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
Sparse Estimation

\[ y = \begin{bmatrix} -4 \\ -5 \\ 3 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 4 & 1 & 1 & 6 \\ -2 & 1 & -4 & 2 & -3 \\ 3 & 3 & 2 & -2 & 1 \end{bmatrix} \]

Want to find an \( x \) that solves
\[ y = A x \]

\[ x = \begin{bmatrix} 4 \\ -1 \\ 3 \\ 5 \\ -2 \end{bmatrix}, \quad x_0 = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \\ -1 \end{bmatrix} \]
Optimization Problem

\[
\min_x \|x\|_0 \quad \text{s.t. } y = A x
\]

\[
\|x\|_0 = \lim_{p \to 0} \sum_i |x_i|^p = \# \text{ of nonzero elements in } x
\]

Equivalent formulation:

\[
\min_x \sum_i \text{rank}(x_i), \quad \text{s.t. } y = \sum_i a_i x_i
\]
Generalized Structure as Rank Minimization

Given $Y$ and linear sensing operators $A_i$, solve:

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

Relaxed version:

$$\min_{\{X_i\}} \left\| Y - \sum_i A_i(X_i) \right\|^2 + \lambda \sum_i \alpha_i \text{rank}[X_i]$$

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:

\[
\min_{\{x_i\}} \left\| Y - \sum_i A_i(X_i) \right\|^2 + \lambda \sum_i \alpha_i \|X_i\|^* 
\]

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: \[ p(Y | \{X_i\}) \propto \exp\left[-\frac{1}{2\lambda} \left\| Y - \sum_i A_i(X_i) \right\|^2 \right] \]

- Prior distribution: \[ p(\{X_i\}) = \prod_i p(X_i) \] (favors low rank)

- Inference: \[ \{\hat{X}_i\} = E_{p(\{X_i\}|Y)}[\{X_i\}], \quad p(\{X_i\}|Y) = \frac{p(Y|\{X_i\})p(\{X_i\})}{\int p(Y|\{X_i\})p(\{X_i\})d\{X_i\}} \]

- Approximate inference:

\[
\hat{p}(\{X_i\}|Y) = \arg\min_{q(\{X_i\}) \in \Omega} KL[q(\{X_i\})||p(\{X_i\}|Y)]
\]

\[
\{\hat{X}_i\} = E_{\hat{p}(\{X_i\}|Y)}[\{X_i\}],
\]

[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:

\[
\min_{\{X_i\}} \left\| Y - \sum_i A_i(X_i) \right\|^2 + \lambda g_{SB}(\{X_i\};\{A_i\})
\]

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

$$\min_{X} \ \text{rank}(X), \quad \text{s.t. } y_i = \text{trace}(\Phi_i^T X), \quad \forall i = 1, \ldots, p$$

Special Case: Matrix completion:

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

[Candès and Recht, 2008; Mohan and Fazel, 2012; Liu et al., 2014]
Existing Algorithms
(blind to true rank)

Solve:

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

Problem: For any possible function $f$, recovery can fail if measurement matrices $\Phi_i$ are “structured.”

[Candès and Recht, 2008; Mohan and Fazel, 2012; Liu et al., 2014]
Contributions of Semi-Bayesian Framework

Solve:
\[
\min_X g_{SB}(X;\{\Phi_i}\}), \quad \text{s.t. } y_i = \text{trace}(\Phi_i^T X), \quad \forall i = 1, \ldots p
\]

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

\[
\# \text{ of d.o.f. in } X = p
\]
Visualization of Minima in 1D Feasible Subspace

\[ X = X_0 + \eta V \]

feasible solution \hspace{1cm} optimal solution \hspace{1cm} \in \text{null}(\{\Phi_i\})

\[ \|X\|_* \quad \sum_i \log(\sigma_i [X]^2 + \gamma) \quad g_{SB}(X;\{\Phi_i\}) \]
Empirical Results: Matrix Completion

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

[Candès and Recht, 2008; Liu et al., 2014; Mohan and Fazel, 2012]
Empirical Results: Correlated Measurements

\[ X \in \mathbb{R}^{100 \times 100}, \quad y = A \text{vec}[X] \]

Sampling matrix

\[ A = \sum_{i=1}^{1000} \frac{1}{\sqrt{i}} u_i v_i^T \]

\[ u_i, v_i \leftarrow N(0,1) \]

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

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

**Problem**: NP-hard optimization, so approximate methods are required.
Convex Relaxation
[Candès et al. 2011]

\[
\min_{X,S} \|Y - X - S\|_F^2 + \lambda_1 \|X\|_* + \lambda_2 \|S\|_1
\]

- 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:

\[
\begin{align*}
(P1) & \quad \min_{X,S} \quad \text{rank}[X] + \frac{1}{n} \|S\|_0 \quad \text{s.t.} \quad Y = X + S \\
(P2) & \quad \min_{X,S} \quad g_{SB}(X, S) \quad \text{s.t.} \quad Y = X + S
\end{align*}
\]
Result

- P1 and P2 have the same global optimum.

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

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

that also globally optimizes P1 for all Y.
Intuition

\[ Y = X + S \in \mathbb{R}^{n \times n} \]

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

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

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

\[ \text{rank}[X] + \|s_j\|_0 < n \]
Simulation Example

• Generate:

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

\[ Y = X + S \]

• 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

Non-Convex, Semi-Bayesian

Convex, known missing entries
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