Thresholding Based Efficient Outlier Robust PCA

Yeshwanth Cherapanamjeri, Prateek Jain, Praneeth Netrapalli
{t-yecher,prajain,praneeth}@microsoft.com
Microsoft Research India

COLT 2017 Microsoft Research

Goal

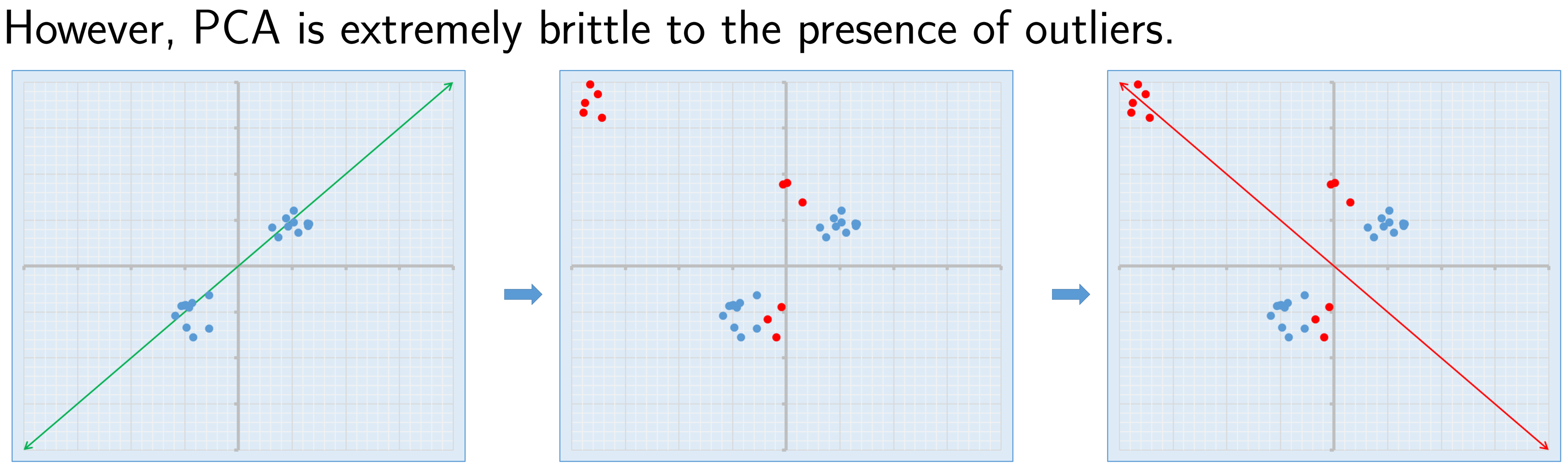
Design an efficient algorithm for Outlier Robust PCA

PCA and its Brittleness

Given data matrix $X^* = \{x_1, x_2, \ldots, x_n\}$, compute the rank-$r$ subspace that best descries the data:

$$U = \underset{U \in \mathbb{R}^{d \times r}}{\arg\min} \left\|(I - UU^\top)X^*\right\|_F$$

However, PCA is extremely brittle to the presence of outliers.



(a) No Outliers (b) Outlier Points (c) Standard PCA

Question

Given sparsely-corrupted data matrix $M^* = X^* + C^*$, recover $PCA(X^*)$.

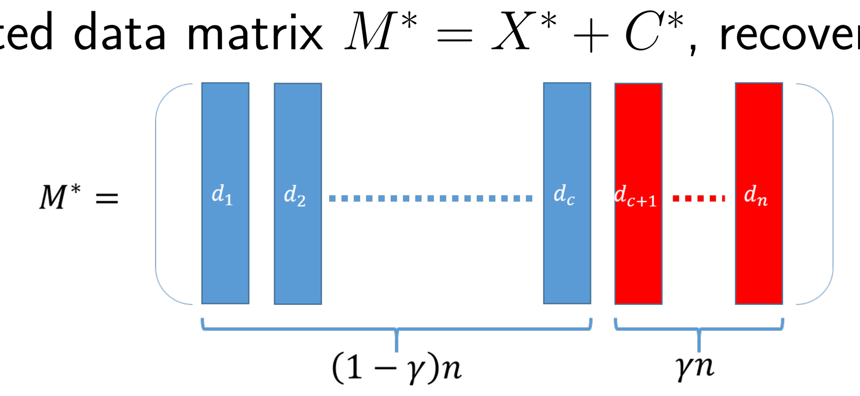


Figure: Data Points with Corruptions

Existing Work

Existing work suffer from high computational cost and weak recovery guarantees.

Therefore, is there an estimator which can:

1. Match the running time of Vanilla PCA and
2. Obtain strong recovery guarantees?

Algorithm - TORP

Algorithm 1 TORP

Input: M, r, ρ, T
Initialize $\mathcal{CS} \leftarrow \{\}$
for $t = 0$ to $t = T$ **do**
  $[U, \Sigma, V] \leftarrow \mathcal{SVD}_r(M_{\setminus \mathcal{CS}})$
  $e_i \leftarrow \left\|\Sigma^{-1} U^\top M_i\right\|$
  $r_i \leftarrow \left\|(I - UU^\top) M_i\right\|$
  $\mathcal{CS} \leftarrow \{\text{Top } \rho n\ e_i\} \cup \{\text{Top } \rho n\ r_i\}$
end for
$[U, \Sigma, V] \leftarrow \mathcal{SVD}_r(M_{\setminus \mathcal{CS}})$
Return: U

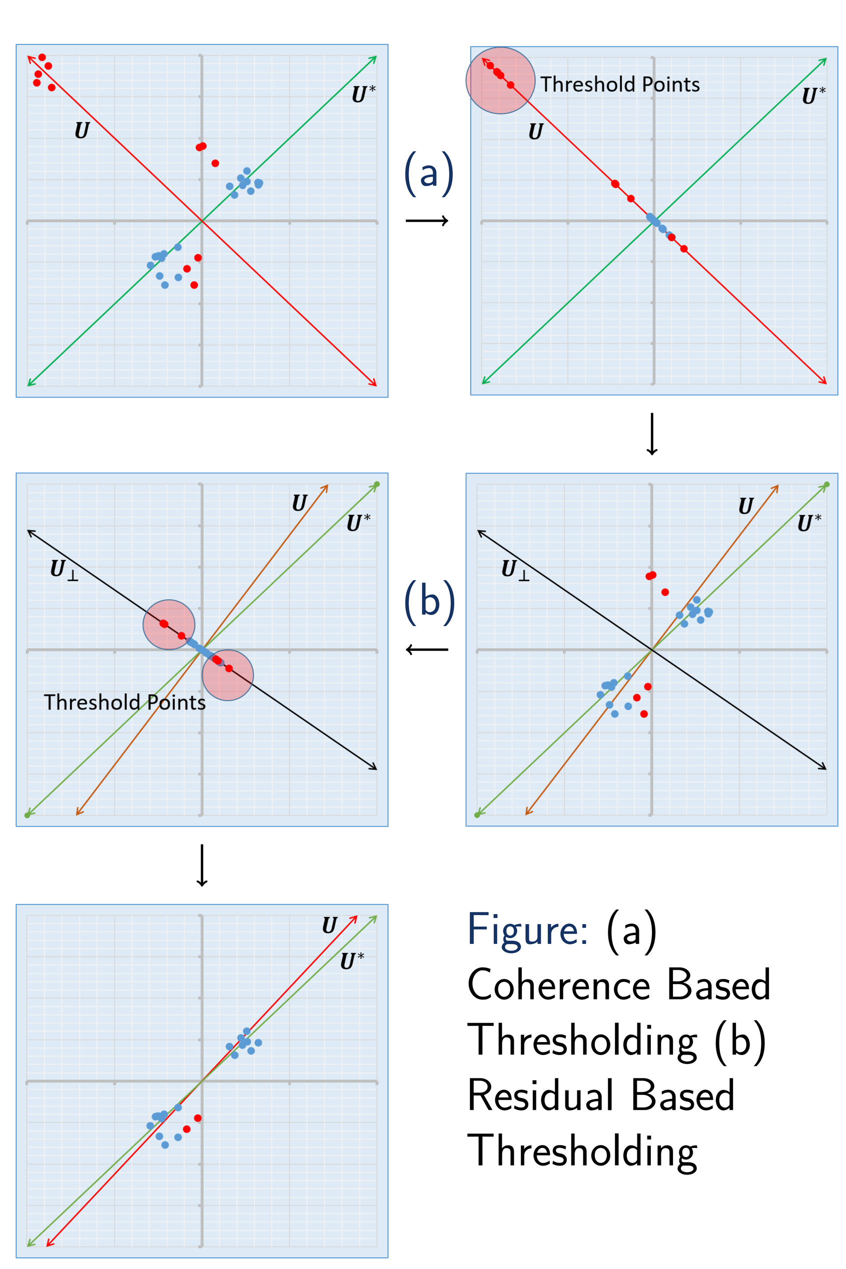


Figure: (a) Coherence Based Thresholding (b) Residual Based Thresholding

Three Settings

Recall that we are given $M^* = X^* + C^*$. Let $X^* = L^* + N^*$ where L^* is the rank-r approximation of X^*. We consider three settings based on our assumptions on N^*:

- Noiseless Setting: $N^* = 0$,
- General Noise Setting: N^* is arbitrary,
- Gaussian Noise: N^* is Gaussian.

We also assume to ensure uniqueness of the decomposition:

- L^* is μ-incoherent, i.e $[U^*, \Sigma^*, V^*] = \mathcal{SVD}(L^*)$ with $\left\|e_i^\top V^*\right\| \le \mu\sqrt{r}/\sqrt{n}$.
- Fraction of corruptions, γ satisfies $\gamma \le \mathcal{O}\left(\frac{1}{\mu^2 r}\right)$.

Noiseless Case

Theorem

TORP run with $\rho = \frac{1}{128\mu^2 r}$ and $T = \log\frac{20\|M^*\|n}{\epsilon}$, returns U satisfying:

$$\left\|\mathcal{P}_\perp^U(L^*)\right\|_F \le \epsilon$$

in at most $\mathcal{O}(ndr\log\frac{1}{\epsilon})$ computational steps.

- Exact recovery of the principal components.
- Time complexity almost matches that of Vanilla PCA.
- [Xu et al, 2010] match the recovery guarantee but have large runtime $\left(\mathcal{O}\left(\frac{n^2 d}{\epsilon^2}\right)\right)$ while [Xu et al, 2013] cannot obtain exact recovery guarantees.

General Noise Case

We, now, allow the noise matrix N^* to be arbitrary.

Theorem

TORP run with $\rho = \frac{1}{128\mu^2 r}$ and $T = \log\frac{20\|M^*\|n}{\epsilon}$, returns U satisfying:

$$\left\|\mathcal{P}_\perp^U(L^*)\right\|_F \le 60\sqrt{r}\left\|N^*\right\|_F + \epsilon$$

in at most $\mathcal{O}\left(ndr\log\frac{1}{\epsilon}\right)$ computational steps.

- Recovery guarantee optimal upto a factor of $\mathcal{O}(\sqrt{r})$.
- [Xu et al, 2010] obtain recovery upto $\mathcal{O}(\sqrt{n}\left\|N^*\right\|_F)$.

Gaussian Noise Case

Improved recovery guarantees can be obtained in special cases.
When each column $i \in [n]$ satisfies $N_i^* \sim \mathcal{N}(0, \sigma I_d)$, we obtain:

Theorem

TORP-G returns U satisfying:

$$\left\|\mathcal{P}_\perp^U(L^*)\right\|_F \le 9\sqrt{r\log d}\left\|N^*\right\|_2 + \epsilon$$

w.h.p when $n = \Omega(d^2)$ in at most $\mathcal{O}(n^2 dr\log\frac{1}{\epsilon})$ computational steps.

- $\mathcal{O}(\sqrt{d})$ improvement over TORP and $\mathcal{O}(\sqrt{nd})$ over [Xu et al, 2010].
- Computational cost: $\mathcal{O}(n^2 dr)$, Sample complexity: $\Omega(d^2)$.
- Also holds when N_i^* obeys a Sub-Gaussian distribution with parameter σ.