Communication-efficient distributed eigenspace estimation with arbitrary node failures

Vasilis Charisopoulos (Cornell ORIE)

Joint work with Anil Damle (Cornell CS)

NeurIPS 2022

Problem setting

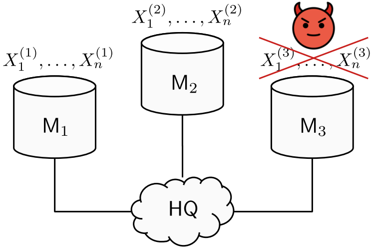
Motivating example: principal component analysis (PCA).

- Given samples $X_1, \ldots, X_n \in \mathbb{R}^d$, reduce dimension to $r \ll d$.
- Solution: top-r eigenspace of *empirical covariance* matrix:

$$\Sigma_n := \frac{1}{n} \sum_{j=1}^{n} X_j X_j^{\mathsf{T}} = V \Lambda V^{\mathsf{T}} + V_{\perp} \Lambda_{\perp} V_{\perp}^{\mathsf{T}}, \qquad V \in O(d, r).$$

Challenges and desiderata:

1. What if data are **distributed**? (*communication-efficiency*)
2. What if some machines are **compromised**? (*robustness*)



Problem setting

General setting and assumptions:

- **Unknown** symmetric matrix $A \in \mathbb{R}^{d \times d}$ with decomposition

$$A = V \Lambda V^{\mathsf{T}} + V_{\perp} \Lambda_{\perp} V_{\perp}^{\mathsf{T}}, \quad V \in O(d, r).$$

- **Eigengap**: we have $\delta_r := \lambda_r(A) - \lambda_{r+1}(A) > 0$.
- **Local errors**: machine i observes symmetric $A^{(i)} \in \mathbb{R}^{d \times d}$ such that

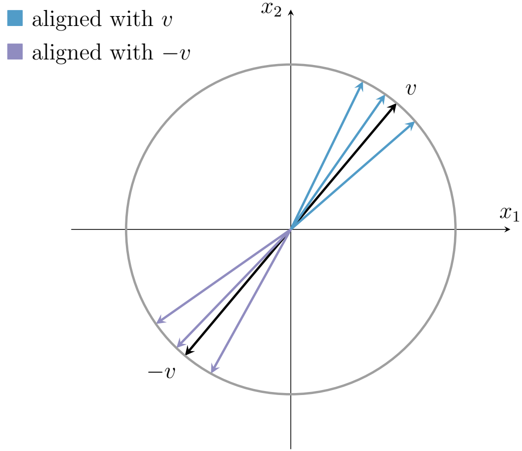
$$\left\| A^{(i)} - A \right\|_2 \leq \frac{\delta_r}{8}, \quad i = 1, \dots, m, \quad (m := \text{number of machines.})$$

Quality of approximation measured in ℓ_2-*subspace distance*:

$$\mathrm{dist}_2(V, U) := \left\| (I - VV^{\mathsf{T}})U \right\|_2 = \left\| (I - UU^{\mathsf{T}})V \right\|_2.$$

Challenge: averaging local solutions with symmetries

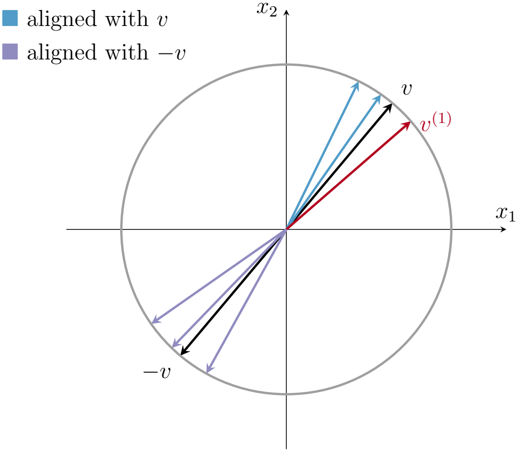
Problem (for $r = 1$): $v^{(i)}$ only defined up to sign.



Question: can we fix the sign ambiguity?

Challenge: averaging local solutions with symmetries

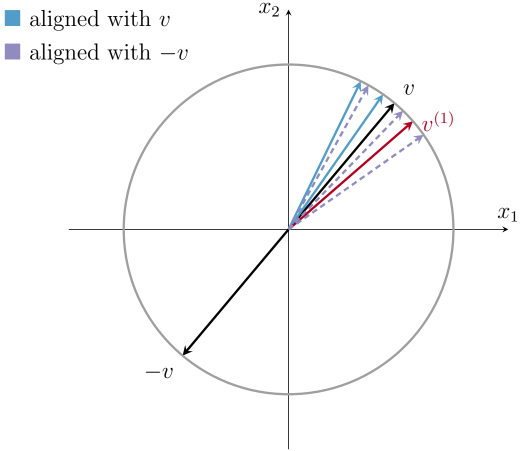
Problem (for $r = 1$): $v^{(i)}$ only defined up to sign.



Algorithm: $v^{(1)}$

Challenge: averaging local solutions with symmetries

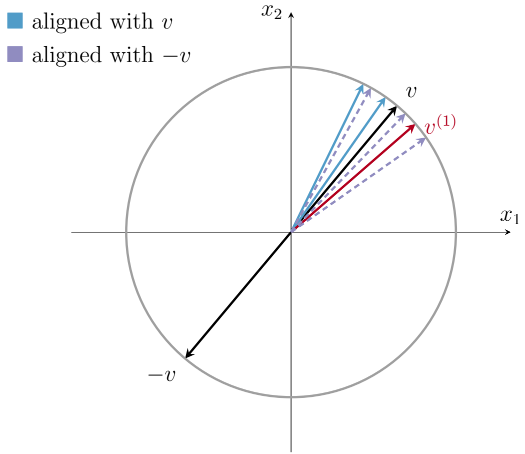
Problem (for $r = 1$): $v^{(i)}$ only defined up to sign.



Algorithm: $\mathrm{sign}(\langle v^{(i)}, v^{(1)} \rangle) \cdot v^{(i)}$

Challenge: averaging local solutions with symmetries

Problem (for $r = 1$): $v^{(i)}$ only defined up to sign.



Algorithm: $\tilde{v} := \frac{1}{m} \sum_{i=1}^{m} \text{sign}(\langle v^{(i)}, v^{(1)} \rangle) \cdot v^{(i)}$

Proposed method

Algorithm (without node failures): average **aligned eigenvectors** of $A^{(i)}$.

Algorithm Procrustes Fixing

Input: local eigenvector matrices $V^{(i)} \in O(d, r)$, $i = 1, \dots, m$,
for $i = 1, \dots, m$ **do**
  $Z_i := \operatorname{argmin}_{U \in O(r)} \left\| V^{(i)} U - V^{(1)} \right\|_{\mathrm{F}}$ ▷ $V^{(1)}$ acts as "reference"
  $\tilde{V}^{(i)} := V^{(i)} Z_i$ ▷ Procrustes alignment
return $\frac{1}{m} \sum_{i=1}^{m} \tilde{V}^{(i)}$

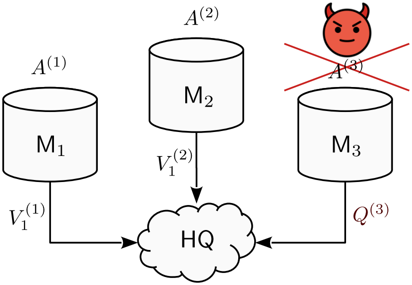
- Without corruptions, achieves approximation error [1]:

$$\operatorname{dist}_2 \left(\frac{1}{m} \sum_{i=1}^{m} \tilde{V}^{(i)}, V \right) \lesssim \frac{1}{m} \sum_{i=1}^{m} \left(\frac{\|A^{(i)} - A\|_2}{\delta} \right)^2 + \frac{1}{\delta} \left\| \frac{1}{m} \sum_{i=1}^{m} A^{(i)} - A \right\|_2$$

- **Key issue**: not robust to nodes that may respond adversarially.

[1] **C.**, Benson & Damle, 2021

Node failures



Corruption model: unknown index set $\mathcal{I}_{\mathsf{bad}} \subset [m]$ such that:

- $|\mathcal{I}_{\mathsf{bad}}| \leq \alpha m$, for $\alpha \in (0, 1/2)$.
- All nodes $i \in \mathcal{I}_{\mathsf{bad}}$ return *arbitrary*, but *structurally valid* $Q^{(i)} \in O(d, r)$.

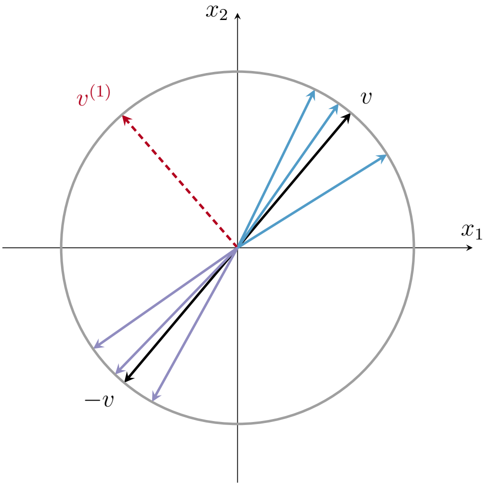
Sources of corruption:

- *Silent* / *soft* errors (e.g., insufficient eigensolver tolerance);
- Outliers / corrupted data (e.g., corrupted data source in some machines);
- Adversarial responses.

A robust algorithm

Strategy: "robustify" Procrustes-fixing algorithm.

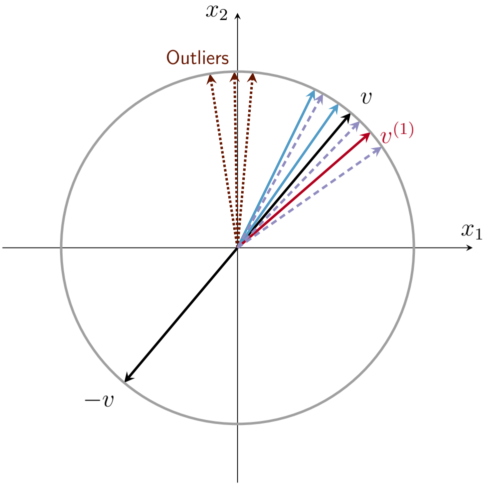
Challenge I: reference solution could be chosen among outliers.



A robust algorithm

Strategy: "robustify" Procrustes-fixing algorithm.

Challenge II: Even with "good" reference, we could average over outliers.



A robust algorithm

Strategy: "robustify" Procrustes-fixing algorithm.

Algorithm Robust Procrustes fixing

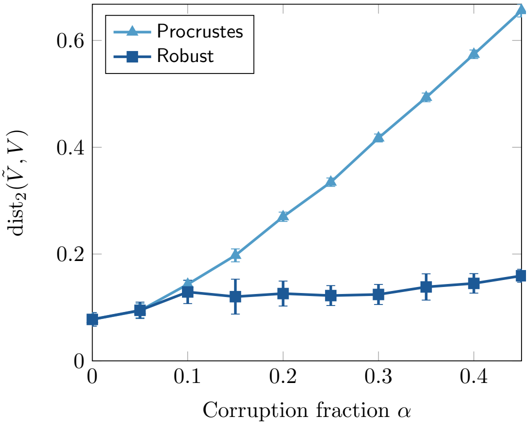
1: **Input**: responses $\left\{\widehat{V}^{(i)} \mid i \in [m]\right\}$, corruption fraction α.
2: $V_{\mathsf{ref}} := \texttt{RobustReferenceEstimator}(\left\{\widehat{V}^{(i)}\right\}_{i=1}^{m})$
3: $\left\{\widetilde{V}^{(i)}\right\}_{i=1}^{m} := \texttt{ProcrustesFixing}(\left\{\widehat{V}^{(i)}\right\}_{i=1}^{m}, V_{\mathsf{ref}})$
4: $\tilde{V} := \texttt{RobustMeanEstimation}(\left\{\widetilde{V}^{(i)}\right\}_{i=1}^{m}, \alpha)$.

Key differences:

1. Instead of picking $\widehat{V}^{(1)}$ as reference, choose it robustly.
2. Instead of averaging Procrustes-fixed responses, compute their robust mean.

Experiment

Setup: distributed PCA with $\lfloor \alpha m \rfloor$ responses replaced by a $V_{\mathsf{adv}} \in O(d, r)$.



✗ "Baseline" solution almost orthogonal to V as $\alpha \to 1/2$.

✓ Robust solution: natural breakdown point at $\alpha = 1/2$.

Thank you!

Full paper: `arXiv:2206.00127`