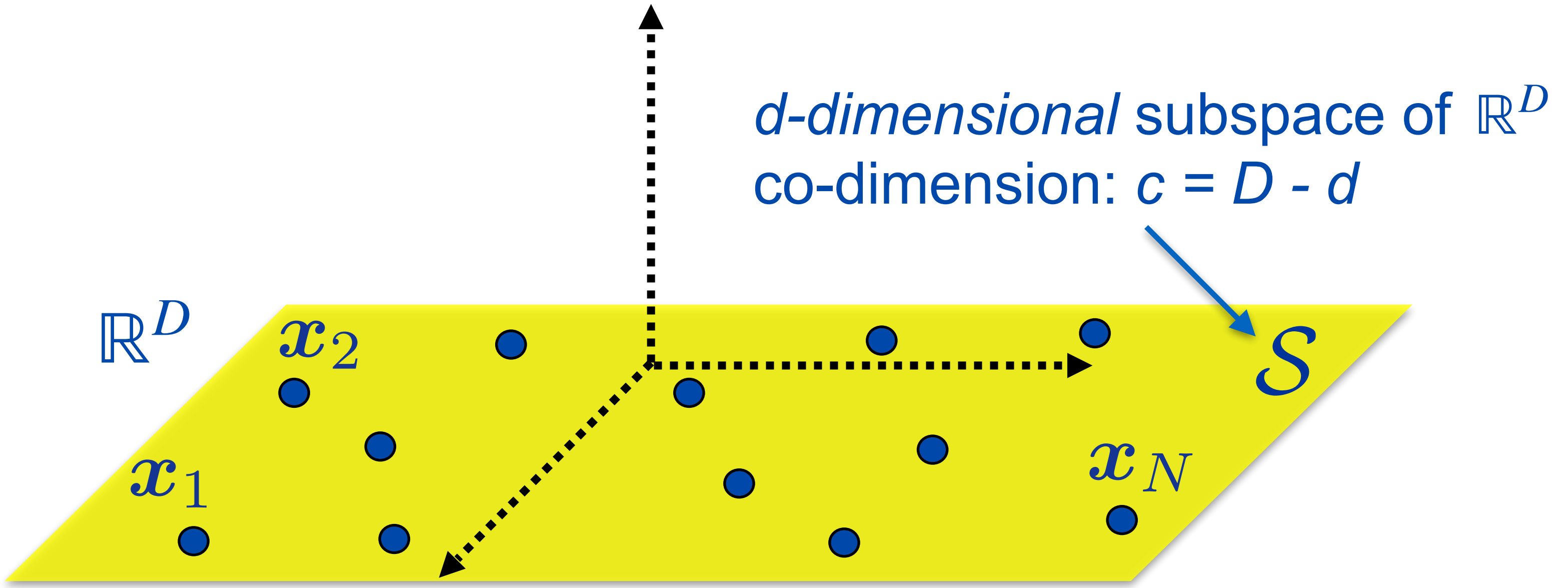
# Dual Principal Component Pursuit

**René Vidal**

Director of the Mathematical Institute for Data Science and Herschel Seder Professor of Biomedical Engineering at Johns Hopkins University, Amazon Scholar, and Chief Scientist at NORCE

# Principal Component Analysis (PCA)

- **Problem**: fit subspace S of dimension d to the data



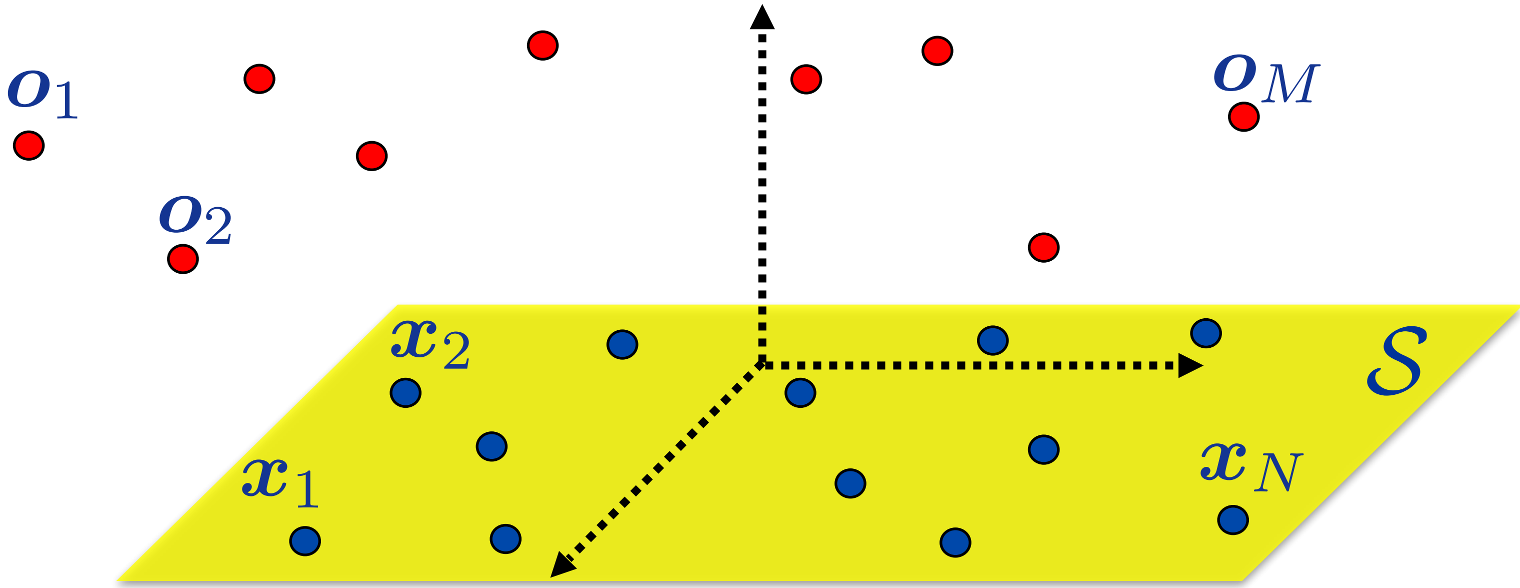

- **Solution**: can be obtained from SVD of the data matrix

$$U\Sigma V^\top = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_N \end{bmatrix} \in \mathbb{R}^{D \times N}$$

Beltrami'1873, Jordan'1874, Hotelling'1933, Eckart-Householder-Young'1936

# Robust Principal Component Analysis (RPCA)

- **Problem**: detect outliers and fit subspace to inliers




$$\tilde{\boldsymbol{X}} = [\underbrace{\boldsymbol{X}}_{D \times N} \quad \underbrace{\boldsymbol{O}}_{D \times M}]\boldsymbol{\Gamma} = [\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_L]$$

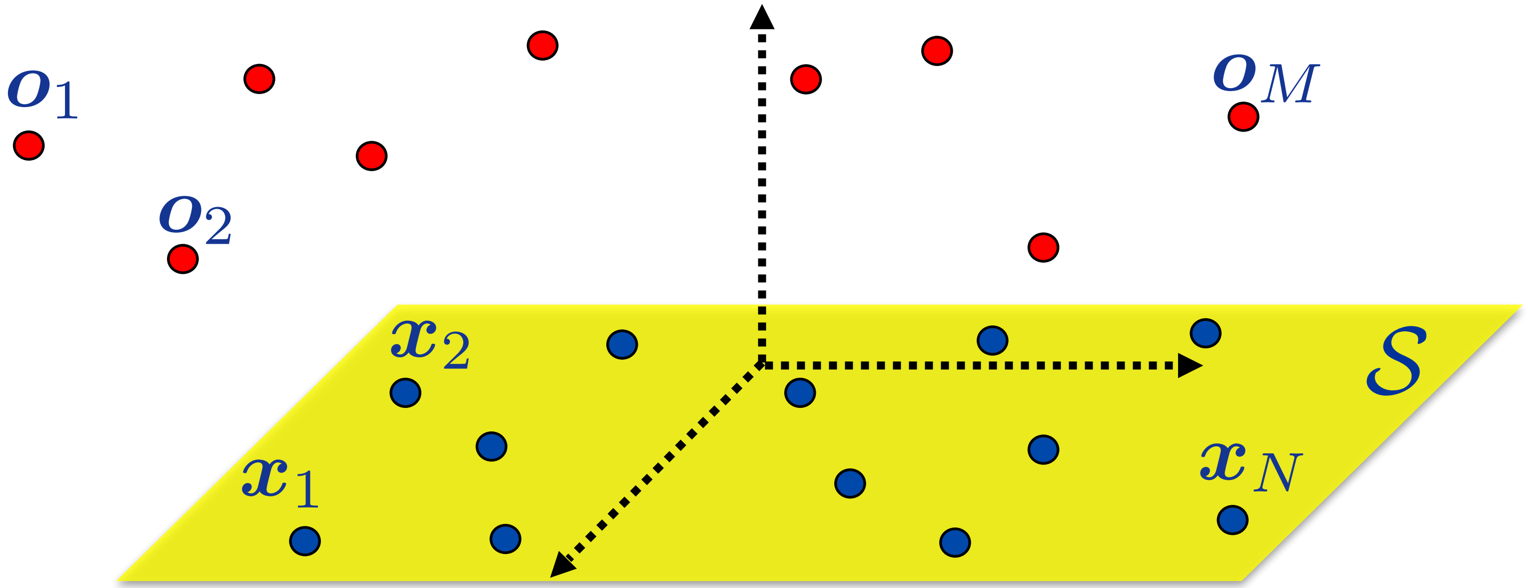
inliers (→ $\boldsymbol{X}$), outliers (→ $\boldsymbol{O}$), unknown permutation (→ $\boldsymbol{\Gamma}$)

- **Challenge**: SVD-based solution is very sensitive to outliers


[1] Fisher-Bolles '1981
[2] Xu-Caramanis-Sanghavi '2010
[3] Soltanokotabi-Candes '2012
[4] Lerman-McCoy-Tropp-Zhang '2015

# RPCA Algorithms

- **RANSAC [1]**: works very well if d and M/(N+M) are small




$N = \#\text{inliers}$ $\quad d = \text{subspace dimension}$

$M = \#\text{outliers}$ $\quad D = \text{ambient dimension}$


[1] Fisher-Bolles '1981
[2] Xu-Caramanis-Sanghavi '2010
[3] Soltanokotabi-Candes '2012
[4] Lerman-McCoy-Tropp-Zhang '2015

# RPCA Algorithms

- **RANSAC [1]**: works very well if d and M/(N+M) are small
- **L21-RPCA [2]**: perfect recovery if d/D & M/(N+M) are small, and if L is incoherent w.r.t. set of column-sparse matrices

$$\min_{L,E} \|L\|_* + \lambda \|E\|_{2,1} \quad \text{s.t.} \quad X = L + E$$

- **SE-RPCA [3]**: perfect recovery if d/D is small and inliers/outliers are well distributed

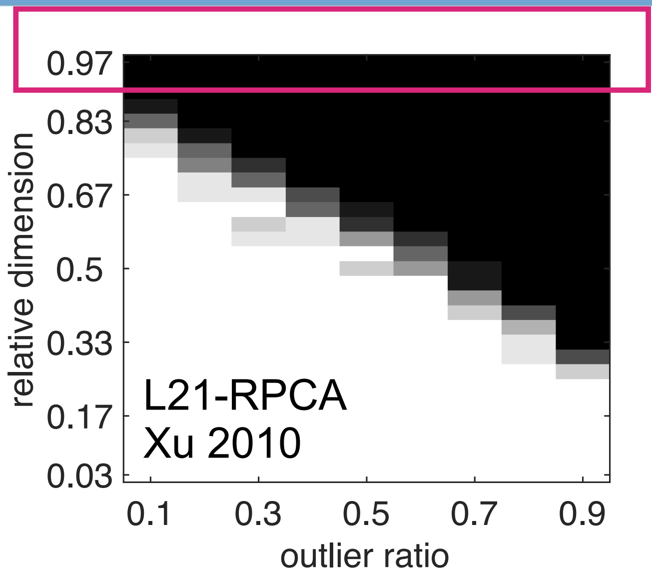
$$\min_{C} \|C\|_1 \quad \text{s.t.} \quad X = XC, \ \operatorname{diag}(C) = 0$$

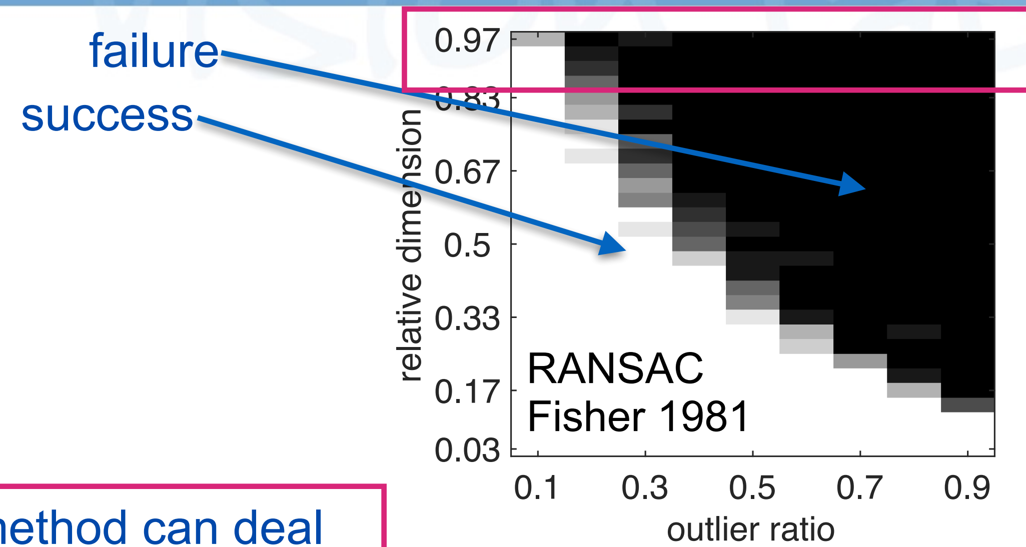
- **REAPER [4]**: perfect recovery if d/D and M/N are small, and inliers/outliers are well distributed

$$\min_{P} \|(I - P)X\|_{2,1} \quad \text{s.t.} \quad 0 \preceq P \preceq I, \operatorname{trace}(P) = d$$

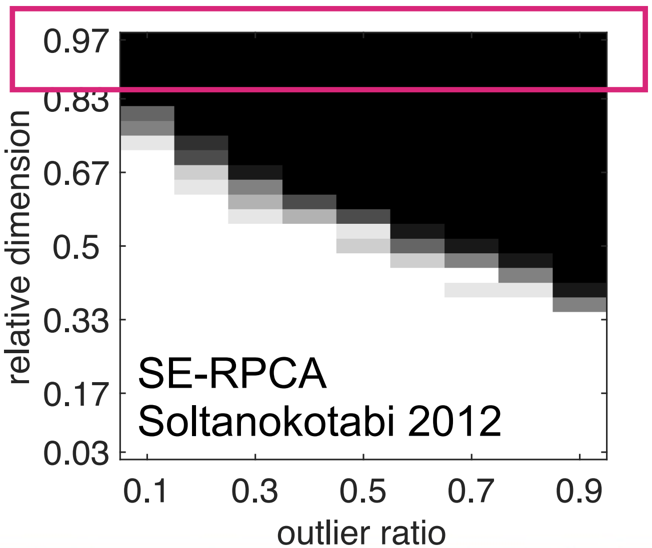

[1] Fisher-Bolles '1981
[2] Xu-Caramanis-Sanghavi '2010
[3] Soltanokotabi-Candes '2012
[4] Lerman-McCoy-Tropp-Zhang '2015

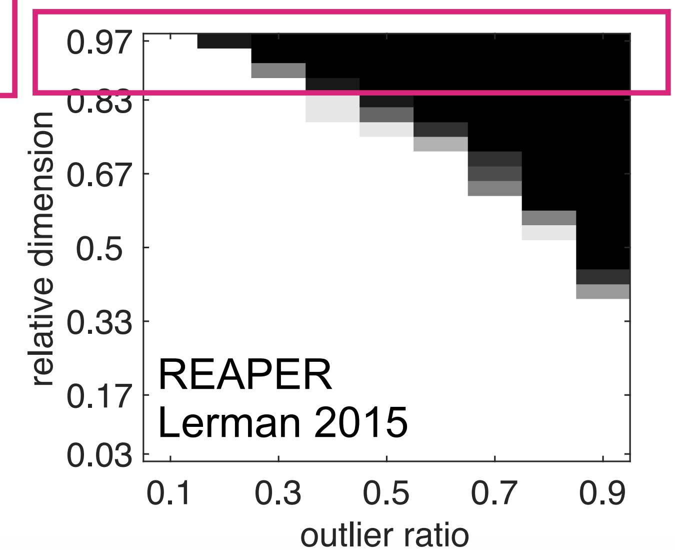
# Performance of RPCA Algorithms




No method can deal with hyperplanes for more than 10% outliers.



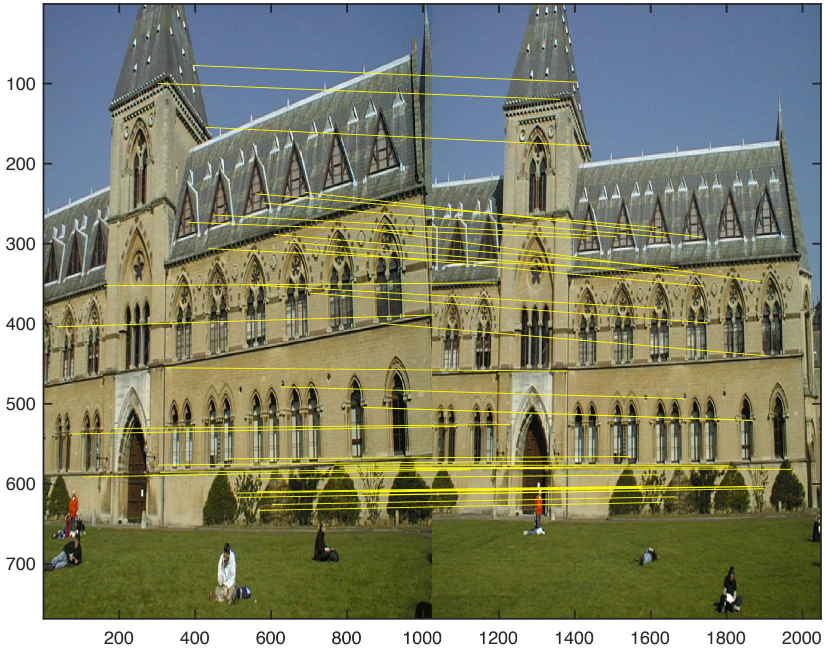




Ambient dimension (D) = 30 Number of inliers (N) = 300

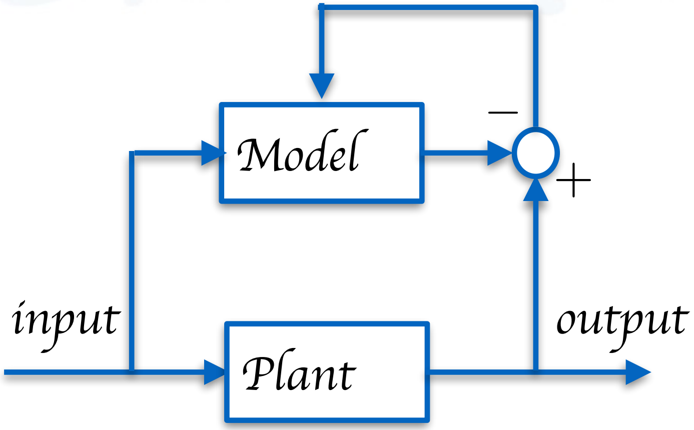
# Why High Relative Dimensions (d/D)?

3D point cloud analysis

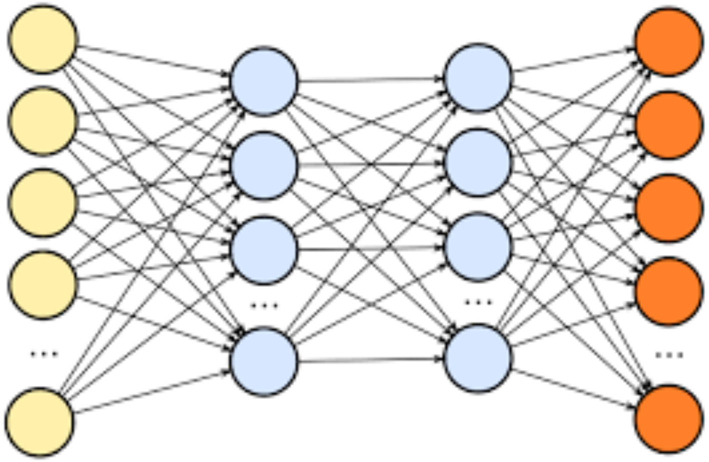
multiple view geometry



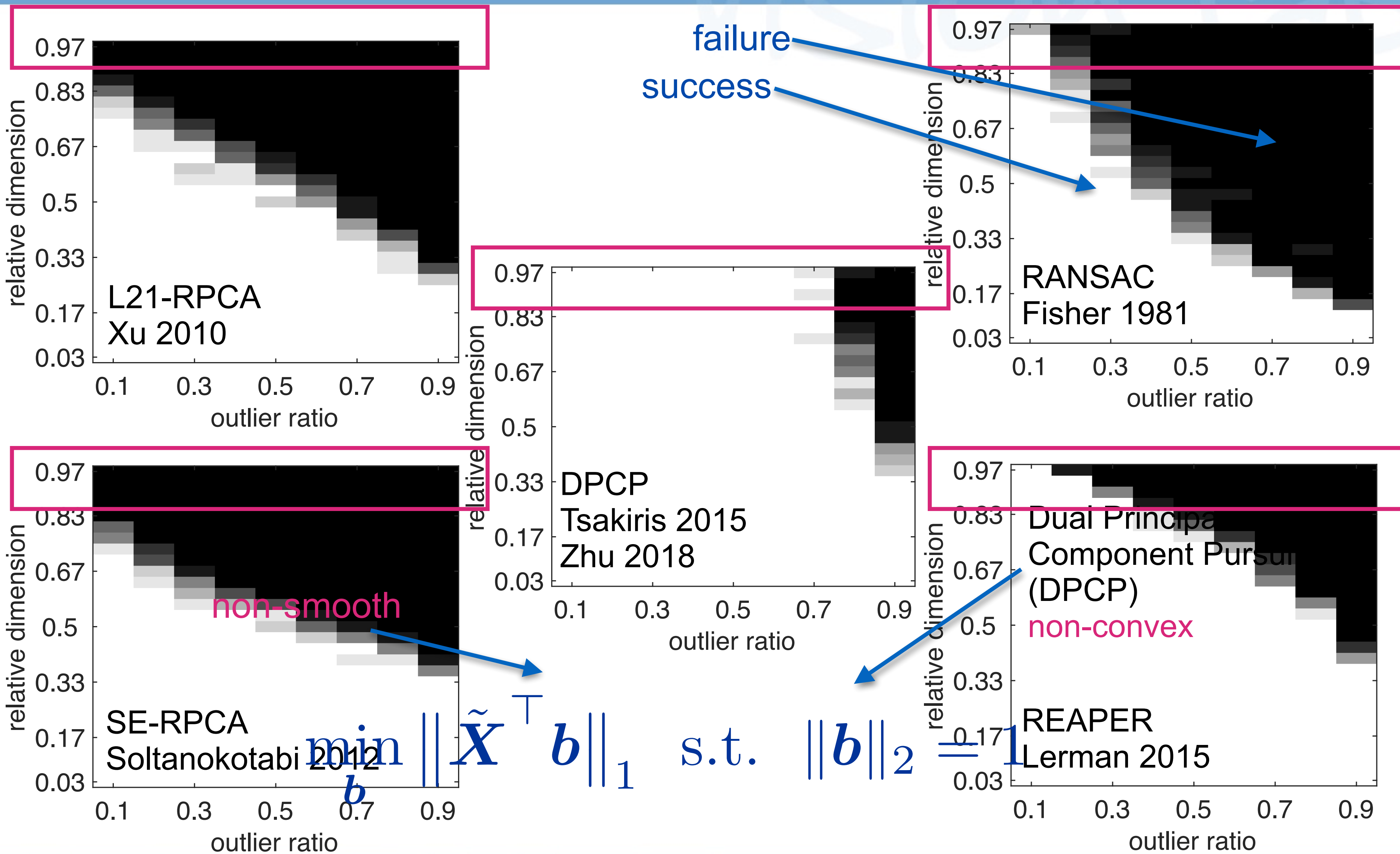

system identification




deep representations



# Dual Principal Component Pursuit (DPCP)




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

Ambient dimension (D) = 30    Number of inliers (N) = 300

# Talk Outline

- **Analysis**: are global minima orthogonal to the subspace?
  - **Geometric analysis**: if both inliers and outliers are well distributed, then any global minimum is orthogonal to the subspace
  - **Probabilistic analysis**: # of outliers $\lesssim$ (# of inliers)$^2$, any global minimum is orthogonal to the subspace with high probability

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

- **Algorithms**: do algorithms converge to global minima?
  - Projected sub-gradient method converges to normal vector at a linear rate if inliers/outliers are well distributed and $\boldsymbol{b}_0$ is far from subspace

$$\boldsymbol{b}_{k+1} \leftarrow \hat{\boldsymbol{b}}_k - \mu_k \tilde{\boldsymbol{X}} \operatorname{sign}(\tilde{\boldsymbol{X}}^\top \hat{\boldsymbol{b}}_k); \; \hat{\boldsymbol{b}}_{k+1} \leftarrow \frac{\boldsymbol{b}_{k+1}}{\|\boldsymbol{b}_{k+1}\|}$$

- **Experiments**: road plane detection on KITTI dataset


[1] Tsakiris and Vidal, Dual Principal Component Pursuit, ICCVW 2015, JMLR 2018.
[2] Tsakiris and Vidal, Hyperplane Clustering via Dual Principal Component Pursuit, ICML 2017.
[3] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, DPCP: Improved Analysis and Efficient Algorithms, NeurIPS 2018.
[4] Zhu, Ding, Robinson, Tsakiris, Vidal. A linearly convergent method for non-smooth non-convex optimization on the Grassmannian. NeurIPS, 2019.
[5] Ding, Zhu, Ding, Yang, Robinson, Tsakiris, Vidal. Noisy Dual Principal Component Pursuit. ICML 2019.
[6] Ding, Yang, Zhu, Robinson, Vidal, Kneip, Tsakiris. Robust Homography Estimation via Dual Principal Component Pursuit, CVPR 2020.

# Geometric Analysis of DPCP

JHU vision lab

THE DEPARTMENT OF BIOMEDICAL ENGINEERING
The Whitaker Institute at Johns Hopkins

JOHNS HOPKINS
MATHEMATICAL INSTITUTE
for DATA SCIENCE

# DPCP Problem Formulation

$$\tilde{\boldsymbol{X}} = [\underbrace{\boldsymbol{X}}_{D\times N} \quad \underbrace{\boldsymbol{O}}_{D\times M}]\boldsymbol{\Gamma} = \left[\tilde{\boldsymbol{x}}_1, \dots, \tilde{\boldsymbol{x}}_L\right], \; L = N + M$$

inliers

outliers

unknown permutation

- Compute distance from data points to hyperplane

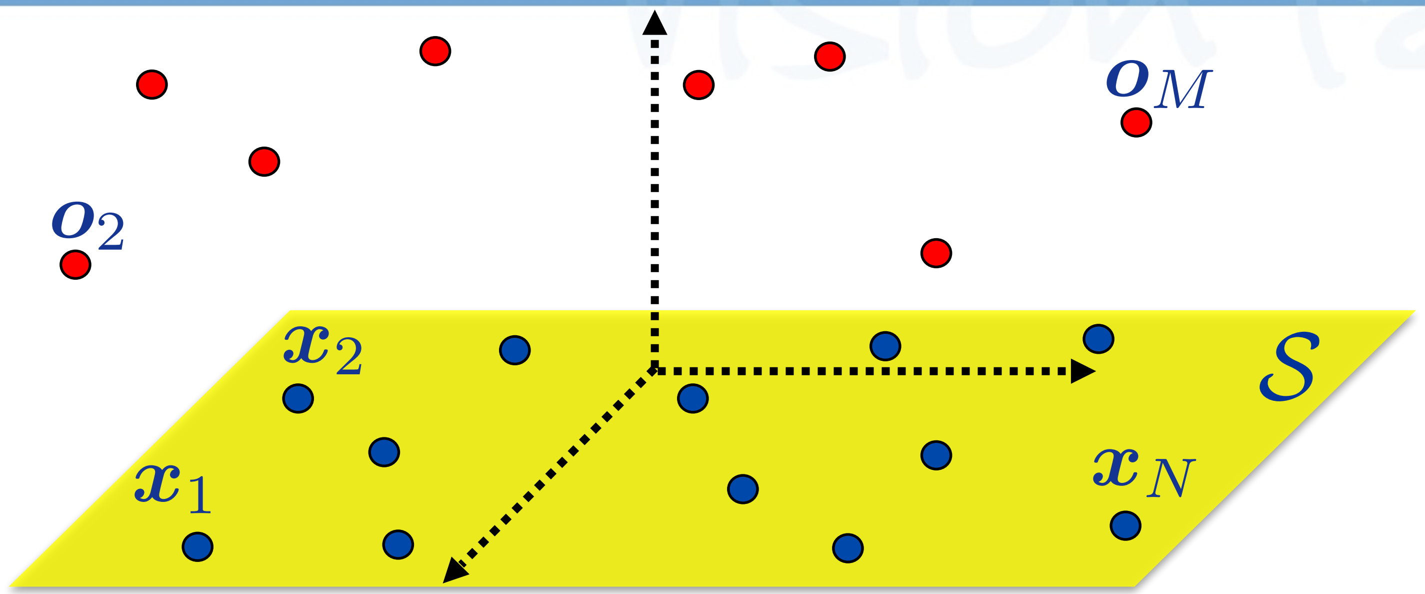
$$\tilde{\boldsymbol{X}}^\top \boldsymbol{b} = \left[\tilde{\boldsymbol{x}}_1^\top \boldsymbol{b}, \dots, \tilde{\boldsymbol{x}}_L^\top \boldsymbol{b}\right]^\top$$

- Find a hyperplane that contains as many points as possible

$$\min_{\boldsymbol{b}} \left\|\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\right\|_0 \quad \text{s.t.} \quad \boldsymbol{b} \neq \boldsymbol{0}$$

[1] Tsakiris and Vidal, Dual Principal Component Pursuit, ICCVW 2015, JMLR 2018.

# DPCP Problem Formulation




- Find vector that is orthogonal to as many points as possible

$$\min_{\boldsymbol{b}} \underbrace{\#\{i : |\tilde{\boldsymbol{x}}_i^\top \boldsymbol{b}| > 0\}}_{\|\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\|_0} \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

- DPCP replaces 0-norm by tightest convex relaxation

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

non-smooth
non-convex


[1] Tsakiris and Vidal, Dual Principal Component Pursuit, ICCVW 2015, JMLR 2018.

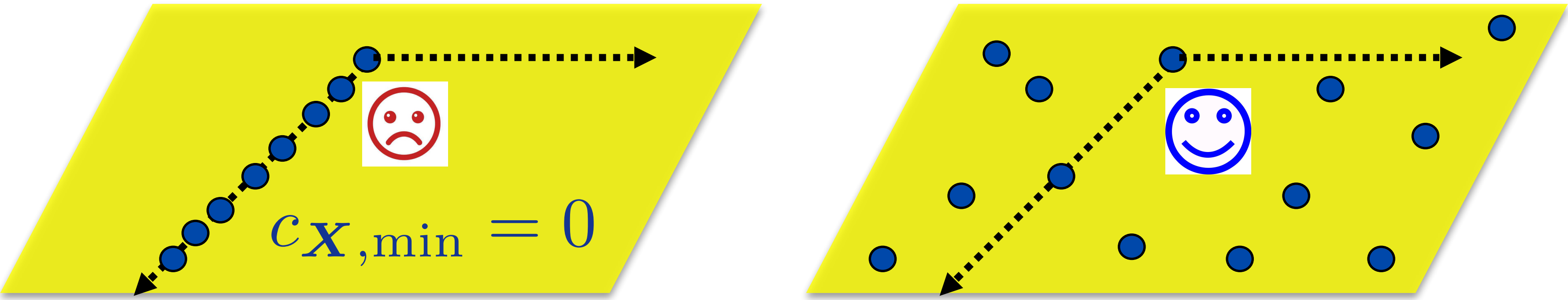
# Geometric Analysis: Measure for Inliers

inliers → $\left\|X^\top b\right\|_1$; outliers → $\left\|O^\top b\right\|_1$

$$\left\|\tilde{X}^\top b\right\|_1 = \left\|X^\top b\right\|_1 + \left\|O^\top b\right\|_1$$

- Inlier permeance

$$c_{X,\min} = \frac{1}{N} \min_{b \in \mathcal{S} \cap \mathbb{S}^{D-1}} \left\|X^\top b\right\|_1$$



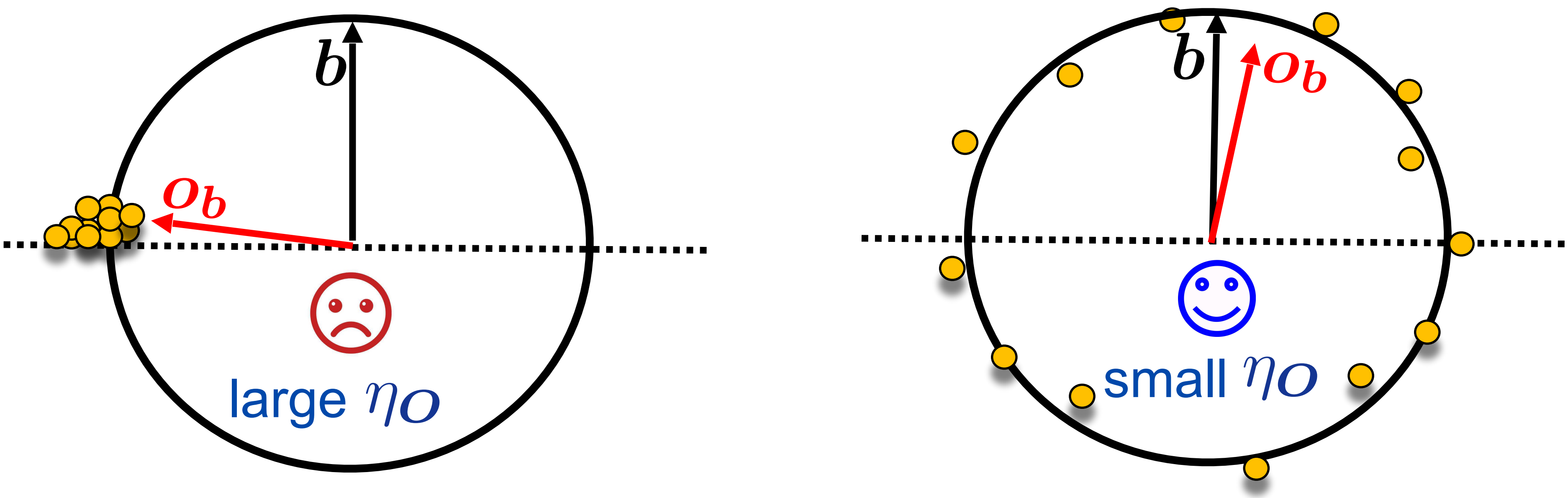

- The better the inliers are distributed, the larger $c_{X,\min}$


[1] Lerman, McCoy, Tropp, Zhang. Robust computation of linear models by convex relaxation, 2015.

# Geometric Analysis: Measure for Outliers

Riemannian sub-gradient

$$\eta_{\mathcal{O}} = \max_{\boldsymbol{b}\in\mathbb{S}^{D-1}} \Big\| \overbrace{(\mathbf{I} - \boldsymbol{b}\boldsymbol{b}^\top)\underbrace{\frac{1}{M}\mathcal{O}\,\mathrm{sign}(\mathcal{O}^\top \boldsymbol{b})}_{\boldsymbol{o_b}}} \Big\|_2$$

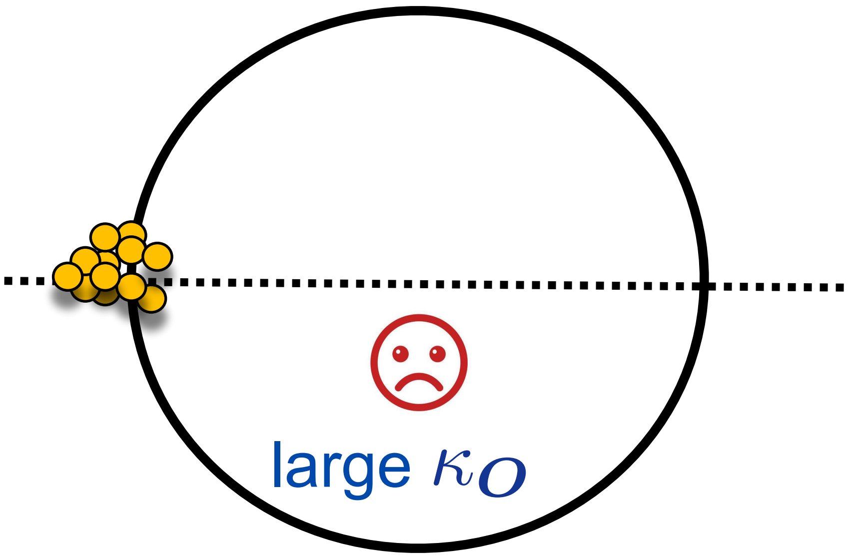


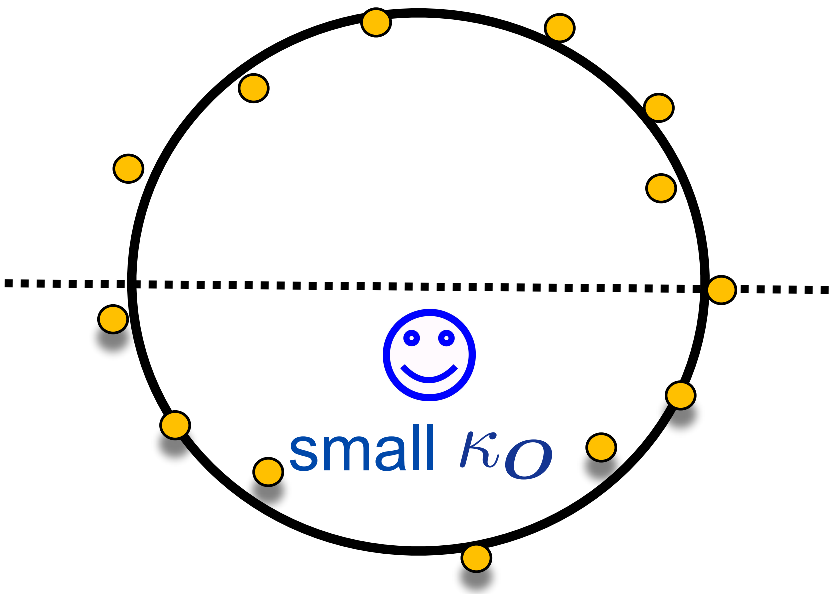

- The better the outliers are distributed, the smaller $\eta_O$


[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

# Geometric Analysis: Measure for Outliers

$$\kappa_{\boldsymbol{O}} = \max_{\boldsymbol{b}\in\mathbb{S}^{D-1}} \|\boldsymbol{O}^\top \boldsymbol{b}\|_1 - \min_{\boldsymbol{b}\in\mathbb{S}^{D-1}} \|\boldsymbol{O}^\top \boldsymbol{b}\|_1$$




- The better the outliers are distributed, the smaller $\kappa_{\boldsymbol{O}}$

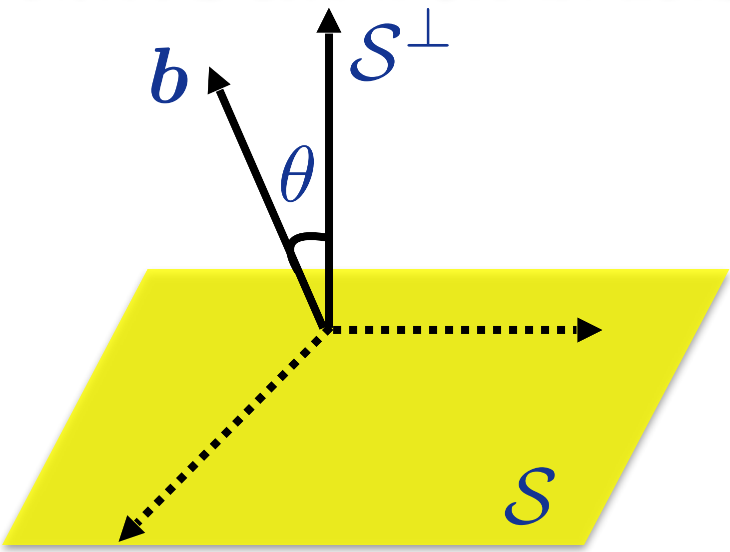

[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

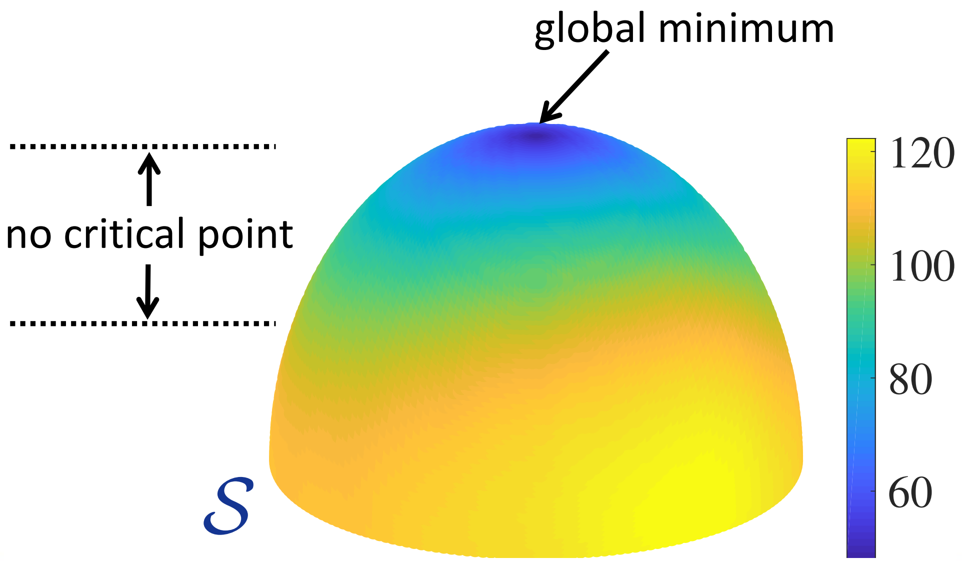
# Distribution of Critical Points

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

- **Lemma:** Any critical point of the DPCP problem must either lie in $\mathcal{S}^\perp$, or have a principal angle

$$\theta \geq \theta^\natural := \arccos\left(\frac{M\eta_{\mathcal{O}}}{N c_{\boldsymbol{X},\min}}\right)$$





[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

# Distribution of Critical Points

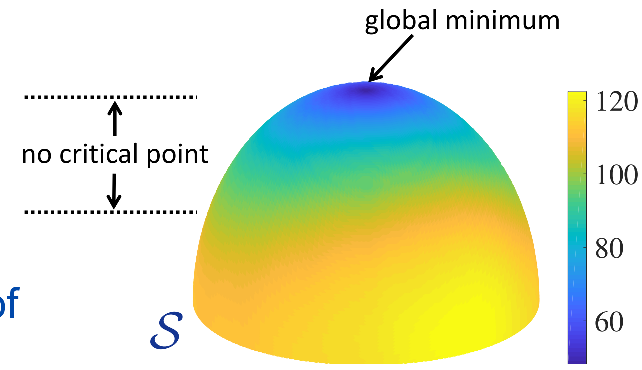
$$\min_{\boldsymbol{b}} \left\|\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

- **Lemma:** Any critical point of the DPCP problem must either lie in $\mathcal{S}^\perp$, or have a principal angle

$$\theta \geq \theta^\natural := \arccos\left(\frac{M\eta_{\boldsymbol{O}}}{N c_{\boldsymbol{X},\min}}\right)$$

| | Data more well-distributed | Data well-distributed, $N, M \to \infty$ |
|---|---|---|
| $c_{\boldsymbol{X},\min}$ | ↑ | $c_d$ |
| $\eta_{\boldsymbol{O}}$ | ↓ | **0** |
| $\theta^\natural$ | ↑ | $\frac{\pi}{2}$ if $\frac{M}{N}$ is fixed |

Average height of unit hemisphere





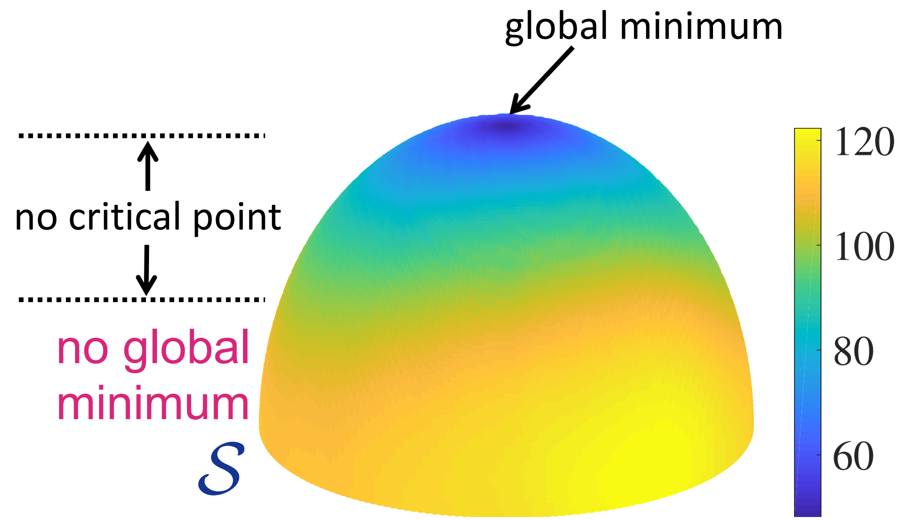
[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

# Global Optimality of DPCP

- **Theorem:** Any global solution to the DPCP problem must lie in $\mathcal{S}^{\perp}$ if

$$\left(\frac{M\eta_{\boldsymbol{O}}}{Nc_{\boldsymbol{X},\min}}\right)^2 + \left(\frac{M\kappa_{\boldsymbol{O}}}{Nc_{\boldsymbol{X},\min}}\right)^2 < 1$$

| | Data more well-distributed | Data well-distributed, $N, M \to \infty$ |
|---|---|---|
| $c_{\boldsymbol{X},\min}$ | ↑ | $c_d$ |
| $\eta_{\boldsymbol{O}}$ | ↓ | 0 |
| $\theta^{\natural}$ | ↑ | $\frac{\pi}{2}$ if $\frac{M}{N}$ is fixed |
| $\kappa_{\boldsymbol{O}}$ | ↓ | 0 |




- This holds for any fixed $\frac{M}{N}$ with $N \to \infty$ if data well distributed


[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

# Probabilistic Analysis of DPCP

# Probabilistic Analysis

- Geometric condition for optimality depends on data $\boldsymbol{X}$ and $\boldsymbol{O}$

$$\left(\frac{M\eta_{\boldsymbol{O}}}{Nc_{\boldsymbol{X},\min}}\right)^2 + \left(\frac{M\kappa_{\boldsymbol{O}}}{Nc_{\boldsymbol{X},\min}}\right)^2 < 1$$

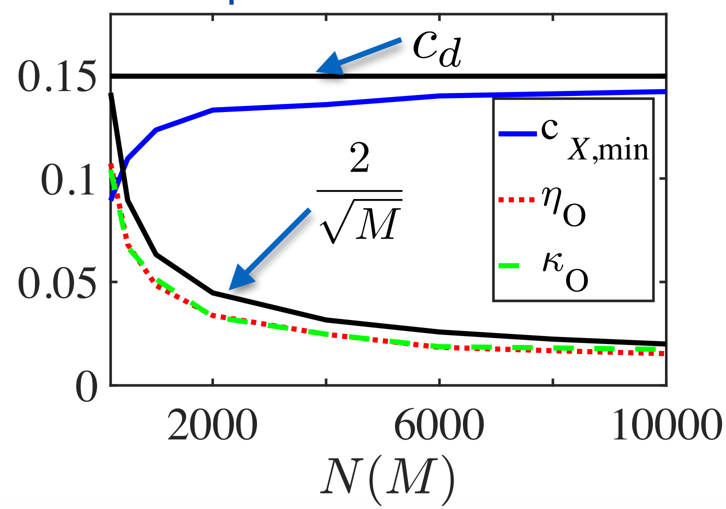
- Inliers are drawn uniformly from $\mathbb{S}^{D-1} \cap \mathcal{S}$

$$c_{\boldsymbol{X},\min} \gtrsim \frac{1}{\sqrt{d}}$$

- Outliers drawn uniformly from $\mathbb{S}^{D-1}$

$$\eta_{\boldsymbol{O}} \lesssim \frac{\sqrt{D}\log D}{\sqrt{M}} \qquad \kappa_{\boldsymbol{O}} \lesssim \frac{1}{\sqrt{M}}$$

Ambient dimension = 30
Subspace dimension = 29




[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

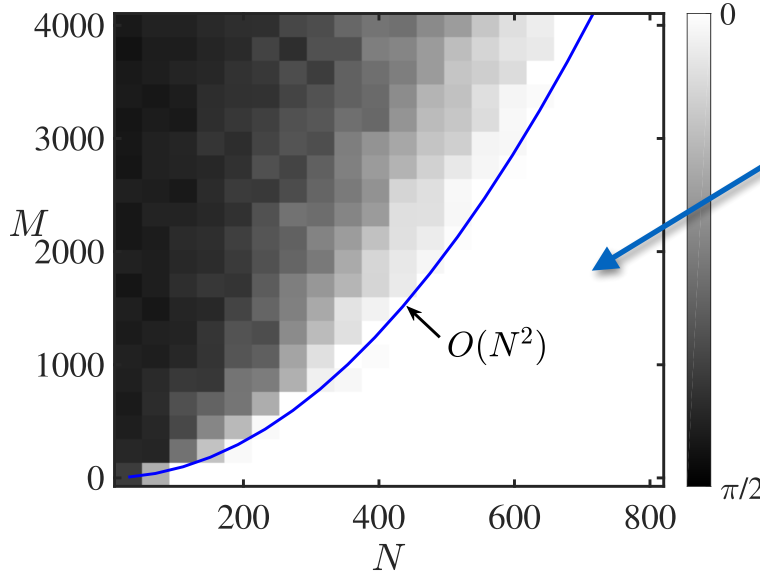
# Probabilistic Analysis

- **Theorem [1]:** Any global solution to the DPCP problem must lie in $\mathcal{S}^\perp$ with probability exceeding $1 - O(1/N^2)$ if

$$M \lesssim \frac{1}{d \max\left(D \log^2 D, \log N\right)} N^2$$

$D = 30, d = 29$

Principal angle of the solution to DPCP




Normal vector to the subspace

[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

# Probabilistic Analysis

- **Theorem [1]:** Any global solution to the DPCP problem must lie in $\mathcal{S}^\perp$ with probability exceeding $1 - O(1/N^2)$ if

$$M \lesssim \frac{1}{d \max\left(D \log^2 D, \log N\right)} N^2$$

- Comparison with other methods

| Geodesic Gradient Descent | REAPER [3] | L21-RPCA [2] | Coherence Pursuit [4] |
|---|---|---|---|
| $M \lesssim \frac{\sqrt{D(D-d)}}{d} N$ | $M \lesssim \frac{D}{d} N,$ $d \leq \frac{D-1}{2}$ | $M \lesssim \frac{1}{d \max\left(1, \frac{\log(M+N)}{d}\right)} N$ | $M \lesssim \frac{D-d^2}{d} N,$ $d < \sqrt{D}$ |


[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018
[2] Xu, Caramanis & Sanghavi, Robust PCA via outlier pursuit, 2010.
[3] Lerman, McCoy, Tropp & T. Zhang, Robust computation of linear models by convex relaxation, 2015.
[4] Rahmani & Atia, Coherence pursuit: Fast, simple, and robust principal component analysis, 2016.
[5] Manu, Zhang & Lerman, A well-tempered landscape for non-convex robust subspace recovery, 2017.
[6] Lerman & Maunu, An overview of robust subspace recovery, 2018.

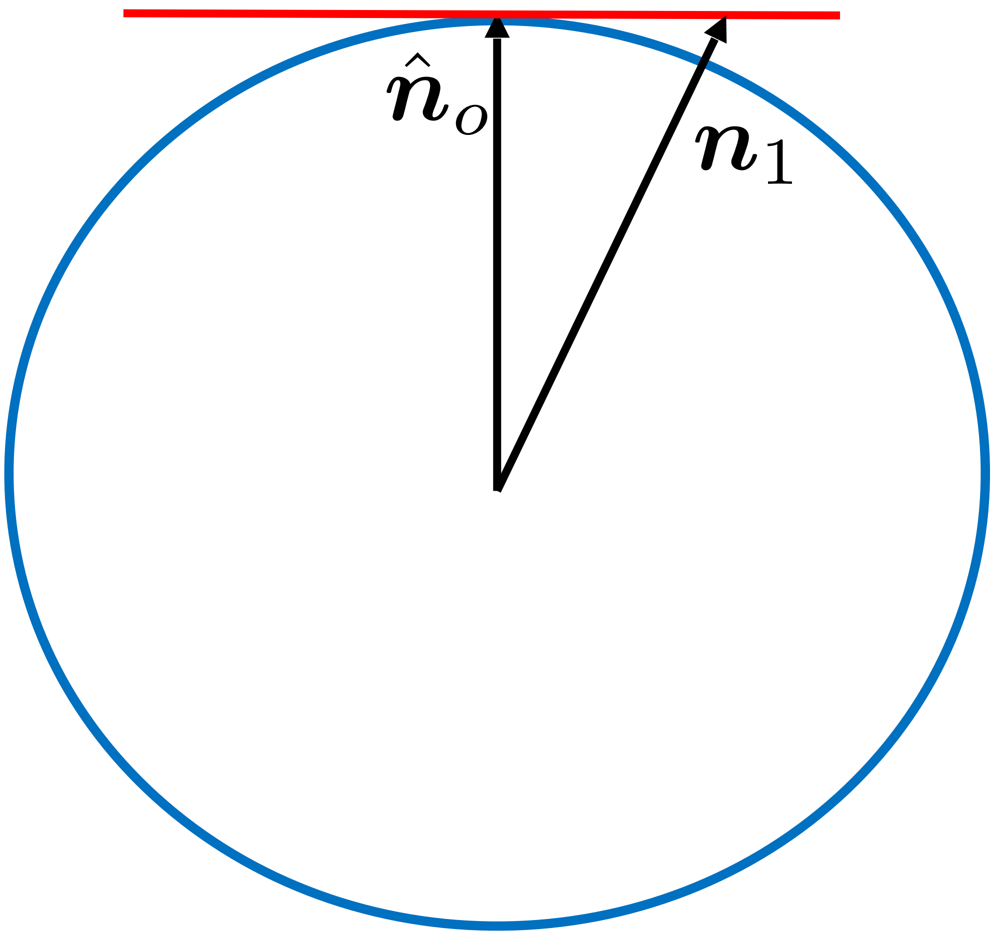
# Analysis of DPCP Algorithms

JHU vision lab

THE DEPARTMENT OF BIOMEDICAL ENGINEERING
The Whitaker Institute at Johns Hopkins
JOHNS HOPKINS
MATHEMATICAL INSTITUTE
for DATA SCIENCE

# Alternating Linear Programs (ALP)

$$\left\|\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\right\|_1$$



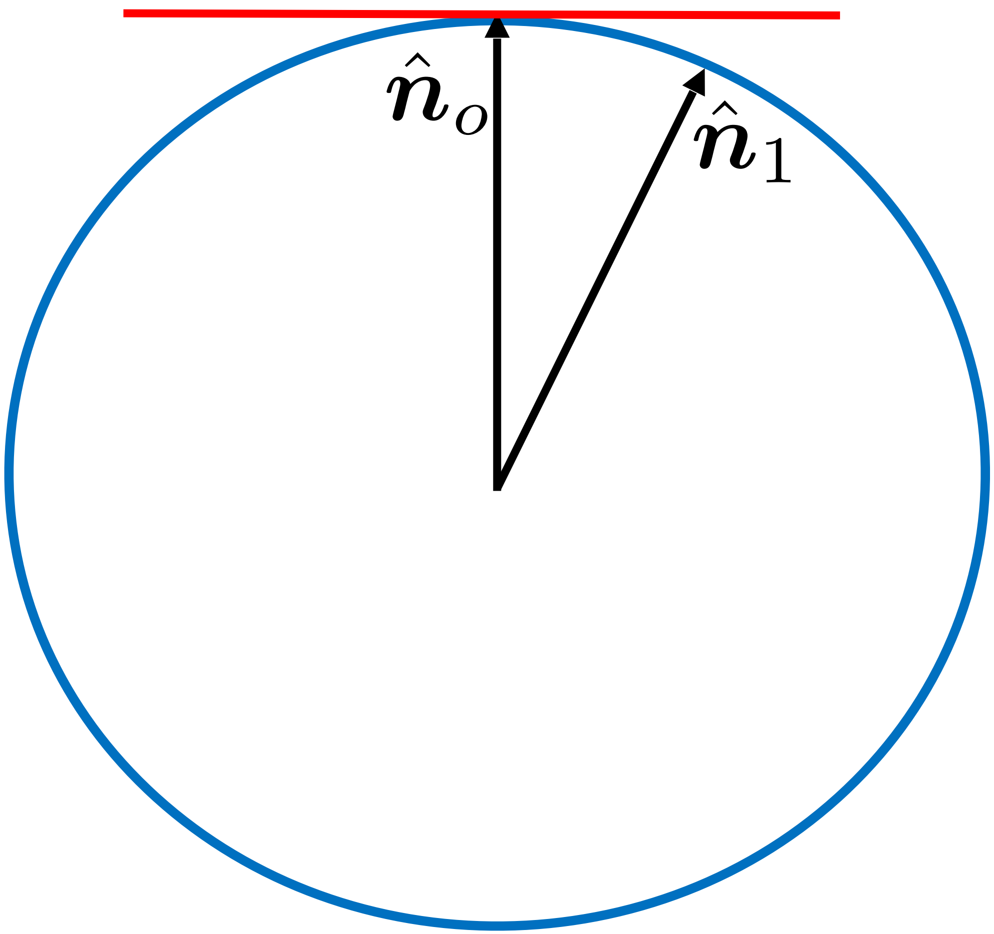

$$\boldsymbol{b}^\top \hat{\boldsymbol{n}}_o = 1$$

$$\boldsymbol{b}^\top \boldsymbol{b} = 1$$


[1] H. Spath and G.A. Watson. On orthogonal linear L1 approximation. Numerische Mathematik, 51(5):531–543, 1987.
[2] Tsakiris and Vidal, Dual Principal Component Pursuit, ICCVW 2015, JMLR 2018.

# Alternating Linear Programs (ALP)

$$\left\| \tilde{\boldsymbol{X}}^\top \boldsymbol{b} \right\|_1$$



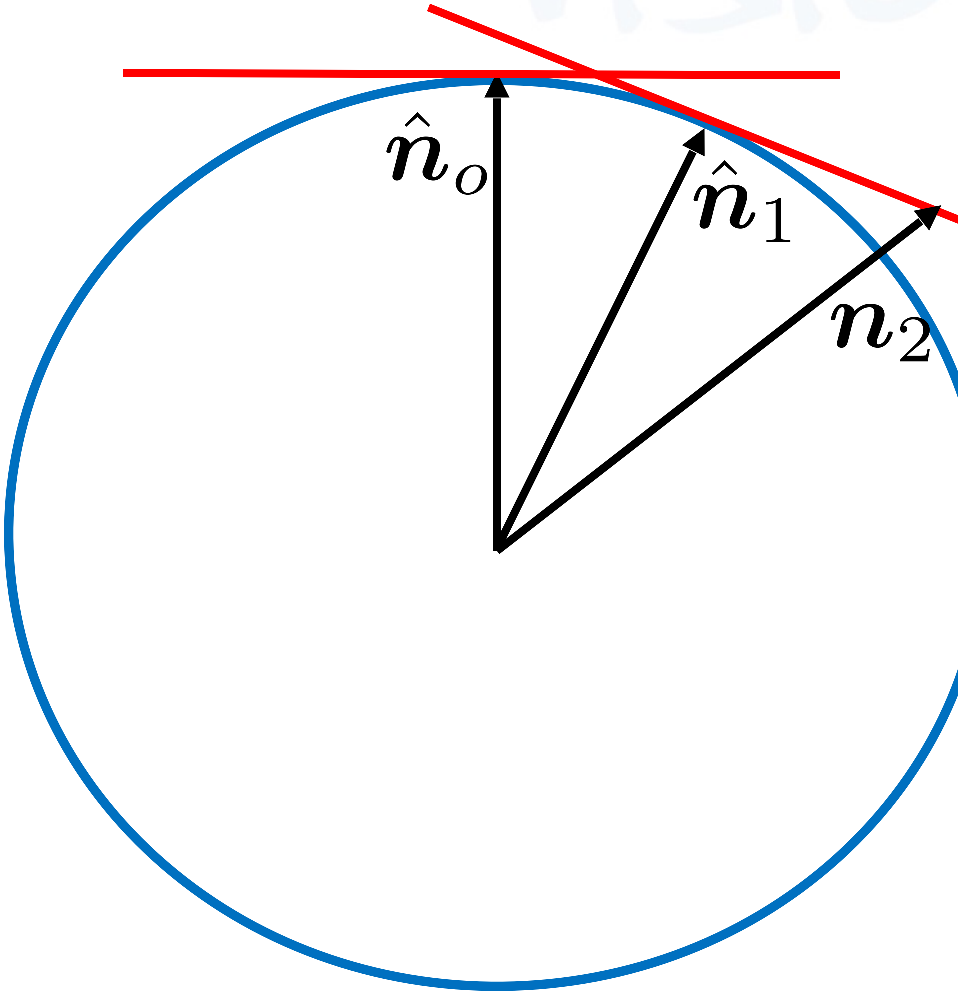

$$\boldsymbol{b}^\top \hat{\boldsymbol{n}}_o = 1$$

$$\boldsymbol{b}^\top \boldsymbol{b} = 1$$

[1] H. Spath and G.A. Watson. On orthogonal linear L1 approximation. Numerische Mathematik, 51(5):531–543, 1987.
[2] Tsakiris and Vidal, Dual Principal Component Pursuit, ICCVW 2015, JMLR 2018.

# Alternating Linear Programs (ALP)

$$\left\|\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\right\|_1$$




$$\boldsymbol{b}^\top \hat{\boldsymbol{n}}_o = 1$$

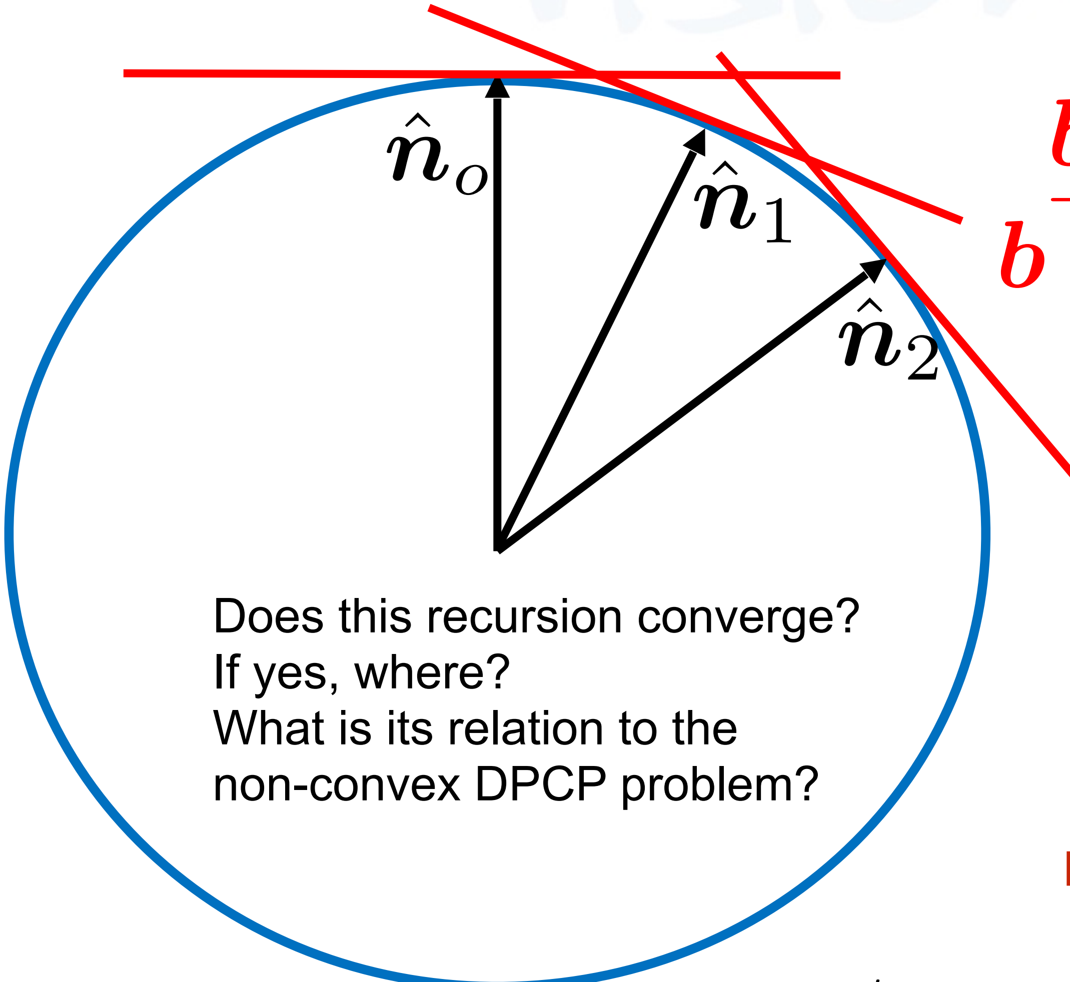
$$\boldsymbol{b}^\top \hat{\boldsymbol{n}}_1 = 1$$

$$\boldsymbol{b}^\top \boldsymbol{b} = 1$$


[1] H. Spath and G.A. Watson. On orthogonal linear L1 approximation. Numerische Mathematik, 51(5):531–543, 1987.
[2] Tsakiris and Vidal, Dual Principal Component Pursuit, ICCVW 2015, JMLR 2018.

# Alternating Linear Programs (ALP)

$$\left\| \tilde{\boldsymbol{X}}^\top \boldsymbol{b} \right\|_1$$




Does this recursion converge?
If yes, where?
What is its relation to the non-convex DPCP problem?

$$\boldsymbol{b}^\top \boldsymbol{b} = 1$$

DPCP-relaxed

$$\boldsymbol{n}_{k+1} \leftarrow \min_{\boldsymbol{b}^\top \hat{\boldsymbol{n}}_k = 1} \left\| \tilde{\boldsymbol{X}}^\top \boldsymbol{b} \right\|_1$$

[1] H. Spath and G.A. Watson. On orthogonal linear L1 approximation. Numerische Mathematik, 51(5): 531–543, 1987.
[2] Tsakiris and Vidal, Dual Principal Component Pursuit, ICCVW 2015, JMLR 2018.

# ALP Converges in Finite #Iterations

$$\boldsymbol{n}_{k+1} \leftarrow \min_{\boldsymbol{b}^\top \hat{\boldsymbol{n}}_k = 1} \left\| \tilde{\boldsymbol{X}}^\top \boldsymbol{b} \right\|_1$$

- **Theorem:** If the initialization is sufficiently far from the inlier hyperplane, then the recursion of L1 problems converges in a finite number of iterations to the normal of the inlier hyperplane.

**Finite convergence to the global minimum of the non-convex DPCP problem**

[1] Tsakiris and Vidal, Dual Principal Component Pursuit, ICCVW 2015, JMLR 2018.

# Iteratively Reweighted Least-Squares (IRLS)

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

$$\min_{\boldsymbol{b}} \sum_{j=1}^{L} \left| \tilde{\boldsymbol{x}}^\top \boldsymbol{b} \right|, \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

$$\min_{\boldsymbol{b}} \sum_{j=1}^{L} \frac{1}{\left| \tilde{\boldsymbol{x}}_j^\top \boldsymbol{b} \right|} \left| \tilde{\boldsymbol{x}}_j^\top \boldsymbol{b} \right|^2, \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

$$\min_{\boldsymbol{b}} \sum_{j=1}^{L} \frac{1}{\left| \tilde{\boldsymbol{x}}_j^\top \hat{\boldsymbol{n}}_0 \right|} \left| \tilde{\boldsymbol{x}}_j^\top \boldsymbol{b} \right|^2, \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1$$

$\hat{\boldsymbol{n}}_1$ solution via SVD

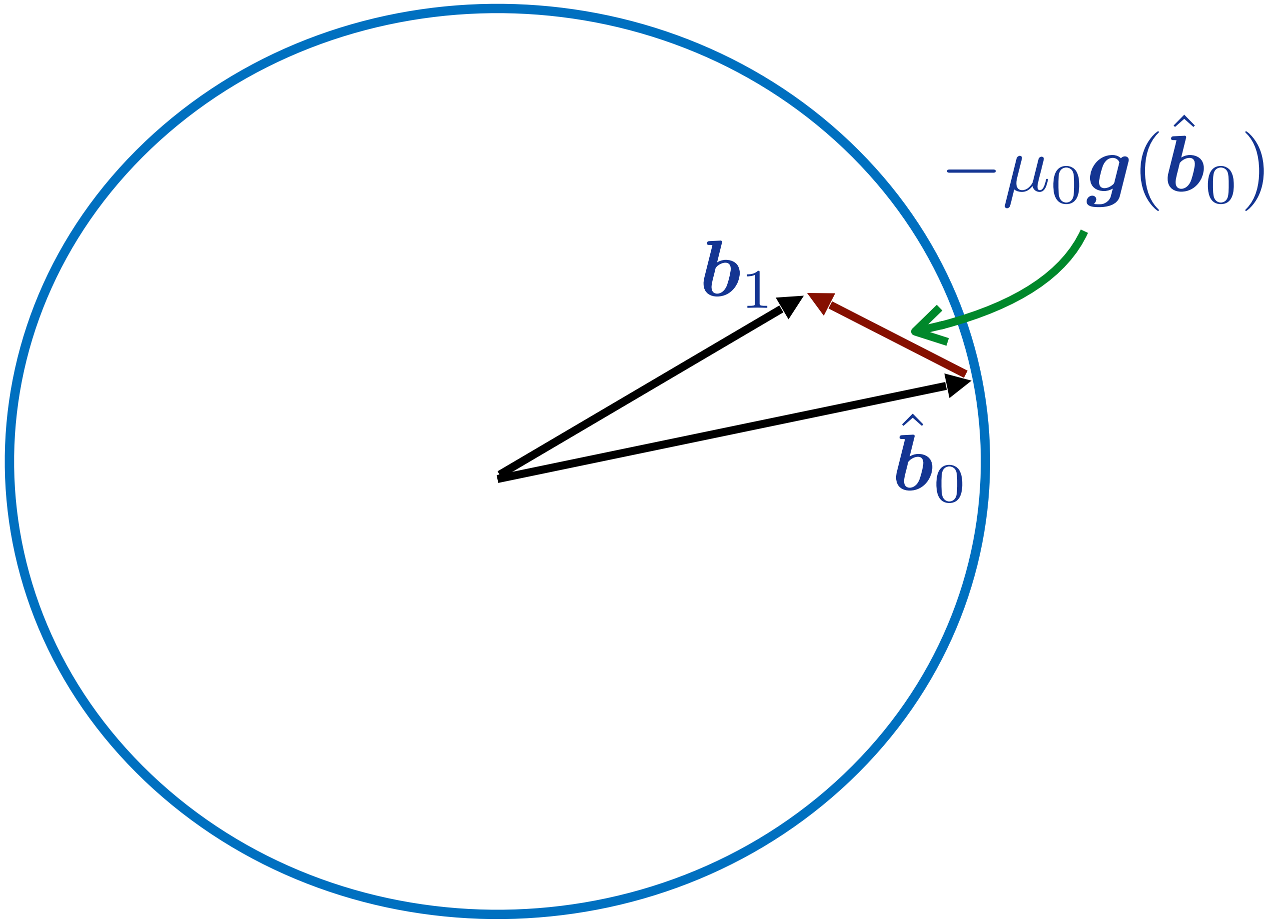
DPCP-IRLS

1. Efficient: 6,000 points of 1000 dimensions, in 1 min
2. Strong experimental evidence
3. Need to establish convergence theory


Daubechies et al., Iteratively reweighted least-squares for sparse recovery, 2010
Chartrand & Yin et al., Iteratively reweighted algorithms for compressive sensing, 2008

# Projected Sub-Gradient Method (PSGM)

- Subgradient of $\left\|\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\right\|_1$ is $\boldsymbol{g}(\boldsymbol{b}) = \tilde{\boldsymbol{X}}\operatorname{sign}\left(\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\right)$

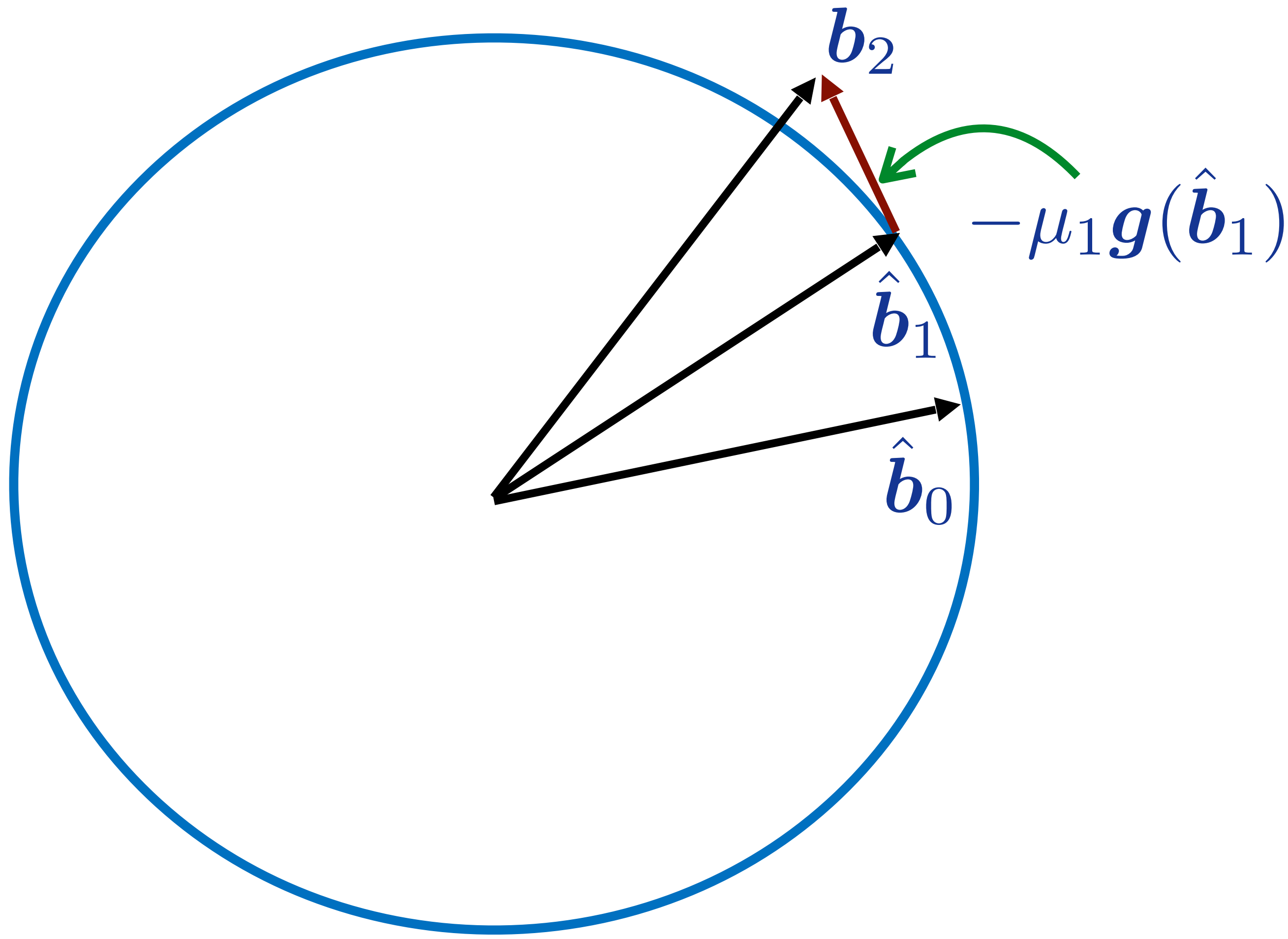



$\boldsymbol{b}^\top \boldsymbol{b} = 1$


[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

# Projected Sub-Gradient Method (PSGM)

- Subgradient of $\left\|\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\right\|_1$ is $\boldsymbol{g}(\boldsymbol{b}) = \tilde{\boldsymbol{X}}\operatorname{sign}\left(\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\right)$

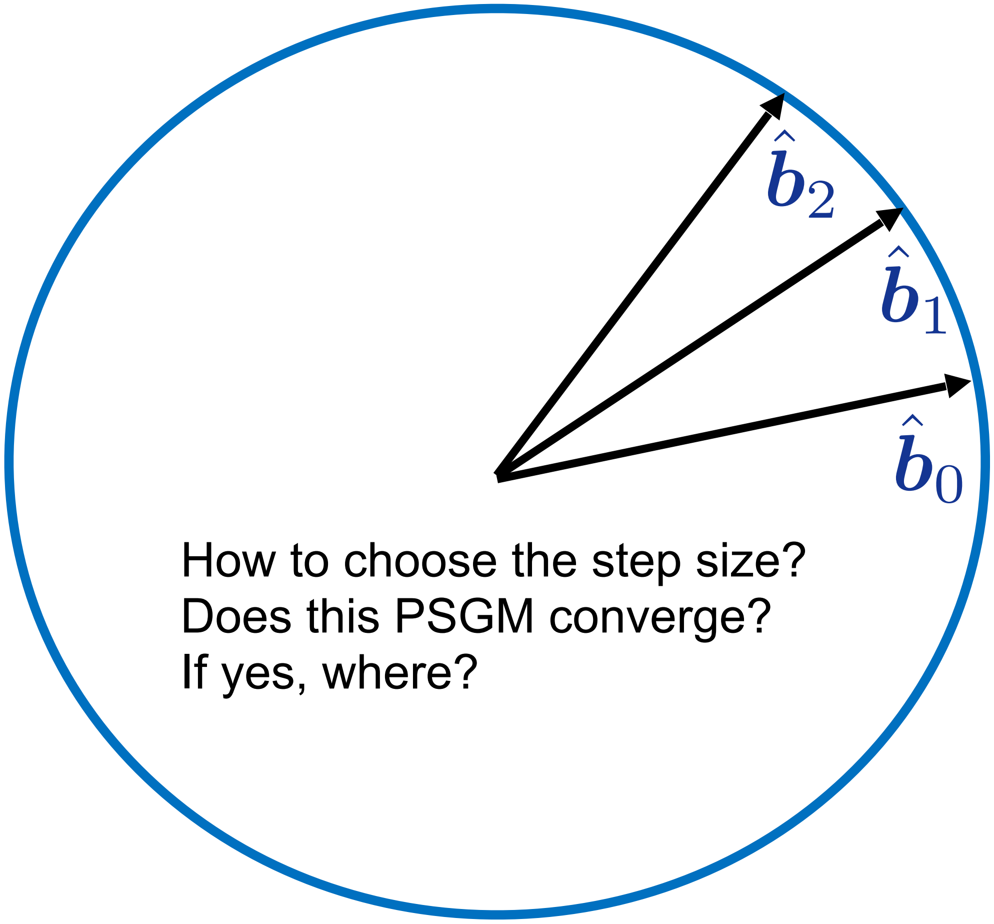



$\boldsymbol{b}^\top \boldsymbol{b} = 1$


[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

# Projected Sub-Gradient Method (PSGM)

- Subgradient of $\|\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\|_1$ is $\boldsymbol{g}(\boldsymbol{b}) = \tilde{\boldsymbol{X}}\,\mathrm{sign}(\tilde{\boldsymbol{X}}^\top \boldsymbol{b})$



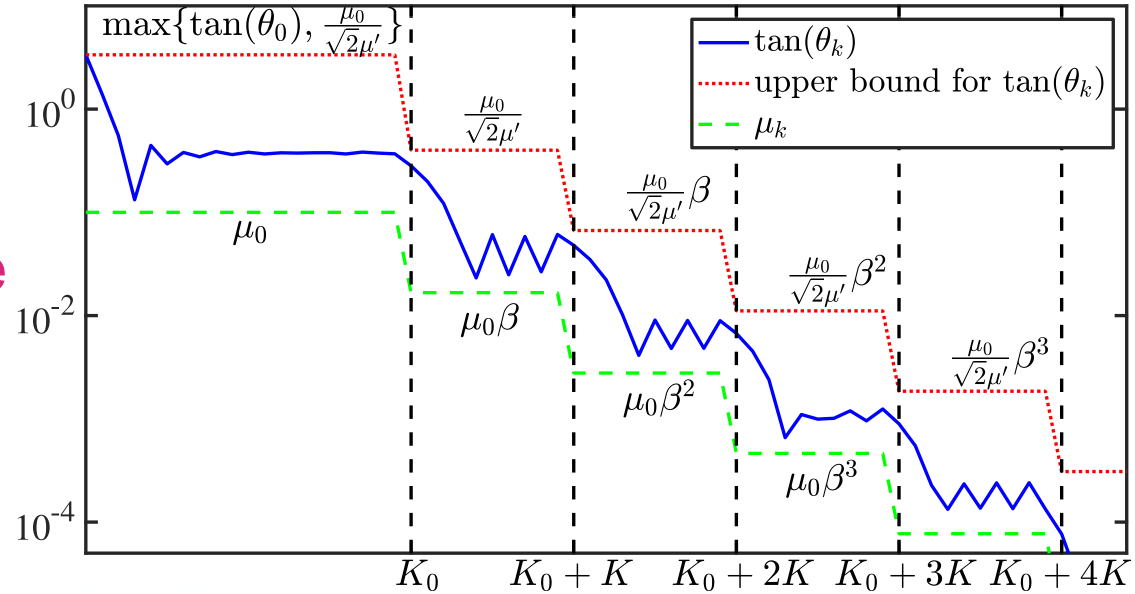

$$\boldsymbol{b}^\top \boldsymbol{b} = 1$$

[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

# Projected Sub-Gradient Method (PSGM)

- **Spectral initialization**: $\hat{\boldsymbol{b}}_0 = \mathrm{argmin}_{\boldsymbol{b}\in\mathbb{S}^{D-1}} \left\|\tilde{\boldsymbol{X}}^\top \boldsymbol{b}\right\|_2$
- **PSGM update**: $\boldsymbol{b}_{k+1} \leftarrow \hat{\boldsymbol{b}}_k - \mu_k \tilde{\boldsymbol{X}} \mathrm{sign}(\tilde{\boldsymbol{X}}^\top \hat{\boldsymbol{b}}_k)$; $\hat{\boldsymbol{b}}_{k+1} \leftarrow \frac{\boldsymbol{b}_{k+1}}{\|\boldsymbol{b}_{k+1}\|}$
- **Step size**: piecewise geometrically diminishing $\mu_k = \mu_0 \beta^{\lfloor (k-K_0)/K \rfloor}$
- **Theorem**: If $\theta_0 < \theta^\natural$, PSGM has a piecewise linear convergence rate for $\tan(\theta_k)$, i.e.,

$$\tan(\theta_k) \lesssim \beta^{\lfloor (k-K_0)/K \rfloor}$$




[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

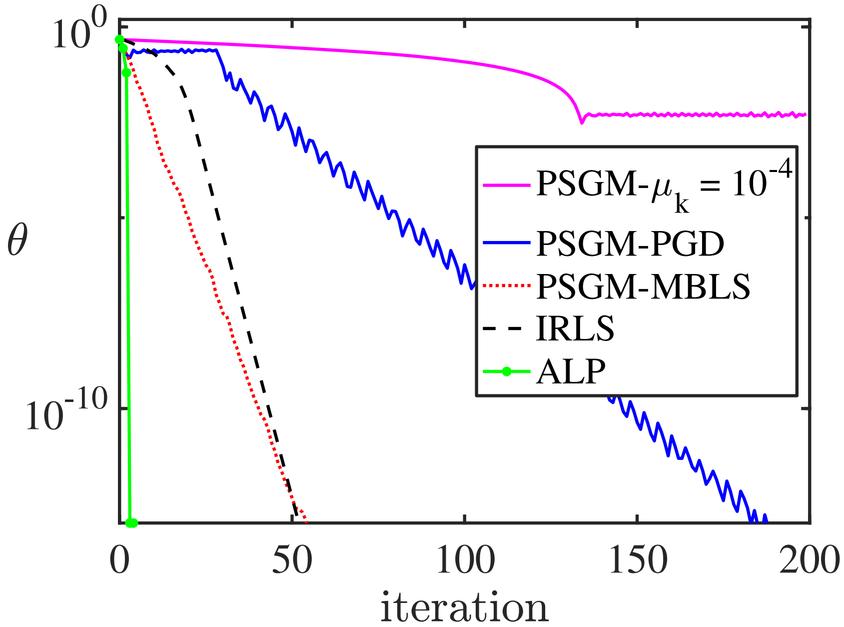
# Comparison of DPCP Algorithms

- ALP: Alternating Linear Projection
- IRLS: Iterative Reweighted Least Squares
- PSGM: Projected Sub Gradient Method

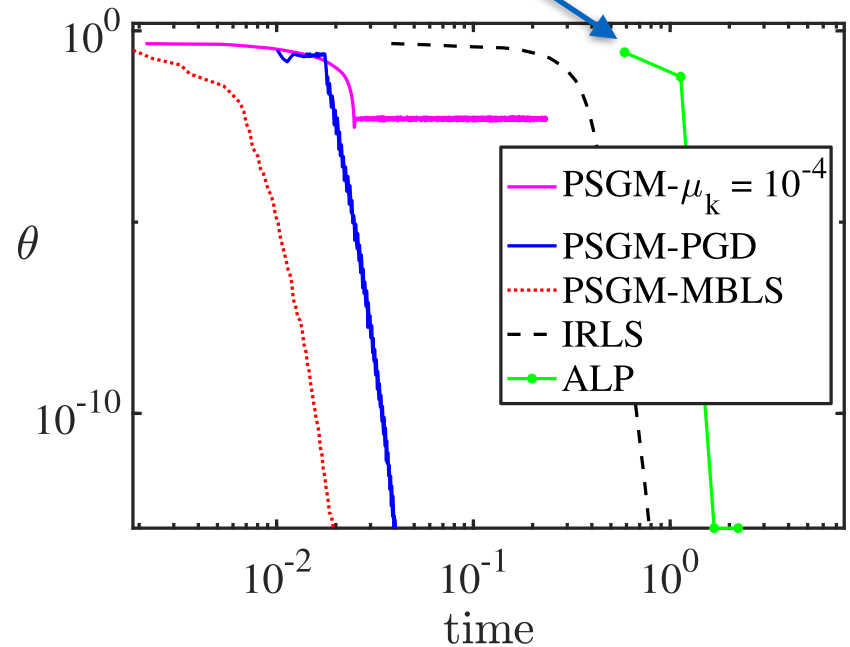
| | Convergence guarantee | Scalable (#points) |
|---|---|---|
| ALP | ✓ | ✗ $(\leq 10^3)$ |
| IRLS | ✗ | ✓ $(\leq 10^4)$ |
| PSGM | ✓ | ✓ $(\leq 10^5)$ |


[1] Tsakiris and Vidal, Dual Principal Component Pursuit, ICCVW 2015, JMLR 2018.
[2] M. C. Tsakiris and R. Vidal, Hyperplane Clustering via Dual Principal Component Pursuit, ICML 2017
[3] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018

# Comparison of DPCP Algorithms




Solving one time of linear programming

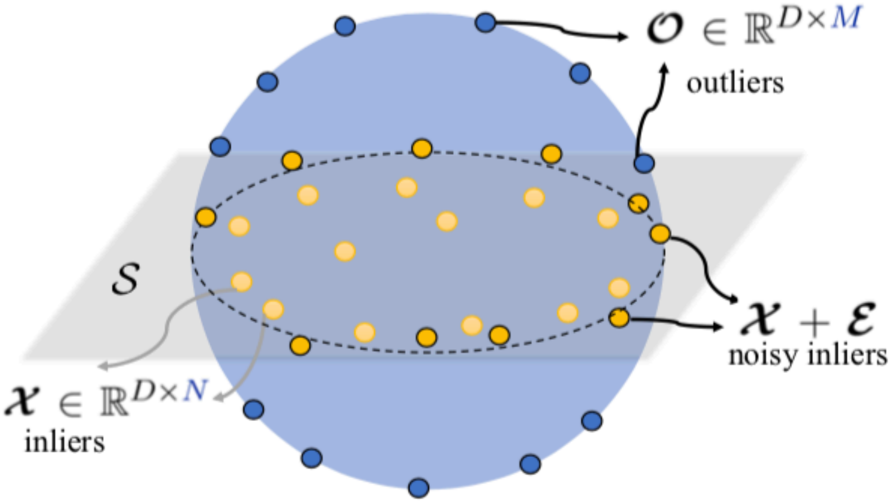



- PSGM-PGD: Piecewise Geometrically Diminishing
- PSGM-MBLS: Modified Backtracking Line Search

Ambient dimension = 30, Subspace dimension = 29
Number of inliers (N) = 500, Outlier ration M/(M+N) = 0.7

# Extensions of DPCP

# Noisy DPCP: Problem Formulation



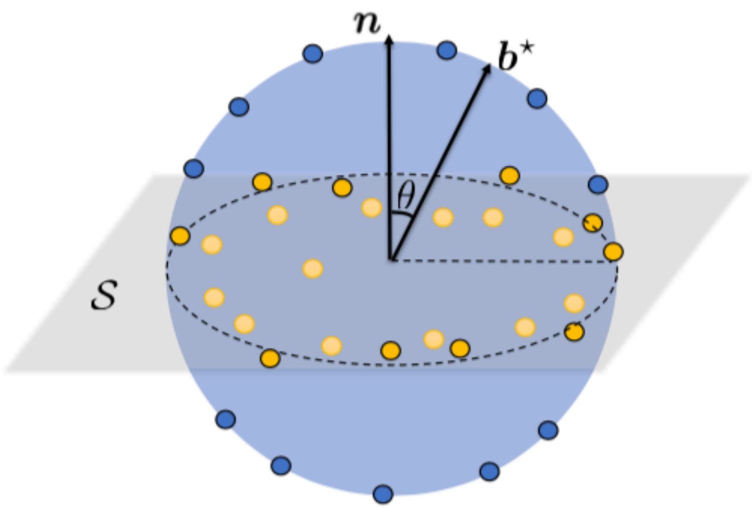

- outliers $\mathcal{O}$ are drawn uniformly random from the unit sphere
- noisy inliers $\mathcal{X} + \mathcal{E}$
  - inliers $\mathcal{X}$ are drawn within the $\mathcal{S}$
  - add Gaussian nose $\mathcal{E}$ on the inliers
  - normalize $\mathcal{X} + \mathcal{E}$ to unit norm (on sphere)
- SNR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

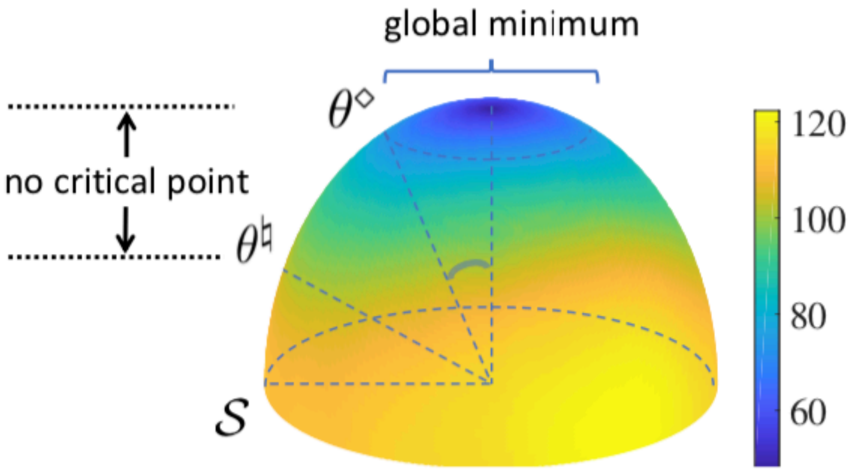
$$\min_{\boldsymbol{b}} \left\|\widetilde{\mathcal{X}}^\top \boldsymbol{b}\right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{b}\|_2 = 1 \tag{1}$$

T Ding, Z Zhu, T Ding, Y Yang, DP Robinson, MC Tsakiris, R Vidal. Noisy Dual Principal Component Pursuit. ICML, 2019

# Noisy DPCP: Geometric 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, $\boldsymbol{b}^\star = \boldsymbol{n}$


T Ding, Z Zhu, T Ding, Y Yang, DP Robinson, MC Tsakiris, R Vidal. Noisy Dual Principal Component Pursuit. ICML, 2019

# Noisy DPCP: Statistical Analysis

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

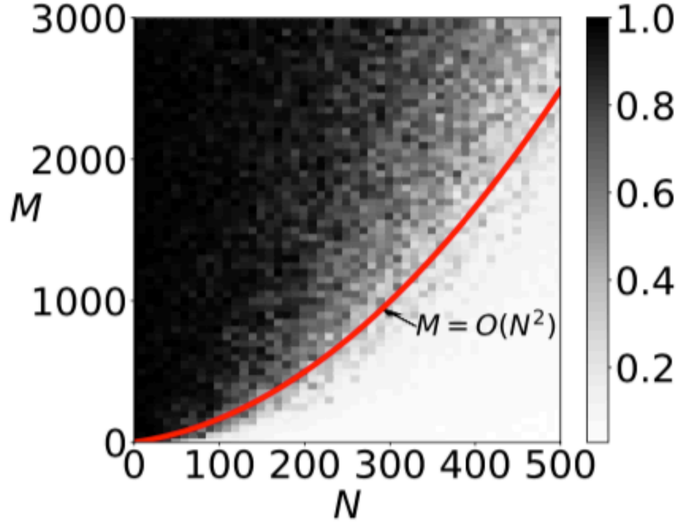
**Theorem**

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

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

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

# Noisy DPCP: PSGM

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

- Spectral initialization:

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

- Piecewise diminishing stepsize:

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

- PSG update:

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

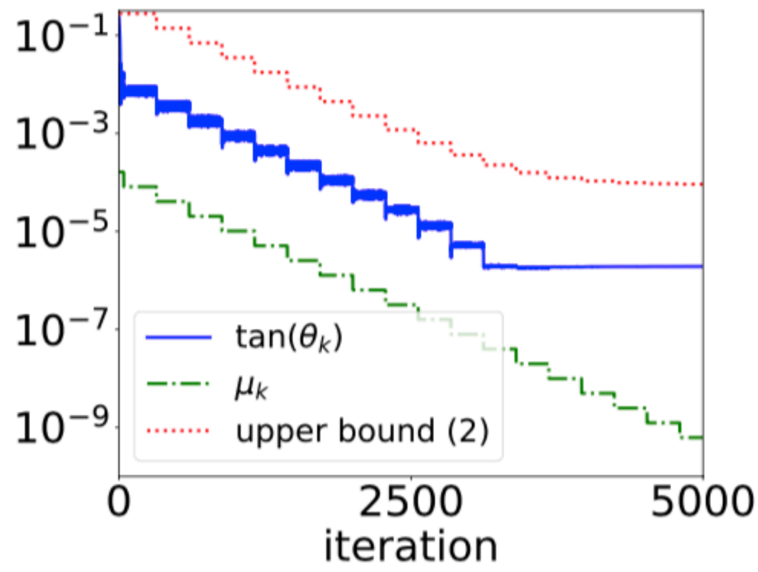
- Projection:

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

**Theorem**

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

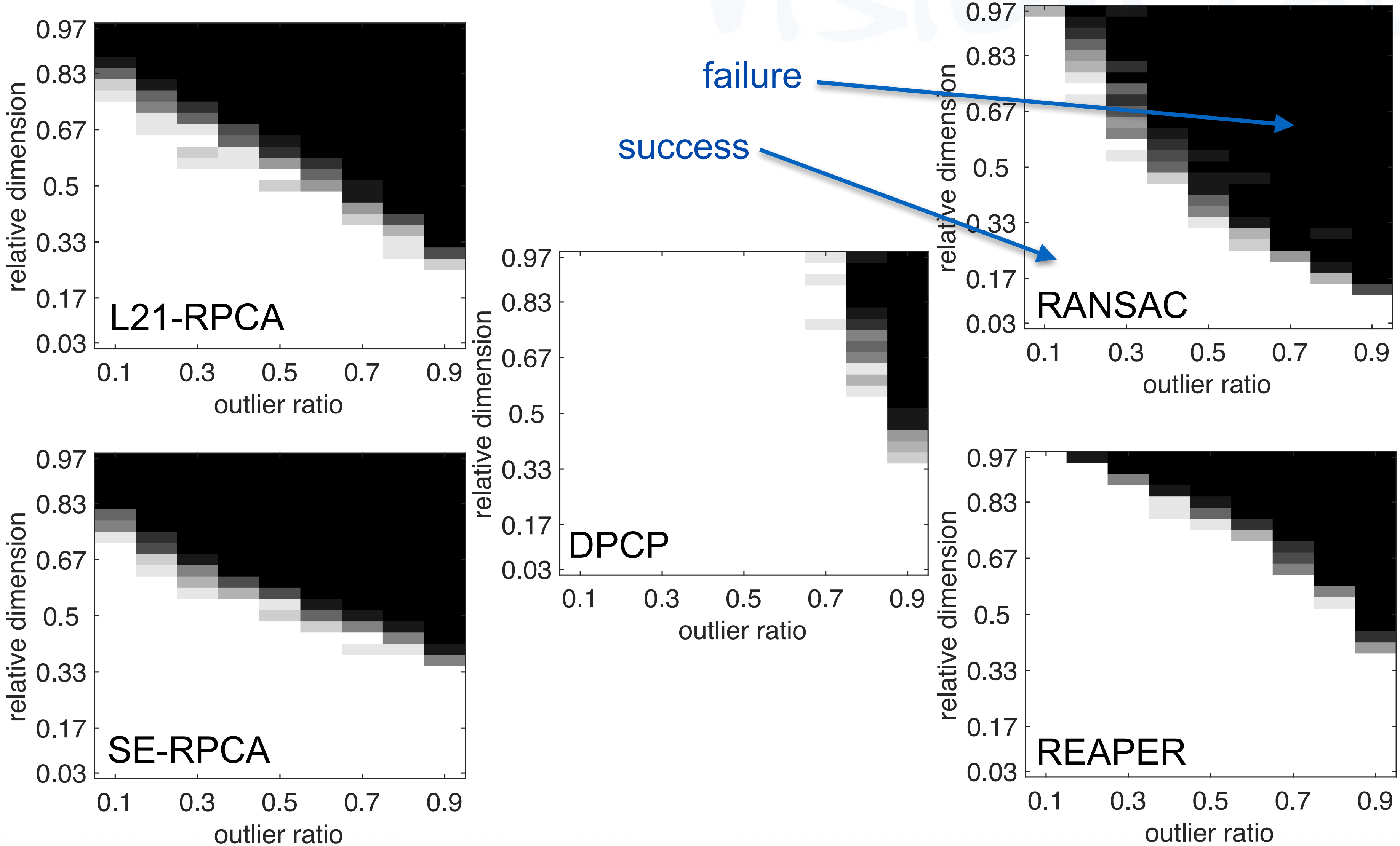
$$\tan(\theta_k) \lesssim \beta^{\lfloor (k-K_0)/K \rfloor} + \frac{\sqrt{\sigma}}{\sqrt{1-2\sqrt{\sigma}}}. \qquad (2)$$




T Ding, Z Zhu, T Ding, Y Yang, DP Robinson, MC Tsakiris, R Vidal. Noisy Dual Principal Component Pursuit. ICML, 2019

# Experiments

# Dual Principal Component Analysis



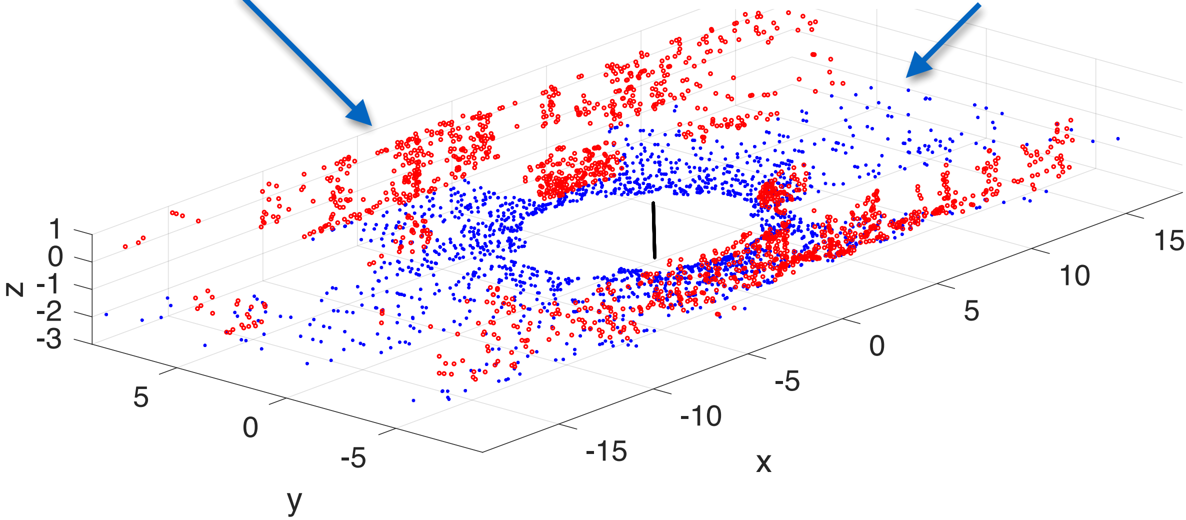

Ambient dimension (D) = 30    Number of inliers (N) = 300


JOHNS HOPKINS
MATHEMATICAL INSTITUTE
for DATA SCIENCE

# Experiments on 3D Point Cloud Road Data

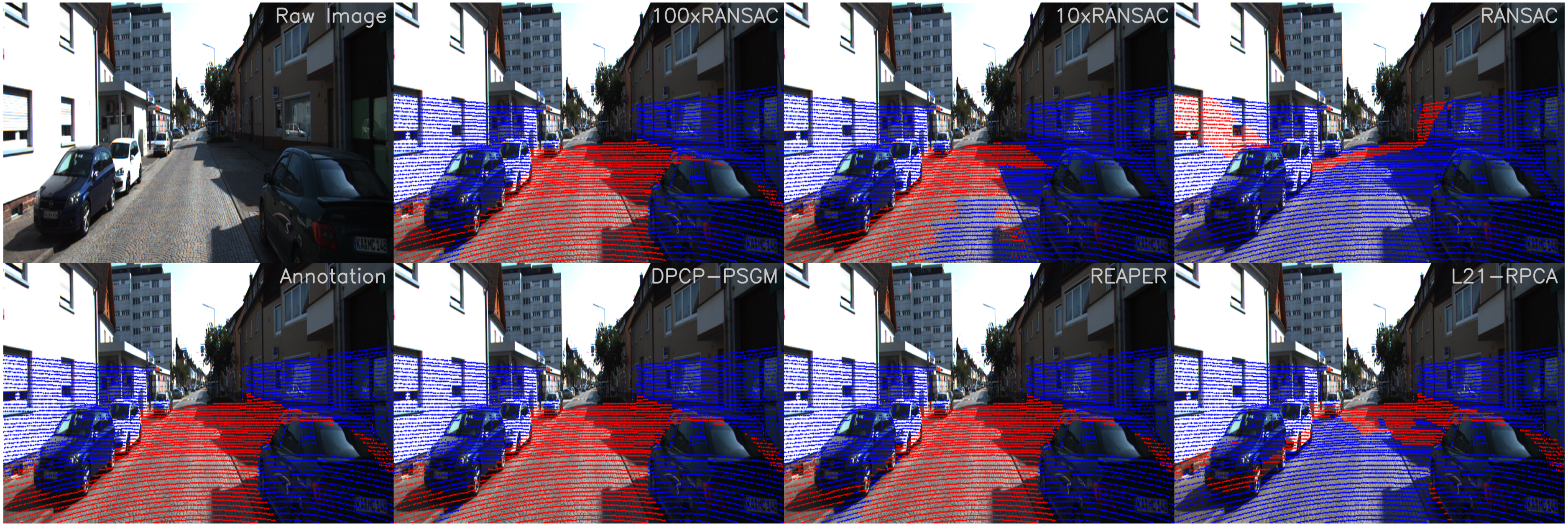

Outliers (53%)

Inliers




Geiger, Lenz, Stiller & Urtasun, Vision meets robotics: The KITTI dataset, 2013.

# Experiments on 3D Point Cloud Road Data




F1 measures: DPCP-PSGM (0.933), REAPER (0.890), L21-RPCA (0.248), RANSAC (0.023), 10xRANSAC (0.622), and 100xRANSAC (0.824)

- Each point cloud has ~ 100,000 points including 50% outliers


[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018
[2] Geiger, Lenz, Stiller & Urtasun, Vision meets robotics: The KITTI dataset, 2013.

# Experiments on 3D Point Cloud Road Data

- Area under ROC

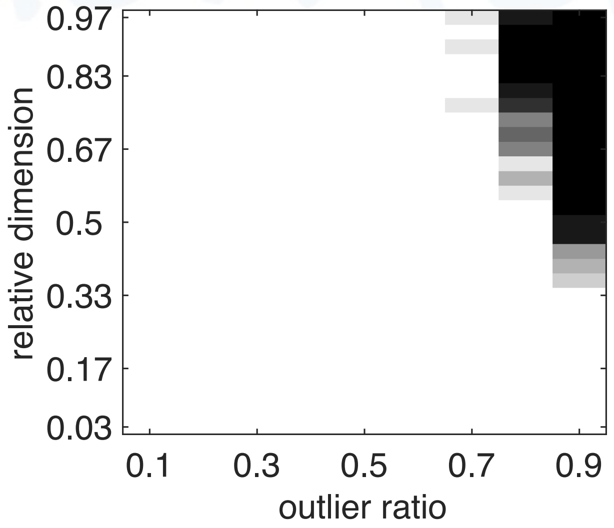
Percentage of outliers

| Methods | KITTY-CITY-5 1(37%) | 45(38%) | 120(53%) | 137(48%) | 153(67%) | KITTY-CITY-48 0(56%) | 21(57%) |
|---|---|---|---|---|---|---|---|
| DPCP-PSGM | **0.998** | **0.999** | 0.868 | **1.000** | 0.749 | **0.994** | **0.991** |
| REAPER | 0.998 | 0.998 | 0.839 | 0.999 | 0.749 | **0.994** | 0.982 |
| $\ell_{2,1}$-RPCA | 0.841 | 0.953 | 0.610 | 0.925 | 0.575 | 0.836 | 0.837 |
| RANSAC | 0.596 | 0.592 | 0.569 | 0.551 | 0.521 | 0.534 | 0.531 |
| 10xRANSAC | 0.911 | 0.773 | 0.717 | 0.654 | 0.624 | 0.757 | 0.598 |
| 100xRANSAC | 0.991 | 0.983 | **0.965** | 0.955 | **0.849** | 0.974 | 0.902 |

[1] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, Dual Principal Component Pursuit: Improved Analysis and Efficient Algorithms, NIPS 2018
[2] Geiger, Lenz, Stiller & Urtasun, Vision meets robotics: The KITTI dataset, 2013.

# Conclusions

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




- DPCP: Nonconvex L1 formulation for robust subspace learning
- Works well for both low and high relative dimensions
- Global minimum is orthogonal to the subspace
- Tolerate as many outliers as the square of the number of inliers
- PSGM finds global minimum at a piecewise linear rate


[1] Tsakiris and Vidal, Dual Principal Component Pursuit, ICCVW 2015, JMLR 2018.
[2] Tsakiris and Vidal, Hyperplane Clustering via Dual Principal Component Pursuit, ICML 2017.
[3] Zhu, Wang, Robinson, Naiman, Vidal and Tsakiris, DPCP: Improved Analysis and Efficient Algorithms, NeurIPS 2018.
[4] Zhu, Ding, Robinson, Tsakiris, Vidal. A linearly convergent method for non-smooth non-convex optimization on the Grassmannian. NeurIPS, 2019.
[5] Ding, Zhu, Ding, Yang, Robinson, Tsakiris, Vidal. Noisy Dual Principal Component Pursuit. ICML 2019.
[6] Ding, Yang, Zhu, Robinson, Vidal, Kneip, Tsakiris. Robust Homography Estimation via Dual Principal Component Pursuit, CVPR 2020.

# Acknowledgements

Manolis Tsakiris
ShanghaiTech

Zhihui Zhu
U Denver

Tianyu Ding
JHU

Daniel Robinson
Lehigh

Rene Vidal
JHU

Research supported by NSF grant 1704458

**Thank You!**