Exploring Algorithmic Limits of Matrix Rank Minimization under Affine Constraints

David Wipf

Microsoft Research

Joint work with

Bo Xin, Peking Univ and **Tae-Hyun Oh**, KAIST

SPARS Workshop, Cambridge 2015

Sparse Estimation

$$\mathbf{y} = \begin{bmatrix} -4 \\ -5 \\ 3 \end{bmatrix}, \qquad A = \begin{bmatrix} 1 & 4 & 1 & 1 & 6 \\ -2 & 1 & -4 & 2 & -3 \\ 3 & 3 & 2 & -2 & 1 \end{bmatrix}$$

Want to find an $\mathbf{x}$ that solves

$$\mathbf{y} = A\,\mathbf{x}$$

non-sparse

$$\mathbf{x} = \begin{bmatrix} 4 \\ -1 \\ 3 \\ 5 \\ -2 \end{bmatrix}$$

sparse

$$\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \\ -1 \end{bmatrix}$$

Optimization Problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \;\; \mathbf{y} = A\,\mathbf{x}$$

$$\|\mathbf{x}\|_0 \;=\; \lim_{p\to 0} \sum_{\mathrm{i}} |x_i|^p \;=\; \#\ \text{of nonzero elements in } \mathbf{x}$$

Equivalent formulation:

$$\min_{\mathbf{x}} \sum_{\mathrm{i}} \operatorname{rank}(x_i), \quad \text{s.t.} \;\; \mathbf{y} = \sum_i \mathbf{a}_i x_i$$

Generalized Structure as Rank Minimization

Given Y and linear sensing operators A_i , solve:

$$\min_{\{X_i\}} \sum_i \alpha_i \operatorname{rank}[X_i], \quad \text{s.t. } Y = \sum_i \mathrm{A}_i(X_i)$$

arbitrary matrices

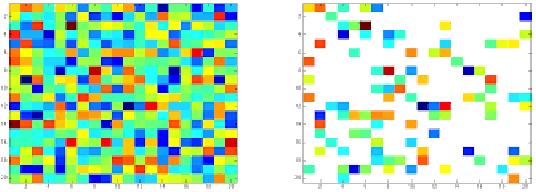
Relaxed version:

$$\min_{\{X_i\}} \left\| Y - \sum_i \mathrm{A}_i(X_i) \right\|^2 + \lambda \sum_i \alpha_i \operatorname{rank}[X_i]$$

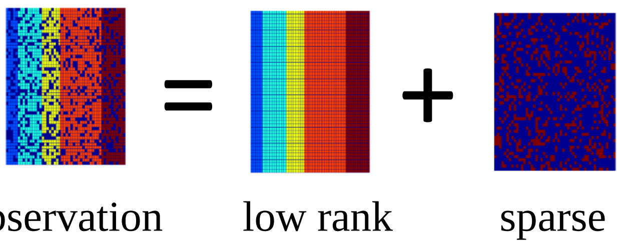
Flexible low-rank matrix estimation problem

Special Cases

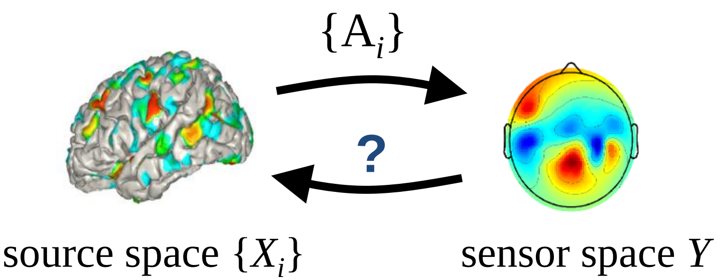
1. Matrix recovery/completion:



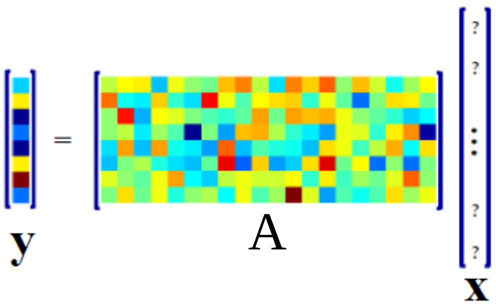
2. Robust PCA:



3. Source localization:



4. Compressive Sensing:



Many others …

Practical Estimation Issues

- Problem is NP-hard (… even for special cases).

- Convex approximation:

$$\min_{\{X_i\}} \left\| Y - \sum_i \mathrm{A}_i(X_i) \right\|^2 + \lambda \sum_i \alpha_i \|X_i\|_*$$

convex nuclear norm

- **Problem:** Performance suffers unless strong conditions on measurement process hold [Candès et al., 2011; Candès and Recht, 2008].

- Alternative solutions …?

Bayesian Alternatives

- Likelihood function: $p(Y \mid \{X_i\}) \propto \exp\left[-\frac{1}{2\lambda}\left\| Y - \sum_i \mathrm{A}_i(X_i)\right\|^2\right]$
- Prior distribution: $p(\{X_i\}) = \prod_i p(X_i)$ ⇐ favors low rank
- Inference: $\{\hat{X}_i\} = E_{p(\{X_i\}|Y)}[\{X_i\}]$,

Bayes Rule

$$p(\{X_i\}|Y) = \frac{p(Y \mid \{X_i\})p(\{X_i\})}{\int p(Y \mid \{X_i\})p(\{X_i\})d\{X_i\}}$$

- **Approximate inference:**

$$\boxed{\begin{aligned} \hat{p}(\{X_i\}|Y) &= \arg\min_{q(\{X_i\})\in\Omega} KL[q(\{X_i\}) \| p(\{X_i\}|Y)] \\ \{\hat{X}_i\} &= E_{\hat{p}(\{X_i\}|Y)}[\{X_i\}], \end{aligned}}$$

[Attias, 1999; Bishop, 2006]

Semi-Bayesian (SB) Proposal

- **Problem:** Bayesian algorithms not well-understood … can have inconsistent performance, and no theoretical support.

- **Solution:** Convert challenging Bayesian problems to equivalent regularized regression form:

$$\min_{\{X_i\}} \left\| Y - \sum_i \mathrm{A}_i(X_i) \right\|^2 + \lambda g_{SB}\left(\{X_i\};\{\mathrm{A}_i\}\right)$$

penalty handles structure
in measurement process

- Build algorithms and theory on top of this novel reformulation.

- Leads to state-of-the-art results in many domains …

Example I: Matrix Recovery

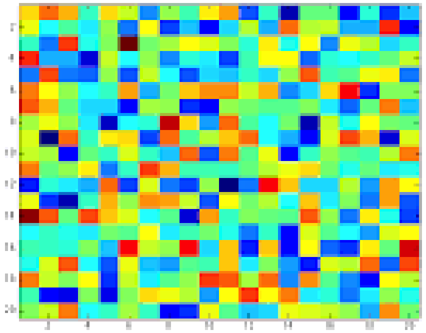
Goal: Recover low-rank X from p affine measurements

$$\min_{X} \ \operatorname{rank}(X), \quad \text{s.t. } y_{\mathrm{i}} = \operatorname{trace}\left(\Phi_i^T X\right), \ \forall i = 1, \dots p$$

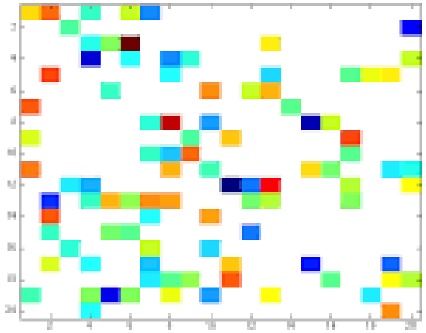
Special Case: Matrix completion:

$$\Phi_i \ \text{all zeros and a single one} \ \forall i = 1, \dots p$$

X



observed entries



[Candès and Recht, 2008; Mohan and Fazel, 2012; Liu et al., 2014]

Existing Algorithms
(blind to true rank)

Solve:

$$\min_X \sum_i f(\sigma_i), \quad \text{s.t. } y_i = \operatorname{trace}\left(\Phi_i^T X\right), \ \forall i = 1, \ldots p$$

singular values of X

[Candès and Recht, 2008; Mohan and Fazel, 2012; Liu et al., 2014]

Problem: For any possible function *f*, recovery can fail if measurement matrices Φ_i are "structured."

Contributions of Semi-Bayesian Framework

Solve:

$$\min_{X} \; g_{SB}\left(X;\{\Phi_i\}\right), \quad \text{s.t. } y_i = \operatorname{trace}\left(\Phi_i^T X\right), \; \forall i = 1,\ldots p$$

Theory: Global optimum has minimal rank, but fewer bad local minima (in certain conditions provably none ...).

Practice: Empirically successful right up to the theoretical limit *of any possible algorithm*

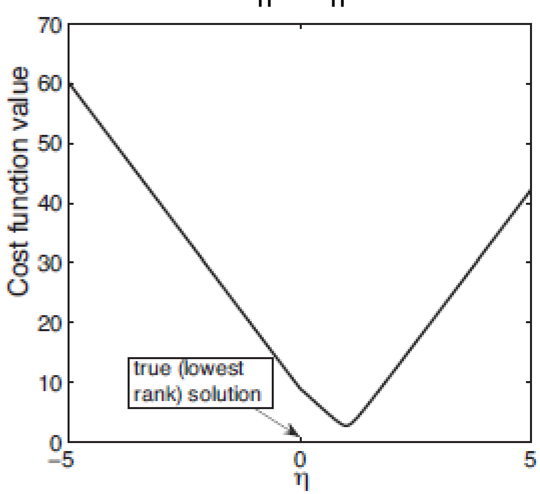
$$\Rightarrow \quad \#\text{ of d.o.f. in } X = p$$

Visualization of Minima in 1D Feasible Subspace

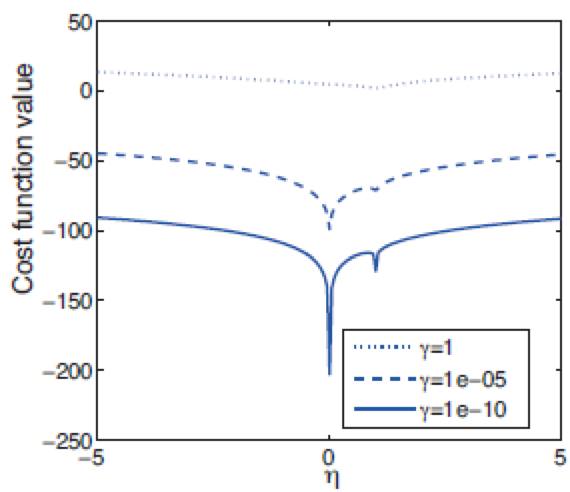
$$X = X_0 + \eta V$$

feasible solution | optimal solution | $\in \mathrm{null}(\{\Phi_i\})$

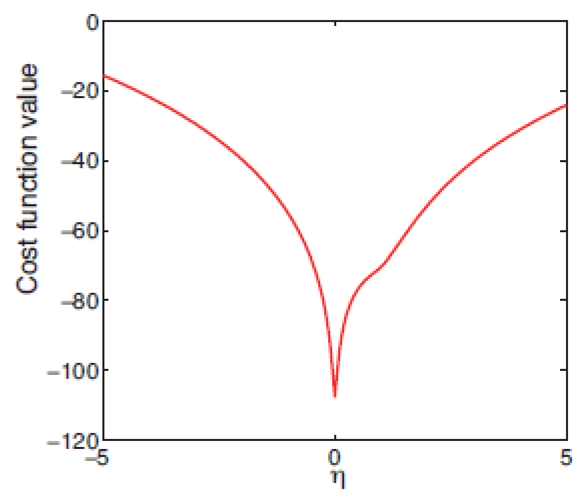
$\|X\|_*$



$\sum_i \log\left(\sigma_i[X]^2 + \gamma\right)$

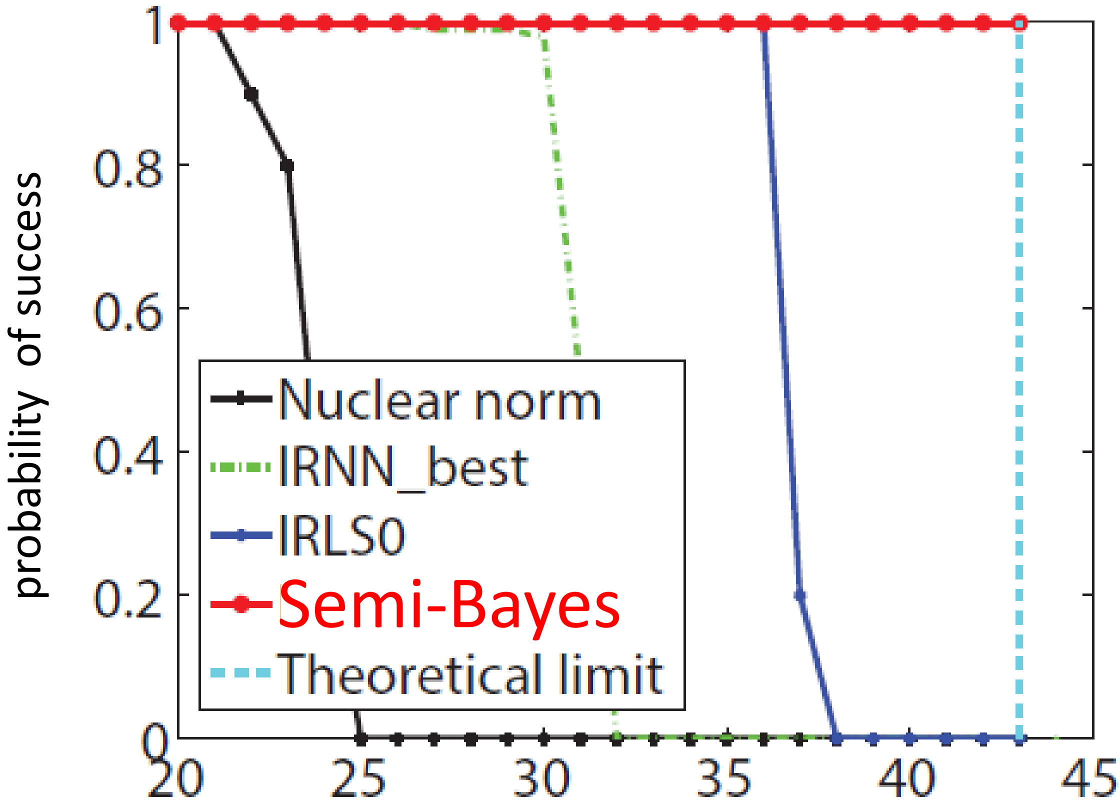


$g_{SB}(X; \{\Phi_i\})$



Empirical Results: Matrix Completion

50% randomly observed entries of $X \in \Re^{150 \times 150}$



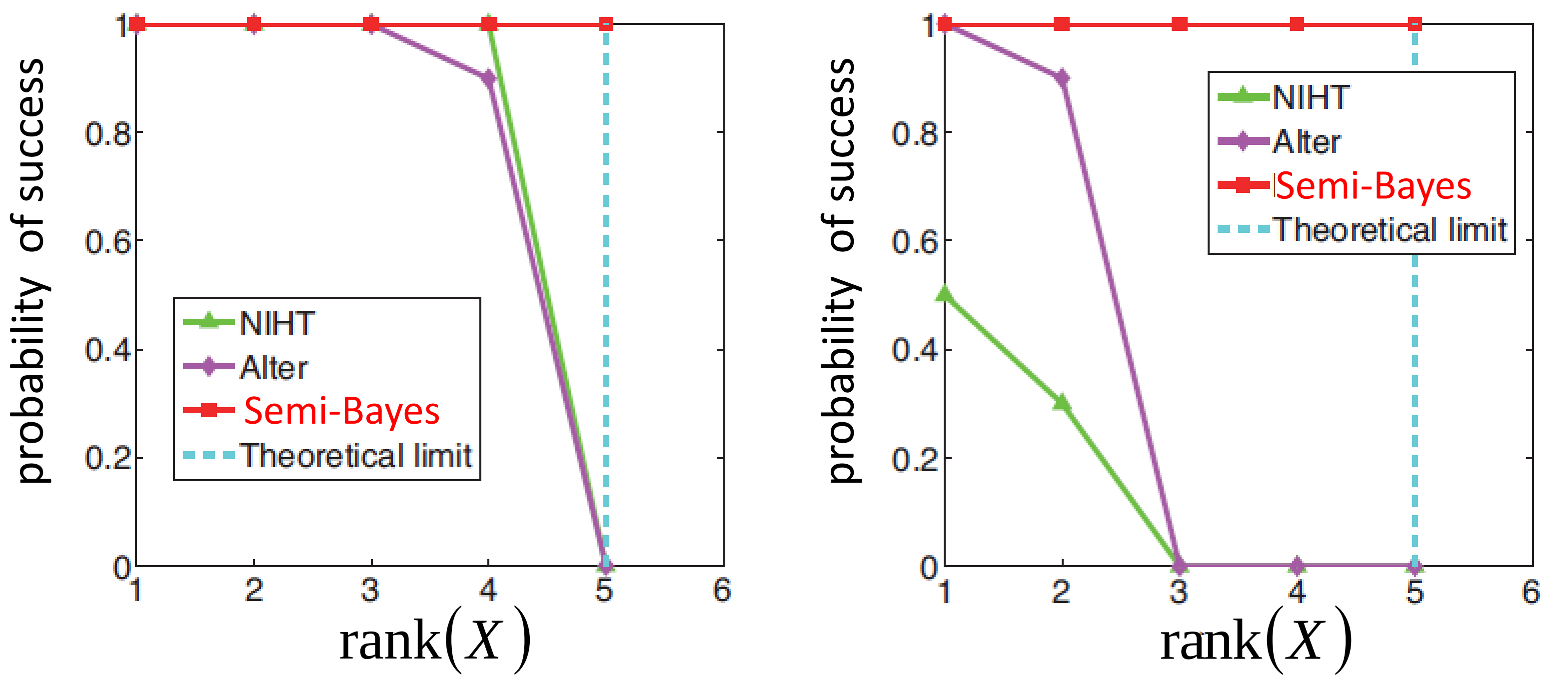
$\mathrm{rank}(X)$

[Candès and Recht, 2008; Liu et al., 2014; Mohan and Fazel, 2012]

Empirical Results: Correlated Measurements

$$X \in \Re^{100\times 100}, \quad \mathbf{y} = A\,\mathrm{vec}[X]$$

sampling matrix $\quad A = \sum_{i=1}^{1000} \frac{1}{\sqrt{i}} \mathbf{u}_i \mathbf{v}_i^T \qquad \mathbf{u}_i, \mathbf{v}_i \leftarrow N(0,1)$



NIHT [Tanner and Wei, 2013] and ***Alternating Minimization*** [Jain et al., 2013] both have knowledge of true rank

Image Rectification: Easy Case

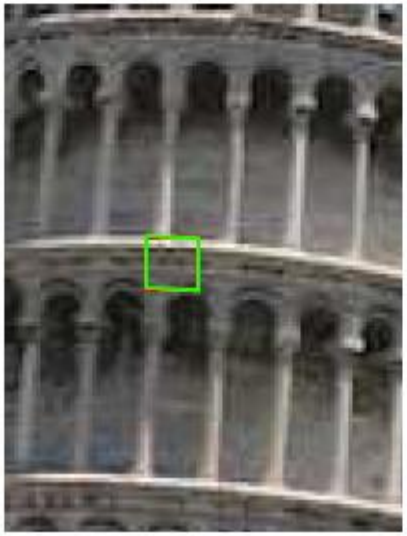
Convex Tilt Algorithm

Semi-Bayesian

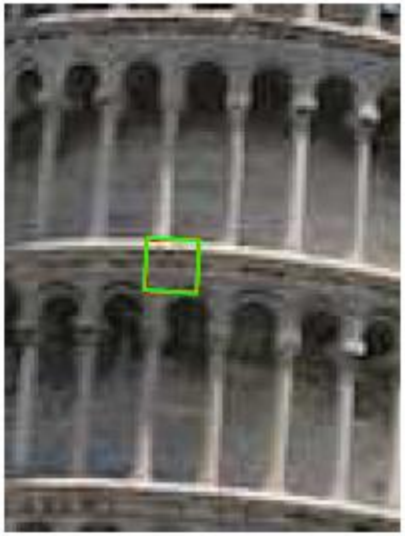
[Zhang et al. 2010]

Image Rectification: Easy Case

Convex Tilt Algorithm



Semi-Bayesian



[Zhang et al. 2010]

Example II: Robust PCA

$$\min_{X,S} \|Y - X - S\|_F^2 + \lambda_1 \operatorname{rank}[X] + \lambda_2 \|S\|_0$$

sparse outlier term

Problem: NP-hard optimization, so approximate methods are required.

Convex Relaxation

[Candès et al. 2011]

$$\min_{X,S} \left\| Y - X - S \right\|_F^2 + \lambda_1 \left\| X \right\|_* + \lambda_2 \left\| S \right\|_1$$

- Efficient minimization algorithms, e.g., *principal component pursuit* (PCP).
- **Problem**: Estimation guarantees exist, but require strong assumptions.
- Semi-Bayesian framework offers dramatic improvement (… both in *theory* and *practice*).

Semi-Bayesian Alternative

- Objective function:

$$g_{SB}(X, S) \neq g_1(X) + g_2(S)$$

- Assume $E = 0$ (canonical R-PCA problem).
- Now consider the following:

$$\text{(P1)} \qquad \min_{X,S} \ \operatorname{rank}[X] + \frac{1}{n}\|S\|_0 \qquad \text{s.t. } Y = X + S$$

$$\text{(P2)} \qquad \min_{X,S} \ g_{SB}(X, S) \qquad \text{s.t. } Y = X + S$$

Result

- P1 and P2 have the same global optimum.

- P2 is smoother (fewer local minima) than any possible problem of the separable form

$$\min_{X,S} \; g_1(\mathrm{X}) + g_2(\mathrm{S}) \qquad \text{s.t. } Y = X + S$$

that also globally optimizes P1 for all Y.

Intuition

$$Y \;=\; \mathrm{X} \;+\; \mathrm{S} \;\in\; \Re^{n\times n}$$

- With penalty of the form $g_1(\mathrm{X}) + g_2(\mathrm{S})$ we enter a local minima whenever

$$\operatorname{rank}[X] < n \quad \text{or} \quad \|s_j\|_0 < n$$

- In contrast, with $g_{SB}(\mathrm{X}, \mathrm{S}; \lambda = 0)$ we enter a local minima whenever

$$\operatorname{rank}[X] + \|s_j\|_0 < n$$

Simulation Example

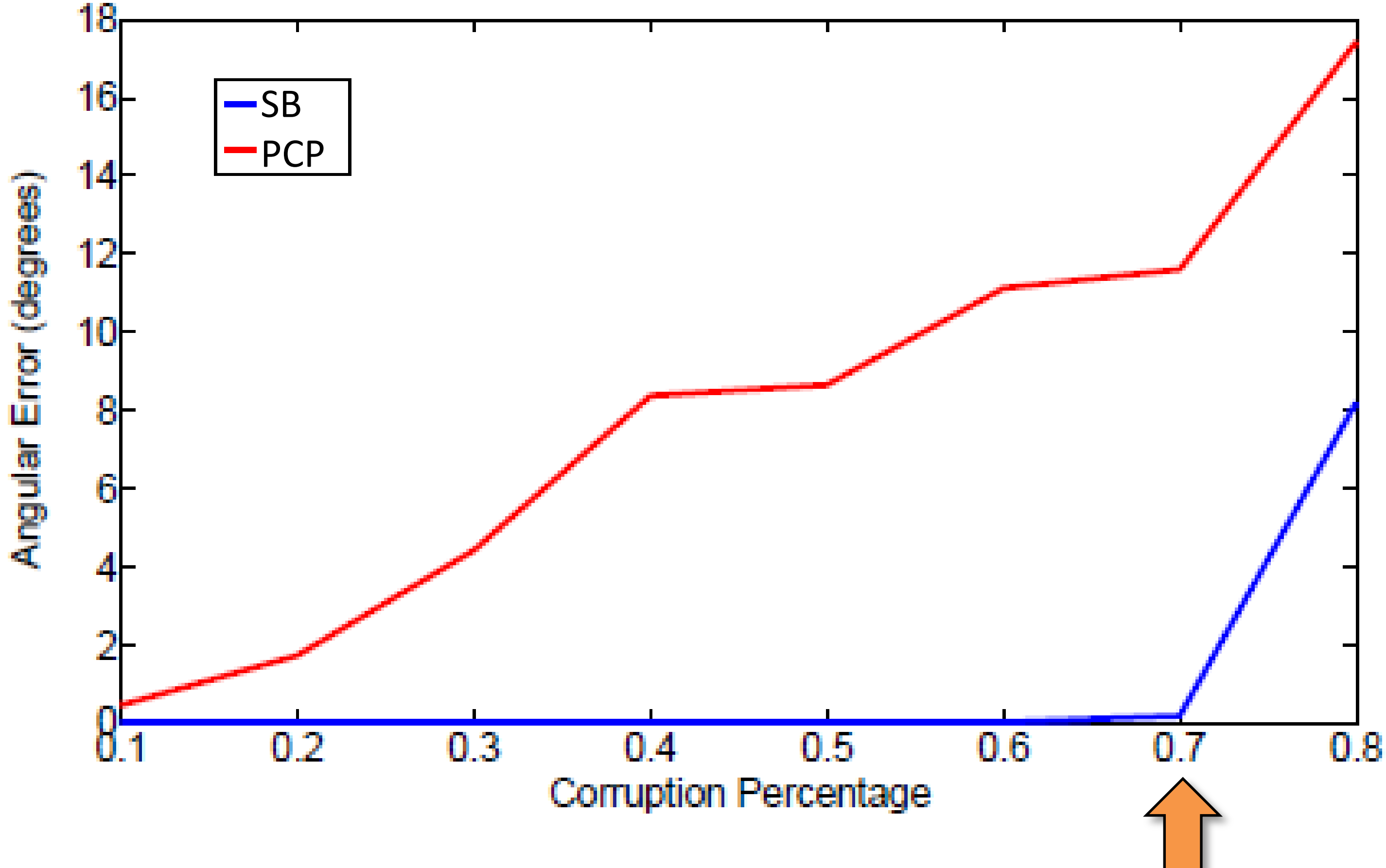
- Generate:

$$X \quad : \quad 20 \ \text{by} \ 10^4 \ \text{matrix, 20\% of full rank}$$
$$S \quad : \quad \text{sparse outlier matrix}$$

$$Y = X + S$$

- Given Y, estimate X and S using:
  - *Convex PCP algorithm* [Candès et al., 2011]
  - *Semi-Bayesian approach*.

Estimated X Subspace Angular Error (1000 trial avg.)

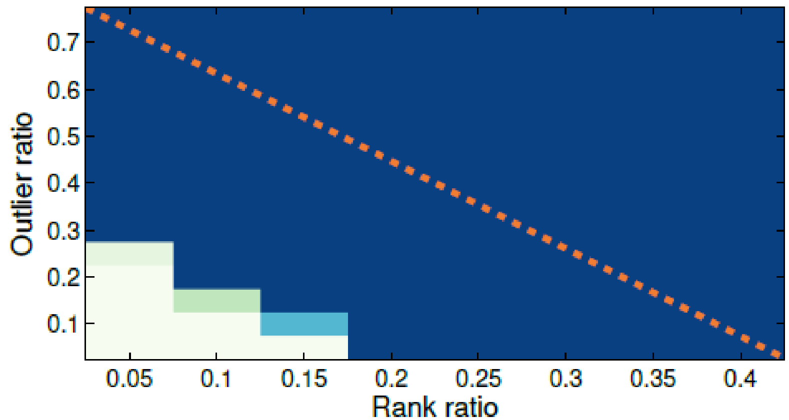


Even with 70% outliers,
semi-Bayes is nearly perfect

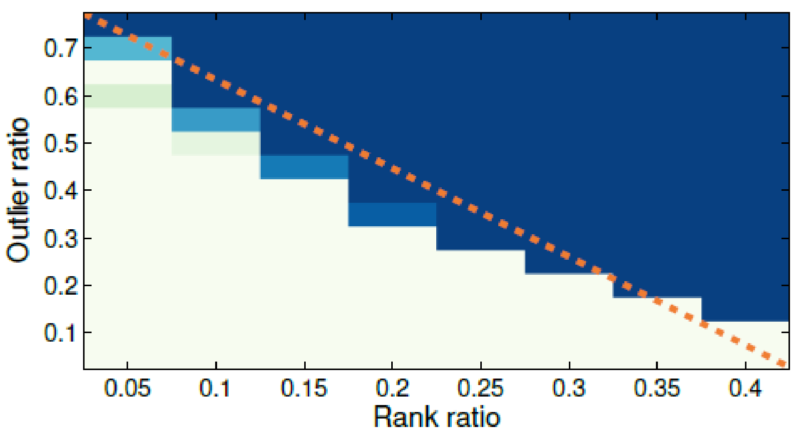
Phase Transition Plots

(100 X 100 Case)

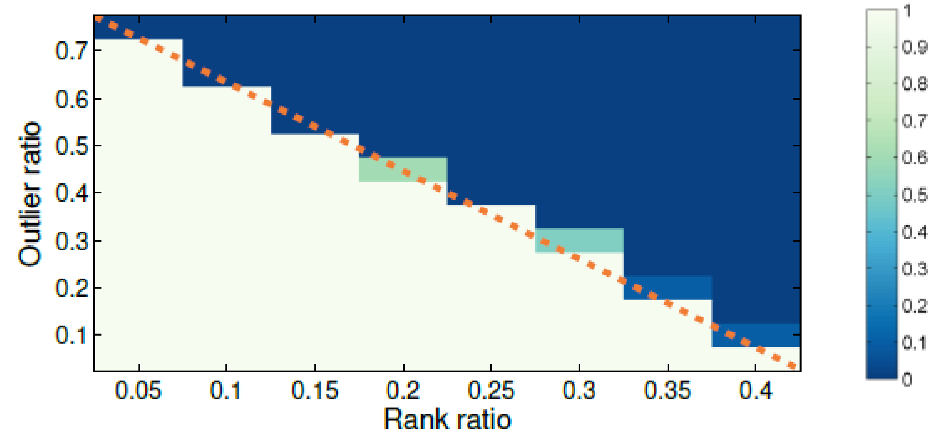
Convex Robust PCA



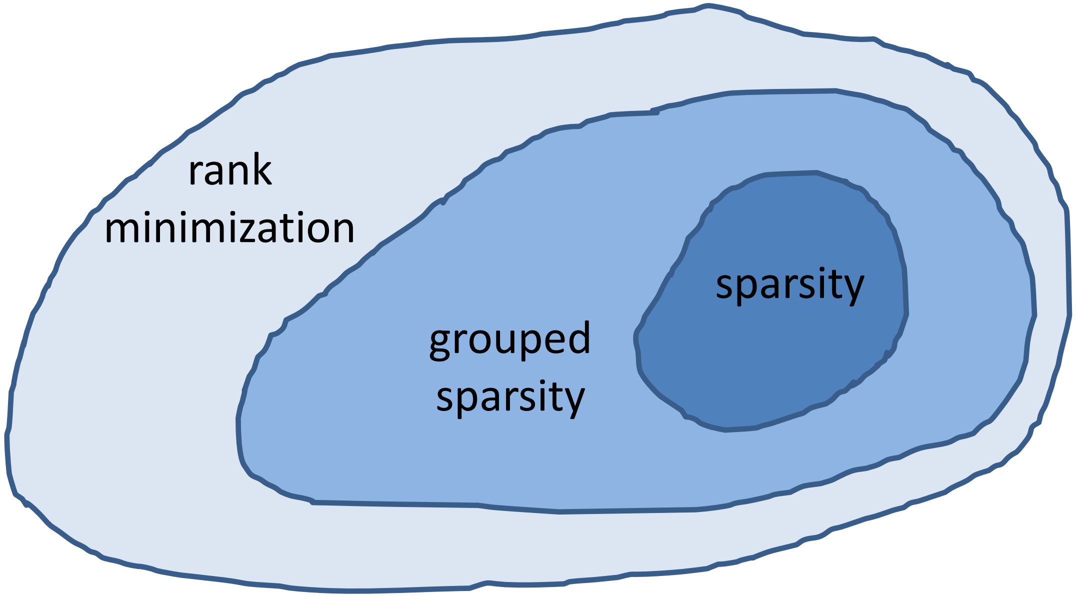
Convex, known missing entries



Non-Convex, Semi-Bayesian



Concluding Remarks



- Semi-Bayesian regression is a natural fit for all.
- Many additional possibilities …

Thank You

References:

Wipf, UAI, 2012
Xin and Wipf, ICML, 2015
Oh and Wipf, under review, 2015
Wipf, under review, 2015