbb{P}(X = X') \geq 1 - \eta$. Denoting $\|X\|_{L^2} := \sqrt{\mathbb{E}[X^2]}$, we have

$$\begin{aligned}
\|\mathbb{E}_{P_1}[X] - \mathbb{E}_{P_2}[X']\|_2 &= \sup_{u,v \in \mathcal{B}} \langle u, (\mathbb{E}_{P_1}[X] - \mathbb{E}_{P_2}[X'])v\rangle \\
&= \sup_{u,v \in \mathcal{B}} \mathbb{E}\left[\langle u, (X - X')v\rangle \mathbf{1}\{X \neq X'\}\right] \\
&\leq \mathbb{E}\left[\mathbf{1}\{X \neq X'\}^2\right]^{1/2} \cdot \sup_{u,v \in \mathcal{B}} \mathbb{E}\left[\langle u, (X - X')v\rangle^2\right]^{1/2} \\
&\leq \sqrt{\eta} \cdot \sup_{u,v \in \mathcal{B}} \|\langle u, (X - X')v\rangle\|_{L^2} .
\end{aligned} \tag{16}$$

Let $\mu_1 := \mathbb{E}_{P_1}[X]$ and $\mu_2 = \mathbb{E}_{P_2}[X]$. Since $\|\cdot\|_{L^2}$ is a norm, the triangle inequality implies that

$$\begin{aligned}
\sup_{u,v \in \mathcal{B}} \|\langle u, (X - X')v\rangle\|_{L^2} &= \sup_{u,v \in \mathcal{B}} \|\langle u, (X - \mu_1 + \mu_1 - \mu_2 + \mu_2 - X')v\rangle\|_{L^2} \\
&\leq \sup_{u,v \in \mathcal{B}} \|\langle u, (X - \mu_1)v\rangle\|_{L^2} + \sup_{u,v \in \mathcal{B}} \|\langle u, (X' - \mu_2)v\rangle\|_{L^2} \\
&\quad + \sup_{u,v \in \mathcal{B}} \|\langle u, (\mu_1 - \mu_2)v\rangle\|_{L^2} .
\end{aligned} \tag{17}$$

We now upper bound the remaining terms. For the first one, we have

$$\begin{aligned}
\sup_{u,v\in\mathcal{B}} \|\langle u,(X-\mu_1)v\rangle\|_{L^2} &= \sup_{u,v\in\mathcal{B}} \mathbb{E}\left[\langle u,(X-\mu_1)v\rangle^2\right]^{1/2} \\
&= \sup_{u,v\in\mathcal{B}} \mathbb{E}[\mathrm{Tr}(u^\mathsf{T}(X-\mu_1)\underbrace{vv^\mathsf{T}}_{\preceq I_d}(X-\mu_1)^\mathsf{T}u)]^{1/2} \\
&\le \sup_{u\in\mathcal{B}} \mathbb{E}\left[\mathrm{Tr}\left(u^\mathsf{T}(X-\mu_1)(X-\mu_1)^\mathsf{T}u\right)\right]^{1/2} \\
&= \left(\sup_{u\in\mathcal{B}} \langle u, \mathbb{E}\left[(X-\mu_1)(X-\mu_1)^\mathsf{T}\right]u\rangle\right)^{1/2} \\
&= \|\Sigma_{P_1}\|_2^{1/2}, \qquad (18)
\end{aligned}$$

where the penultimate equality uses linearity of the trace operator and the last equality is the definition of the spectral norm for symmetric positive semidefinite matrices. Similar arguments also yield

$$\sup_{u,v\in\mathcal{B}} \|\langle u,(X'-\mu_2)v\rangle\|_{L^2} \le \|\Sigma_{P_2}\|_2^{1/2}, \qquad (19)$$

$$\sup_{u,v\in\mathcal{B}} \|\langle u,(\mu_1-\mu_2)v\rangle\|_{L^2} \le \|\mathbb{E}_{P_1}[X]-\mathbb{E}_{P_2}[X']\|_2. \qquad (20)$$

Plugging Eqs. (17) to (20) back into Eq. (16) and rearranging yields the expected result:

$$\|\mathbb{E}_{P_1}[X]-\mathbb{E}_{P_2}[X']\|_2 \le \frac{\sqrt{\eta}}{1-\sqrt{\eta}}\left(\|\Sigma_{P_1}\|_2^{1/2}+\|\Sigma_{P_2}\|_2^{1/2}\right).$$

□

Lemma B.4. *Let $G\subset S\subset[m]$. Moreover, let μ_S and μ_G be their respective empirical means, and let v be the leading eigenvector of Σ_S so that the outlier scores satisfy*

$$\tau_i = \left\langle v,(X_i-\mu_S)(X_i-\mu_S)^\mathsf{T}v\right\rangle, \quad \forall i\in S.$$

Moreover, define $\eta := 1-{}^{|G|}/_{|S|}$ and fix a $\gamma\in(0,{}^1/_\eta)$. Then, we have the implication

$$\|\Sigma_S\|_2 \ge (1-\eta)^2\left(\frac{\gamma}{1-\gamma\eta}\right)\|\Sigma_G\|_2 \implies \sum_{j\in G}\tau_j \le \frac{1}{\gamma}\sum_{i\in S}\tau_i. \qquad (21)$$

Proof. Recall that the (normalized) sum of outlier scores over the set G is given by

$$\begin{aligned}
\frac{1}{|G|}\left\langle v,\sum_{i\in G}(X_i-\mu_S)(X_i-\mu_S)^\mathsf{T}v\right\rangle &= \frac{1}{|G|}\left\langle v,\sum_{i\in G}(X_i-\mu_G)(X_i-\mu_G)^\mathsf{T}v\right\rangle \\
&\quad + \left\langle v,(\mu_S-\mu_G)(\mu_S-\mu_G)^\mathsf{T}v\right\rangle \\
&= \langle v,\Sigma_G v\rangle + \left\langle v,(\mu_S-\mu_G)(\mu_S-\mu_G)^\mathsf{T}v\right\rangle. \qquad (22)
\end{aligned}$$

We now simplify the second term. Indeed, we have

$$\begin{aligned}
\mu_S-\mu_G &= \frac{1}{|S|}\sum_{i\in G}X_i + \frac{1}{|S|}\sum_{i\in S\setminus G}X_i - \frac{1}{|G|}\sum_{i\in G}X_i \\
&= \left(1-\frac{|G|}{|S|}\right)(\mu_{S\setminus G}-\mu_G) \qquad (23)
\end{aligned}$$

For brevity, denote $\eta := \frac{|S\setminus G|}{|S|}$. Plugging (23) back into (22), we obtain

$$\frac{1}{|G|}\sum_{j\in G}\tau_j = \langle v,\Sigma_G v\rangle + \eta^2\left\langle v,(\mu_{S\setminus G}-\mu_G)(\mu_{S\setminus G}-\mu_G)^\mathsf{T}v\right\rangle \qquad (24)$$

We now bound the second term in (24). From [3, Lemma 2.4], it follows that

$$\langle v,\Sigma_S v\rangle = (1-\eta)\langle v,\Sigma_G v\rangle + \eta\langle v,\Sigma_{S\setminus G}v\rangle + \eta(1-\eta)\left\langle v,(\mu_{S\setminus G}-\mu_G)(\mu_{S\setminus G}-\mu_G)^\mathsf{T}v\right\rangle$$

Rearranging and multiplying by $\eta/(1-\eta)$ gives

$$\begin{aligned}\eta^2 \left\langle v, (\mu_{S\setminus G} - \mu_G)(\mu_{S\setminus G} - \mu_G)^\mathsf{T} v \right\rangle &= \frac{\eta}{1-\eta} \left\langle v, \Sigma_S v \right\rangle - \eta \left\langle v, \Sigma_G v \right\rangle - \frac{\eta^2}{(1-\eta)} \left\langle v, \Sigma_{S\setminus G} v \right\rangle \\ &\leq \frac{\eta}{1-\eta} \left\langle v, \Sigma_S v \right\rangle - \eta \left\langle v, \Sigma_G v \right\rangle .\end{aligned}$$

Plugging back into Eq. (24) and using the fact that $|G| = |S|(1-\eta)$, we obtain

$$\begin{aligned}\sum_{j\in G} \tau_j &\leq |G|(1-\eta) \left\langle v, \Sigma_G v \right\rangle + \frac{|G|\,\eta}{1-\eta} \left\langle v, \Sigma_S v \right\rangle \\ &\leq |G|(1-\eta) \left\|\Sigma_G\right\|_2 + (|S| - |G|)\,|S| \left\|\Sigma_S\right\|_2 \end{aligned} \tag{25}$$

Finally, replacing $|G| = |S|(1-\eta)$ in (25) and rearranging, we obtain

$$\left\|\Sigma_G\right\|_2 \leq (\gamma^{-1} - \eta) \frac{\left\|\Sigma_S\right\|_2}{(1-\eta)^2} \implies \sum_{j\in G} \tau_j \leq \frac{1}{\gamma} \sum_{i\in S} \tau_i.$$

□

Lemma B.5. *Suppose that (10) is true. Then the following holds for any $k \leq T_p$:*

$$\frac{|S_k \setminus G_k|}{|S_k|} \leq \frac{4\alpha}{3}.$$

Proof. Recall that $B_k = S_k \setminus G_k$ and notice that

$$\frac{|B_k|}{|S_k|} = \frac{|B_k|}{|S_0|}\frac{|S_0|}{|S_k|} \leq \frac{|B_0|}{|S_0|}\frac{|S_0|}{|S_0| - T_p} = \alpha \cdot \frac{1}{1 - (3\alpha + \frac{18\log(1/p)}{m})} \leq \frac{4\alpha}{3},$$

where the first inequality follows from the fact that $|B_k| \leq |B_0|$. □

Lemma B.6. *For any integer k, the sets G_k and G_0 satisfy*

$$\left\|\Sigma_{G_k}\right\|_2 \leq \frac{|G_0|}{|G_k|} \left\|\Sigma_{G_0}\right\|_2 \tag{26}$$

Proof. We expand the definition of Σ_{G_0} and rewrite:

$$\begin{aligned}\Sigma_{G_0} &= \frac{1}{|G_0|} \sum_{i\in G_0} (X_i - \mu_{G_0})(X_i - \mu_{G_0})^\mathsf{T} \\ &= \underbrace{\frac{1}{|G_0|} \sum_{i\in G_k} (X_i - \mu_{G_0})(X_i - \mu_{G_0})^\mathsf{T}}_{T_1} + \underbrace{\frac{1}{|G_0|} \sum_{i\in G_0\setminus G_k} (X_i - \mu_{G_0})(X_i - \mu_{G_0})^\mathsf{T}}_{T_2}\end{aligned}$$

We now rewrite the first term in the above sum using

$$\begin{aligned}T_1 &= \frac{1}{|G_0|} \sum_{i\in G_k} (X_i - \mu_{G_k} + \mu_{G_k} - \mu_{G_0})(X_i - \mu_{G_k} + \mu_{G_k} - \mu_{G_0})^\mathsf{T} \\ &= \frac{|G_k|}{|G_0|}\Sigma_{G_k} + \frac{|G_k|}{|G_0|}\left(\frac{1}{|G_k|}\sum_{i\in G_k}(X_i - \mu_{G_k})\right)(\mu_{G_k} - \mu_{G_0})^\mathsf{T} \\ &\quad + \frac{|G_k|}{|G_0|}(\mu_{G_k} - \mu_{G_0})\left(\frac{1}{|G_k|}\sum_{i\in G_k} X_i - \mu_{G_k}\right)^\mathsf{T} + \frac{|G_k|}{|G_0|}(\mu_{G_k} - \mu_{G_0})(\mu_{G_k} - \mu_{G_0})^\mathsf{T} \\ &= \frac{|G_k|}{|G_0|}\left(\Sigma_{G_k} + (\mu_{G_k} - \mu_{G_0})(\mu_{G_k} - \mu_{G_0})^\mathsf{T}\right)\end{aligned}$$

Letting $v \in \mathbb{S}^{d-1}$ and using the fact that T_2 is positive semidefinite, we arrive at

$$\left\langle v, \Sigma_{G_0} v \right\rangle = \frac{|G_k|}{|G_0|}\left(\left\langle v, \Sigma_{G_k} v \right\rangle + \left\|(\mu_{G_k} - \mu_{G_0})^\mathsf{T} v\right\|^2\right) + \left\langle v, T_2 v \right\rangle \geq \frac{|G_k|}{|G_0|} \left\langle v, \Sigma_{G_k} v \right\rangle \tag{27}$$

Finally, taking suprema over both sides yields the desired inequality. □

Lemma B.7. *Suppose that* (10) *is true and that the following inequality holds for some index $k \le T_p$:*

$$\sum_{i \in G_k} \tau_i \ge \frac{1}{\gamma} \sum_{j \in S_k} \tau_j. \tag{28}$$

Then the empirical means satisfy

$$\left\| \mathbb{E}_{\mathrm{Unif}(G_0)}[X] - \mathbb{E}_{\mathrm{Unif}(S_k)}[X] \right\|_2 \le 18 \left(5\alpha + \frac{20 \log(1/p)}{m} \right)^{1/2} \left\| \Sigma_{G_0} \right\|_2^{1/2}.$$

Proof. Let $P_1 := \mathrm{Unif}(G_0)$ and $P_2 := \mathrm{Unif}(S_k)$. From Lemma B.3, it follows that

$$\left\| \mathbb{E}_{P_1}[X] - \mathbb{E}_{P_2}[X] \right\|_2 \le \frac{\sqrt{d_{\mathrm{TV}}(P_1, P_2)}}{1 - \sqrt{d_{\mathrm{TV}}(P_1, P_2)}} \cdot \left(\left\| \Sigma_{G_0} \right\|_2^{1/2} + \left\| \Sigma_{G_k} \right\|_2^{1/2} \right). \tag{29}$$

Since (28) is the reverse of (21), we obtain

$$\begin{aligned} \left\| \Sigma_{S_k} \right\|_2 &\le (1-\eta)^2 \frac{\gamma}{1 - \gamma\eta} \left\| \Sigma_{G_k} \right\|_2 \\ &\le \frac{3}{1 - 6\alpha} \left\| \Sigma_{G_k} \right\|_2 \\ &\le 6 \cdot \frac{|G_0|}{|G_k|} \left\| \Sigma_{G_0} \right\|_2, \end{aligned}$$

where the first inequality follows from the contrapositive of Lemma B.4, the second inequality from $\gamma = 3$ and Lemma B.5, and the last inequality follows by our assumption on α. Now, let $K \le T_p$ be the number of samples in G_0 that were removed by the algorithm. We have

$$\frac{|G_0|}{|G_k|} = \frac{m - |B_0|}{m - |B_0| - K} \le \frac{m - |B_0|}{m - |B_0| - T_p} \le \frac{m - |B_0|}{m - 18\log(1/p) - 4|B_0|} = \frac{1 - \alpha}{1 - 4\alpha - \frac{18 \log(1/p)}{m}}$$

From (10), we additionally have that

$$1 - \left(4\alpha + \frac{18\log(1/p)}{m}\right) \ge 1 - \frac{4}{3}\left(3\alpha + \frac{18\log(1/p)}{m}\right) \ge \frac{1}{3} \implies \left\| \Sigma_{S_k} \right\|_2 \le 18 \left\| \Sigma_{G_0} \right\|_2.$$

Substituting the above into (29) and using Lemma B.8 yields the desired bound:

$$\begin{aligned} \left\| \mathbb{E}_{P_1}[X] - \mathbb{E}_{P_2}[X] \right\|_2 &\le \frac{\left(5\alpha + \frac{20\log(1/p)}{m}\right)^{1/2}}{1 - \left(5\alpha + \frac{20\log(1/p)}{m}\right)^{1/2}} \left(\left\| \Sigma_{G_0} \right\|_2^{1/2} + \sqrt{18} \left\| \Sigma_{G_0} \right\|_2^{1/2} \right) \\ &\le 18 \left(5\alpha + \frac{20\log(1/p)}{m} \right)^{1/2} \left\| \Sigma_{G_0} \right\|_2^{1/2}. \end{aligned}$$

□

Lemma B.8. *Suppose $k \le T_p$ and* (10) *holds. Then we have that*

$$d_{\mathrm{TV}}(\mathrm{Unif}(S_k), \mathrm{Unif}(G_0)) \le 5\alpha + \frac{20\log(1/p)}{m}. \tag{30}$$

Proof. We let $P_1 := \mathrm{Unif}(S_k)$, $P_2 := \mathrm{Unif}(G_0)$ and $P_3 := \mathrm{Unif}(G_k)$, and write $K \le k \le T_p$ for the number of samples originally in G_0 that were removed by the `Filter` algorithm by the k^{th} step. From the triangle inequality, it follows that

$$\begin{aligned} d_{\mathrm{TV}}(P_1, P_2) &\le d_{\mathrm{TV}}(P_1, P_3) + d_{\mathrm{TV}}(P_2, P_3) \\ &= \frac{|S_k| - |G_k|}{|S_k|} + \frac{|G_0| - |G_k|}{|G_0|} \\ &= \frac{m - k - (m - |B_0| - K)}{m - k} + \frac{K}{m - |B_0|} \end{aligned}$$

$$
\begin{aligned}
&= \frac{|B_0| + (K-k)}{m-k} + \frac{K}{m-|B_0|} \\
&\leq \frac{|B_0|}{m-k} + \frac{T_p}{m-|B_0|} \\
&\leq \frac{|B_0|}{m-T_p} + \frac{T_p}{m-|B_0|},
\end{aligned}
$$

where the second line follows from Lemma B.9 and the last two inequalities follow from $K \leq m$ and $m \leq T_p$. Finally, using Lemma B.2 and Eq. (10), we obtain

$$
\begin{aligned}
\frac{|B_0|}{m-T_p} + \frac{T_p}{m-|B_0|} &= \frac{\alpha}{1-\frac{T_p}{m}} + \frac{\frac{T_p}{m}}{1-\alpha} \\
&\leq \frac{\alpha}{1-\frac{18\log(1/p)}{m} - 3\alpha} + \frac{\frac{18\log(1/p)}{m} + 3\alpha}{1-\alpha} \\
&\leq \frac{4\alpha}{3} + \frac{\frac{18\log(1/p)}{m} + 3\alpha}{1-\alpha} \\
&\leq \frac{4\alpha}{3} + \frac{\frac{18\log(1/p)}{m} + 3\alpha}{1-\frac{1}{12}} \\
&\leq 5\alpha + \frac{20\log(1/p)}{m}.
\end{aligned}
$$

□

Lemma B.9. *Consider a pair of discrete sets S, S' such that $S' \subset S$. We have:*

$$
d_{\mathrm{TV}}(\mathrm{Unif}(S), \mathrm{Unif}(S')) = \frac{|S| - |S'|}{|S|}. \tag{31}
$$

Proof. Using the fact that $d_{\mathrm{TV}}(p,q) = \frac{1}{2}\|p-q\|_1$, we have:

$$
\begin{aligned}
d_{\mathrm{TV}}(\mathrm{Unif}(S), \mathrm{Unif}(S')) &= \frac{1}{2}\left(\sum_{x \in S \cap S'} \left|\frac{1}{|S|} - \frac{1}{|S'|}\right| + \sum_{x \in S \setminus S'} \frac{1}{|S|}\right) \\
&= \frac{1}{2}\left(1 - \frac{|S'|}{|S|} + \frac{|S| - |S'|}{|S|}\right) \\
&= 1 - \frac{|S'|}{|S|}.
\end{aligned}
$$

□

B.5 Proof of Theorem 4

We now present the proof of the main theorem on distributed PCA. We first recall that

$$
A_i = \frac{1}{n}\sum_{j=1}^{n} X_j^{(i)}(X_j^{(i)})^{\mathsf{T}}; \quad i \in \mathcal{I}_{\mathsf{good}},
$$

where $X_j^{(i)} \overset{\text{iid}}{\sim} \mathcal{P}$, and that the responses $V_i \in \mathbb{O}_{d,r}$ span the leading r-dimensional eigenspace of A_i. Under this model, the local errors $E_i := A_i - A$ as well as the error of the empirical average over the inliers are bounded with high probability. We will condition on the following events for the remainder of this section:

$$
\begin{aligned}
\mathcal{E}_1 &= \left\{ \max_{i \in \mathcal{I}_{\mathsf{good}}} \|A_i - A\|_2 \leq \min\left\{\frac{\delta}{8}, C_1\|A\|_2\sqrt{\frac{r_\star + \log(m/p)}{n}}\right\}\right\}, \\
\mathcal{E}_2 &= \left\{ \left\|\frac{1}{|\mathcal{I}_{\mathsf{good}}|}\sum_{i \in \mathcal{I}_{\mathsf{good}}} A_i - A\right\|_2 \leq C_2\|A\|_2\sqrt{\frac{r_\star + \log(n)}{|\mathcal{I}_{\mathsf{good}}|\, n}}\right\}.
\end{aligned} \tag{32}
$$

Lemma B.10. *Suppose that $n \gtrsim \kappa^2 \cdot (r_\star + \log(mn/p))$. Then the following hold:*

$$\mathbb{P}\left(\mathcal{E}_1\right) \leq p, \quad \mathbb{P}\left(\mathcal{E}_2\right) \leq \frac{2}{n}. \tag{33}$$

Proof. The bound on $\mathbb{P}\left(\mathcal{E}_2\right)$ in Eq. (33) follows from an application of [8, Exercise 9.2.5] and the assumed lower bound on n. On the other hand, the same result yields

$$\mathbb{P}\left(\|A_i - A\|_2 \geq C_1 \|A\|_2 \left(\sqrt{\frac{r_\star + \log(m/p)}{n}} + \frac{r_\star + \log(m/p)}{n}\right)\right) \leq \frac{p}{m},$$

for any fixed $i \in \mathcal{I}_{\mathsf{good}}$. From the lower bound on n, it follows that

$$\frac{r_\star + \log(m/p)}{n} \leq \sqrt{\frac{r_\star + \log(m/p)}{n}} \quad \text{and} \quad C_1 \|A\|_2 \sqrt{\frac{r_\star + \log(m/p)}{n}} < \frac{\delta}{8}.$$

Finally, taking a union bound over $\mathcal{I}_{\mathsf{good}}$ recovers the bound on $\mathbb{P}\left(\mathcal{E}_1\right)$. $\square$

An immediate corollary is a bound on the error of `RobustReferenceEstimator`.

Corollary B.1. *There is a universal constant C_{ref} such that the output of Alg. 2 satisfies*

$$\mathrm{dist}(\widehat{V}_{\mathsf{ref}}, V) \leq C_{\mathsf{ref}} \kappa \cdot \sqrt{\frac{r_\star + \log(m/p)}{n}}.$$

Proof. From the bound $\alpha < \frac{1}{2}$ and the conditioning on $\mathcal{E}_1$, we deduce the existence of an index set S' such that $|S'| > \frac{m}{2}$, and

$$\mathrm{dist}(\widehat{V}_i, V) \leq \frac{\|A_i - A\|_2}{\delta - \frac{\delta}{4}} \leq \frac{2C_1 \|A\|_2}{\delta} \sqrt{\frac{r_\star + \log(m/p)}{n}}, \quad \text{for all } i \in S',$$

where the first bound on $\mathrm{dist}(\widehat{V}_i, V)$ follows from the Davis-Kahan theorem [2, Theorem 2.7] and the fact that $\|A_i - A\|_2 \leq \frac{\delta}{8}$ for any $i \notin \mathcal{I}_{\mathrm{bad}}$. From Proposition 1 in the main text, it follows that

$$\mathrm{dist}(\widehat{V}_{\mathsf{ref}}, V) \leq \underbrace{6C_1}_{C_{\mathsf{ref}}} \frac{\|A\|_2}{\delta} \sqrt{\frac{r_\star + \log(m/p)}{n}}.$$

$\square$

The next Proposition instantiates the bounds of Lemma B.1 for for the case of distributed PCA.

Proposition B.1. *In the setting of Lemma B.1, the matrix $\Sigma_{\mathcal{I}_{\mathsf{good}}}$ satisfies*

$$\|\Sigma_{\mathcal{I}_{\mathsf{good}}}\|_2 \lesssim \kappa \sqrt{\frac{r(r_\star + \log(n))}{(1-\alpha)mn}} + \kappa^2 \cdot \frac{\sqrt{r}(r_\star + \log(n))}{n} + \kappa^4 \cdot \frac{r_\star + \log(m/p)}{n}. \tag{34}$$

Proof. From Lemma B.1, it follows that

$$\|\Sigma_{\mathcal{I}_{\mathsf{good}}}\|_2 \leq \left\| \frac{1}{|\mathcal{I}_{\mathsf{good}}|} \sum_{i \in \mathcal{I}_{\mathsf{good}}} V_i V_i^{\mathsf{T}} - V V^{\mathsf{T}} \right\|_2 + 2 \cdot \left\| \frac{1}{|\mathcal{I}_{\mathsf{good}}|} \sum_{i \in \mathcal{I}_{\mathsf{good}}} \widetilde{V}_i - V \right\|_2$$

From Proposition 2 in the main text and conditioning $\mathcal{E}_1$ and $\mathcal{E}_2$, we have

$$\begin{aligned} \left\| \frac{1}{|\mathcal{I}_{\mathsf{good}}|} \sum_{i \in \mathcal{I}_{\mathsf{good}}} \widetilde{V}_i - V \right\|_2 \lesssim \frac{1}{\delta} \left\| \frac{1}{|\mathcal{I}_{\mathsf{good}}|} \sum_{i \in \mathcal{I}_{\mathsf{good}}} A_i - A \right\|_2 & \\ + \left(\frac{\|A\|_2}{\delta}\right)^2 \max\left(C_1^2, C_{\mathsf{ref}}^2 \left(\frac{\|A\|_2}{\delta}\right)^2\right) \cdot \frac{r_\star + \log(m/p)}{n} & \\ \leq C_2 \kappa \sqrt{\frac{r_\star + \log(n)}{(1-\alpha)mn}} + \kappa^2 \max\left(C_1^2, C_{\mathsf{ref}}^2 \kappa^2\right) \cdot \frac{r_\star + \log(m/p)}{n} & \end{aligned}$$

$$\lesssim \kappa \sqrt{\frac{r_\star + \log(n)}{(1-\alpha)mn}} + \kappa^4 \cdot \frac{r_\star + \log(m/p)}{n}. \tag{35}$$

On the other hand, using [4, Theorem 2], we have that

$$\left\| \frac{1}{\mathcal{I}_{\mathsf{good}}} \sum_{i \in \mathcal{I}_{\mathsf{good}}} V_i V_i^\mathsf{T} - V V^\mathsf{T} \right\|_2 \lesssim \kappa \sqrt{\frac{r(r_\star + \log(n))}{(1-\alpha)mn}} + \kappa^2 \frac{\sqrt{r}(r_\star + \log(n))}{n}.$$

Putting all the bounds together yields (34). □

We now invoke Proposition 3 and recall that ϱ is defined as

$$\varrho := \kappa \sqrt{\frac{r(r_\star + \log(n))}{(1-\alpha)mn}} + \kappa^2 \cdot \frac{\sqrt{r}(r_\star + \log(n))}{n} + \kappa^4 \cdot \frac{r_\star + \log(m/p)}{n} \tag{36}$$

From that and Proposition B.1, it follows that Alg. 5 from the main text invoked with $\lambda_{\mathsf{lb}} = \omega := \sqrt{1/mn}$ and $\lambda_{\mathsf{ub}} = 6$ outputs an estimate satisfying

$$\begin{aligned} \left\| \bar{V} - \frac{1}{|\mathcal{I}_{\mathsf{good}}|} \sum_{i \in \mathcal{I}_{\mathsf{good}}} \widetilde{V}_i \right\|_2 &\lesssim \sqrt{\max\{\varrho, \omega\}} \cdot \left(\alpha + \frac{\log(1/p)}{m}\right)^{1/2} &(37)\\ &= \sqrt{\varrho} \cdot \left(\alpha + \frac{\log(1/p)}{m}\right)^{1/2} &(38) \end{aligned}$$

with failure probability at most $2\log_2(6/\omega)p$. Finally, from Eqs. (35) and (38) it follows that

$$\begin{aligned} \|\bar{V} - V\|_2 &\le \left\| \bar{V} - \frac{1}{|\mathcal{I}_{\mathsf{good}}|} \sum_{i \in \mathcal{I}_{\mathsf{good}}} \widetilde{V}_i \right\|_2 + \left\| \frac{1}{|\mathcal{I}_{\mathsf{good}}|} \sum_{i \in \mathcal{I}_{\mathsf{good}}} \widetilde{V}_i - V \right\|_2 \\ &\lesssim \sqrt{\varrho}\left(\alpha + \frac{\log(1/p)}{m}\right)^{1/2} + \kappa \sqrt{\frac{r(r_\star + \log(n))}{(1-\alpha)mn}} + \kappa^4 \frac{\sqrt{r}(r_\star + \log(m/p))}{n}. \end{aligned}$$

In particular, the success probability is at least (given that ω is set as $\sqrt{1/mn}$):

$$1 - p - \frac{2}{n} - 2\log_2\left(\frac{6}{\omega}\right) p \ge 1 - \frac{2}{n} - 2\log_2(6mn)p.$$

Letting $p := \frac{p'}{2\log_2(6mn)}$ and relabeling $p' \to p$ yields the result.

C Experiment details

The numerical experiments in the main text were coded in `Julia` and run on a machine with Intel(R) Core(TM) i7-7700 CPU, 16GB of RAM and a GNU/Linux environment. The code is attached as part of the supplementary material and consists of a library called `RobustDistributedPCA.jl` and a subfolder `scripts/` that contains the scripts reproducing the experiments in the manuscript. We refer the reader to the included `README.md` file for installation and usage instructions.

Notes on implementation. Our implementations of the `Filter` and `AdaptiveFilter` algorithms deviate from the theory in the following ways:

1. In `Filter`, we remove the point with the largest outlier score by default:

   $$Z := \operatorname*{argmax}_{j \in S} \tau_j.$$

   The randomized selection can be enabled by passing `randomized = true` to the appropriate function. Please see the attached `README.md` file for details.

2. The error proxy $f(\lambda; p, \alpha)$ we use in `AdaptiveFilter` has been simplified to

$$f(\lambda; p, \alpha) \equiv f(\lambda; \alpha) = \sqrt{\lambda\alpha}.$$

The reason is twofold: on one hand, we suspect that many of the constants involved in the original definition of f are artifacts of our proof and are much lower in practice; on the other hand, since the default behavior of `Filter` is not randomized, the term owing to failure probability in f can be removed.

3. Finally, we change the stopping condition of `AdaptiveFilter` to $\lambda \leq \lambda_{\mathsf{ub}}$ in order to simplify constants. As the value of λ_{ub} is determined adaptively, this does not affect the output of the algorithm.

References

[1] Vasileios Charisopoulos, Austin R. Benson, and Anil Damle. Communication-efficient distributed eigenspace estimation. *SIAM Journal on Mathematics of Data Science*, 3(4):1067–1092, 2021.

[2] Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, 14(5):566–806, 2021.

[3] Ilias Diakonikolas and Daniel M. Kane. Recent Advances in Algorithmic High-Dimensional Robust Statistics. *arXiv e-prints*, page arXiv:1911.05911, 2019, 1911.05911.

[4] Jianqing Fan, Dong Wang, Kaizheng Wang, and Ziwei Zhu. Distributed estimation of principal eigenspaces. *Ann. Statist.*, 47(6):3009–3031, 12 2019.

[5] Pravesh K. Kothari and David Steurer. Outlier-robust moment-estimation via sum-of-squares. *arXiv e-prints*, page arXiv:1711.11581, 2017, 1711.11581.

[6] Adarsh Prasad, Sivaraman Balakrishnan, and Pradeep Ravikumar. A Unified Approach to Robust Mean Estimation. *arXiv e-prints*, page arXiv:1907.00927, 2019, 1907.00927.

[7] G. W. Stewart. Smooth local bases for perturbed eigenspaces. Technical Report TR-5010, University of Maryland, Institute for Advanced Computer Studies, 2012.

[8] Roman Vershynin. *High-Dimensional Probability: An introduction with applications in data science*, volume 47 of *Cambridge Series in Statistical & Probabilistic Mathematics*. Cambridge University Press, 2018.