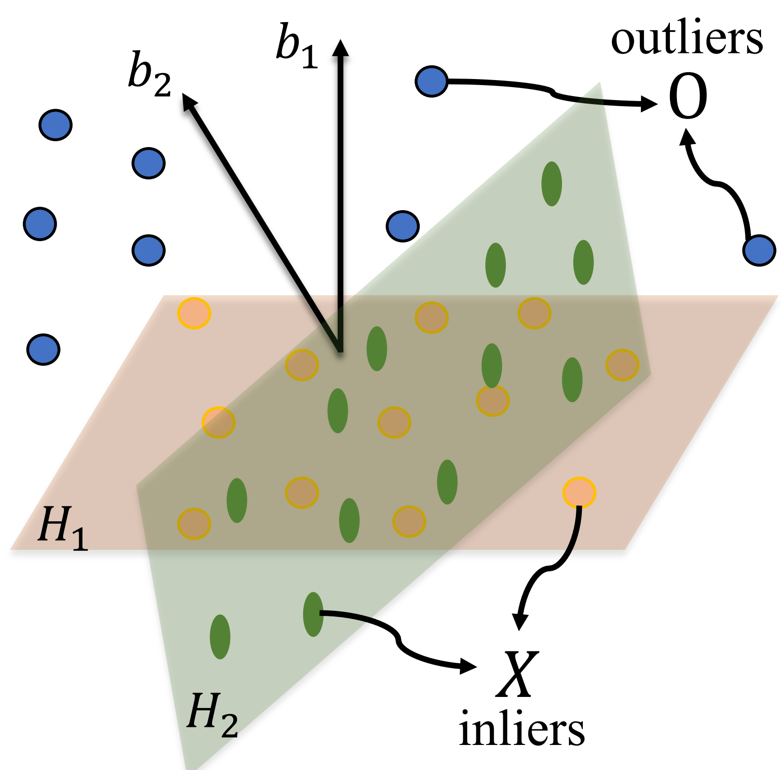
Dual Principal Component Pursuit for Learning a Union of Hyperplanes: Theory and Algorithms

Tianyu Ding, Zhihui Zhu, Manolis Tsakiris, Rene Vidal, Daniel Robinson

AISTATS 2021

Motivation

Problem: Clustering data points to a union of hyperplanes (UoH)



- Clustering low dimensional subspaces
  - Self-expressive methods [Elhamifar & Vidal 13]
  - Low-rank methods [Liu et al. 10]

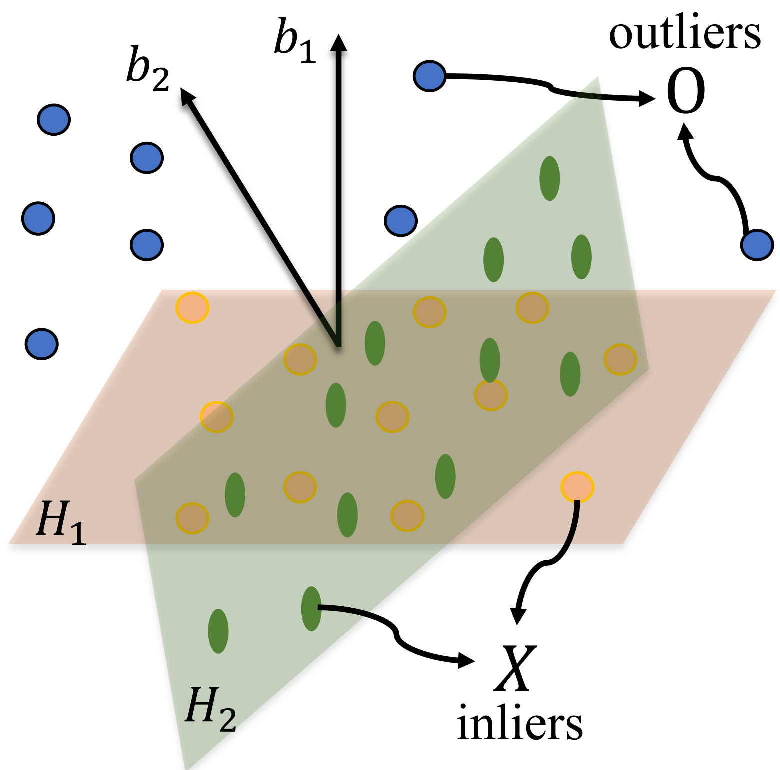
The theory and algorithms
do not apply to a UoH setting

- Clustering a union of hyperplanes
  - DPCP [Tsakiris & Vidal 17]

The analysis is difficult to interpret, and it lacks
a scalable algorithm with convergence guarantee

**We provide both geometric and probabilistic analysis;
and a scalable algorithm with linear convergence guarantee.**

Dual Principal Component Pursuit (DPCP)



- $X \in \mathbb{R}^{D \times N}$ are N inlier points that lie in $\cup_{k=1}^{K} H_k$, each of which has unit normal vectors $\mathbf{b}_k$
- $O \in \mathbb{R}^{D \times M}$ are M outlier points that lie in $\mathbb{R}^D$
- Dataset: $\tilde{X} = [X, O]$

DPCP problem formulation

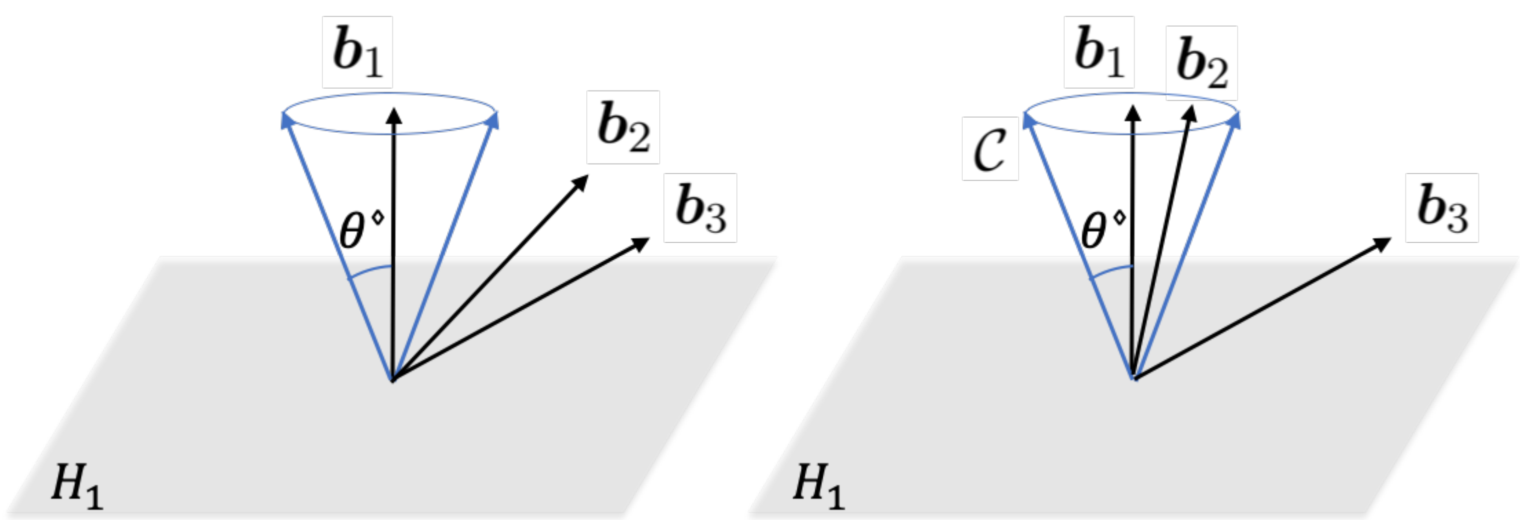
$$\min_{\mathbf{b}:\|\mathbf{b}\|_2=1} f(\mathbf{b}) := \|\tilde{X}^\top \mathbf{b}\|_1$$

Geometric Global Optimality Analysis

$$\min_{\mathbf{b}:\|\mathbf{b}\|_2=1} f(\mathbf{b}) := \|\tilde{X}^\top \mathbf{b}\|_1$$

- Informally, H_1 is a geometrically dominant hyperplane if:
  - it has a large enough number of points
  - the data points are well-distributed
  - the other hyperplanes are well-separated from each other

- **Lemma:** Any critical point is either a normal vector of H_1 or close to H_1.



- **Theorem:** Any global solution is a normal vector of H_1 given it is sufficiently geometrically dominant.

Probabilistic Analysis

$$\min_{\mathbf{b}:\|\mathbf{b}\|_2=1} f(\mathbf{b}) := \|\tilde{X}^\top \mathbf{b}\|_1$$

Consider a random spherical model:

- M outliers are drawn uniformly from the unit sphere $\mathbb{S}^{D-1}$ in $\mathbb{R}^D$
- N_k inliers to H_k are drawn uniformly from $H_k \cap \mathbb{S}^{D-1}$ with $N_1 + \cdots + N_K = N$

- **Theorem:** Any global solution must be a normal vector of H_1 with high probability if

$$M \leq C \cdot \left(N_1 - \sum_{k \neq 1} N_k\right)^2$$

- Our analysis allows a sampling of only $N + M = \Omega(D^3)$ points to establish a high probable recovery while [Lerman & Zhang] need a sampling of $\Omega(D^{18} \log D)$ points

Projected Riemannian SubGradient Method

$$\min_{\mathbf{b}:\|\mathbf{b}\|_2=1} f(\mathbf{b}) := \|\tilde{X}^\top \mathbf{b}\|_1$$

Initialization: $\tilde{X}$, $\{\mu_0, \beta\} \subseteq (0,1)$ and $k \leftarrow 0$

Spectral initialization: set $\widehat{\mathbf{b}}_0 \leftarrow \arg \min_{\|\mathbf{b}\|_2=1} \|\tilde{X}^\top \mathbf{b}\|_2$

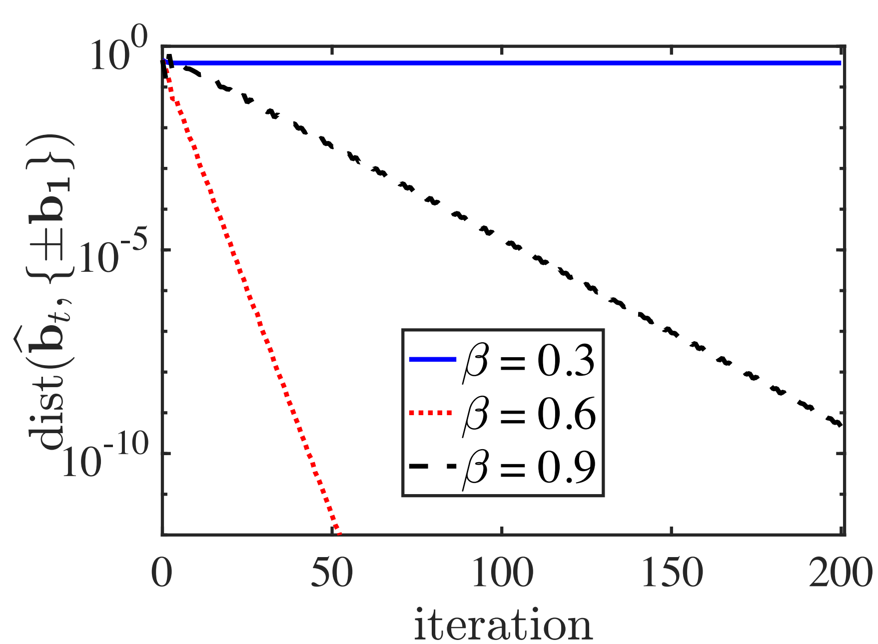
Geometrically diminishing step size: $\mu_t \leftarrow \mu_0 \beta^t$

Compute a Riemannian subgradient: $\mathcal{G}(\widehat{\mathbf{b}}_t) \leftarrow \left(\mathbf{I} - \widehat{\mathbf{b}}_t \widehat{\mathbf{b}}_t^\top\right) \tilde{X} \operatorname{sign}(\tilde{X}^\top \widehat{\mathbf{b}}_t)$

Update the iterate as: $\widetilde{\mathbf{b}}_{t+1} \leftarrow \widehat{\mathbf{b}}_t - \mu_t \mathcal{G}(\widehat{\mathbf{b}}_t)$; $\widehat{\mathbf{b}}_{t+1} \leftarrow \widetilde{\mathbf{b}}_{t+1} / \|\widetilde{\mathbf{b}}_{t+1}\|_2$

- **Theorem:** The iterates converge linearly:

$$\operatorname{dist}\left(\widehat{\mathbf{b}}_t, \{\pm \mathbf{b}_1\}\right) \lesssim \beta^k$$

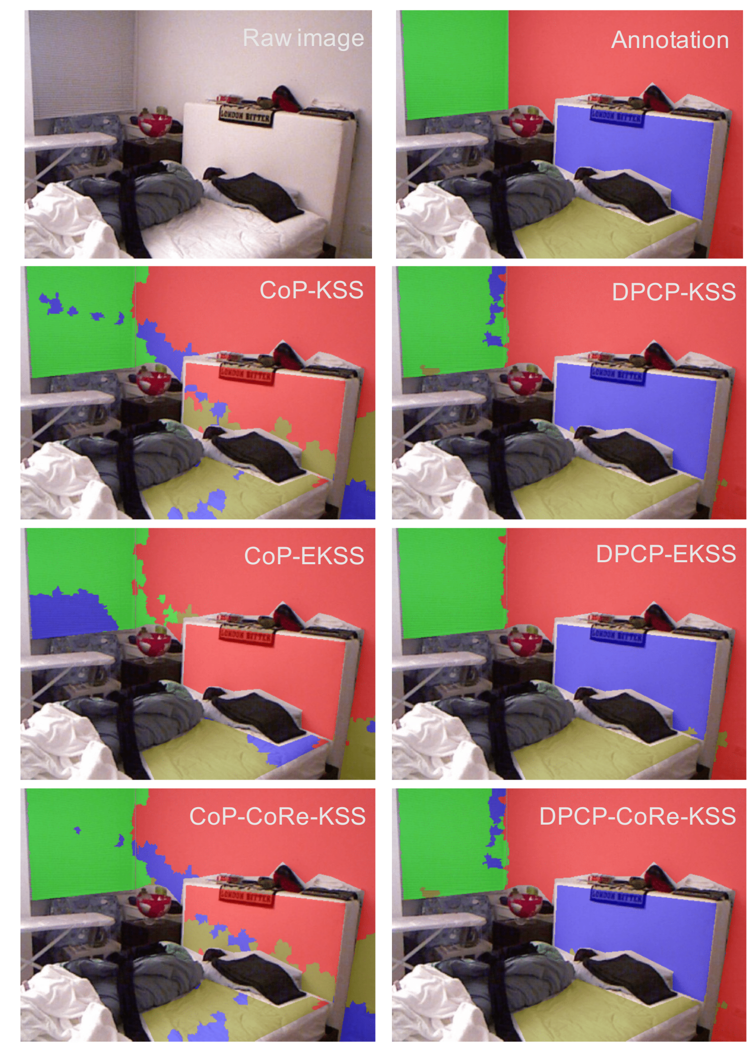


Hyperplane Clustering with DPCP

- We integrate DPCP into KSS (DPCP-KSS) by using it to estimate the dominant hyperplane for each cluster as a substitute for PCA or CoP [Rahmani & Atia]

	$D=4$			
	$K=2$	$K=3$	$K=4$	$K=5$
MKF	0.7937	0.6263	0.5548	0.4643
SCC	0.9445	0.9209	0.9093	0.8784
EnSC	0.7011	0.4912	0.3913	0.3254
SSC-ADMM	0.6801	0.4810	0.3795	0.3175
SSC-OMP	0.5707	0.4134	0.3291	0.2747
DPCP-KSS	**0.9834**	**0.9463**	0.8985	0.8103
CoP-KSS	0.9614	0.8747	0.8300	0.7630
PCA-KSS	0.9601	0.8623	0.8142	0.7461
DPCP-EKSS	**0.9889**	0.8807	**0.9778**	**0.9489**
CoP-EKSS	0.8278	0.8393	0.8772	0.7938
PCA-EKSS	0.8278	0.8274	0.8517	0.7542
DPCP-CoRe-KSS	0.9832	**0.9715**	**0.9561**	**0.9599**
CoP-CoRe-KSS	0.9612	0.8992	0.9065	0.8907
PCA-CoRe-KSS	0.9603	0.8981	0.8769	0.8586

Clustering accuracy for KSS variants and methods designed for clustering low dimensional subspaces



Visualization in clustering four hyperplanes from a 3D point cloud of NYUdepthV2