RANSAC Revisited: An Improved Algorithm for Robust Subspace Recovery under Adversarial and Noisy Corruptions

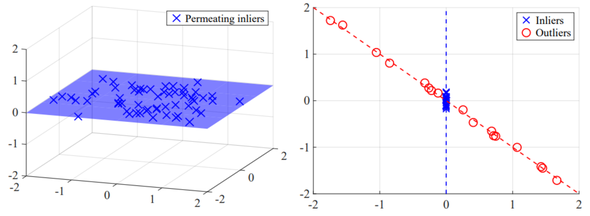
Guixian Chen, Jianhao Ma, Salar Fattahi

INDUSTRIAL & OPERATIONS ENGINEERING, UNIVERSITY OF MICHIGAN

MICHIGAN ENGINEERING, UNIVERSITY OF MICHIGAN

Introduction

Robust Subspace Recovery (RSR) Task:
Given a collection of n points in $\mathbb{R}^d$—some of which may be corrupted by adversarial contamination, and all of which are perturbed by small noise—how can we efficiently and provably identify a subspace $\mathcal{S}^\star$ of dimension $\dim(\mathcal{S}^\star) = r^\star \ll d$ that nearly contains most of the uncorrupted samples?



Adversarial and Noisy Contamination Model

Given a clean distribution $\mathcal{P}$, a corruption parameter ϵ, and a noise covariance Σ_ξ, (ϵ, Σ_ξ)-corrupted samples are generated as follows:

(i) Draw n samples independently from $\mathcal{P}$.

(ii) The adversary adds i.i.d. **noise** to each sample, drawn from a zero-mean **Gaussian** distribution with covariance Σ_ξ.

(iii) An arbitrarily powerful adversary **inspects** the samples, **removes** $\lfloor \epsilon n \rfloor$ of them, and **replaces** them with arbitrary points.

Applications

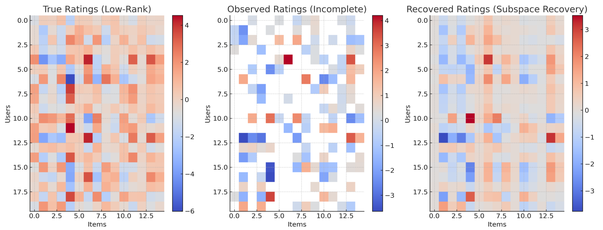


Figure 1: Recover missing values in recommender system by subspace recovery.

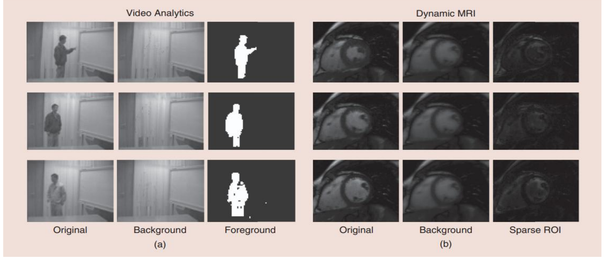


Figure 2: [Vaswani et al., 2018] Foreground-background separation tasks: (a) A video analytics application; (b) Accelerated dynamic MRI.

Comparison on A Toy Example

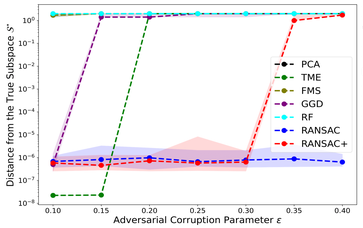


Figure 3:
TME: Tyler's M-estimator
FMS: Fast Median Subspace
GGD: Geodesic Gradient Descent
RF: Randomized-Find
RANSAC: Classic RANSAC
RANSAC+: Our Algorithm
(See References for details)

The setting of the toy example:

- The ambient dimension $d = 100$.
- The sample size $n = 500$, the clean samples are drawn from a Gaussian distribution $N(0, \Sigma^\star)$, where the covariance $\mathrm{rank}(\Sigma^\star) = 10$.
- An adversary randomly **replaces** ϵ-fraction of the clean samples with samples independently drawn from $N(0, \Sigma^{\text{outlier}})$, where $\mathrm{rank}(\Sigma^{\text{outlier}}) = 2$, and the eigenvectors corresponding to nonzero eigenvalues of Σ^{outlier} are **orthogonal** to those of $\Sigma^\star$.
- All nonzero eigenvalues of $\Sigma^\star$ are set to 1, and all nonzero eigenvalues of Σ^{outlier} are all set to 10.
- Only adversarial corruption but no Gaussian noise in this example.

Figure 4: The search dimension r for RANSAC is overestimated by one, while RANSAC+ operates without any prior information of $r^\star$.

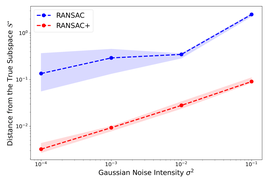
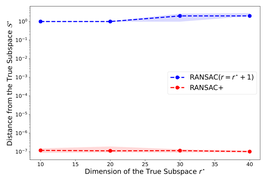


Figure 5: An additional Gaussian noise with mean zero and variance σ^2/d is added to all the samples.



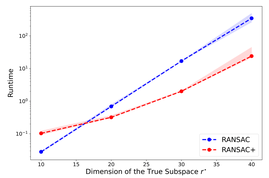


Figure 6: Runtime of RANSAC and RANSAC+ for different subspace dimensions $r^\star$.

RANSAC Algorithm

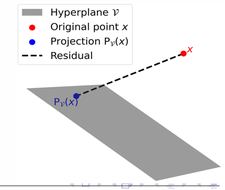
Algorithm RANSAC [Fischler and Bolles, 1981]

Require: The adversarial corrupted sample set $\mathcal{X} = \{x_1, \ldots, x_n\}$, and $r^\star$.
$t \leftarrow 0$, $T \leftarrow \mathcal{O}(e^{r^\star})$, $\eta_{\text{thresh}} \leftarrow 10^{-4}$, and $N_{\text{consensus}} = n/2$.
while $t < T$ **do**
  Randomly draws samples until they span a $r^\star$-dimensional subspace. Denote the collection of samples by $\mathcal{X}_B$ and the subspace by $\mathcal{V}_B$.
  Calculate **Consensus**$(\mathcal{V}_B) \leftarrow \#\{x \in \mathcal{X} : \|x - \mathrm{P}_{\mathcal{V}_B}(x)\| \le \eta_{\text{thresh}}\}$.
  if Consensus$(\mathcal{V}_B) \ge N_{\text{consensus}}$ **then**
    break
  end if
end while
return $\mathcal{V}_B$.

First Stage Algorithm

Algorithm First Stage: Coarse-grained Estimation

Require: The (ϵ, Σ_ξ)-corrupted sample set $\mathcal{X} = \{x_1, \ldots, x_n\}$, $\mathrm{tr}(\Sigma_\xi)$, $\|\Sigma_\xi\|$.
$B \leftarrow 2$ and $\eta_{\text{thresh}} \leftarrow \mathcal{O}\left(\sqrt{r^\star \|\Sigma_\xi\|} + \sqrt{\mathrm{tr}(\Sigma_\xi)}\right)$.
while $B < d$ **do**
  Randomly select B samples from $\mathcal{X}$ and form a matrix $X_B \in \mathbb{R}^{d \times B}$.
  Compute the basis $V \in \mathbb{R}^{d \times B}$ for **col**(X_B).
  MedRes$(V) \leftarrow \mathrm{median}\left(\left\{\left\|x - VV^\top x\right\| : x \in \mathcal{X} \backslash \mathcal{X}_B\right\}\right)$.
  if MedRes$(V) \le \eta_{\text{thresh}}$ **then**
    break
  else
    $B \leftarrow 2B$.
  end if
end while
Set $\hat{r} = B$, and subspace $\mathcal{V} =$ **col**(X_B).
return $V \in O(d, \hat{r})$ and $\mathcal{V}$.



Theorem 1

Given an (ϵ, Σ_ξ)-corrupted sample set with a sample size $n = \Omega(r^\star \log(r^\star))$ and Gaussian noise covariance satisfying $\|\Sigma_\xi\| = \mathcal{O}(\gamma^\star_{\min})$ (here $\gamma^\star_{\min}$ is the smallest eigenvalue of $\Sigma^\star$), First Stage Algorithm runs in $\mathcal{O}(ndr^\star \log(r^\star))$ time, and with an overwhelming probability, generates a subspace $\mathcal{V}$ with dimension $\hat{r} = \mathcal{O}(r^\star)$ such that, for each $x \sim \mathcal{P}$:

$$\|(I_d - \mathrm{P}_{\mathcal{V}})\, x\| \le \mathcal{O}\left(\left(\sqrt{\frac{r^\star \|\Sigma_\xi\|}{\gamma^\star_{\min}}} + \sqrt{\frac{\mathrm{tr}(\Sigma_\xi)}{\gamma^\star_{\min}}}\right) \|x\|\right).$$

In the noiseless setting ($\Sigma_\xi = 0$), we have $\mathcal{V} \supseteq \mathcal{S}^\star$ with high probability.

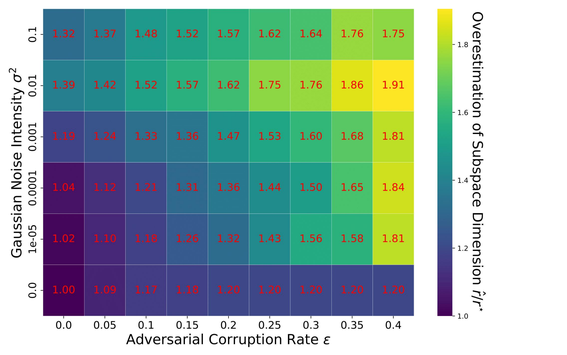


Figure 7: Overestimation of the subspace dimension, measured by $\hat{r}/r^\star$, for First Stage Algorithm under varying corruption rate ϵ and Gaussian noise levels σ^2.

References

- Martin Fischler and Robert Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 1981.
- Moritz Hardt and Ankur Moitra. Algorithms and hardness for robust subspace recovery. In Conference on Learning Theory, PMLR, 2013.
- Gilad Lerman and Tyler Maunu. Fast, robust and non-convex subspace recovery. Information and Inference: A Journal of the IMA, 2018.

Second Stage Algorithm

Algorithm Second Stage: Fine-grained Estimation

Require: Failure probability δ, corruption rate ϵ, projected sample set $\widehat{\mathcal{X}} = \left\{V^\top x_1, \ldots, V^\top x_n\right\}$, basis matrix $V \in O(d, \hat{r})$ produced by First Stage.
$B \leftarrow C' \cdot \max\left\{\hat{r}, \log\left(\frac{3}{\delta}\log\left(\frac{1}{\delta}\right)\right)\right\}$ for sufficiently large constant $C' > 0$.
$T \leftarrow \left(\frac{1}{1-1.1\epsilon}\right)^B \log\left(\frac{1}{\delta}\right)$.
while $j < T$ **do**
  Randomly select B samples from $\widehat{\mathcal{X}}$ and form a matrix $\widehat{X}_j \in \mathbb{R}^{\hat{r} \times B}$.
  Compute the singular values of $\widehat{X}_j$ as $\hat{\sigma}_1^{(j)} \ge \hat{\sigma}_2^{(j)} \ge \cdots \ge \hat{\sigma}_{\hat{r}}^{(j)}$.
end while
Compute $\hat{\sigma}_i := \min_{j \in [T]} \hat{\sigma}_i^{(j)}$, for $i \in [\hat{r}]$.
Find the smallest index $\tilde{r} \in [\hat{r}]$ such that $\hat{\sigma}^2_{\tilde{r}+1} \le C' \|\Sigma_\xi\|$ for $C' > 0$.
Find $k = \mathrm{argmin}_{j \in [T]} \hat{\sigma}^{(j)}_{\tilde{r}+1}$.
Compute the top-$\tilde{r}$ left singular vectors of $\widehat{X}_k \in \mathbb{R}^{\hat{r} \times B}$ as $\widehat{U}_k \in \mathbb{R}^{\hat{r} \times \tilde{r}}$.
return $V\widehat{U}_k \in O(d, \tilde{r})$ and $\widehat{\mathcal{S}} =$ **col**$\left(V\widehat{U}_k\right)$.

Theorem 2

Suppose that the conditions of Theorem 1 are satisfied and First Stage Algorithm succeeds. Given $\widehat{\mathcal{X}} = \left\{V^\top x_1, \ldots, V^\top x_n\right\} \subset \mathbb{R}^{\hat{r}}$, Second Stage Algorithm runs in $\mathcal{O}(r^{\star 3} e^{r^\star})$ time and, with probability $1 - \delta$, generates a subspace $\widehat{\mathcal{S}}$ with dimension $r^\star$ that satisfies

$$\left\|\mathrm{P}_{\widehat{\mathcal{S}}} - \mathrm{P}_{\mathcal{S}^\star}\right\| \le \mathcal{O}\left(\sqrt{\frac{r^\star \|\Sigma_\xi\|}{\gamma^\star_{\min}}} + \sqrt{\frac{\mathrm{tr}(\Sigma_\xi)}{\gamma^\star_{\min}}}\right).$$

In the noiseless setting ($\Sigma_\xi = 0$), we have $\widehat{\mathcal{S}} = \mathcal{S}^\star$ with high probability.

Spectrum	X_1	X_2	X_3	$\cdots$	X_{T-1}	X_T	$\star$
σ_1	10.0	10.1	9.9	$\cdots$	100.0	10.2	9.9
σ_2	9.2	9.3	9.5	$\cdots$	9.3	9.4	9.2
σ_3	8.3	8.4	8.6	$\cdots$	8.1	8.9	8.1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\sigma_{r^\star}$	0.9	1.1	1.0	$\vdots$	1.3	1.1	0.9
$\sigma_{r^\star+1}$	0.7	0.0	0.6	$\vdots$	0.9	0.8	0.0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\sigma_{\hat{r}}$	0.0	0.0	0.0	$\cdots$	0.0	0.0	0.0

Table 1: A concrete example for the uniform lower spectrum across all the batches, for Second Stage Algorithm in the noiseless case ($\Sigma_\xi = 0$) .

References

- Gilad Lerman and Tyler Maunu. An overview of robust subspace recovery. Proceedings of the IEEE, 2018.
- Tyler Maunu and Gilad Lerman. Robust subspace recovery with adversarial outliers. arXiv preprint, 2019.
- Tyler Maunu, Teng Zhang, Gilad Lerman. A well-tempered landscape for non-convex robust subspace recovery. Journal of Machine Learning Research, 2019.
- Namrata Vaswani, Thierry Bouwmans, Sajid Javed, Praneeth Narayanamurthy. Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery. IEEE signal processing magazine, 2018.
- Teng Zhang. Robust subspace recovery by Tyler's M-estimator. Information and Inference: A Journal of the IMA, 2016.