Compressed sensing of low-rank plus sparse matrices

Simon Vary

simonvary.github.io/slides/SeLMA21.pdf

ICTEAM, Université catholique de Louvain

joint work with Jared Tanner & Andrew Thompson
Mathematical Institute, University of Oxford, United Kingdom

EOS SeLMA Seminar, 15/6/2021

Overview

Plan for the talk:

1. Models using rank or sparsity
2. The set of low-rank plus sparse matrices is not closed
3. Theory: (i) Closing the set, (ii) Restricted isometry constants, (iii) Convergence guarantees
4. Numerical experiments

Publications covered:

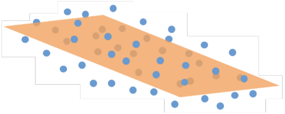
Matrix rigidity and the ill-posedness of Robust PCA and matrix completion,
SIAM Journal on Mathematics of Data Science, 1(3), 537–554, 2019.
Jared Tanner, Andrew Thompson, Simon Vary

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 $M = [y_1, \ldots, y_n]$ with mean-centered samples y_i



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

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

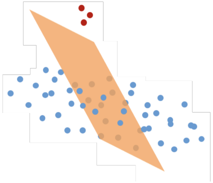
$$\min_{X \in \mathbb{R}^{m \times n}} \|\mathcal{A}(X) - b\|_F, \quad \text{s.t.} \quad \operatorname{rank}(X) \leq r$$

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

What if the low-rank structure is corrupted?

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

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



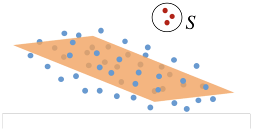
Define the following set

$$\mathrm{LS}_{m,n}(r,s) = \left\{L + S \in \mathbb{R}^{m\times n} : \operatorname{rank}(L) \le r,\ \|S\|_0 \le s\right\}.$$

Robust PCA (Candès et al., 2011)

For a given matrix M, find X

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



Low-rank plus sparse matrix sensing (Tanner & Vary, 2020)

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

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

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

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 $\mu \in \left[1, \sqrt{mn}\right]$

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

Convex problem

$$\min_{L,S \in \mathbb{R}^{n \times n}} \|L\|_* + \lambda \|S\|_1, \quad \text{s.t.} \quad M = L + S$$

- Solves also the non-convex problem
- $\|\cdot\|_*$ requires the <u>full</u> SVD $\Rightarrow \mathcal{O}(n^3)$
- PCP (Candès *et al.*, 2011)
  IALM (Lin *et al.*, 2010)
- Optimal rate: $s = \mathcal{O}\left(1/(\mu^2 r^2)\right)$

Non-convex problem

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

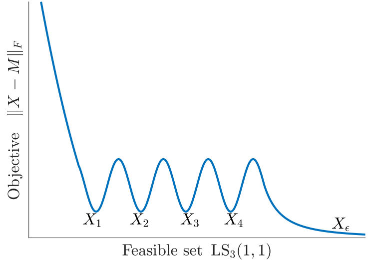
- $L = YZ^T$ for $Y, Z \in \mathbb{R}^{n \times r}$
- Alternating projections for L, S
- Per iteration complexity:
  AltProj, $\mathcal{O}(r^2 n^2)$ (Netrapalli *et al.*, 2014)
  FastGD, $\mathcal{O}(rn^2)$ (Yi *et al.*, 2016)
- Best proved rate: $s = \mathcal{O}\left(1/(\mu^2 r^2 m \log(m))\right)$

Simple example of non-closedness

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

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

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



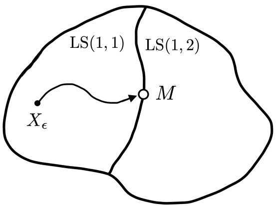
- As $\varepsilon \to 0$, the error $\|X_\varepsilon - M\|_F = 2\varepsilon \to 0$.
- However, X_ε converges to M which is outside of the feasible set $\mathrm{LS}_{3,3}(1,1)$.
- As $\varepsilon \to 0$ $\|L_\varepsilon\|_F$ and $\|S_\varepsilon\|_F$ become unbounded.

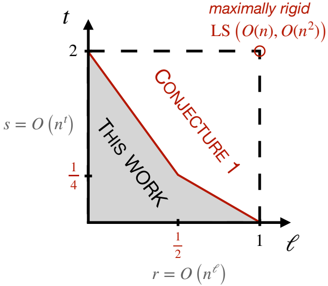
Generalization of the previous example

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

The set of low-rank plus sparse matrices $\mathrm{LS}_n(r,s)$ *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.



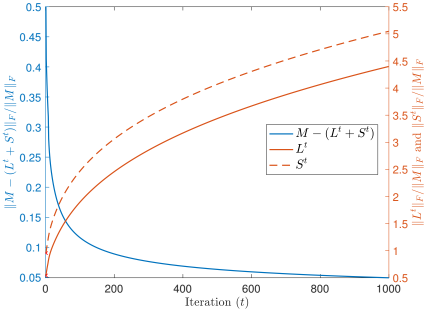


[1] Tanner, Thompson & Vary. (2019). Matrix rigidity and the ill-posedness of Robust PCA and matrix completion

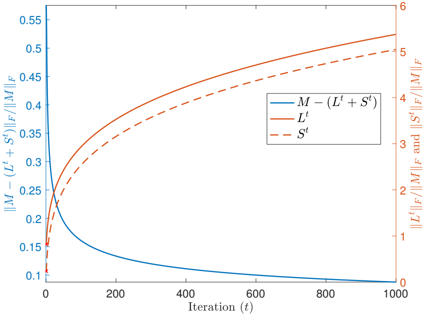
Divergence of non-convex Robust PCA

Solve $\min_X \|X - M\|_F$, s.t. $X \in \mathrm{LS}_3(1,1)$ 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 (Zhou & Tao, 2011) for $M^{(2)}$

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

Restrict the norm of one of the components

The following set

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

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

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

Lemma 1 (Subset relation between the LS sets)

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

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

where $\tau = \left(1 - \gamma^2_{r,s,\mu}\right)^{-1/2}$. *As a consequence,* $\mathrm{LS}_{m,n}(r,s,\mu)$ *is a closed set when* $\mu < \sqrt{mn}/(r\sqrt{s})$.

Definition 1 (Restricted isometry constants for $\mathrm{LS}_{m,n}(r,s,\mu)$)

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

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

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

Theorem 2 (Upper bound on the RICs for $\mathrm{LS}_{m,n}(r,s,\mu)$)

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 $\mathcal{A}: \mathbb{R}^{m\times n} \to \mathbb{R}^p$ there exist constants $c_0, c_1 > 0$ such that the RIC for $\mathrm{LS}_{m,n}(r,s,\mu)$ is upper bounded with $\Delta_{r,s,\mu} \le \Delta$ provided

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

with probability at least $1-\exp(-c_1 p)$, where c_0, c_1 are constants that depend only on Δ and $\gamma = \mu\frac{r\sqrt{s}}{\sqrt{mn}}$.

Part 3: Compressed sensing of low-rank plus sparse matrices

Convex recovery

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

$$(L^*, S^*) = \underset{L,S\in\mathbb{R}^{m\times n}}{\arg\min} \|L\|_* + \lambda\|S\|_1,$$

$$\text{s.t.} \quad \mathcal{A}(L+S) = b.$$

Nonconvex recovery

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

Algorithm 1 NAHT (Tanner & Vary, 2020)

while not converged **do**
  Compute the residual $R_L^j = \mathcal{A}^*\left(\mathcal{A}(X^j) - b\right)$
  Set $V^j = L^j - \alpha_j^L R_L^j$
  Set $L^{j+1} = \mathrm{HT}(V^j;\, r)$
  Set $X^{j+\frac{1}{2}} = L^{j+1} + S^j$
  Compute the residual $R_S^j = \mathcal{A}^*\left(\mathcal{A}(X^{j+\frac{1}{2}}) - b\right)$
  Set $W^j = S^j - \alpha_j^S R_S^j$
  Set $S^{j+1} = \mathrm{HT}(W^j;\, s)$
  Set $X^{j+1} = L^{j+1} + S^{j+1}$
  $j = j + 1$
end while

Theorem 3 (Guaranteed convex recovery)
Let $b = \mathcal{A}(X_0)$ and suppose that $r, s \in \mathbb{N}$ and $\mu < \sqrt{mn}/\left(4r\sqrt{3s}\right)$ 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 $X_* = X_0$.

Theorem 4 (Guaranteed recovery by NAHT)
Suppose that $r, s \in \mathbb{N}$ and $\mu < \sqrt{mn}/\left(3r\sqrt{3s}\right)$ 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

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

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

Lemma 2 (Rank-sparsity correlation bound)

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. $\mu < \frac{\sqrt{mn}}{r\sqrt{s}}$.

Lemma 3 (Upper bound on the correlation in the subsampled space)

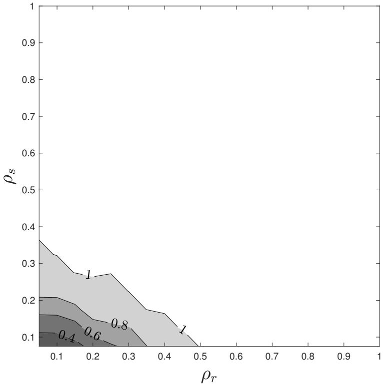
For an operator $\mathcal{A}(\cdot)$ which has RICs bounded by $\Delta_{2r,2s,\mu} < \Delta_2$ and two incoherent low-rank plus sparse matrices $X_1 = L_1 + S_1 \in \mathrm{LS}_{m,n}(r, s, \mu)$, $X_2 = L_2 + S_2 \in \mathrm{LS}_{m,n}(r, s, \mu)$ that have orthogonal components $\langle L_1, L_2\rangle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

$$\left|\langle \mathcal{A}(X_1), \mathcal{A}(X_2)\rangle\right| \leq \left(\Delta_2 + \frac{2\gamma_2}{1-\gamma_2^2}\right) \|X_1\|_F \|X_2\|_F, \tag{3}$$

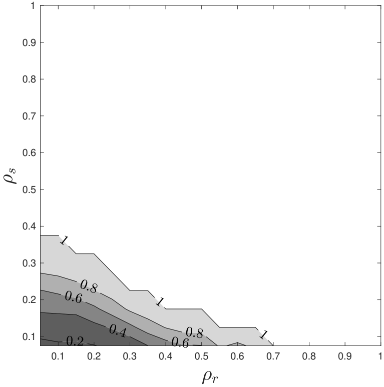
where $\gamma_2 = \mu \frac{2r\sqrt{2s}}{\sqrt{mn}}$.

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

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

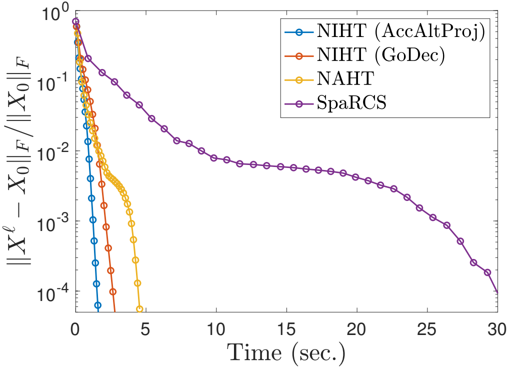


(a) Convex recovery for 30 × 30 matrix, $\mu \approx 3$.

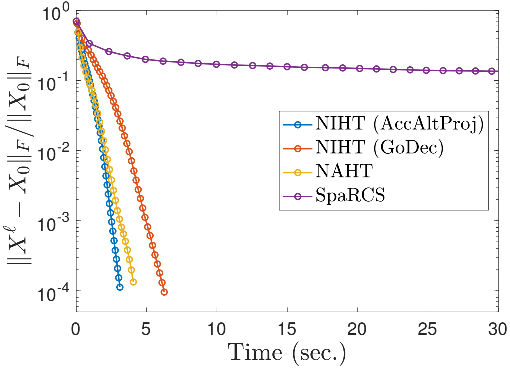


(b) NAHT recovery for 30 × 30 matrix, $\mu \approx 3$.

Linear convergence of non-convex recovery



(a) $\rho_r = \rho_s = 0.05$



(b) $\rho_r = \rho_s = 0.1$

Figure 3: Relative error in the approximate $\|X^\ell\|$ as a function of time for synthetic problems with $m = n = 100$ and $p = (1/2)100^2$, $\delta = 1/2$ for Gaussian linear measurements $\mathcal{A}$, and $\mu \approx 3$. In (b), SpaRCS converged in 171 sec. (45 iterations).

Subsampled dynamic-foreground/static-background seperation

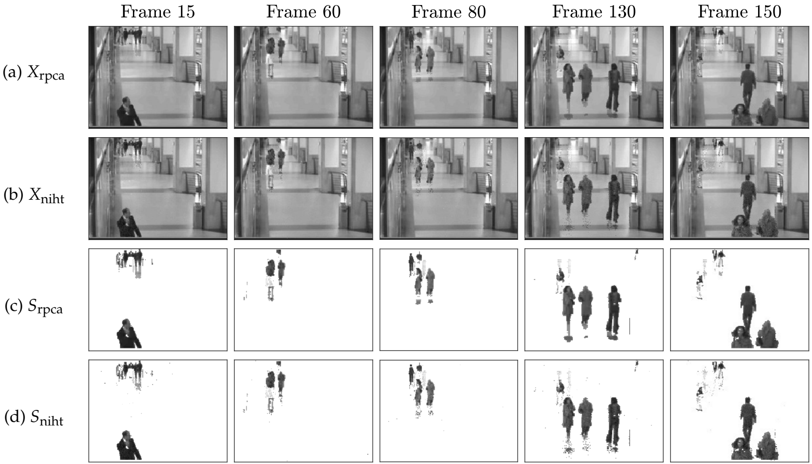


Figure 4: Recovery of a $190 \times 140 \times 150$ video sequence compared to the approximation of the complete video by Robust PCA. The video is shaped into 26600×150 and recovered using FJLT from $\delta = 1/3$ using $r = 1$ and $s = 197505$. Recovery from the subsampled data achieves PSNR of 34.5 dB and the Robust PCA 35.5 dB.

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, Zhouchen, Chen, Minming, & Ma, Yi. 2010.
The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices.
sep.

Netrapalli, Praneeth, Niranjan, U N, Sanghavi, Sujay, Anandkumar, Animashree, & Jain, Prateek. 2014.
Non-convex robust PCA.
In: Advances in neural information processing systems 27.

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.

TANNER, JARED, & VARY, SIMON. 2020.
Compressed Sensing of Low-rank Plus Sparse Matrices.
jul, 1–35.

YI, XINYANG, PARK, DOHYUNG, CHEN, YUDONG, & CARAMANIS, CONSTANTINE. 2016.
Fast algorithms for robust PCA via gradient descent.
In: Advances in neural information processing systems 29.

ZHOU, TIANYI, & TAO, DACHENG. 2011.
GoDec: Randomized low-rank & sparse matrix decomposition in noisy case.
Proceedings of the 28th international conference on machine learning, **35**(1), 33–40.