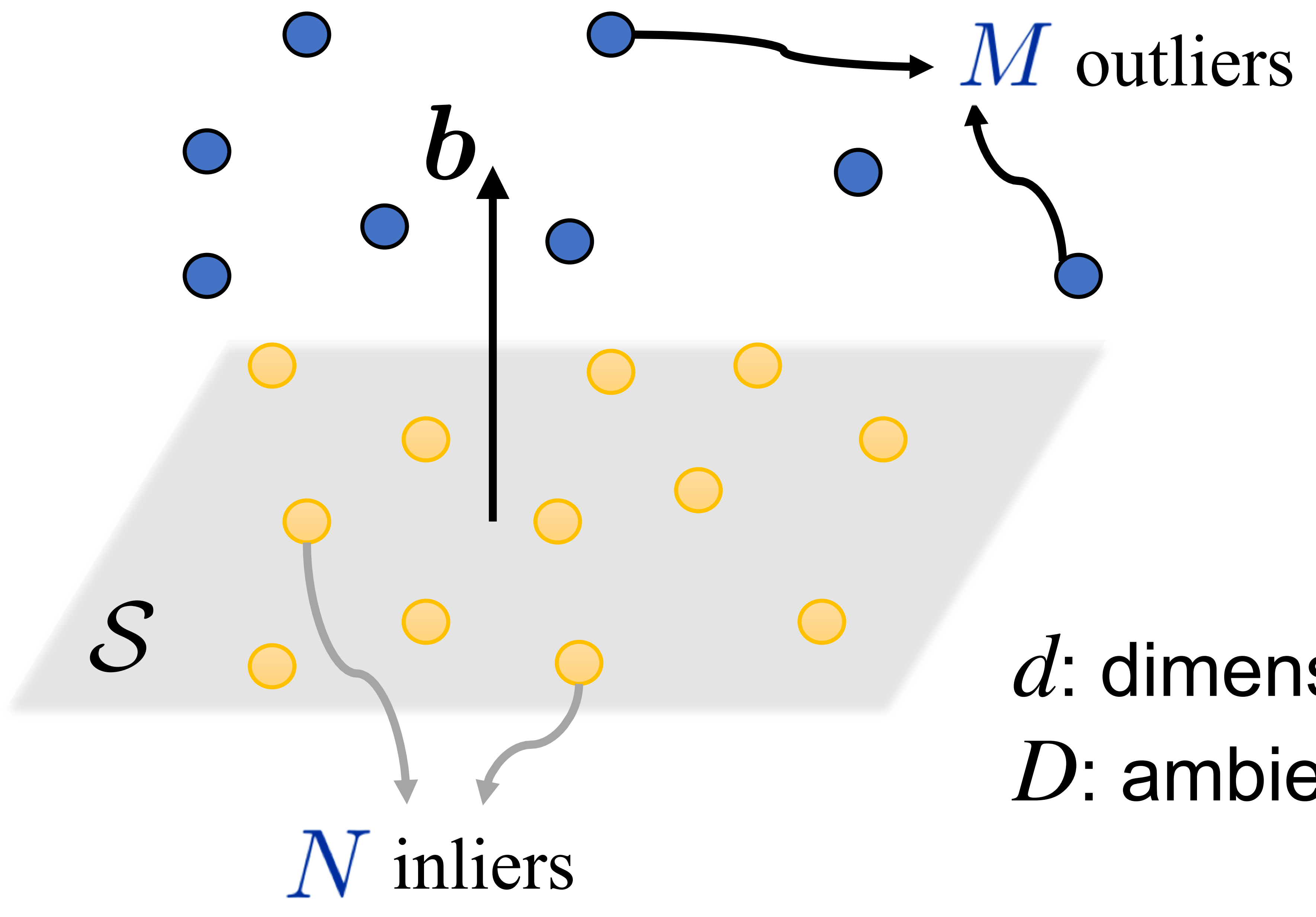
ICML | 2021
Thirty-eighth International Conference on Machine Learning

Dual Principal Component Pursuit for Robust Subspace Learning: Theory and Algorithms for a Holistic Approach

Tianyu Ding[1]
Rene Vidal[1]
Zhihui Zhu[2]
Daniel P. Robinson[3]

Johns Hopkins University
University of Denver
Lehigh University

Problem: Learning a Subspace from Corrupted Data



d: dimension of $\mathcal{S}$

D: ambient dimension

- *Low relative dimension* $(d/D \approx 0)$

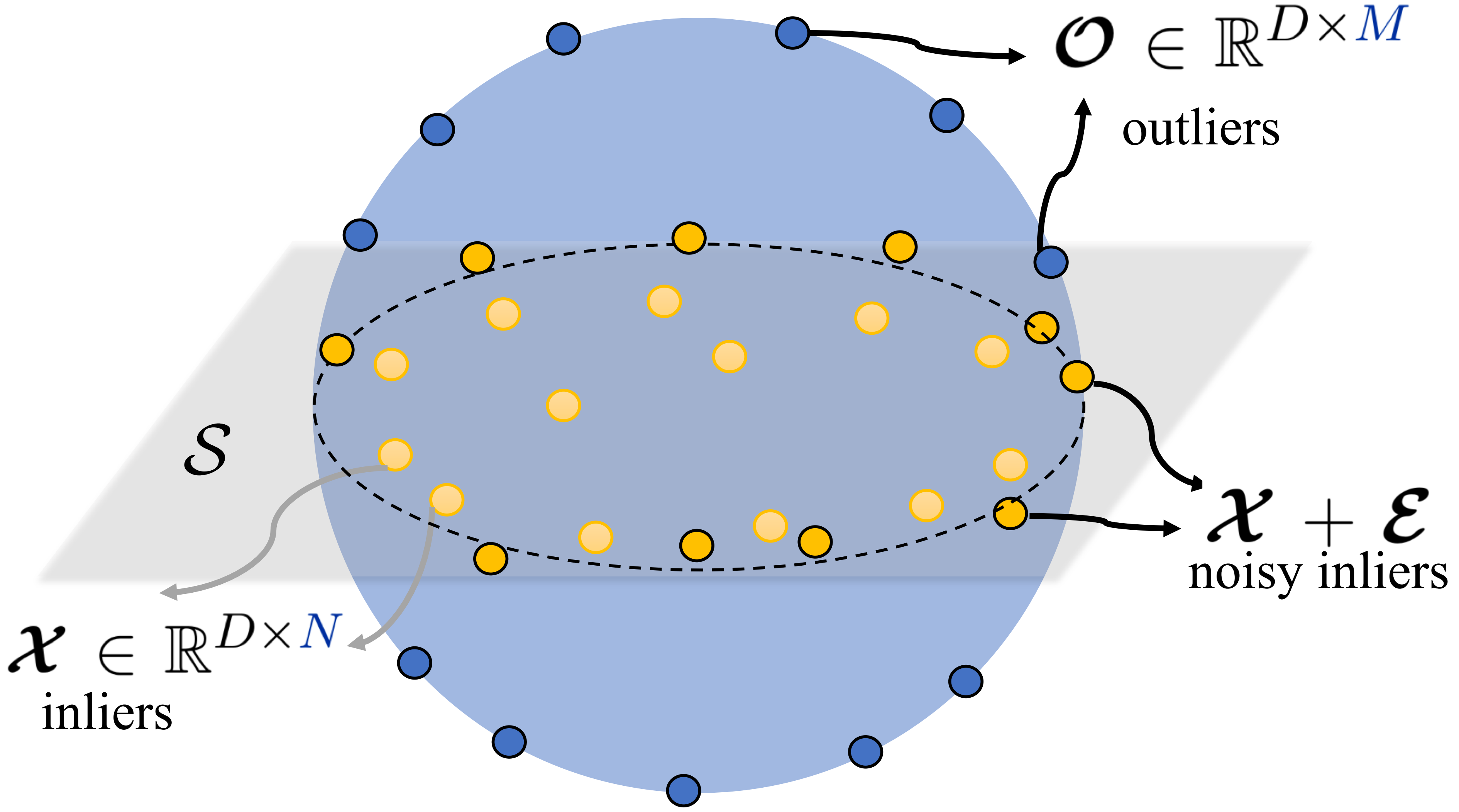
Method	# of outliers it can tolerate
Outlier Pursuit [Xu et al. 10]	$M = O(N)$
REAPER [Lerman et al. 15]	$M = O(N)$
CoP [Rahmani and Aita, 16]	$M = O(N)$
GGD [Maunu et al. 19]	$M = O(N)$

- *High relative dimension* $(d/D \approx 1)$
  - DPCP [Tsakiris and Vidal 18, Zhu et al. 18]
    $M = O(N^2)$ outliers

Recursively finds a new basis element $\boldsymbol{b}$ of $\mathcal{S}^\perp$

This paper considers a holistic DPCP approach for simultaneously computing the entire basis of $\mathcal{S}^\perp$

Dual Principal Component Pursuit (DPCP): A Holistic Approach



- Outliers $\mathcal{O}$ are drawn uniformly random from the unit sphere
- Noisy inliers $\mathcal{X} + \mathcal{E}$
  - inliers $\mathcal{X} \subset \mathcal{S}$, Gaussian noise $\mathcal{E}$ with std σ
  - normalize $\mathcal{X} + \mathcal{E}$ to unit norm
- SNR is $\mathbb{E}[\|\mathcal{X}\|_F]/\mathbb{E}[\|\mathcal{E}\|_F] = 1/\sigma$
- Dataset $\widetilde{\mathcal{X}} = [\mathcal{X} + \mathcal{E},\ \mathcal{O}]$

d: dimension of $\mathcal{S}$
D: ambient dimension
$c = D - d$: codimension

$$\boxed{\min_{\boldsymbol{B}\in\mathbb{R}^{D\times c}} \left\|\widetilde{\mathcal{X}}^\top \boldsymbol{B}\right\|_{1,2} := \sum_j \left\|\widetilde{\boldsymbol{x}}_j^\top \boldsymbol{B}\right\|_2 \quad \text{s.t.} \quad \boldsymbol{B}^\top\boldsymbol{B} = \mathbf{I}}$$

- [Tsakiris and Vidal 18, Zhu et al. 18, Ding et al. 19] analyzed the problem with $c = 1$
- [Zhu et al. 19] proposed an algorithm solves the holistic problem with noiseless data

the recursive approach is inefficient

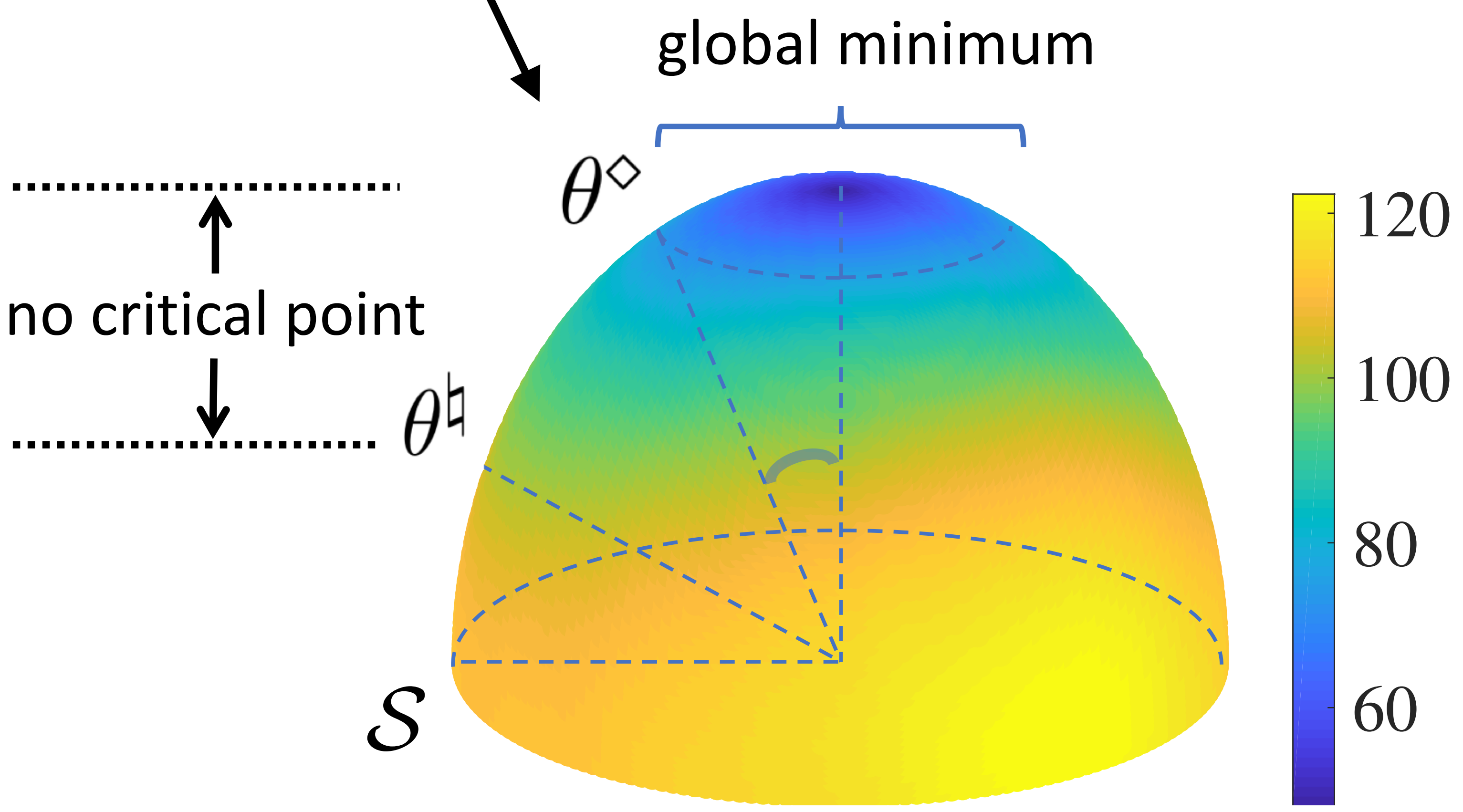
no landscape analysis with noisy data

We provide a landscape analysis and
extend the algorithm under the noisy setting

Deterministic and Probabilisitic Analysis

$$\min_{\boldsymbol{B}\in\mathbb{R}^{D\times c}} \left\|\widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{B}\right\|_{1,2} := \sum_j \left\|\widetilde{\boldsymbol{x}}_j^\top \boldsymbol{B}\right\|_2 \quad \text{s.t.} \quad \boldsymbol{B}^\top\boldsymbol{B} = \mathbf{I}$$

subspace angle between $\mathrm{Span}(\boldsymbol{B})$ and $\mathcal{S}^\perp$



Theorem 1: $\boldsymbol{B}^*$ is close to $\mathcal{S}^\perp$:

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

- In the noiseless case,

$$\mathrm{Span}(\boldsymbol{B}^*) = \mathcal{S}^\perp$$

Theorem 2: $\boldsymbol{B}^*$ is close to $\mathcal{S}^\perp$:

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

with high probability if

$$M \lesssim N^2.$$

- Cf. state-of-the-art: existing robust subspace recovery method can only handle at most $M = O(N)$ outliers in theory

Projected Riemannian SubGradient Method

$$\min_{\boldsymbol{B}\in\mathbb{R}^{D\times c}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{B} \right\|_{1,2} := \sum_j \left\| \widetilde{\boldsymbol{x}}_j^\top \boldsymbol{B} \right\|_2 \quad \text{s.t.} \quad \boldsymbol{B}^\top \boldsymbol{B} = \mathbf{I}$$

- Spectral initialization:

$$\boldsymbol{B}_0 \leftarrow \arg \min_{\boldsymbol{B}\in\mathbb{R}^{D\times c}, \boldsymbol{B}^\top \boldsymbol{B} = \mathbf{I}} \left\| \widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{B} \right\|_F^2$$

- Compute a Riemannian SubGradient:

$$\mathcal{G}(\boldsymbol{B}_k) \leftarrow (\mathbf{I} - \boldsymbol{B}_k \boldsymbol{B}_k^\top) \sum_j \widetilde{\boldsymbol{x}}_j \operatorname{sign}(\widetilde{\boldsymbol{x}}_j^\top \boldsymbol{B}_k)$$

- Geometrically diminishing step size:

$$\mu_k \leftarrow \mu_0 \beta^k$$

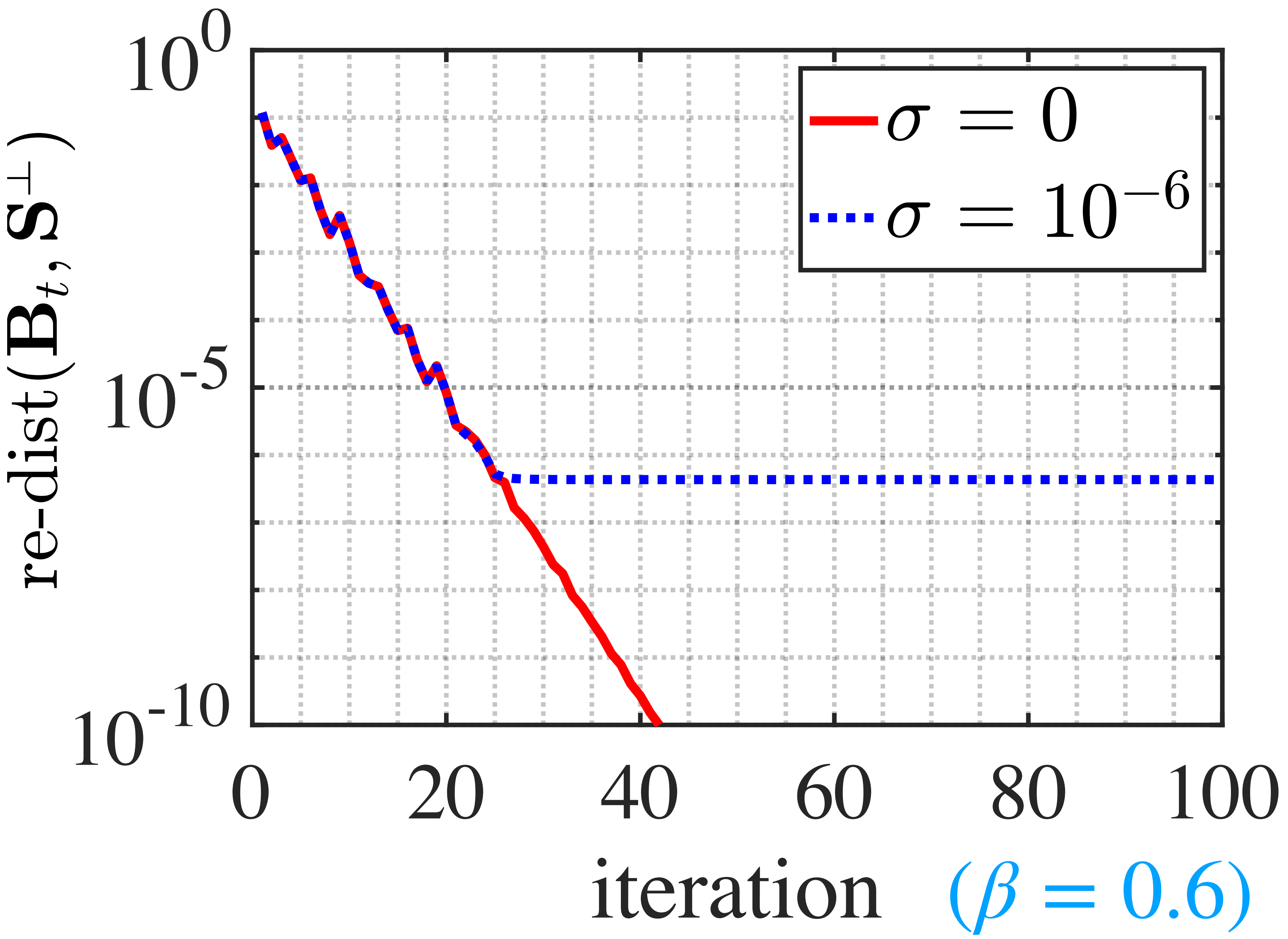
- Update the iterate:

$$\widehat{\boldsymbol{B}}_{k+1} \leftarrow \boldsymbol{B}_k - \mu_k \mathcal{G}(\boldsymbol{B}_k)$$

$$\boldsymbol{B}_{k+1} \leftarrow \operatorname{orth}(\widehat{\boldsymbol{B}}_{k+1})$$

Theorem 3: $\boldsymbol{B}_k$ converges linearly to $\mathcal{S}^\perp$:

$$\operatorname{dist}(\boldsymbol{B}_k, \mathcal{S}^\perp) \lesssim \operatorname{dist}(\boldsymbol{B}_0, \mathcal{S}^\perp)\beta^k + \sqrt{\sigma}$$



$(\beta = 0.6)$

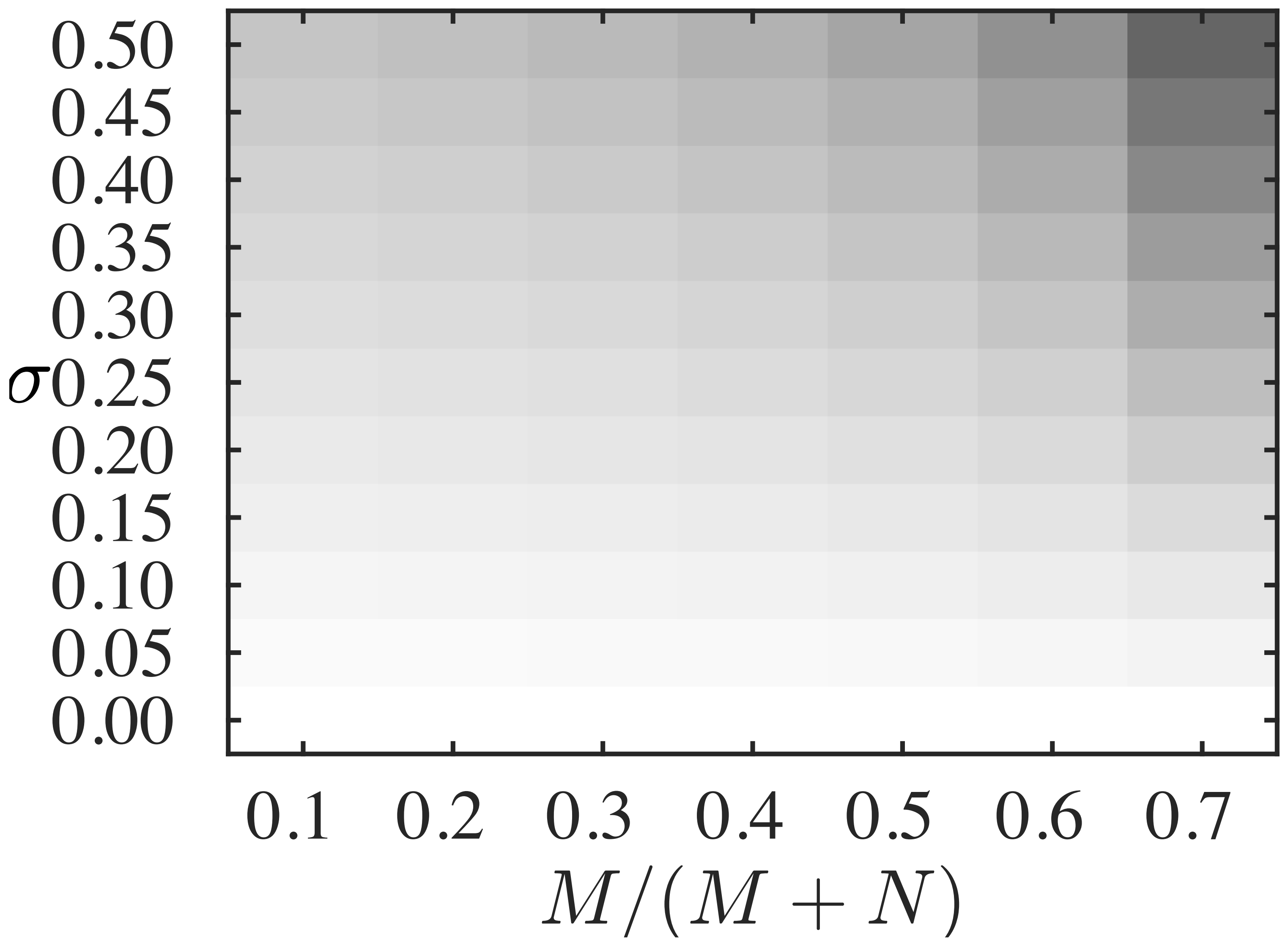
Experiments

Phase transition of dist(computed basis, ground-truth basis)
when varying outlier ratio $M/(M+N)$ and noise level σ

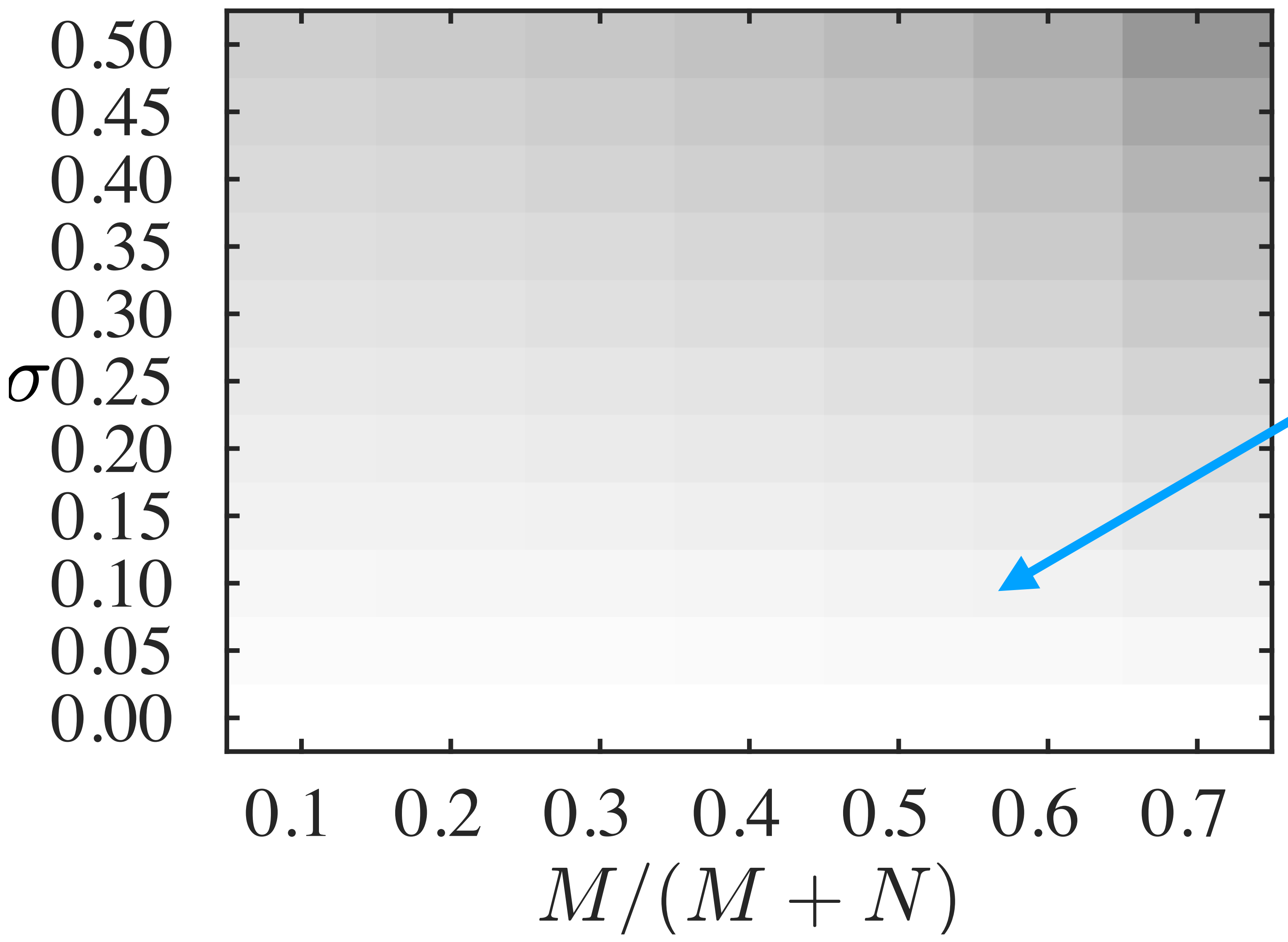
recursively learns one basis element at a time

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

$$\min_{\boldsymbol{B}\in\mathbb{R}^{D\times c}} \left\|\widetilde{\boldsymbol{\mathcal{X}}}^\top \boldsymbol{B}\right\|_{1,2} := \sum_j \left\|\widetilde{\boldsymbol{x}}_j^\top \boldsymbol{B}\right\|_2 \quad \text{s.t.} \quad \boldsymbol{B}^\top \boldsymbol{B} = \mathbf{I}$$



DPCP-recursive (79.74s)



DPCP-holistic (1.78s)

the lighter the better