Noisy Dual Principal Component Pursuit

Tianyu Ding[1], Zhihui Zhu[1], Tianjiao Ding[2], Yunchen Yang[2], René Vidal[1], Manolis Tsakiris[2], Daniel Robinson[1]

[1]Johns Hopkins University [2]ShanghaiTech University

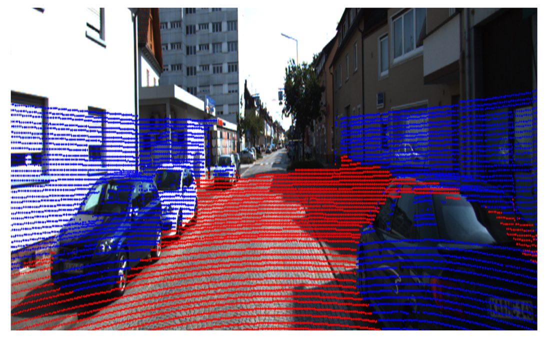
Johns Hopkins Mathematical Institute for Data Science — Vision Lab @ JHU http://www.vision.jhu.edu

ICML | 2019 — Thirty-sixth International Conference on Machine Learning

Motivation

- Problem: Fit a subspace to data contaminated with noise and outliers
- Prior work: Sparse & low-rank methods on low dimensional subspace
- Challenges: Many applications require subspace to be of high relative dimension; not obvious to establish guarantees in presence of noise

Estimating the geometry of a room from its depth map by fitting multiple planes

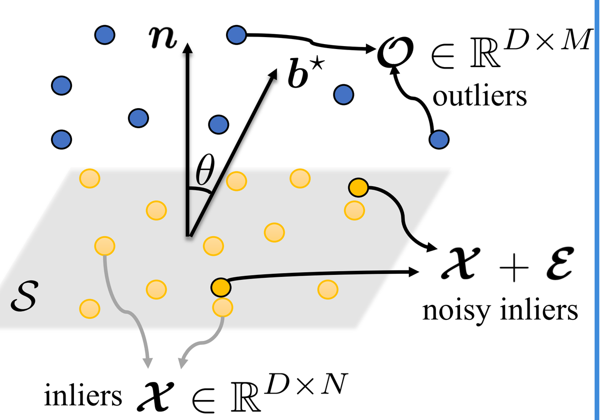


Estimating the road plane from points collected by a laser scanner

Dual Principal Component Pursuit (DPCP) in Noisy Setting

- Inliers $\boldsymbol{\mathcal{X}}$ span a d-dimensional subspace $\mathcal{S}$
- Outliers $\boldsymbol{\mathcal{O}}$ lie in ambient space $\mathbb{R}^D$
- Ambient noise $\boldsymbol{\mathcal{E}}$ are additive on inliers
- Given dataset $\widetilde{\boldsymbol{\mathcal{X}}} = [\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{E}}\ \boldsymbol{\mathcal{O}}]$, DPCP computes solution $\boldsymbol{b}^\star$ (ideally $\boldsymbol{b}^\star \perp \mathcal{S}$) by

$$(1) \qquad \min_{\boldsymbol{b}\in\mathbb{S}^{D-1}\text{: unit sphere}} \left\|\widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{b}\right\|_1$$



- Noisy DPCP assumes $\boldsymbol{\mathcal{E}} \neq \boldsymbol{0}$, which is opposite with the regular noiseless DPCP that based on $\boldsymbol{\mathcal{E}} = \boldsymbol{0}$

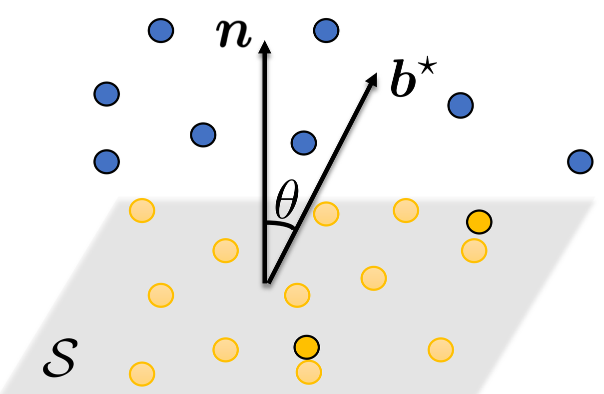
Main Contributions

- When $\boldsymbol{\mathcal{E}} = \boldsymbol{0}$ (noiseless), [Tsakiris & Vidal], [Zhu et al.] derived geometric conditions under which $\boldsymbol{b}^\star \perp \mathcal{S}$ and proposed efficient algorithms
- This paper extends the global optimality and convergence theory of DPCP to the case of data corrupted by noise

	Main Contributions	
Theory	Geometric Deterministic Analysis	$\text{dist}(\boldsymbol{b}^\star, \mathcal{S}^\perp) \propto$ noise level under some conditions
	Geometric Probabilistic Analysis	$\text{dist}(\boldsymbol{b}^\star, \mathcal{S}^\perp) \propto$ noise level with high probability if $\#\text{outliers} = O((\#\text{inliers})^2)$
Algorithm	Projected SubGradient Method (PSGM)	piecewise linear convergence: $\text{dist}(\boldsymbol{b}_k, \mathcal{S}^\perp) = O(\beta^k) +$ const, $\beta < 1$, const $\propto$ noise level

Deterministic Global Optimality Analysis

- Assume the data is given and fixed
- When $\boldsymbol{\mathcal{E}} \neq \boldsymbol{0}$, $\boldsymbol{b}^\star$ will be perturbed away from $\mathcal{S}^\perp$ by an angle θ
- $\text{dist}(\boldsymbol{b}^\star, \mathcal{S}^\perp)$ depends on noise level



Lemma: Any critical point of (1) must have its principal angle θ satisfy:

$$\theta \le \theta^\circ \quad \text{or} \quad \theta \ge \theta^\natural,$$

where $0 \le \theta^\circ \le \theta^\natural \le \pi/2$ are closely realted to the two nonnegative roots of a certain quartic equation.

Remark: Any critical point is either close to $\mathcal{S}^\perp$ or close to $\mathcal{S}$

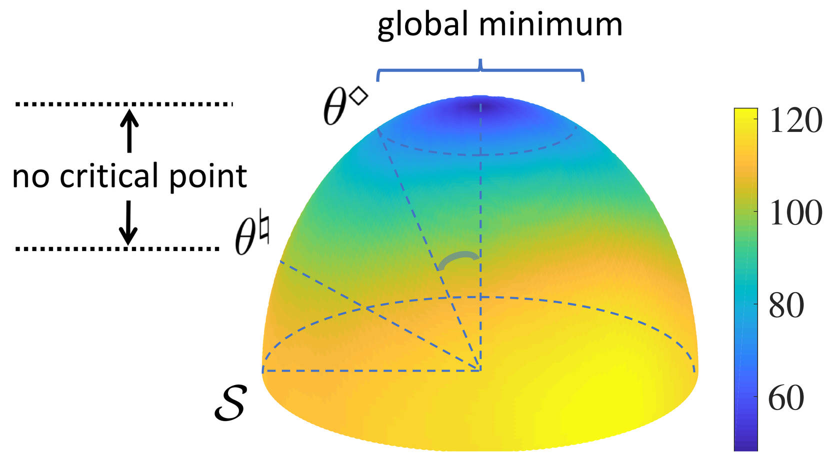
Theorem: Any global solution $\boldsymbol{b}^\star$ to (1) must be close to $\mathcal{S}^\perp$ such that $\theta \in [0, \theta^\circ]$ as long as the outlier ratio and the noise level is small.

Remark: Any global solution $\boldsymbol{b}^\star$ is close to $\mathcal{S}^\perp$ such that $\theta \le \theta^\circ$

- $\theta^\circ \propto$ the effective noise level

	θ°	$\theta^\natural$
smaller $\|\boldsymbol{\mathcal{E}}\|$, smaller $\|\boldsymbol{\mathcal{O}}\|$	$\downarrow 0$	$\uparrow \frac{\pi}{2}$

- $\boldsymbol{\mathcal{O}} = \boldsymbol{0} \Rightarrow \theta^\natural = \pi/2$
- $\boldsymbol{\mathcal{E}} = \boldsymbol{0} \Rightarrow \theta^\circ = 0$, which means solution $\boldsymbol{b}^\star$ is a normal vector of $\mathcal{S}$



Landscape of DPCP problem (1)

Probabilistic Analysis

- Columns of outliers $\boldsymbol{\mathcal{O}}$ are drawn uniformly from the sphere $\mathbb{S}^{D-1}$
- Columns of noisy inliers $\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{E}}$ are drawn from $\mathbb{S}^{D-1}$ by normalizing i.i.d. $\mathcal{N}(\boldsymbol{0}, \frac{1}{d}\boldsymbol{\Pi}_\mathcal{S} + \frac{\sigma^2}{D}\boldsymbol{I}_D)$-distributed points to have unit ℓ_2 norm
- Under this random model, the SNR is $\mathbb{E}[\|\boldsymbol{\mathcal{X}}\|_F]/\mathbb{E}[\|\boldsymbol{\mathcal{E}}\|_F] = 1/\sigma$

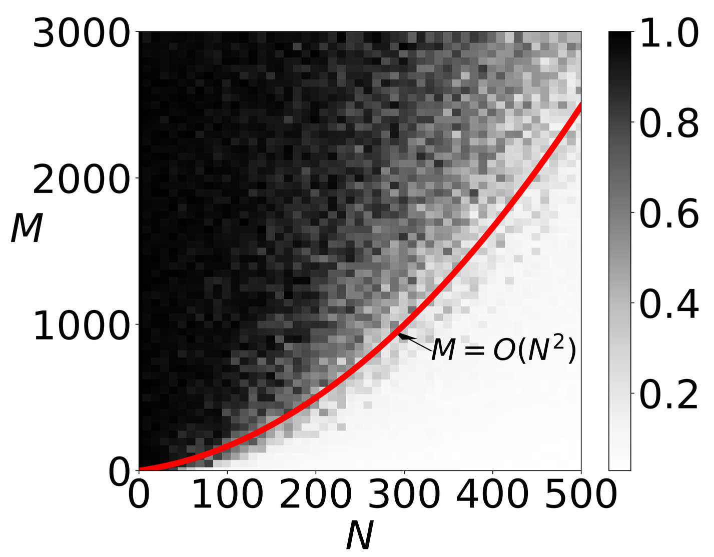
Theorem: Any global solution to (1) must lie in a neighborhood of $\mathcal{S}^\perp$ such that $\sin(\theta) \lesssim \sqrt{\sigma}/(1-\sqrt{\sigma})$ with probability exceeding $1 - O(\frac{1}{N^2})$ if

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

- C is a constant depends on σ, D and d

	Smaller σ	Larger D, d
C	$\uparrow$	$\downarrow$

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



Plot of $\sin(\theta)$, $\sigma = 0.1$

Projected SubGradient Method (PSGM)

Input: $\widetilde{\boldsymbol{\mathcal{X}}}, \{\mu_0, K_0, K\} > 0, \beta < 1$

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

Piecewise geometrically diminishing stepsize: $\mu_k = \mu_0 \beta^{\lfloor (k-K_0)/K \rfloor}$

PSGM update: $\boldsymbol{b}_{k+1} \leftarrow \widehat{\boldsymbol{b}}_k - \mu_k \boldsymbol{g}_k$; $\boldsymbol{g}_k$ is the sub-gradient

Projected back to sphere: $\widehat{\boldsymbol{b}}_{k+1} \leftarrow \boldsymbol{b}_{k+1}/\|\boldsymbol{b}_{k+1}\|_2$

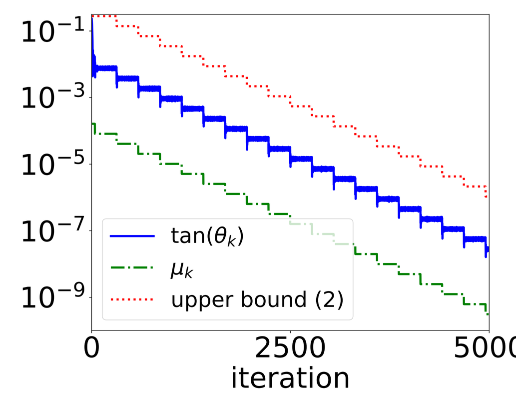
Theorem: If $\theta_0 < \theta^\natural$, $\tan(\theta_k)$ has a piecewise linear convergence rate:

$$(2) \qquad \tan(\theta_k) \le \beta^{\lfloor (k-K_0)/K \rfloor} + \tan(\theta'), \ \forall k \ge K_0.$$

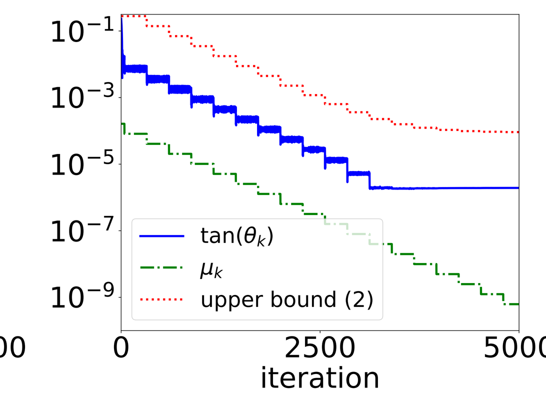
- θ' is a critical angle proportional to the effective noise level, which also depends on M, N, and the distribution of the data

	$\boldsymbol{\mathcal{E}} = \boldsymbol{0}$	M/N fixed, larger M, N, smaller noise level more well-distributed data
θ'	0	$\downarrow$

- Principal angle θ_k decays in a piecewise linear rate until θ' (noise effect); when noiseless $\theta' = 0 \Rightarrow \theta_k \to 0$
- $D = 30, d = 29, N = 1000, \frac{M}{M+N} = 0.7$



$\sigma = 0$

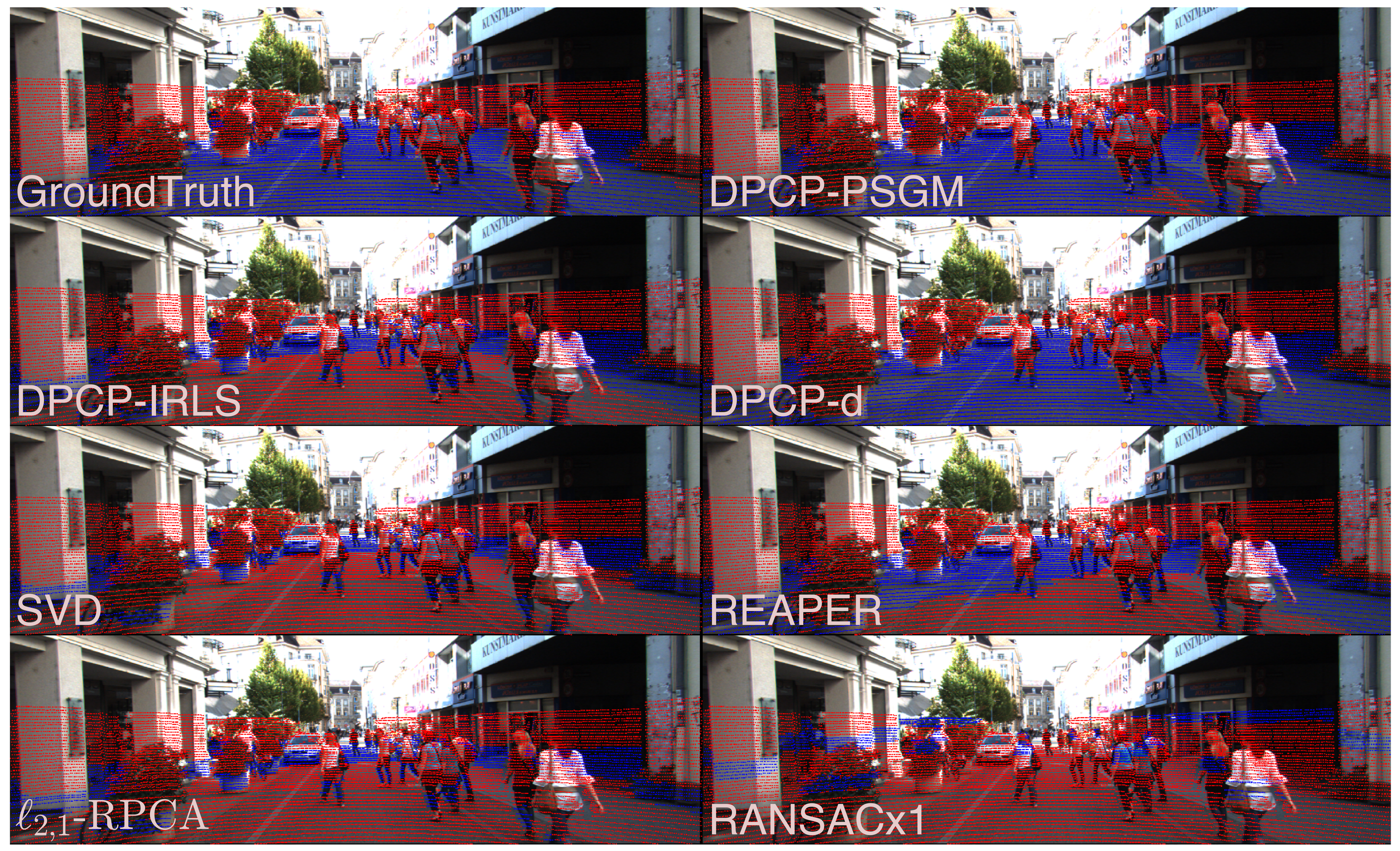


$\sigma = 10^{-6}$

Experiments on 3D Point Cloud Road Data

Task: Given a 3D point cloud of a road scene, the goal is to learn an affine plane as a model for the road

- 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

Supported by NSF grant 1704458.