Dual Principal Component Pursuit for Learning a Union of Hyperplanes

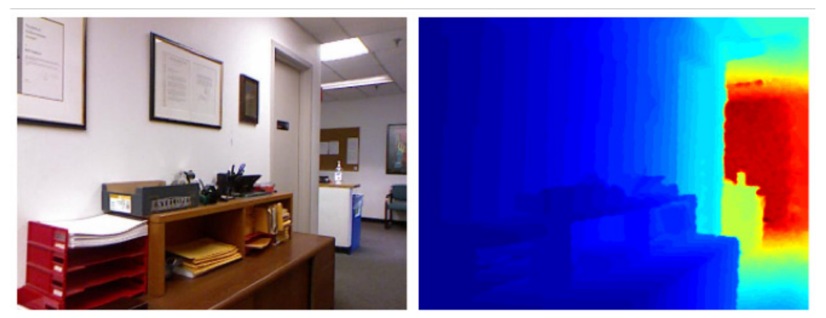
Tianyu Ding[1], Zhihui Zhu[2], Manolis Tsakiris[3], Rene Vidal[1], Daniel Robinson[4]

[1]Johns Hopkins University [2]University of Denver [3]Shanghai Tech University [4]Lehigh University

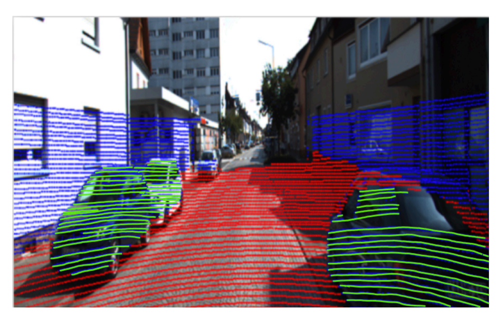
The 24th International Conference on Artificial Intelligence and Statistics

Introduction

- Motivation: Many applications in computer vision can be reduced to hyperplane clustering problems (corrupted with outliers)
- Challenges: Existing subspace clustering methods are based on sparse or low-rank approaches, whose theory and algorithms for low-dimensional subspaces do not apply to a union of hyperplanes



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



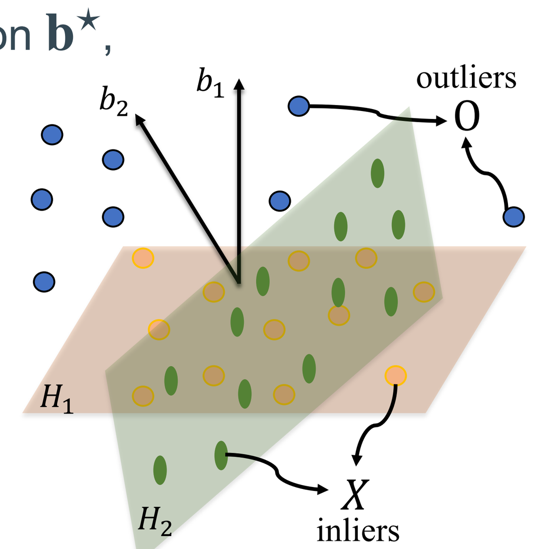
Estimating the road condition from points collected by a laser scanner

Dual Principal Component Pursuit (DPCP)

- Inliers $X \in \mathbb{R}^{D\times N}$ are N inlier points that lie in the union of K hyperplanes $\{H_k\}_{k=1}^K$ of $\mathbb{R}^D$ with unit normal vectors $\{\mathbf{b}_k\}_{k=1}^K$
- Outliers $O \in \mathbb{R}^{D\times M}$ are M outlier points that lie in ambient space $\mathbb{R}^D$
- Given dataset $\tilde{X} = [X, O]$, DPCP computes a solution $\mathbf{b}^\star$, ideally a normal to one specific hyperplane, by

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

- Problem (1) is challenging:
  - X contains inliers from a union of hyperplanes
  - Analysis of learning a single hyperplane cannot be applied here since inliers from other planes also exhibit certain linear structures



Main Contributions

- Derive both *geometric* and *probabilistic* conditions under which $\mathbf{b}^\star$ is a normal vector to a geometrically dominant hyperplane, denoted by H_1
- Prove a Projected Riemannian SubGradient Method converges linearly to $\mathbf{b}_1$

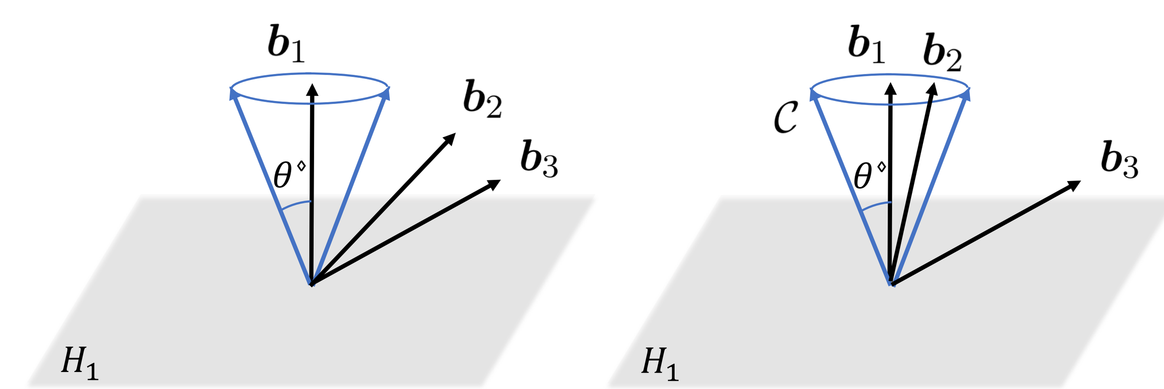
		Main Contributions	
Theory	Deterministic Analysis	$\mathbf{b}^\star \perp H_1$ under some conditions	[Tsakiris & Vidal]: a more transparent analysis that explicitly captures data dist.
	Probabilistic Analysis	$\mathbf{b}^\star \perp H_1$ with high probability if #outliers = O((#inliers of H_1 − #remaining inliers)^2)	[Lerman & Zhang]: a new probabilistic guarantee with a mild sample complexity: $\Omega(D^3)$ v.s. $\Omega(D^{18}\log D)$
Algorithm	Projected Riemannian SubGradient Method	linear convergence of $\hat{\mathbf{b}}_t$ to $\mathbf{b}_1$	[Tsakiris & Vidal]: a scalable method that has a convergence guarantee

Deterministic Global Optimality Analysis

- Informally, H_1 is the 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
- ζ_1 quantifies the dominance level of H_1: more dominance of H_1 leads to larger ζ_1

Lemma: Any critical point $\mathbf{b}^\star$ of (1) is either a normal vector of H_1 ($\mathbf{b}^\star \in \{\pm\mathbf{b}_1\}$), or has a principal angle θ from $\mathbf{b}_1$ such that $\theta \ge \theta^\circ := \arccos(1/\zeta_1)$.

- Remark 1: greater dominance of H_1 implies a more restricted location for $\mathbf{b}^\star$
- Remark 2: normal vectors of other hyperplanes within a θ°-neighborhood of $\mathbf{b}_1$ cannot be critical points of (1) → Inspires the convergence of an algorithm to $\mathbf{b}_1$



(Left) $\mathbf{b}_2$, $\mathbf{b}_3$ could be critical points

(Right) $\mathbf{b}_2$ cannot be a critical point

Theorem: Any global solution $\mathbf{b}^\star$ of (1) satisfies $\mathbf{b}^\star \in \{\pm\mathbf{b}_1\}$ as long as H_1 is sufficiently geometrically dominant.

Probabilistic Analysis

Consider a random spherical model:

- M outliers are drawn uniformly from the unit sphere in $\mathbb{R}^D$
- N_k inliers to H_k are drawn uniformly from the intersection of H_k and the unit sphere in $\mathbb{R}^D$, where $N_1 + \cdots + N_K = N$

Theorem: Global solutions of (1) are normal vectors of H_1 with high probability if

$$M \le C \cdot \Big(N_1 - \sum_{k\ne 1} N_k\Big)^2$$

- C is a decreasing function of D
- Remark 3: the Theorem implies $N_1 > N_2 + \cdots + N_K$, and the non-convex DPCP approach can roughly tolerate #outliers on the order of the square of #inliers
- [Lerman & Zhang] analyzes the same problem while requires a total sampling of $N + M = \Omega(D^{18}\log D)$ points to make the probability overwhelming
- Our analysis allows a sampling of only $N + M = \Omega(D^3)$ points to establish a high probable recovery

Projected Riemannian SubGradient Method

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

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

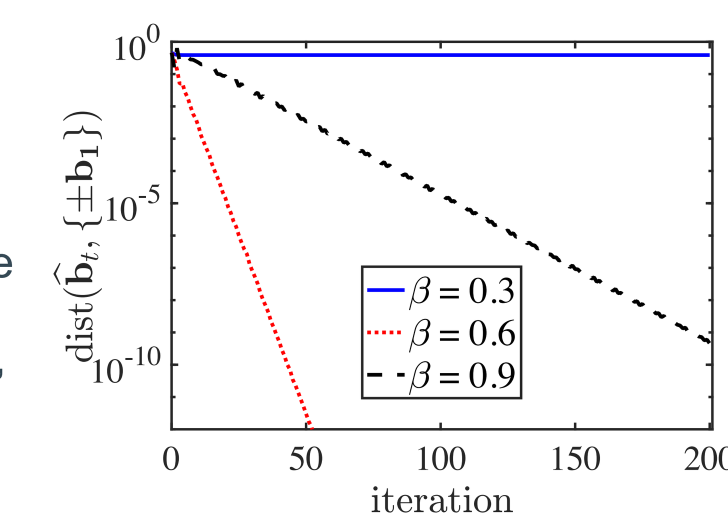
Geometrically diminishing step size: $\mu_t \leftarrow \mu_0 \beta^t$

Compute a Riemannian subgradient: $\mathscr{G}(\hat{\mathbf{b}}_t) \leftarrow \big(\mathbf{I} - \hat{\mathbf{b}}_t\hat{\mathbf{b}}_t^\top\big)\tilde{X}\text{sign}(\tilde{X}^\top\hat{\mathbf{b}}_t)$

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

Theorem: Let $\hat{\theta}_t$ be the principal angle between $\hat{\mathbf{b}}_t$ and $\mathbf{b}_1$. If $\hat{\theta}_0 < \theta^\circ$, then with appropriate initial step size μ_0 and diminishing factor β, we have $\sin(\hat{\theta}_t) \lesssim \beta^k$.

- Remark 4: the Theorem indicates the above algorithm has a linear convergence rate to $\mathbf{b}_1$
- Convergence may fail if β is too small, while convergence may be slow when β is too large
- In the right figure, $D = 9$, $K = 3$, $N = 1200$, $N_3 = 0.8N_2 = 0.8^2 N_1$, the outlier ratio $M/(M+N) = 0.3$, and $\mu_0 = 0.01$



Hyperplane Clustering with DPCP

- K-subspaces (KSS) [Bradley & Mangasarian] is a clustering method that alternates between assigning data to clusters and estimating a subspace for each cluster using PCA, which is known to be sensitive to outliers
- We integrate DPCP into KSS (DPCP-KSS) by using DPCP to estimate the geometrically dominant hyperplane for each cluster due to its robustness in fitting a hyperplane under a UoH model
- We also leverage an ensemble of DPCP-KSS via the EKSS [Lipor et al.] and CoRe [Lane et al.] frameworks to potentially further boost the performance

	$D = 4$			
	$K = 2$	$K = 3$	$K = 4$	$K = 5$
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

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

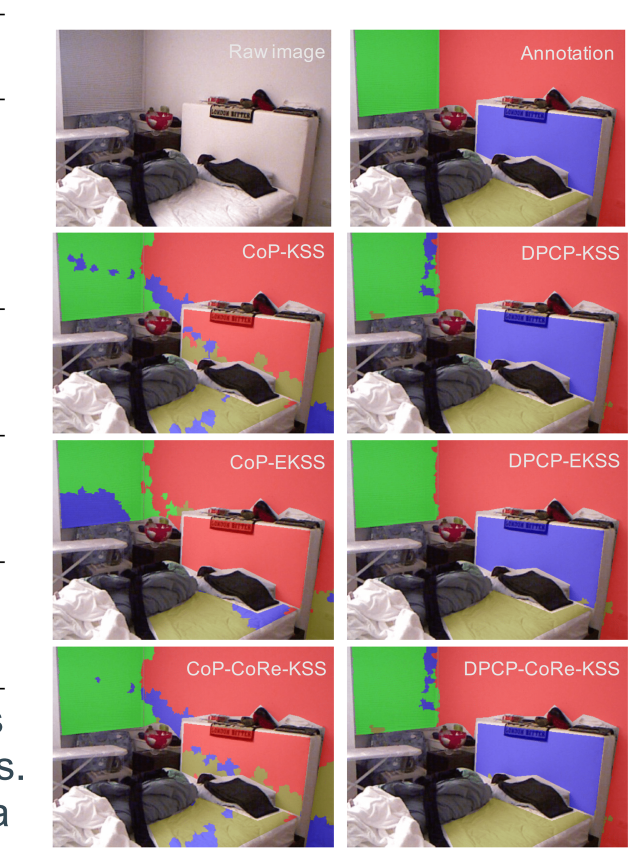


Figure: Visualization in clustering four hyperplanes from a 3D point cloud of NYUdepthV2.