Matrix ALPS: Accelerated Low Rank and Sparse Matrix Reconstruction

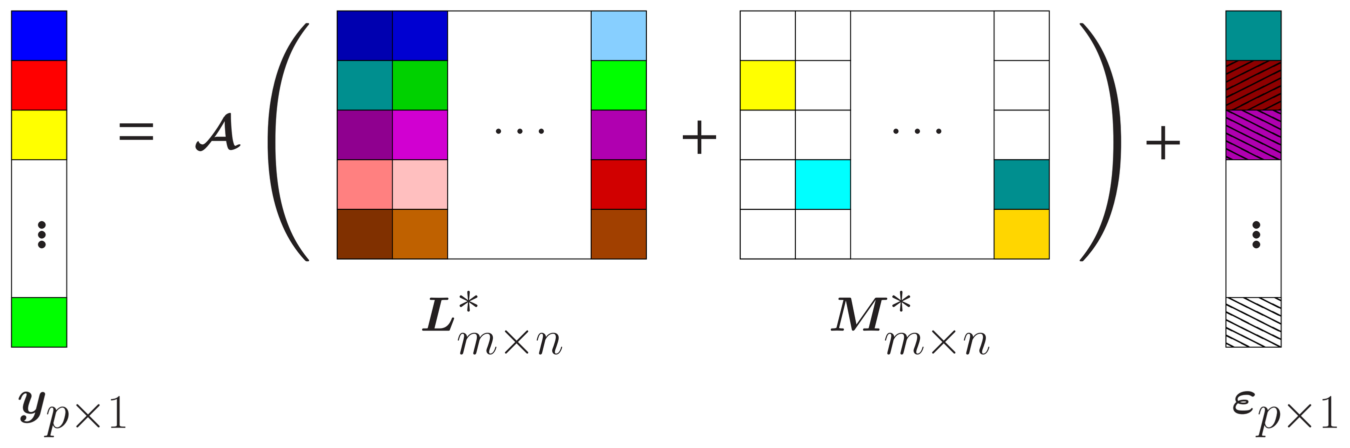
Anastasios Kyrillidis Volkan Cevher
Laboratory for Information and Inference Systems (LIONS), EPFL
{anastasios.kyrillidis, volkan.cevher}@epfl.ch

Problem statement

PROBLEM. *Given a linear operator $\boldsymbol{\mathcal{A}} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ and a set of observations $\boldsymbol{y} \in \mathbb{R}^p$ ($p \ll m \times n$): $\boldsymbol{y} = \boldsymbol{\mathcal{A}}(\mathbf{X}^*) + \boldsymbol{\varepsilon}$, where $\mathbf{X}^* := \mathbf{L}^* + \mathbf{M}^* \in \mathbb{R}^{m\times n}$ is the superposition of a rank-k $\mathbf{L}^*$ and a s-sparse $\mathbf{M}^*$ component, find a minimizer such that:*

$$\{\widehat{\mathbf{L}}, \widehat{\mathbf{M}}\} = \underset{\mathbf{L},\,\mathbf{M}:\ \mathrm{rank}(\mathbf{L})\le k,\ \|\mathbf{M}\|_0\le s}{\arg\min} \big\|\boldsymbol{y} - \boldsymbol{\mathcal{A}}(\mathbf{L}+\mathbf{M})\big\|_2.$$

- **Assumptions on $\boldsymbol{\mathcal{A}}$:**
  1. Sparse Restricted Isometry Property: $(1-\delta_s) \le \|\boldsymbol{\mathcal{A}}(\mathbf{X})\|_2/\|\mathbf{X}\|_F \le (1+\delta_s)$, $\forall \mathbf{X}$ s.t. $\|\mathbf{X}\|_0 \le s$,
  2. Rank Restricted Isometry Property: $(1-\delta_k) \le \|\boldsymbol{\mathcal{A}}(\mathbf{X})\|_2/\|\mathbf{X}\|_F \le (1+\delta_k)$, $\forall \mathbf{X}$ s.t. $\mathrm{rank}(\mathbf{X}) \le k$.



- **Special problem instances:**
  1. Compressed sensing (CS).
  2. Affine rank minimization (ARM).
  3. Matrix Completion (MC).
  4. Robust PCA (RPCA).
- **Paper overview:**
  1. We provide **better RIP guarantees** compared to state-of-the-art methods.
  2. We introduce **MATRIX ALPS, an accelerated, memory-based algorithm**.

Previous work

	Convex-based
Solution via	Convex relaxation $\|\cdot\|_* + \|\cdot\|_{1}$,...
Criteria example	$\min_{\|\mathbf{y}-\mathcal{A}(\mathbf{L}+\mathbf{M})\|_2\le\sigma} \|\mathbf{L}\|_* + \lambda\|\mathbf{M}\|_1$
Algorithms	(S)PCP[1,2,3], CPCP[1,2,3,4], SVT[1,3], ...
	Greedy-based
Solution via	Non-convex projections, ...
Criteria example	$\min_{\mathrm{rank}(\mathbf{L})\le k, \|\mathbf{M}\|_0\le s} \|\mathbf{y}-\mathcal{A}(\mathbf{L}+\mathbf{M})\|_2^2$
Algorithms	SpaRCS[1,2,3,4], GoDec[1,2], SVP[1,3], ...
	Manifold-based
Solution via	Manifold Trust regions, subspace identification, ...
Criteria example	$\min_{\mathrm{rank}(\mathbf{US})\le k, \|\mathbf{M}\|_0\le s} \|\mathbf{y}-\mathcal{A}(\mathbf{US}+\mathbf{M})\|_2^2$
Algorithms	RTRMC[1], GROUSE[1], GRASTA[1], ...

[1]MC, [2]RPCA, [3]ARM, [4]handles CS data

- **SpaRCS [1]:** covers MC, ARM, RPCA problem cases, handles CS data and considers general $\boldsymbol{\mathcal{A}}$ satisfying sparse- and rank-RIP.

[1] A. Waters, A. Sankaranarayanan, and R. Baraniuk, SpaRCS: Recovering low-rank and sparse matrices from CS measurements.

The SpaRCS algorithm

Input: $\boldsymbol{y}, \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{A}}^*$, Tolerance η, MaxIterations
Initialize: $\{\mathbf{L}_0, \mathbf{M}_0\} \leftarrow 0$, $\{\mathcal{L}_0, \mathcal{M}_0\} \leftarrow \{\emptyset\}$, $i \leftarrow 0$
repeat
  $\mathcal{S}_i^{\mathcal{L}} \leftarrow \mathcal{D}_i^{\mathcal{L}} \cup \mathcal{L}_i$ where $\mathcal{D}_i^{\mathcal{L}} \leftarrow \mathrm{ortho}\left(\mathcal{P}_k(\nabla f(\mathbf{X}_i))\right)$
  $\mathcal{S}_i^{\mathcal{M}} \leftarrow \mathcal{D}_i^{\mathcal{M}} \cup \mathcal{M}_i$ where $\mathcal{D}_i^{\mathcal{M}} \leftarrow \mathrm{supp}\left(\mathcal{P}_{\Sigma_s}(\nabla f(\mathbf{X}_i))\right)$
  Low rank matrix estimation:
    $\mathbf{V}_i^{\mathcal{L}} \leftarrow \arg\min_{\mathbf{V}:\mathbf{V}\in\mathrm{span}(\mathcal{S}_i^{\mathcal{L}})} \|\boldsymbol{y} - \boldsymbol{\mathcal{A}}(\mathbf{V}+\mathbf{M}_i)\|_2^2$
    $\mathbf{L}_{i+1} \leftarrow \mathcal{P}_k(\mathbf{V}_i^{\mathcal{L}})$ with $\mathcal{L}_{i+1} \leftarrow \mathrm{ortho}(\mathbf{L}_{i+1})$
  Sparse matrix estimation:
    $\mathbf{V}_i^{\mathcal{M}} \leftarrow \arg\min_{\mathbf{V}:\mathbf{V}\in\mathrm{supp}(\mathcal{S}_i^{\mathcal{M}})} \|\boldsymbol{y} - \boldsymbol{\mathcal{A}}(\mathbf{V}+\mathbf{L}_i)\|_2^2$
    $\mathbf{M}_{i+1} \leftarrow \mathcal{P}_{\Sigma_s}(\mathbf{V}_i^{\mathcal{M}})$ with $\mathcal{M}_{i+1} \leftarrow \mathrm{supp}(\mathbf{M}_i)$
  $\mathbf{X}_{i+1} \leftarrow \mathbf{L}_{i+1} + \mathbf{M}_{i+1}$
  $i \leftarrow i+1$
until $\|\mathbf{X}_i - \mathbf{X}_{i-1}\|_2 \le \eta\|\mathbf{X}_i\|_2$ or MaxIterations.

Improving SpaRCS

THEOREM 1. *Assume $\boldsymbol{\mathcal{A}}$ satisfies the sparse-RIP and rank-RIP for $\delta_{4s}(\boldsymbol{\mathcal{A}}) \le 0.075$, $\delta_{4k}(\boldsymbol{\mathcal{A}}) \le 0.04$ and $\delta_{2s+3k}(\boldsymbol{\mathcal{A}}) \le 0.07$. Then, SpaRCS satisfy:*

$$\|\mathbf{L}^* - \mathbf{L}_{i+1}\|_F \le \rho_1^{\mathcal{L}}\|\mathbf{L}^* - \mathbf{L}_i\|_F + \rho_1^{\mathcal{M}}\|\mathbf{M}^* - \mathbf{M}_i\|_F + \gamma_1\|\boldsymbol{\varepsilon}\|_2$$
$$\|\mathbf{M}^* - \mathbf{M}_{i+1}\|_F \le \rho_2^{\mathcal{L}}\|\mathbf{L}^* - \mathbf{L}_i\|_F + \rho_2^{\mathcal{M}}\|\mathbf{M}^* - \mathbf{M}_i\|_F + \gamma_2\|\boldsymbol{\varepsilon}\|_2$$

where $\rho_1^{\mathcal{L}} = 0.16$, $\rho_2^{\mathcal{L}} = 0.34$, $\rho_1^{\mathcal{M}} = 0.34$, $\rho_2^{\mathcal{M}} = 0.14$, $\gamma_1 = 4.36$ and, $\gamma_2 = 4.45$.

- **SpaRCS:** $\rho_1^{\mathcal{L}} = 0.48$, $\rho_2^{\mathcal{L}} = 0.47$, $\rho_1^{\mathcal{M}} = 0.47$, $\rho_2^{\mathcal{M}} = 0.32$, $\gamma_1 = 6.68$ and, $\gamma_2 = 6.88$.
- **CAVEAT:** holds iff SpaRCS computes a low-rank + sparse decomposition **at each iteration**.
- **But:** Under mild conditions, a stationary point to a non-convex problem can always be obtained.

The accelerated MATRIX ALPS algorithm and its guarantees

Input: $\boldsymbol{y}, \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{A}}^*$, Tolerance η, MaxIterations, τ_i, $\forall i$
Initialize: $\{\mathbf{Q}_0, \mathbf{M}_0, \mathbf{L}_0\} \leftarrow 0$, $\{\mathcal{L}_0, \mathcal{M}_0\} \leftarrow \{\emptyset\}$, $i \leftarrow 0$
repeat
  Low rank matrix estimation:
    $\mathcal{D}_i^{\mathcal{L}} \leftarrow \mathrm{ortho}\left(\mathcal{P}_k(\nabla f(\mathbf{Q}_i))\right)$
    $\mathcal{S}_i^{\mathcal{L}} \leftarrow \mathcal{D}_i^{\mathcal{L}} \cup \mathcal{L}_i$
    $\mathbf{V}_i^{\mathcal{L}} \leftarrow \mathbf{Q}_i^{\mathcal{L}} - \frac{\mu_i^{\mathcal{L}}}{2}\mathcal{P}_{\mathcal{S}_i^{\mathcal{L}}}\nabla f(\mathbf{Q}_i)$
    $\mathbf{L}_{i+1} \leftarrow \mathcal{P}_k(\mathbf{V}_i^{\mathcal{L}})$ with $\mathcal{L}_{i+1} \leftarrow \mathrm{ortho}(\mathbf{L}_{i+1})$
    $\mathbf{Q}_{i+1}^{\mathcal{L}} \leftarrow \mathbf{L}_{i+1} + \tau_i(\mathbf{L}_{i+1} - \mathbf{L}_i)$
    $\mathbf{Q}_{i} \leftarrow \mathbf{Q}_{i+1}^{\mathcal{L}} + \mathbf{Q}_i^{\mathcal{M}}$
  Sparse matrix estimation:
    $\mathcal{D}_i^{\mathcal{M}} \leftarrow \mathrm{supp}\left(\mathcal{P}_{\Sigma_s}(\nabla f(\mathbf{Q}_{i+1}))\right)$
    $\mathcal{S}_i^{\mathcal{M}} \leftarrow \mathcal{D}_i^{\mathcal{M}} \cup \mathcal{M}_i$
    $(\mathbf{V}_i^{\mathcal{M}})_{\mathcal{S}_i^{\mathcal{M}}} \leftarrow (\mathbf{Q}_i^{\mathcal{M}})_{\mathcal{S}_i^{\mathcal{M}}} - \frac{\mu_i^{\mathcal{M}}}{2}(\nabla f(\mathbf{Q}_{i+1}))_{\mathcal{S}_i^{\mathcal{M}}}$
    $\mathbf{M}_{i+1} \leftarrow \mathcal{P}_{\Sigma_s}(\mathbf{V}_i^{\mathcal{M}})$ with $\mathcal{M}_{i+1} \leftarrow \mathrm{supp}(\mathbf{M}_{i+1})$
    $\mathbf{Q}_{i+1}^{\mathcal{M}} \leftarrow \mathbf{M}_{i+1} + \tau_i(\mathbf{M}_{i+1} - \mathbf{M}_i)$
    $\mathbf{Q}_{i+1} \leftarrow \mathbf{Q}_{i+1}^{\mathcal{L}} + \mathbf{Q}_{i+1}^{\mathcal{M}}$
  $i \leftarrow i+1$
until $\|\mathbf{X}_i - \mathbf{X}_{i-1}\|_2 \le \eta\|\mathbf{X}_i\|_2$ or MaxIterations.

THEOREM 2. *Assume:*

- $\boldsymbol{\mathcal{A}} : \mathbb{R}^{m\times n}$ *satisfies the rank-RIP and sparse-RIP with constants* $\delta_{4k}(\boldsymbol{\mathcal{A}}) \le 0.09$ *and* $\delta_{4s}(\boldsymbol{\mathcal{A}}) \le 0.095$.
- *Noiseless case:* $\boldsymbol{y} = \boldsymbol{\mathcal{A}}(\mathbf{X}^*)$ *with constant momentum term* $\tau := \tau_i = 1/4$, $\forall i$.

Then, MATRIX ALPS *satisfies the following second-order linear system:*

$$\mathbf{x}(i+1) \le (1+\tau)\boldsymbol{\Delta}\mathbf{x}(i) + \tau\boldsymbol{\Delta}\mathbf{x}(i-1),\ \text{where } \mathbf{x}(i) := \begin{bmatrix} \|\mathbf{L}_i - \mathbf{L}^*\|_F \\ \|\mathbf{M}_i - \mathbf{M}^*\|_F \end{bmatrix} \text{ and } \boldsymbol{\Delta} := \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}$$

depends on $\delta_{4k}(\boldsymbol{\mathcal{A}})$ and $\delta_{4s}(\boldsymbol{\mathcal{A}})$. Furthermore, the following first-order linear system holds:

$$\mathbf{w}(i+1) \le \underbrace{\begin{bmatrix} (1+\tau)\boldsymbol{\Delta} & \tau\boldsymbol{\Delta} \\ I & \mathbf{0} \end{bmatrix}}_{\widehat{\boldsymbol{\Delta}}}{}^{i}\, \mathbf{w}(0),$$

for $\mathbf{w}(i) := [\mathbf{x}(i+1)\ \ \mathbf{x}(i)]^T$. We observe that $\lim_{i\to\infty} \mathbf{w}(i) = \mathbf{0}$ since $|\lambda_j(\widehat{\boldsymbol{\Delta}})| \le 1$, $\forall j$.

Synthetic and real data results

$m \times n$	k	Relative Error (10^{-3})	Time (sec)
200×400	5	0.134/0.18/0.002/0.78/0.04	2.26/0.27/0.95/0.36/**0.21**
200×400	5	0.127/0.164/0.01/0.76/0.05	2.16/0.26/0.96/0.36/**0.23**
200×400	10	6.7/0.5/0.01/1.2/0.1	36.38/0.45/1.13/0.64/**0.37**
200×400	15	150/0.93/340/2.1/0.15	98.12/0.82/1.29/1.08/**0.68**
1000×5000	10	$-$/0.09/0.008/0.34/0.03	$-$/10.8/27.6/10.2/**5.5**
1000×5000	50	$-$/0.2/0.002/0.73/0.11	$-$/23.4/171.37/35.5/**17.2**
1000×5000	120	$-$/0.52/0.07/1.22/0.077	$-$/139/501/228/**101**

- Comparison table for the Matrix Completion problem. Table depicts median values over 50 Monte-Carlo iterations. The list of algorithms includes: SpaRCS / ALM / GROUSE / SVP / MATRIX ALPS.

- Background subtraction in video sequence. Median execution times over 10 Monte-Carlo iterations. GoDec: 34.8 sec—MATRIX ALPS: 15.8 sec.