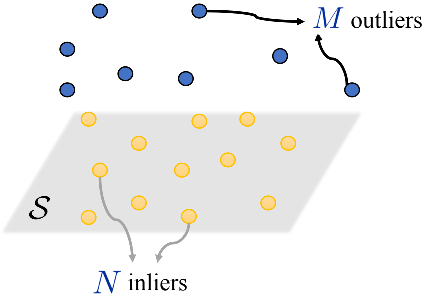
Noisy Dual Principal Component Pursuit

Tianyu Ding, Zhihui Zhu, Tianjiao Ding, Yunchen Yang,
René Vidal, Manolis C. Tsakiris, Daniel P. Robinson

ICML 2019

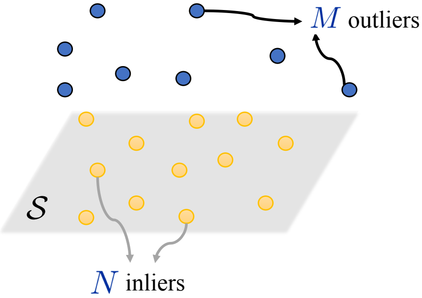
Motivation

Problem: learning a subspace from corrupted data



Motivation

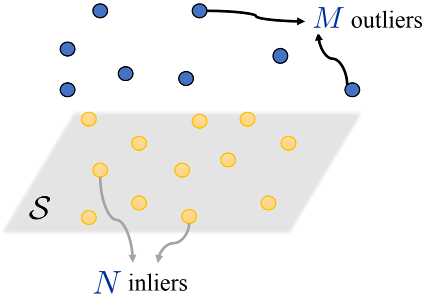
Problem: learning a subspace from corrupted data



- *Low relative dimension* (d/D)
  - Outlier Pursuit [Xu et al. 10]
    $M = O(N)$ outliers

Motivation

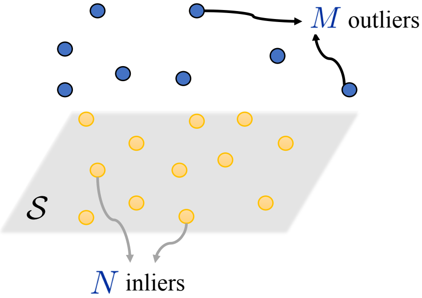
Problem: learning a subspace from corrupted data



- *Low relative dimension* (d/D)
  - Outlier Pursuit [Xu et al. 10]
    $M = O(N)$ outliers
- *High relative dimension* (d/D)
  - REAPER [Lerman and Maunu 18]
    $M = O(N)$ outliers
  - DPCP [Tsakiris 15, Zhu 18]
    $M = O(N^2)$ outliers

Motivation

Problem: learning a subspace from corrupted data

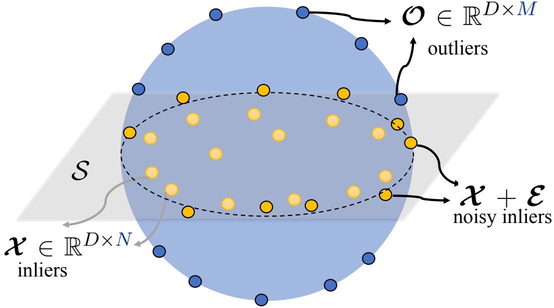


- *Low relative dimension* (d/D)
  - Outlier Pursuit [Xu et al. 10]
    $M = O(N)$ outliers
- *High relative dimension* (d/D)
  - REAPER [Lerman and Maunu 18]
    $M = O(N)$ outliers
  - DPCP [Tsakiris 15, Zhu 18]
    $M = O(N^2)$ outliers

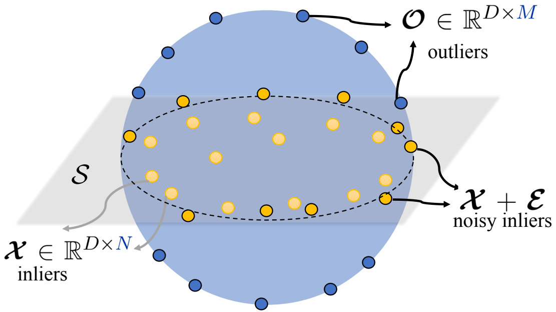
Focus of this paper

Robust subspace learning of **high relative dimension** with **noisy data**

Noisy Dual Principal Component Pursuit (DPCP)

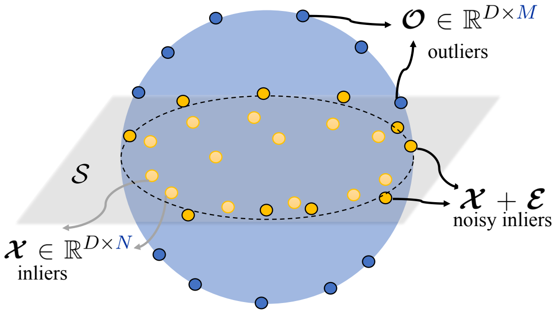


Noisy Dual Principal Component Pursuit (DPCP)



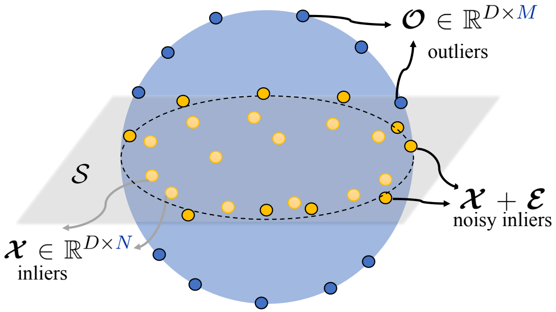
- outliers $\boldsymbol{\mathcal{O}}$ are drawn uniformly random from the unit sphere

Noisy Dual Principal Component Pursuit (DPCP)



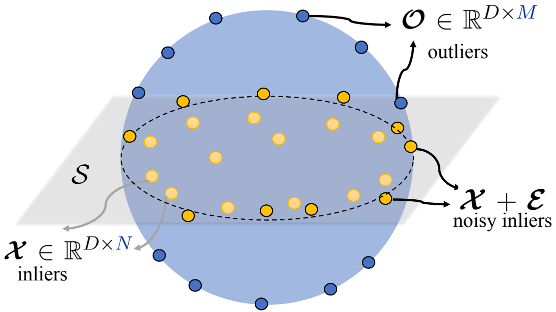
- outliers $\mathcal{O}$ are drawn uniformly random from the unit sphere
- noisy inliers $\mathcal{X} + \mathcal{E}$
  - inliers $\mathcal{X}$ are drawn within the $\mathcal{S}$

Noisy Dual Principal Component Pursuit (DPCP)



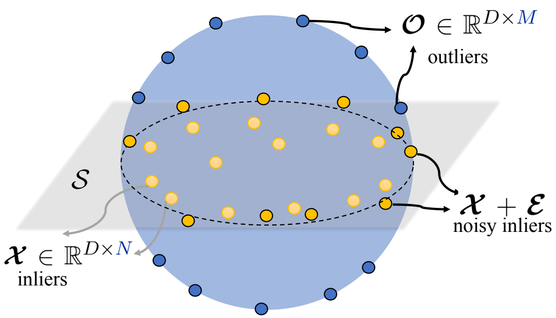
- outliers $\mathcal{O}$ are drawn uniformly random from the unit sphere
- noisy inliers $\mathcal{X} + \mathcal{E}$
    - inliers $\mathcal{X}$ are drawn within the $\mathcal{S}$
    - add Gaussian nose $\mathcal{E}$ on the inliers

Noisy Dual Principal Component Pursuit (DPCP)



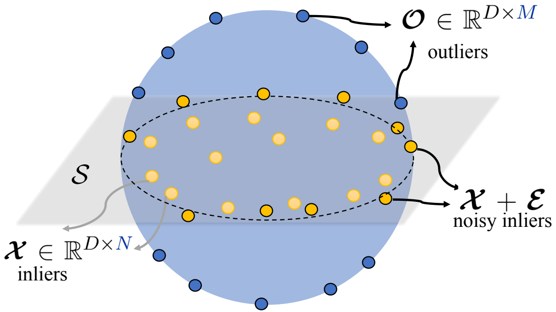
- outliers $\mathcal{O}$ are drawn uniformly random from the unit sphere
- noisy inliers $\mathcal{X} + \mathcal{E}$
  - inliers $\mathcal{X}$ are drawn within the $\mathcal{S}$
  - add Gaussian nose $\mathcal{E}$ on the inliers
  - normalize $\mathcal{X} + \mathcal{E}$ to unit norm (on sphere)

Noisy Dual Principal Component Pursuit (DPCP)



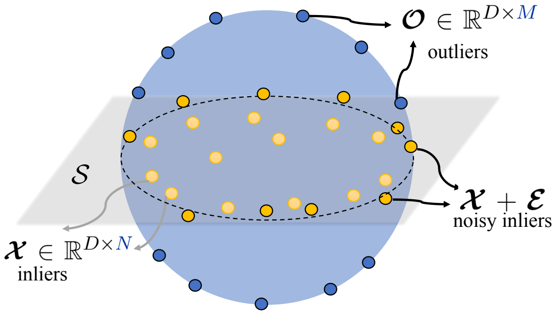
- outliers $\boldsymbol{\mathcal{O}}$ are drawn uniformly random from the unit sphere
- noisy inliers $\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{E}}$
  - inliers $\boldsymbol{\mathcal{X}}$ are drawn within the $\mathcal{S}$
  - add Gaussian nose $\boldsymbol{\mathcal{E}}$ on the inliers
  - normalize $\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{E}}$ to unit norm (on sphere)
- SNR is $\mathbb{E}[\|\boldsymbol{\mathcal{X}}\|_F]/\mathbb{E}[\|\boldsymbol{\mathcal{E}}\|_F] = 1/\sigma$

Noisy Dual Principal Component Pursuit (DPCP)



- outliers $\mathcal{O}$ are drawn uniformly random from the unit sphere
- noisy inliers $\mathcal{X} + \mathcal{E}$
  - inliers $\mathcal{X}$ are drawn within the $\mathcal{S}$
  - add Gaussian nose $\mathcal{E}$ on the inliers
  - normalize $\mathcal{X} + \mathcal{E}$ to unit norm (on sphere)
- SNR is $\mathbb{E}[\|\mathcal{X}\|_F]/\mathbb{E}[\|\mathcal{E}\|_F] = 1/\sigma$
- dataset $\widetilde{\mathcal{X}} = [\mathcal{X} + \mathcal{E},\ \mathcal{O}]$

Noisy Dual Principal Component Pursuit (DPCP)

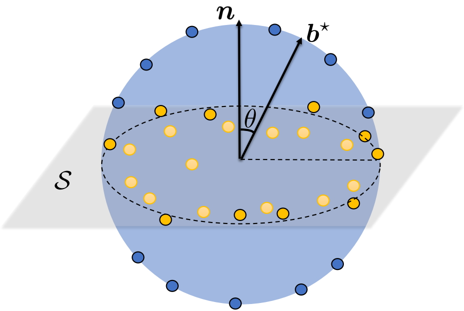


- outliers $\mathcal{O}$ are drawn uniformly random from the unit sphere
- noisy inliers $\mathcal{X} + \mathcal{E}$
  - inliers $\mathcal{X}$ are drawn within the $\mathcal{S}$
  - add Gaussian nose $\mathcal{E}$ on the inliers
  - normalize $\mathcal{X} + \mathcal{E}$ to unit norm (on sphere)
- SNR is $\mathbb{E}[\|\mathcal{X}\|_F]/\mathbb{E}[\|\mathcal{E}\|_F] = 1/\sigma$
- dataset $\widetilde{\mathcal{X}} = [\mathcal{X} + \mathcal{E},\ \mathcal{O}]$

DPCP problem formulation

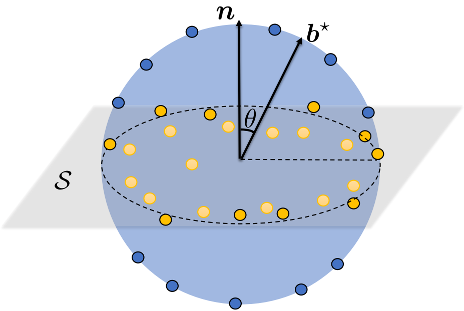
$$\min_{\boldsymbol{b}} \left\|\widetilde{\mathcal{X}}^\top \boldsymbol{b}\right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1 \tag{1}$$

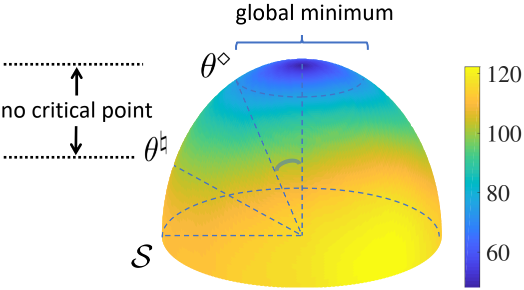
Deterministic Global Optimality Analysis



$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

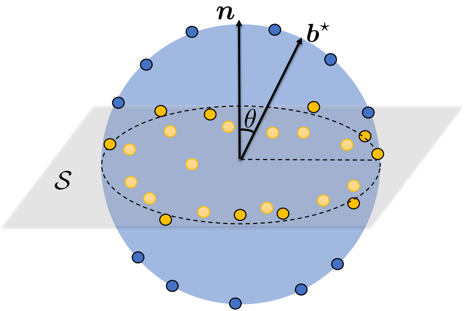
Deterministic Global Optimality Analysis



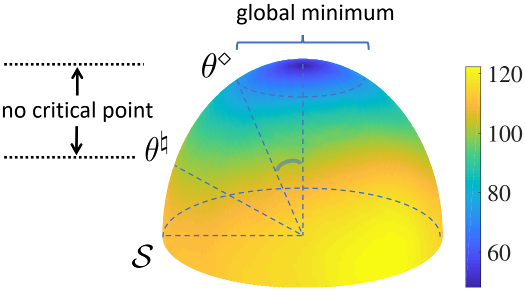


$$\min_{\boldsymbol{b}} \left\|\widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b}\right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

Deterministic Global Optimality Analysis

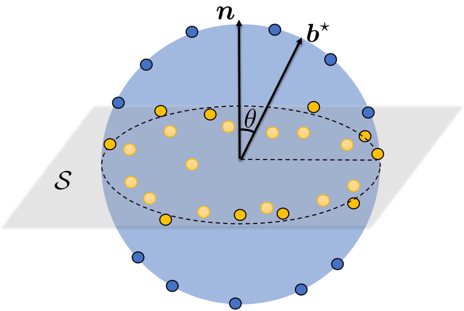


$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

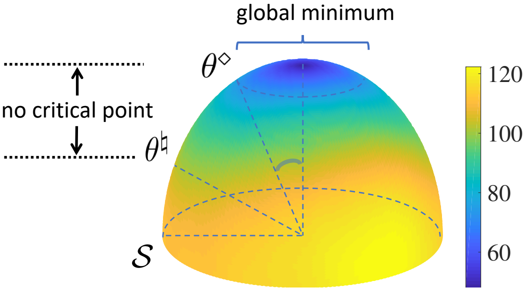


- Lemma: critical point is close to $\boldsymbol{n}$ or is close to $\mathcal{S}$

Deterministic Global Optimality Analysis



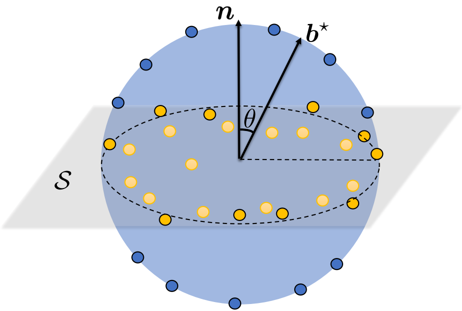
$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$



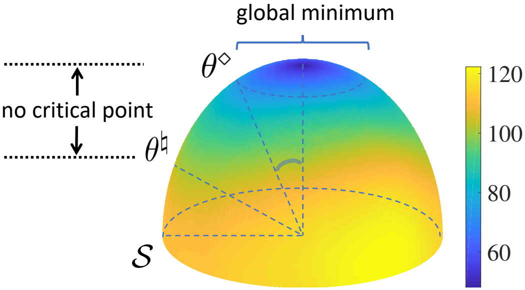
- Lemma: critical point is close to $\boldsymbol{n}$ or is close to $\mathcal{S}$
- Theorem: $\boldsymbol{b}^\star$ is close to $\boldsymbol{n}$:

$$\sin(\theta^\diamond) \lesssim \frac{\text{noise level}}{1 - \text{outlier ratio}}$$

Deterministic Global Optimality Analysis



$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$



- Lemma: critical point is close to $\boldsymbol{n}$ or is close to $\mathcal{S}$
- Theorem: $\boldsymbol{b}^\star$ is close to $\boldsymbol{n}$:

$$\sin(\theta^\diamond) \lesssim \frac{\text{noise level}}{1 - \text{outlier ratio}}$$

- In the noiseless case, $\boldsymbol{b}^\star = \boldsymbol{n}$

Probabilistic Analysis

$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^{\top} \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

Probabilistic Analysis

$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

Theorem

$\boldsymbol{b}^\star$ has a principal angle θ satisfies

$$\sin(\theta) \lesssim \sqrt{\sigma}/(1 - \sqrt{\sigma})$$

with probability exceeding $1 - O\left(\frac{1}{N^2}\right)$ if

$$M \leq C \cdot N^2.$$

Probabilistic Analysis

$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^{\top} \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

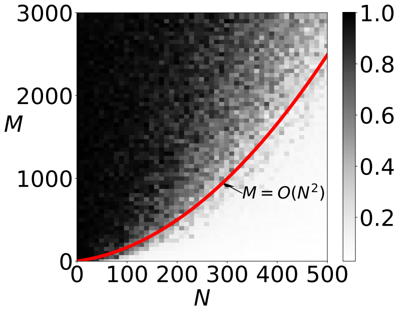
Theorem

$\boldsymbol{b}^{\star}$ has a principal angle θ satisfies

$$\sin(\theta) \lesssim \sqrt{\sigma}/(1-\sqrt{\sigma})$$

with probability exceeding $1 - O\left(\frac{1}{N^2}\right)$ if

$$M \leq C \cdot N^2.$$



Probabilistic Analysis

$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^{\top} \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

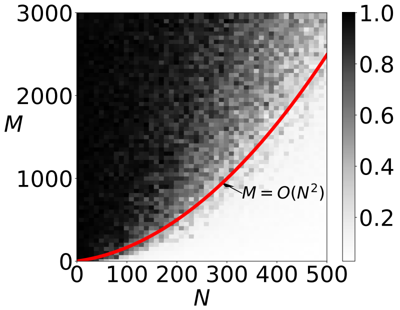
Theorem

$\boldsymbol{b}^\star$ has a principal angle θ satisfies

$$\sin(\theta) \lesssim \sqrt{\sigma}/(1 - \sqrt{\sigma})$$

with probability exceeding $1 - O\left(\frac{1}{N^2}\right)$ if

$$M \leq C \cdot N^2.$$



- Comparison with state-of-the-art: other methods can only handle at most $M = O(N)$ outliers in theory [Lerman and Maunu 18]

Projected SubGradient Method (PSGM)

$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

Projected SubGradient Method (PSGM)

$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

- Spectral initialization:

$$\widehat{\boldsymbol{b}}_0 \leftarrow \arg \min_{\boldsymbol{b} \in \mathbb{S}^{D-1}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_2$$

Projected SubGradient Method (PSGM)

$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

- Spectral initialization:

$$\widehat{\boldsymbol{b}}_0 \leftarrow \arg \min_{\boldsymbol{b} \in \mathbb{S}^{D-1}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_2$$

- Piecewise diminishing stepsize:

$$\mu_k \leftarrow \mu_0 \beta^{\lfloor (k - K_0)/K \rfloor}, \quad \beta < 1$$

Projected SubGradient Method (PSGM)

$$\min_{\boldsymbol{b}} \left\|\widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b}\right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

- Spectral initialization:

$$\widehat{\boldsymbol{b}}_0 \leftarrow \arg \min_{\boldsymbol{b} \in \mathbb{S}^{D-1}} \left\|\widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b}\right\|_2$$

- Piecewise diminishing stepsize:

$$\mu_k \leftarrow \mu_0 \beta^{\lfloor (k-K_0)/K \rfloor}, \quad \beta < 1$$

- PSG update:

$$\boldsymbol{b}_{k+1} \leftarrow \widehat{\boldsymbol{b}}_k - \mu_k \boldsymbol{g}_k$$

Projected SubGradient Method (PSGM)

$$\min_{\boldsymbol{b}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

- Spectral initialization:

$$\widehat{\boldsymbol{b}}_0 \leftarrow \arg \min_{\boldsymbol{b} \in \mathbb{S}^{D-1}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b} \right\|_2$$

- Piecewise diminishing stepsize:

$$\mu_k \leftarrow \mu_0 \beta^{\lfloor (k-K_0)/K \rfloor}, \quad \beta < 1$$

- PSG update:

$$\boldsymbol{b}_{k+1} \leftarrow \widehat{\boldsymbol{b}}_k - \mu_k \boldsymbol{g}_k$$

- Projection:

$$\widehat{\boldsymbol{b}}_{k+1} \leftarrow \boldsymbol{b}_{k+1} / \|\boldsymbol{b}_{k+1}\|_2$$

Projected SubGradient Method (PSGM)

$$\min_{\boldsymbol{b}} \left\|\widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b}\right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

- Spectral initialization:

$$\widehat{\boldsymbol{b}}_0 \leftarrow \arg\min_{\boldsymbol{b}\in\mathbb{S}^{D-1}} \left\|\widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b}\right\|_2$$

- Piecewise diminishing stepsize:

$$\mu_k \leftarrow \mu_0 \beta^{\lfloor (k-K_0)/K \rfloor}, \quad \beta < 1$$

- PSG update:

$$\boldsymbol{b}_{k+1} \leftarrow \widehat{\boldsymbol{b}}_k - \mu_k \boldsymbol{g}_k$$

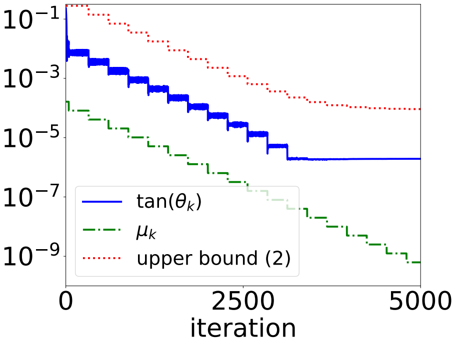
- Projection:

$$\widehat{\boldsymbol{b}}_{k+1} \leftarrow \boldsymbol{b}_{k+1} / \|\boldsymbol{b}_{k+1}\|_2$$

Theorem

$\tan(\theta_k)$ has a piecewise linear convergence rate:

$$\tan(\theta_k) \lesssim \beta^{\lfloor (k-K_0)/K \rfloor} + \frac{\sqrt{\sigma}}{\sqrt{1-2\sqrt{\sigma}}}. \qquad (2)$$

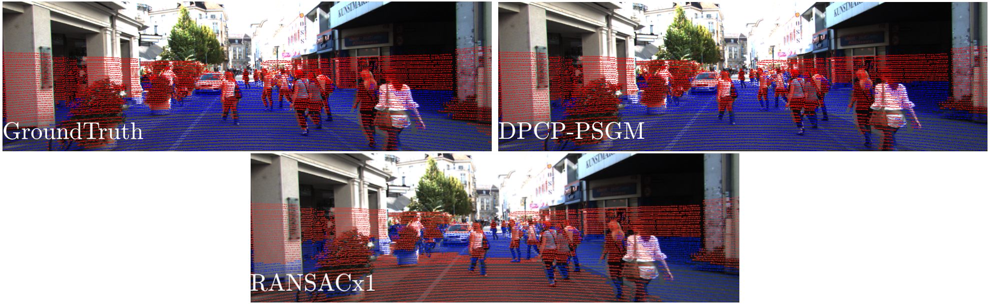


Experiments on 3D Point Cloud Road Data

Task

Learn an affine plane as a model for the road from a 3D point cloud

- Determine points that lie on the plane (inliers) / off the plane (outliers)
- Frame 328 of dataset KITTI-CITY-71, with inliers (blue) / outliers (red)



Contains around 10^5 points with approximately 50% outliers

Thank you!

Poster Session:

Today 06:30 – 09:00 PM
@ Pacific Ballroom #188