Compressed sensing of low-rank plus sparse matrices

Simon Vary

ICTEAM, Université catholique de Louvain

joint work with Jared Tanner & Andrew Thompson

Seminar in Mathematical Engineering 26/10/2021

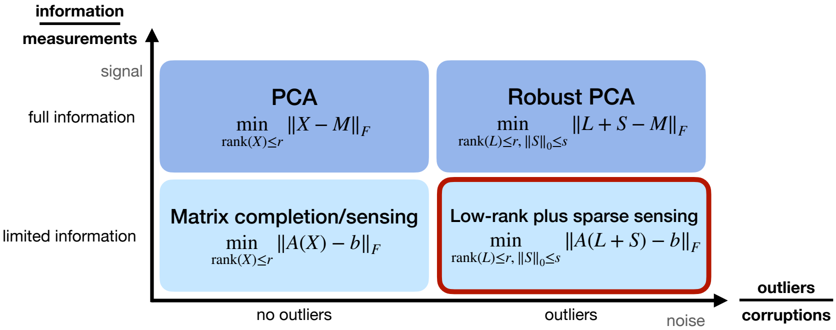
Compressed sensing of low-rank plus sparse matrices

Consider the following problem:

- $M = L_0 + S_0 \in \mathbb{R}^{m \times n}$ such that $\text{rank}(L_0) \le r$ and $\|S_0\|_0 \le s$,
- $b = \mathcal{A}(M) \in \mathbb{R}^p$ be a vector of $p < mn$ linear measurements from $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$.

The question is:

Can we recover L_0 and S_0 only from the subsampled information $b = \mathcal{A}(M)$ and $\mathcal{A}(\cdot)$?



Subsampled dynamic-foreground/static-background seperation

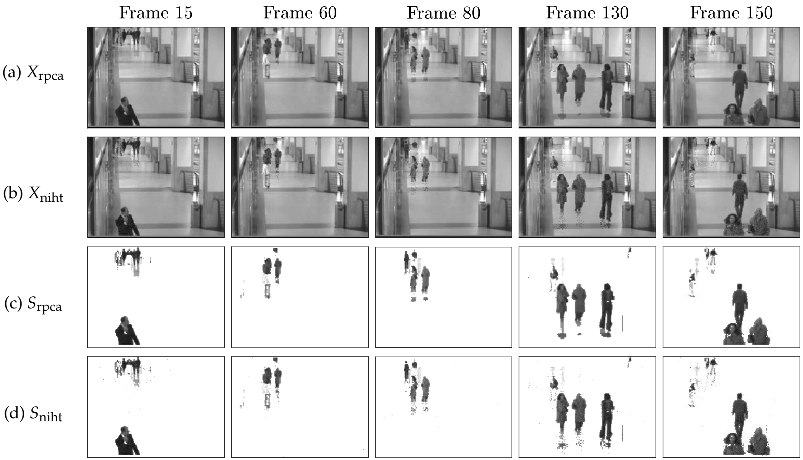


Figure 1: Recovery of a $190 \times 140 \times 150$ video sequence. The video is shaped into 26600×150 and recovered using FJLT from $\delta = 1/3$ using $r = 1$ and $s = 197505$.

Computable solution under conditions on identifiability and recoverability

Under some conditions on

- the structure of the matrix matrix $M = L_0 + S_0$ (idenitifiability) and
- the linear subsampling $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ (recoverability),

we can retrieve L and S from the subsampled measurement vector $b = \mathcal{A}(M)$, either by solving the convex optimization problem

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

or by solving the following non-convex optimization problem

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

where

$$\mathrm{LS}_{m,n}(r,s) = \left\{L + S \in \mathbb{R}^{m\times n} : \ \mathrm{rank}\,(L) \le r, \ \|S\|_0 \le s\right\}. \qquad (1)$$

Identifiability: Which matrices (L, S) can we hope to identify?

We wish to avoid cases of $X_0 = L_0 + S_0$ which is simultaneously low-rank and sparse. In linearized form this requires that the tangent spaces $T(L_0)$ and $\Omega(S_0)$ intersect transversally

$$T(L_0) \cap \Omega(S_0) = \{0\} . \tag{2}$$

Incoherence of the low-rank component (Candès & Recht, 2009):

Correlation of the 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 $\mu \in \left[1, \sqrt{mn/r}\right]$

$$\max_{i\in\{1,\dots,r\}} \left\|U^T e_i\right\|_2 \leq \sqrt{\frac{\mu r}{m}}, \qquad \max_{i\in\{1,\dots,r\}} \left\|V^T e_i\right\|_2 \leq \sqrt{\frac{\mu r}{n}}. \tag{3}$$

Sparsity pattern of the sparse component (Chandrasekaran et al., 2011):

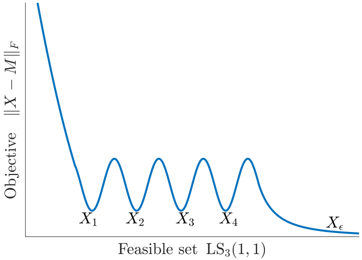
$$\exists\alpha \in [0, 1) : \quad \|S^T e_i\|_0 \leq \alpha n, \qquad \|S e_j\|_0 \leq \alpha m, \tag{4}$$

Non-closedness: a simple example

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

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

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

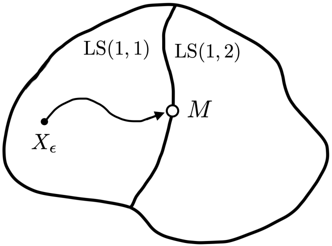


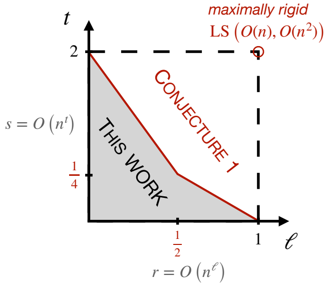
- As $\varepsilon \to 0$, the error $\|X_\varepsilon - M\|_F = 2\varepsilon \to 0$.
- However, X_ε converges to M which is outside of the feasible set $\mathrm{LS}_{3,3}(1,1)$.
- As $\varepsilon \to 0$ $\|L_\varepsilon\|_F$ and $\|S_\varepsilon\|_F$ become unbounded.

Non-closedness generalization

Theorem ($\mathrm{LS}_{n,n}(r, s)$ 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

Closing the set and making the pair (L, S) identifiable

Restrict the incoherence of the low-rank component

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

This also guarantees that $\|L\|_F \leq \left(1 - \gamma_{r,s,\mu}^2\right)^{-1/2} \|X\|_F$ with $\gamma_{r,s,\mu} := \mu \frac{r\sqrt{s}}{\sqrt{mn}}$.
As a consequence the set $\mathrm{LS}_{m,n}(r,s,\mu)$ is closed when $\mu < \sqrt{mn}/(r\sqrt{s}) = 1/(\alpha r)$.

The fully observed case: Robust PCA

The solution to the convex problem $(\star)$ with $\mathcal{A} = \mathrm{Id}$ identifies (L, S) from M when

$$\mu < \mathcal{O}\left(\sqrt{mn}/(r\sqrt{s}) = 1/(\alpha r)\right), \qquad \text{(Hsu et al., 2011)} \tag{5}$$

and there exists an algorithm for the non-convex problem $(\star\star)$ when

$$\mu < \mathcal{O}\left(\sqrt{mn}/(r^{1.5}\sqrt{s}) = 1/(\alpha r^{1.5})\right) \qquad \text{(Yi et al., 2016; Wei et al., 2019).} \tag{6}$$

Recoverability: For which $\mathcal{A}(\cdot)$ can we recover (L, S) from $b = \mathcal{A}(b)$?

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

$$(1 - \Delta_{r,s,\mu}) \|X\|_F^2 \leq \| \mathcal{A}(X)\|_2^2 \leq (1 + \Delta_{r,s,\mu}) \|X\|_F^2, \quad (7)$$

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

Suppose that $\Delta_{2r,2s,\mu}(\mathcal{A}) < 1$ for some integers $r, s \geq 1$ and $\mu \geq 1$

Let $X_0, X_1 \in \text{LS}_{m,n}(r, s, \mu)$ and $b_0 = \mathcal{A}(X_0)$, $b_1 = \mathcal{A}(X_1)$. Then

$$(1 - \Delta_{2r,2s,\mu}) \|X_0 - X_1\|_F^2 \leq \| \mathcal{A}(X_0 - X_1)\|_2^2 \leq (1 + \Delta_{2r,2s,\mu}) \|X_0 - X_1\|_F^2, \quad (8)$$

since $X_0 - X_1 \in \text{LS}_{m,n}(2r, 2s, \mu)$.

Theorem (Bound on the RICs for $\mathrm{LS}_{m,n}(r, s, \mu)$)[2]:

For a given $m, n, p \in \mathbb{N}$, $\Delta \in (0, 1)$, $\mu < \frac{\sqrt{mn}}{r\sqrt{s}}$, 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 $\Delta_{r,s,\mu} \leq \Delta$ provided

$$p > c_0 \left(r(m + n - r) + s \right) \log \left(\left(1 - \gamma^2\right)^{-1/2} \frac{mn}{s} \right), \tag{9}$$

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

[2] Tanner & Vary. (2020). Compressed sensing of low-rank plus sparse matrices

Recovery by the convex relaxation

Recall the convex optimization problem

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

Theorem (Guaranteed convex recovery):

Let $b = \mathcal{A}(X_0)$ and suppose that $r, s \in \mathbb{N}$ and $\mu < 1/\left(4\sqrt{3}r\alpha\right)$ are such that the RICs satisfy

$$\Delta_{4r,3s,\mu} \leq \frac{1}{5} - 12\mu r\alpha, \tag{10}$$

and $X_* = L_* + S_*$ be the solution of the convex relaxation with $\lambda = \sqrt{r/s}$, then $X_* = X_0$.

Recall the non-convex optimization:

$$\min_{X \in \mathsf{LS}_{m,n}(r,s,\mu)} \| \mathcal{A}(X) - b \|_F. \quad (\star\star)$$

Algorithm 1 NAHT

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

Incorrect theorem ((Not yet) 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

$$\Delta_3 := \Delta_{3r,3s,\mu} < \frac{1}{9} - 3\mu\frac{r\sqrt{s}}{\sqrt{mn}}, \tag{11}$$

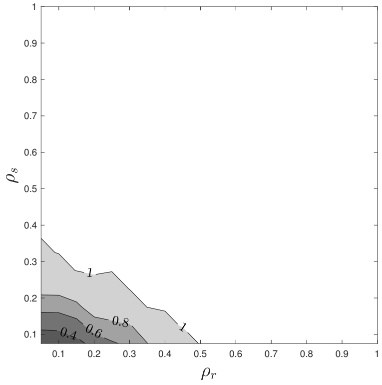
then NAHT applied to $b = \mathcal{A}(X_0)$ as described in NAHT Algorithm will linearly converge to $X_0 = L_0 + S_0$ as

$$\left\|L^{j+1} - L_0\right\|_F + \left\|S^{j+1} - S_0\right\|_F \leq \frac{6\Delta_3 + \frac{9}{2}\gamma_2}{1 - 3\Delta_3 - \frac{9}{2}\gamma_2}\left(\left\|L^{j} - L_0\right\|_F + \left\|S^{j} - S_0\right\|_F\right), \tag{12}$$

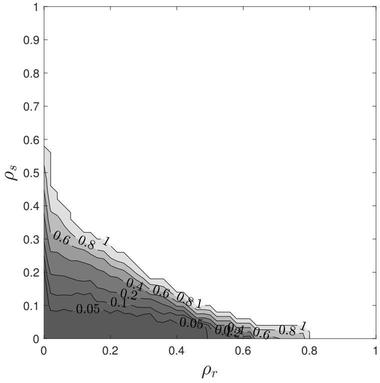
where $\gamma_2 := \frac{2r\sqrt{2s}}{\sqrt{mn}}$.

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 100×100 matrix, $\mu \approx 3$.

Linear convergence of non-convex recovery

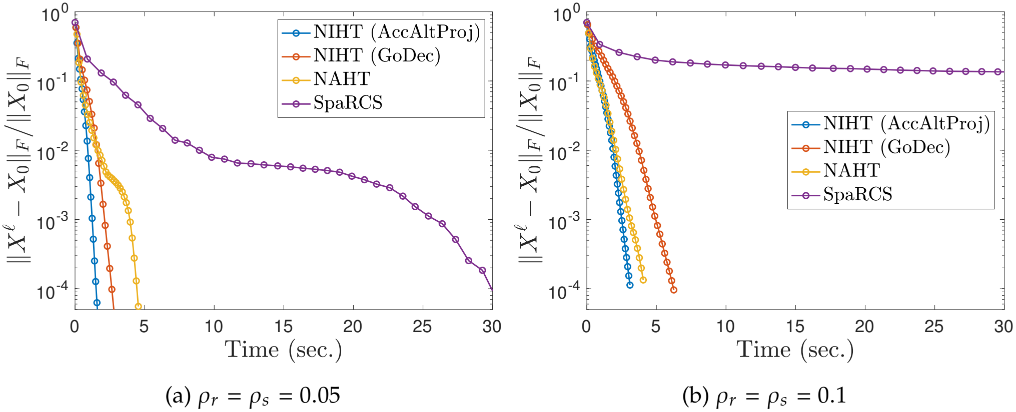


Figure 3: Relative error in the approximate $\|X^\ell\|$ for $m = n = 100$ and $p = (1/2)100^2$, $\delta = 1/2$ and Gaussian $\mathcal{A}$, and $\mu \approx 3$. In (b), SpaRCS (Waters et al., 2011) converged in 171 sec. (45 iterations).

Thank you for your attention.

References

BLANCHARD, JEFFREY D., TANNER, JARED, & WEI, KE. 2015.
CGIHT: Conjugate gradient iterative hard thresholding for compressed sensing and matrix completion.
Information and inference, nov, iav01.

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.

TANNER, JARED, & WEI, KE. 2016.
Low rank matrix completion by alternating steepest descent methods.
Applied and computational harmonic analysis, **40**(2), 417–429.

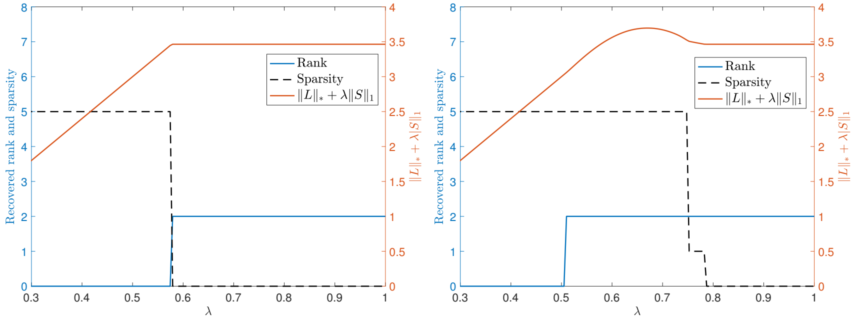
YI, XINYANG, PARK, DOHYUNG, CHEN, YUDONG, & CARAMANIS, CONSTANTINE. 2016.
Fast algorithms for robust PCA via gradient descent.
In: Advances in neural information processing systems 29.

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.

Problems with convex Robust PCA and non-closedness

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

where $\|\cdot\|_*$ is the nuclear norm (sum of the singular values of L) and $\|\cdot\|_1$ denotes the ℓ_1-norm (sum of the absolute values of the entries of S).



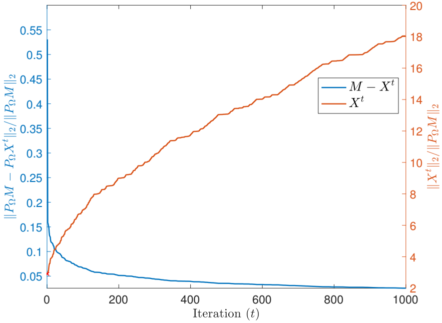
(a) PCP (Candès *et al.*, 2011) for $M^{(1)}$ (b) IALM (Lin *et al.*, 2010) for $M^{(1)}$

Divergence of non-convex low-rank matrix completion

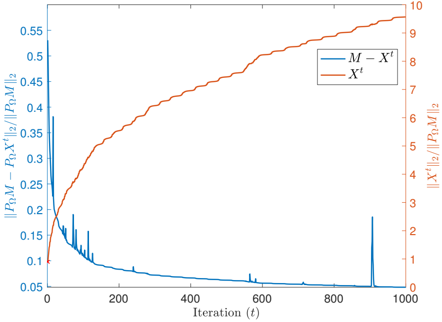
We are given only entries of M at indices Ω in the form of $b = P_\Omega(M)$. Solving

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

recovers M for many r and an entry-wise subsampling operator $P_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^p$.



(a) ASD (Tanner & Wei, 2016) for $M^{(1)}$



(b) CGIHT (Blanchard *et al.*, 2015) for $M^{(1)}$