Compressed sensing of low-rank plus sparse matrices

Simon Vary
ICTEAM, Université catholique de Louvain, Belgium

joint work with Jared Tanner & Andrew Thompson

XXI Householder Symposium 2022, 16/06/2022

Low-rank plus sparse model

Principal Component Analysis (PCA)

Correlation matrix $M = \frac{1}{N} YY^T$ of a mean-centered samples y_i,

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

or from subsamped data $b = \mathcal{A}(M) \in \mathbb{R}^p$

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

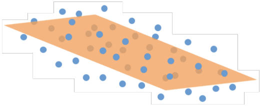
Robust PCA

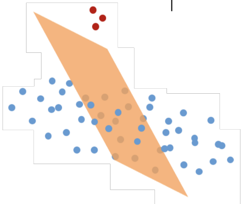
For a given matrix M, find X

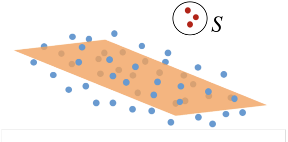
$$\min_{X \in \mathbb{R}^{m \times n}} \|X - M\|_F, \quad \text{s.t.} \quad X \in \mathrm{LS}_{m,n}(r,s).$$

where

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





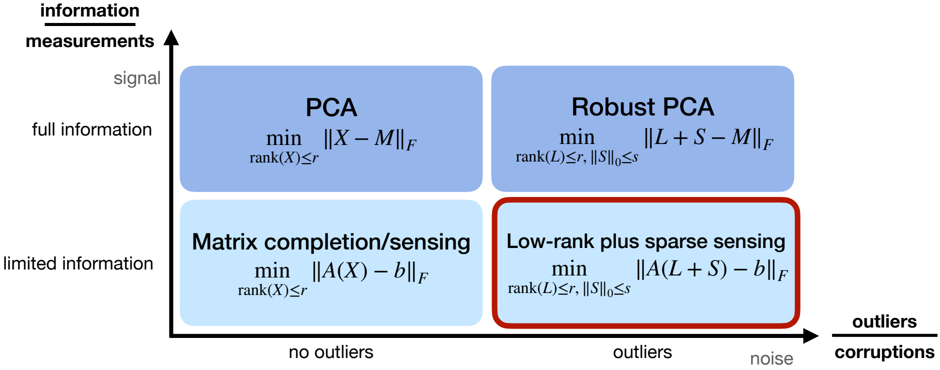


Compressed sensing of low-rank plus sparse matrices

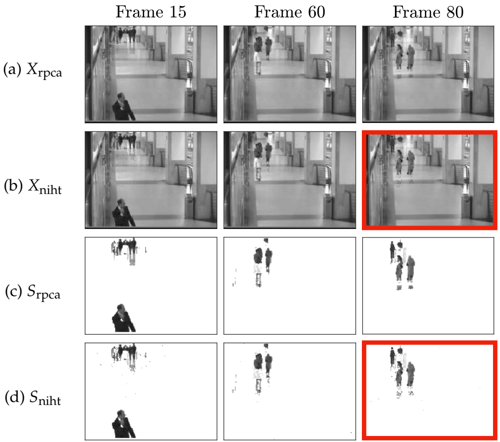
Let $b = \mathcal{A}(L_0 + S_0) \in \mathbb{R}^p$, where

- $\operatorname{rank}(L_0) \le r$ and $\|S_0\|_0 \le s$,
- $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a linear subsampling.

Recover L_0 and S_0 only from $b = \mathcal{A}(M)$ and $\mathcal{A}(\cdot)$?



Subsampled dynamic-foreground/static-background seperation



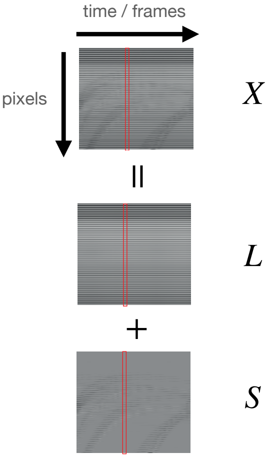


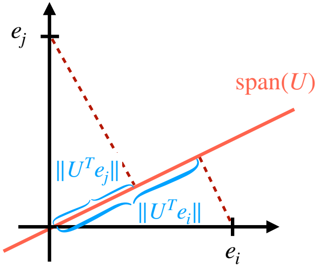
Figure 1: Recovery of a $190 \times 140 \times 150$ video sequence. The video is shaped into 26600×150 and recovered using FJLT from $\delta = 1/3$ using $r = 1$ and $s = 197505$.

Assumptions for Robust PCA recovery

Incoherence of L (Candès & Recht, 2009):

For the truncated SVD of $L = U\Sigma V^T$ we have

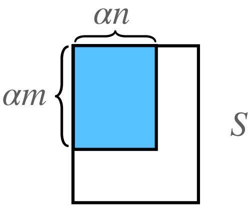
$$\exists \mu \geq 1 : \quad \begin{array}{l} \max_{i \in \{1,\dots,r\}} \left\| U^T e_i \right\|_2 \leq \sqrt{\frac{\mu r}{m}}, \\ \max_{i \in \{1,\dots,r\}} \left\| V^T e_i \right\|_2 \leq \sqrt{\frac{\mu r}{n}}. \end{array}$$



Sparsity pattern of S (Chandrasekaran et al., 2011):

For the sparse component $S \in \mathbb{R}^{m \times n}$

$$\exists \alpha \in [0,1) : \quad \begin{array}{l} \|S^T e_i\|_0 \leq \alpha n \\ \|S e_j\|_0 \leq \alpha m \end{array}$$



Existing recovery guarantees for Robust PCA

Convex relaxation

The solution to the convex problem

$$\underset{L,S\in\mathbb{R}^{m\times n}}{\arg\min} \|L\|_* + \lambda\|S\|_1, \qquad \text{s.t.} \quad L+S=M, \tag{1}$$

identifies (L,S) from M when $s < \mathcal{O}\left(mn/\left(\mu^2 r^2\right)\right)$ (Hsu et al., 2011).

Non-convex algorithms

Provable gradient descent methods for the non-convex problem($\star\star$) when

$$\underset{L+S\in\mathbb{R}^{m\times n}}{\arg\min} \|(L+S)-M\|_F, \quad \text{s.t.} \quad L+S\in \mathrm{LS}_{m,n}(r,s), \tag{2}$$

identifies (L,S) from M when $s < \mathcal{O}\left(mn/\left(\mu^2 r^3\right)\right)$ (Yi et al., 2016; Wei et al., 2019).

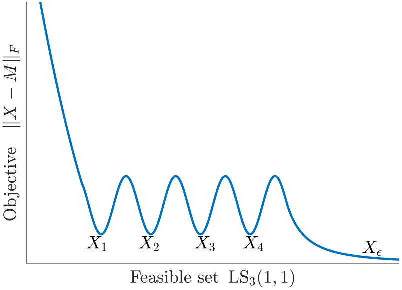
. . . and the support set of S is well spread with $\alpha \leq \sqrt{s/mn}$.

A 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.

Matrix rigidity in complexity theory

Matrix rigidity (Valiant 1977)

The smallest number of entries of M that need to be changed such that the rank becomes at most r:

$$\begin{aligned} Rig(M,r) &= \min\{\|S\|_0 : \operatorname{rank}(M-S) \le r\} = \min\{s : \quad M \in \mathrm{LS}_{m,n}(r,s)\}, \\ &\le (m-r)(n-r). \end{aligned}$$

It can happen that $M_\epsilon \in \mathrm{LS}_{m,n}(r,s) \to M$ and $Rig(M,r') > s$ where $r' \ge r$.

Related to complexity of linear transforms

- Lower bound of the form $Rig(A,\epsilon n) \ge n^{1+\delta}$, for some constants $\epsilon,\delta > 0$, implies that multiplication by $A \in \mathbb{R}^{n\times n}$ cannot be computed in $\mathcal{O}(n \log n)$.

Non-closedness $\iff$ $Rig(\cdot, r)$ is not semicontinuous (Kumar & Lokam 2014)

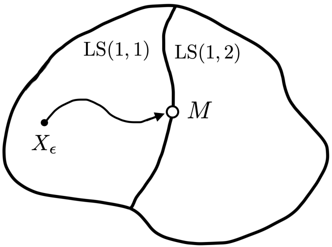
- Small perturbation A_ϵ of A ($\forall \epsilon > 0 : \|A - A_\epsilon\|_F < \epsilon$) may decrease $Rig(A,r) > Rig(A_\epsilon, r)$.

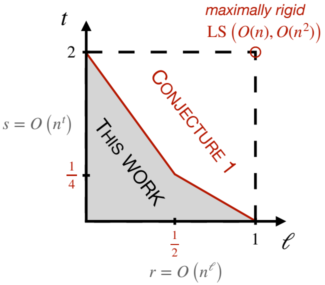
Non-closedness generalization

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 $\mathrm{LS}_{n,n}(r,s)$ 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, V. (2019). Matrix rigidity and the ill-posedness of Robust PCA and matrix completion

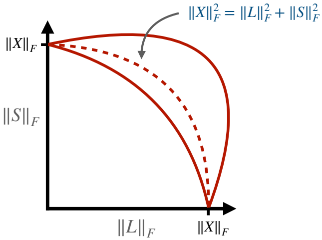
Lemma (Tanner & V., 2020):

Let $s < mn/(\mu^2 r^2)$ and $X = L + S \in \mathrm{LS}_{m,n}(r, s, \mu)$. Then the following holds

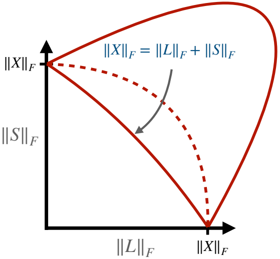
(i) $|\langle L, S\rangle| \leq \mu \frac{r\sqrt{s}}{\sqrt{mn}} \|L\|_F \|S\|_F$,

(ii) $\|L\|_F \leq \left(1 - \mu^2 \frac{r^2 s}{mn}\right)^{-1/2} \|X\|_F$ and $\|S\|_F \leq \left(1 - \mu^2 \frac{r^2 s}{mn}\right)^{-1/2} \|X\|_F$,

(iii) $\mathrm{LS}_{m,n}(r, s, \mu)$ is a closed set.



$\mu \approx 1$



$\mu \gg 1$

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_0 and S_0 from the subsampled measurement vector $b = \mathcal{A}(M)$, either by solving the convex optimization problem

$$(L^*, S^*) = \underset{L,S\in\mathbb{R}^{m\times n}}{\arg\min} \|L\|_* + \lambda\|S\|_1, \qquad \text{s.t.} \quad \|\mathcal{A}(L+S) - b\|_2 \le \epsilon_b, \tag{$\star$}$$

or by solving the following non-convex optimization problem

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

where

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

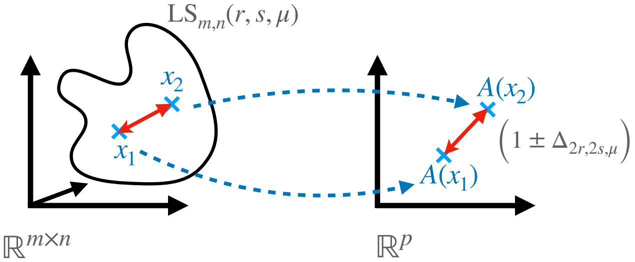
Recoverability via the restricted isometry property

Definition (Restricted isometry property of $\mathcal{A}$ on $\mathrm{LS}_{m,n}(r, s, \mu)$):

For a linear subsampling $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$, there exists $\Delta_{r,s,\mu} \in (0, 1)$ such that

$$(1 - \Delta_{r,s,\mu}) \|X\|_F^2 \leq \| \mathcal{A}(X)\|_2^2 \leq (1 + \Delta_{r,s,\mu}) \|X\|_F^2, \quad (3)$$

for all matrices $X \in \mathrm{LS}_{m,n}(r, s, \mu)$ whose low-rank component has bounded coherence by μ.

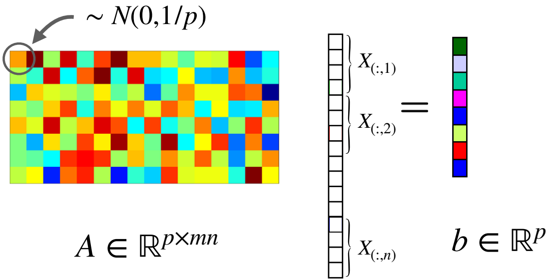


Theorem (Bound on the RICs for $\mathrm{LS}_{m,n}(r, s, \mu)$):

For given $m, n, p \in \mathbb{N}$, $\Delta \in (0, 1)$, $s < mn/(\mu^2 r^2)$, and a Gaussian subsampling $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ there exist constants $c_0, c_1 > 0$ such that $\Delta_{r,s,\mu} \leq \Delta$ when

$$p > c_0(\Delta)\,(r(m + n - r) + s)\log\left(\left(1 - \gamma^2\right)^{-1/2} \frac{mn}{s}\right), \tag{4}$$

with probability at least $1 - \exp(-c_1 p)$, where $\gamma := \mu\frac{r\sqrt{s}}{\sqrt{mn}}$.



Recovery by the convex relaxation

Recall the convex optimization problem

$$(L^*, S^*) = \underset{L,S \in \mathbb{R}^{m \times n}}{\arg\min} \|L\|_* + \lambda \|S\|_1, \qquad \text{s.t.} \quad \|\mathcal{A}(L+S) - b\|_2 \le \epsilon_b. \qquad (\star)$$

Theorem (Guaranteed convex recovery):

Let $b = \mathcal{A}(M)$ and suppose that $r, s \ge 1$ and $s < mn/(32\mu^2 r^2)$ are such that the restricted isometry constant $\Delta_{4r,2s,2\mu}(\mathcal{A}) \le \frac{1}{7} - 2\gamma$ where $\gamma := \mu \frac{4r\sqrt{2s}}{\sqrt{mn}}$. Let $X^* = L^* + S^*$ be the solution of $(\star)$ with $\lambda = \sqrt{2r/s}$, then $\|X^* - M\|_F \le 42\epsilon_b$.

Non-convex algorithm: Normalized Alternating Hard Thresholding

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

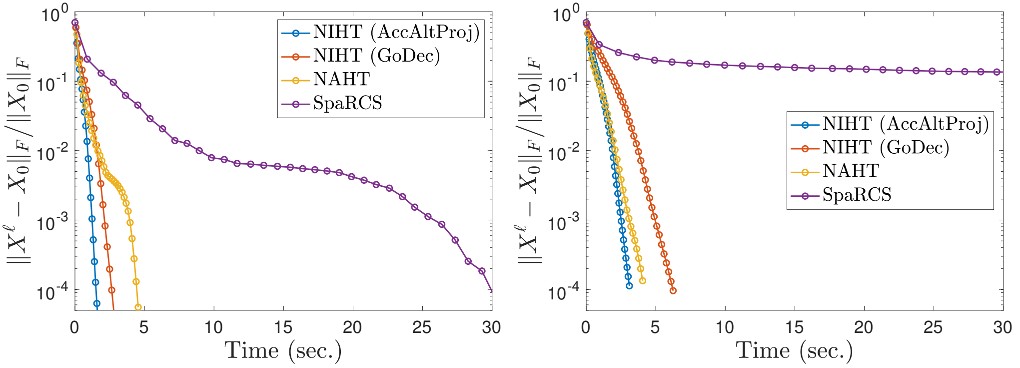
while not converged **do**
  Compute the residual $R_L^j = \mathcal{A}^* \left(\mathcal{A}(X^j) - b \right)$
  Set $V^j = L^j - \alpha_j^L \, R_L^j$
  Set $L^{j+1} = \mathrm{HT}(V^j;\, r)$
  Set $X^{j+\frac{1}{2}} = L^{j+1} + S^j$
  Compute the residual $R_S^j = \mathcal{A}^* \left(\mathcal{A}(X^{j+\frac{1}{2}})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

Non-convex algorithm: Normalized Alternating Hard Thresholding

Theorem (Guaranteed recovery by NAHT):
Suppose that $r, s \in \mathbb{N}$ and $s < mn/\left(8\mu^2 r^2\right)$ are such that the restricted isometry constant $\Delta_3 := \Delta_{3r,3s,\mu}(\mathcal{A}) < \frac{1}{9} - \gamma_2$ where $\gamma_2 := \mu\frac{2r\sqrt{2s}}{\sqrt{mn}}$. Then

$$\left\|L^{j+1} - L_0\right\|_F + \left\|S^{j+1} - S_0\right\|_F \leq \frac{6\Delta_3 + \frac{9}{8}\gamma_2}{1 - 3\Delta_3 - \frac{9}{8}\gamma_2}\left(\left\|L^j - L_0\right\|_F + \left\|S^j - S_0\right\|_F\right). \quad (5)$$

Linear convergence of non-convex recovery



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

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

Figure 2: Relative error in the approximate $\|X^\ell\|_F$ 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).

Conclusions

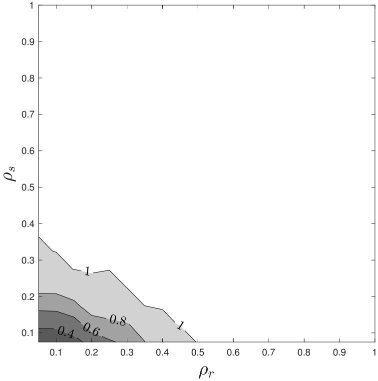
1. Non-convex optimisation problems can have no solutions[1].
2. For $\mathrm{LS}_{m,n}(r, s, \mu)$ to have $s < mn/(\mu^2 r^2)$ closes the set.
3. We do not need structure for the supp(S) in Robust PCA and similar problems.
4. Restricted isometry constants, guaranteed convex and non-convex solution of the subsampled low-rank plus sparse problem[2].

[1] Tanner, Thompson, V. (2019). Matrix rigidity and the ill-posedness of Robust PCA and matrix completion
[2] Tanner & V. (2020). Compressed sensing of low-rank plus sparse matrices

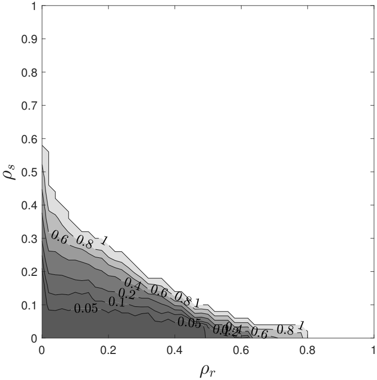
Thank you for your attention.

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, $\mu \approx 3$.



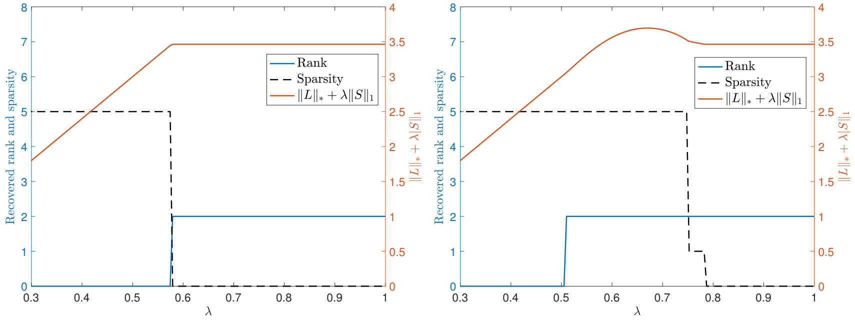
(b) NAHT recovery for 100 × 100 matrix, $\mu \approx 3$.

References i

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 (?) for $M^{(1)}$

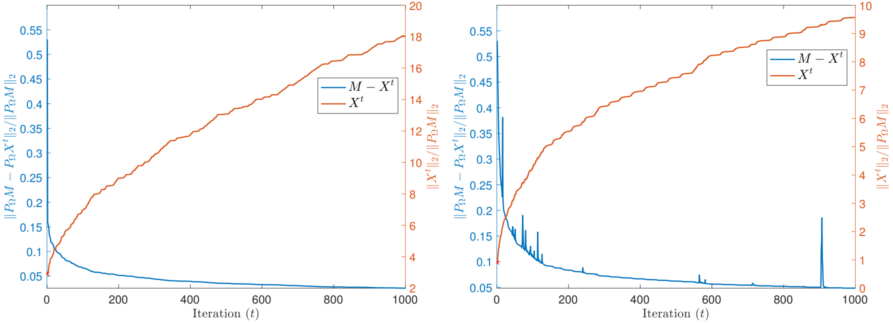
(b) IALM (?) for $M^{(1)}$

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

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

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



(a) ASD (?) for $M^{(1)}$ (b) CGIHT (?) for $M^{(1)}$