

Compressed sensing of low-rank plus sparse matrices

SIMON VARY

simonvary.github.io/siam-la21.pdf

*Mathematical Institute, University of Oxford, United Kingdom
ICTEAM, UCLouvain – Université catholique de Louvain, Belgium*

joint work with Jared Tanner & Andrew Thompson

SIAM Conference on Applied Linear Algebra, New Orleans, 18/5/2021

Plan for the talk:

1. Models using rank or sparsity
2. Non-closedness of the feasible set
3. Theory: (i) Closing the set, (ii) Restricted isometry, (iii) Correlation between the components
4. Algorithms and convergence theorems
5. Numerical phase transition

Publications covered:

Matrix rigidity and the ill-posedness of Robust PCA and matrix completion,

SIAM Journal on Mathematics of Data Science, 1(3), 537–554, 2019.

Jared Tanner, Andrew Thompson, Simon Vary

Compressed sensing of low-rank plus sparse matrices,

preprint, arxiv.org/abs/2007.09457, 2020.

Jared Tanner, Simon Vary

On low-rank plus sparse matrix sensing,

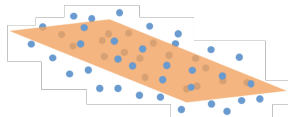
PhD thesis, <http://simonvary.github.io/thesis.pdf>, 2021.

Simon Vary

Principal Component Analysis (PCA)

We have a data matrix $M = [y_1, \dots, y_n]$ with mean-centered samples y_i

$$\min_{X \in \mathbb{R}^{m \times n}} \|X - M\|_F, \quad \text{s.t. } \text{rank}(X) \leq r.$$



Low-rank matrix sensing (Recht *et al.*, 2010)

We are given subsampled information about M in the form of $b = \mathcal{A}(M) \in \mathbb{R}^p$. Solving

$$\min_{X \in \mathbb{R}^{m \times n}} \|\mathcal{A}(X) - b\|_F, \quad \text{s.t. } \text{rank}(X) \leq r$$

recovers M for many r and various linear maps $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p$.

What if the low-rank structure is corrupted?

We instead observe a low-rank structure corrupted by some sparse noise

$$M = \underbrace{\hat{L}}_{\text{low-rank}} + \underbrace{\hat{S}}_{\text{sparse}}.$$

Define the following set

$$\text{LS}_{m,n}(r, s) = \{L + S \in \mathbb{R}^{m \times n} : \text{rank}(L) \leq r, \|S\|_0 \leq s\}.$$

Robust PCA (Candès et al., 2011)

For a given matrix M , find X

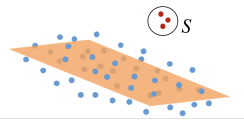
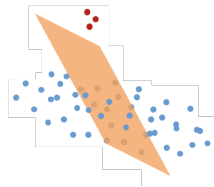
$$\min_{X \in \mathbb{R}^{m \times n}} \|X - M\|_F, \quad \text{s.t.} \quad X \in \text{LS}_{m,n}(r, s).$$

Low-rank plus sparse matrix sensing

We are given subsampled information about M in the form of $b = \mathcal{A}(M) \in \mathbb{R}^p$. Solve

$$\min_{X \in \mathbb{R}^{m \times n}} \|\mathcal{A}(X) - b\|_F, \quad \text{s.t.} \quad X \in \text{LS}_{m,n}(r, s).$$

to recover $M \in \text{LS}_{m,n}(r, s)$ for a wide range of ranks r , support sizes s of the corruption, and linear maps $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p$.



Low-rank plus sparse matrix approximation (Robust PCA)

Incoherence between the low-rank and the sparse component (Candès et al., 2011)

It is necessary to control the correlation of singular vectors of the rank- r matrix $L = U\Sigma V^T \in \mathbb{R}^{m \times n}$ and the canonical basis with the coherence parameter $\mu \in [1, \sqrt{mn}]$

$$\max_{i \in \{1, \dots, r\}} \|U^T e_i\|_2 \leq \sqrt{\frac{\mu r}{m}}, \quad \max_{i \in \{1, \dots, r\}} \|V^T e_i\|_2 \leq \sqrt{\frac{\mu r}{n}}.$$

For each triplet of (r, s, μ) , define the following *correlation coefficient* $\gamma_{r,s,\mu} = \mu \frac{r\sqrt{s}}{\sqrt{mn}}$.

Convex problem

$$\min_{L, S \in \mathbb{R}^{n \times n}} \|L\|_* + \lambda \|S\|_1,$$

$$\text{s.t. } M = L + S$$

- Solves also the non-convex problem
- $\|\cdot\|_*$ requires the full SVD $\Rightarrow \mathcal{O}(n^3)$
- PCP (Candès et al., 2011)
- IALM (Lin et al., 2009)

Non-convex problem

$$\min_{X \in \mathbb{R}^{m \times n}} \|X - M\|_F,$$

$$\text{s.t. } X = L + S \in \text{LS}_{m,n}(r, s)$$

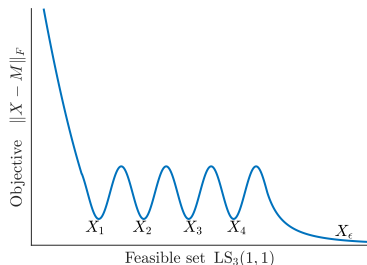
- $L = YZ^T$ for $Y, Z \in \mathbb{R}^{n \times r}$
- Alternating projections for L, S
- Per iteration complexity:
 - AltProj, $\mathcal{O}(r^2 n^2)$ (Netrapalli et al., 2014)
 - FastGD, $\mathcal{O}(rn^2)$ (Yi et al., 2016)

Simple example of non-closedness

Consider the best $\text{LS}_{3,3}(1, 1)$ approximation to M

$$\min_{X \in \mathbb{R}^{3 \times 3}} \|X - M\|_F, \quad \text{s.t. } X \in \text{LS}_{3,3}(1, 1),$$

with $M = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \leftarrow \underbrace{\begin{bmatrix} 0 & 1 & 1 \\ 1 & \varepsilon & \varepsilon \\ 1 & \varepsilon & \varepsilon \end{bmatrix}}_{X_\varepsilon \in \text{LS}_{3,3}(1,1)} = \underbrace{\begin{bmatrix} 1/\varepsilon & 1 & 1 \\ 1 & \varepsilon & \varepsilon \\ 1 & \varepsilon & \varepsilon \end{bmatrix}}_{L_\varepsilon} + \underbrace{\begin{bmatrix} -1/\varepsilon & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{S_\varepsilon}.$



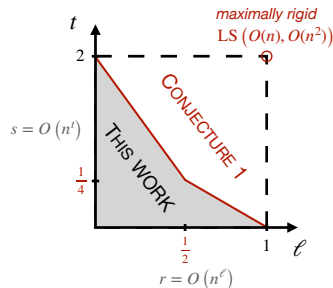
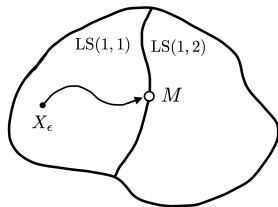
- As $\varepsilon \rightarrow 0$, the error $\|X_\varepsilon - M\|_F = 2\varepsilon \rightarrow 0$.
- However, X_ε converges to M which is outside of the feasible set $\text{LS}_{3,3}(1, 1)$.
- As $\varepsilon \rightarrow 0$ $\|L_\varepsilon\|_F$ and $\|S_\varepsilon\|_F$ become unbounded.

Generalization of the previous example

Theorem $(\text{LS}_n(r, s))$ is not closed for a range of $r, s \in \mathbb{N}$ ¹

The set of low-rank plus sparse matrices $\text{LS}_n(r, s)$ is not closed for $r \geq 1, s \geq 1$ provided $(r+1)(s+2) \leq n$, or provided $(r+2)^{3/2}s^{1/2} \leq n$ where s is of the form $s = p^2 r$ for an integer $p \geq 1$.

As a consequence, there are matrices $M \in \mathbb{R}^{n \times n}$ for which Robust PCA and low-rank matrix completion are ill-posed in the sense that they have no global minimum.

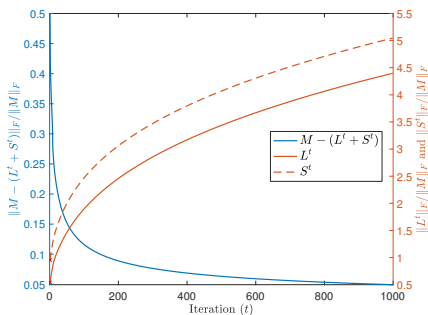


¹Tanner, Thompson & Vary, (2019), Matrix rigidity and the ill-posedness of Robust PCA and matrix completion

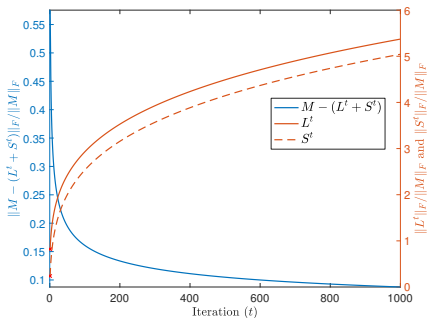
Divergence of non-convex Robust PCA

Solve $\min_X \|X - M\|_F$, s.t. $X \in \text{LS}_3(1, 1)$ for the following matrices

$$M^{(1)} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad M^{(2)} = \begin{bmatrix} 1 & -2 & -2 \\ -2 & 0 & 0 \\ -2 & 0 & 0 \end{bmatrix},$$



(a) FastGD (Yi *et al.*, 2016) for $M^{(1)}$



(b) GoDec (Liu *et al.*, 2013) for $M^{(2)}$

Part 1: Closing the $\text{LS}_{m,n}(r, s)$ set

Restrict the norm of one of the components

The following set

$$\text{LS}_{m,n}^{\tau}(r, s) = \{X = L + S \in \mathbb{R}^{m \times n} : \text{rank}(L) \leq r, \|S\|_0 \leq s, \|L\|_F \leq \tau \|X\|_F\}$$

is the Minkowski sum of a closed and a compact set and therefore it is a **closed** set.

Restrict the incoherence of the low-rank component

$$\text{LS}_{m,n}(r, s, \mu) = \left\{ L + S \in \mathbb{R}^{m \times n} : \begin{array}{l} \text{rank}(L) \leq r, \|S\|_0 \leq s \\ \max_{i \in \{1, \dots, m\}} \|U^T e_i\|_2 \leq \sqrt{\frac{\mu r}{m}} \\ \max_{i \in \{1, \dots, n\}} \|V^T f_i\|_2 \leq \sqrt{\frac{\mu r}{n}} \end{array} \right\}.$$

Subset relations between the sets (Lemma 2.10; Vary 2021)

When $\gamma_{r,s,\mu} := \mu \frac{r\sqrt{s}}{\sqrt{mn}} < 1$ (i.e. $\mu < \sqrt{mn}/(r\sqrt{s})$), we have that

$$\text{LS}_{m,n}(r, s, \mu) \subset \text{LS}_{m,n}^{\tau}(r, s),$$

for $\tau = (1 - \gamma_{r,s,\mu}^2)^{-1/2}$. As a consequence, $\text{LS}_{m,n}(r, s, \mu)$ is a closed set when $\mu < \sqrt{mn}/(r\sqrt{s})$.

Part 2: Restricted isometry constants (RICs) for the additive structure (Tanner & Vary, 2020)

Definition: Restricted isometry constants for $\text{LS}_{m,n}(r, s, \mu)$

For every pair of integers (r, s) and every $1 \leq \mu \leq \sqrt{mn}/r$, define the (r, s, μ) -restricted isometry constant to be the smallest $\Delta_{r,s,\mu} > 0$ such that

$$(1 - \Delta_{r,s,\mu}) \|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1 + \Delta_{r,s,\mu}) \|X\|_F^2,$$

for all matrices $X \in \text{LS}_{m,n}(r, s, \mu)$.

Theorem: Restricted isometry constants for random Gaussian subsampling and $\text{LS}_{m,n}(r, s, \mu)$

For a given $m, n, p \in \mathbb{N}$, $\Delta \in (0, 1)$, $\mu \in \left[1, \frac{\sqrt{mn}}{r\sqrt{s}}\right)$, and a random Gaussian subsampling transform $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p$ there exist constants $c_0, c_1 > 0$ such that the RIC for $\text{LS}_{m,n}(r, s, \mu)$ is upper bounded with $\Delta_{r,s,\mu} \leq \Delta$ provided

$$p > c_0 (r(m + n - r) + s) \log \left(\left(1 - \gamma^2\right)^{-1/2} \frac{mn}{s} \right), \quad (1)$$

with probability at least $1 - \exp(-c_1 p)$, where c_0, c_1 are constants that depend only on Δ .

Part 3: Controlling correlation between the two components

Rank-sparsity correlation bound (Lemma 2.1; Vary, 2021)

If L is a rank- r matrix that is μ -incoherent and S is an s -sparse matrix

$$|\langle L, S \rangle| \leq \gamma_{r,s,\mu} \|L\|_F \|S\|_F,$$

where $\gamma_{r,s,\mu} := \mu \frac{r\sqrt{s}}{\sqrt{mn}}$. The bound is only meaningful, when $\gamma_{r,s,\mu} < 1$, i.e. $\mu < \frac{\sqrt{mn}}{r\sqrt{s}}$.

Upper bound on the correlation in the subsampled space (Lemma 4.3; Vary, 2021)

For an operator $\mathcal{A}(\cdot)$ which has RICs bounded by $\Delta_{2r,2s,\mu} < \Delta_2$ and two incoherent low-rank plus sparse matrices $X_1 = L_1 + S_1 \in \text{LS}_{m,n}(r, s, \mu)$, $X_2 = L_2 + S_2 \in \text{LS}_{m,n}(r, s, \mu)$ that have orthogonal components $\langle L_1, L_2 \rangle = 0$, $\langle S_1, S_2 \rangle = 0$ and have bounded $\mu < \frac{\sqrt{mn}}{2r\sqrt{2s}}$, we have that

$$\left| \langle \mathcal{A}(X_1), \mathcal{A}(X_2) \rangle \right| \leq \left(\Delta_2 + \frac{2\gamma_2}{1 - \gamma_2^2} \right) \|X_1\|_F \|X_2\|_F, \quad (2)$$

where $\gamma_2 = \mu \frac{2r\sqrt{2s}}{\sqrt{mn}}$.

Convex recovery

Let $X_0 = L_0 + S_0 \in \text{LS}(r, s, \tau)$ and $b = \mathcal{A}(X_0)$ where $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p$ is a linear subsampling operator satisfying the RIP for low-rank plus sparse matrices.

$$(L^*, S^*) = \arg \min_{L, S \in \mathbb{R}^{m \times n}} \|L\|_* + \lambda \|S\|_1, \\ \text{s.t. } \mathcal{A}(L + S) = b.$$

Nonconvex recovery

Algorithm 1 NAHT (Tanner & Vary, 2020)

- 1: **while** not converged **do**
 - 2: Compute the residual $R_L^j = \mathcal{A}^* (\mathcal{A}(X^j) - b)$
 - 3: Set $V^j = L^j - \alpha_j^L R_L^j$
 - 4: Set $L^{j+1} = \text{HT}(V^j; r)$
 - 5: Set $X^{j+\frac{1}{2}} = L^{j+1} + S^j$
 - 6: Compute the residual $R_S^j = \mathcal{A}^* (\mathcal{A}(X^{j+\frac{1}{2}}) - b)$
 - 7: Set $W^j = S^j - \alpha_j^S R_S^j$
 - 8: Set $S^{j+1} = \text{HT}(W^j; s)$
 - 9: Set $X^{j+1} = L^{j+1} + S^{j+1}$
 - 10: $j = j + 1$
 - 11: **end while**
-

Guaranteed convex recovery (Theorem 4.2; Tanner & Vary, 2021)

Let $b = \mathcal{A}(X_0)$ and suppose that $r, s \in \mathbb{N}$ and $\mu < \sqrt{mn}/(4r\sqrt{3s})$ are such that the restricted isometry constant $\Delta_{4r,3s,\mu} \leq \frac{1}{5} - 12\mu \frac{r\sqrt{s}}{\sqrt{mn}}$. Let $X_* = L_* + S_*$ be the solution of the convex relaxation with $\lambda = \sqrt{r/s}$, then $X_* = X_0$.

Global convergence of NAHT (Theorem 4.4; Tanner & Vary, 2021)

Suppose that $r, s \in \mathbb{N}$ and $\mu < \sqrt{mn}/(3r\sqrt{3s})$ are such that the restricted isometry constant $\Delta_3 := \Delta_{3r,3s,\mu} < \frac{1}{9} - 3\mu \frac{r\sqrt{s}}{\sqrt{mn}}$. Then NAHT applied to $b = \mathcal{A}(X_0)$ as described in NAHT Algorithm will linearly converge to $X_0 = L_0 + S_0$ as

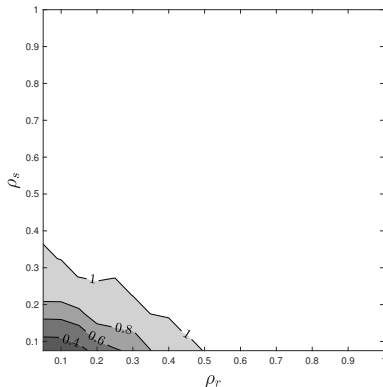
$$\left\|L^{j+1} - L_0\right\|_F + \left\|S^{j+1} - S_0\right\|_F \leq \frac{6\Delta_3 + \frac{9}{2}\gamma_2}{1 - 3\Delta_3 - \frac{9}{2}\gamma_2} \left(\left\|L^j - L_0\right\|_F + \left\|S^j - S_0\right\|_F\right), \quad (3)$$

where $\gamma_2 := \frac{2r\sqrt{2s}}{\sqrt{mn}}$.

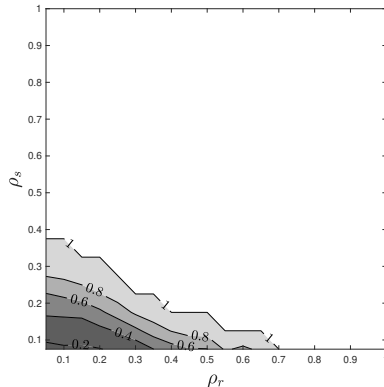
Numerical phase transition: Convex relaxation and NAHT

Phase transition δ^* above which recovery is possible, where

$$\text{subsampling: } \delta = \frac{p}{mn}, \quad \text{rank: } \rho_r = \frac{r(m+n-r)}{p}, \quad \text{sparsity: } \rho_s = s/p$$



(a) Convex recovery for 30×30 matrix, $\mu \approx 3$.



(b) NAHT recovery for 30×30 matrix, $\mu \approx 3$.

Thank you for your attention.

References

CANDÈS, EMMANUEL J., LI, XIAODONG, MA, YI, & WRIGHT, JOHN. 2011.

Robust principal component analysis?

Journal of the acm, **58**(3), 1–37.

LIN, ZHOUCHE, CHEN, M., WU, L., & MA, Y. 2009.

The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices.

Technical report.

LIU, GUANGCAN, LIN, ZHOUCHE, YAN, SHUICHENG, SUN, JU, YU, YONG, & MA, YI. 2013.

Robust recovery of subspace structures by low-rank representation.

IEEE transactions on pattern analysis and machine intelligence, **35**(1), 171–184.

NETRAPALLI, PRANEETH, NIRANJAN, U N, SANGHAVI, SUJAY, ANANDKUMAR, ANIMASHREE, & JAIN, PRATEEK. 2014.

Non-convex robust PCA.

Advances in neural information processing systems.

RECHT, BENJAMIN, FAZEL, MARYAM, & PARRILO, PABLO A. 2010.

Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization.

Siam review, **52**(3), 471–501.

YI, XINYANG, PARK, DOHYUNG, CHEN, YUDONG, & CARAMANIS, CONSTANTINE. 2016.

Fast algorithms for robust PCA via gradient descent.

Advances in neural information processing systems.