Low-Rank Factorization Models for Matrix Completion and Matrix Separation

Zaiwen Wen
Joint work with Wotao Yin, Yin Zhang and Shen Yuan

IPAM, UCLA
Oct. 5, 2010

Low rank minimization problems

- Matrix completion: find a low-rank matrix $W \in \mathbb{R}^{m \times n}$ so that $W_{ij} = M_{ij}$ for some given M_{ij} with all (i, j) in an index set Ω:

$$\min_{W \in \mathbb{R}^{m \times n}} \operatorname{rank}(W), \text{ s.t. } W_{ij} = M_{ij}, \ \forall (i, j) \in \Omega$$

- Matrix separation: find a low-rank $Z \in \mathbb{R}^{m \times n}$ and a sparse matrix $S \in \mathbb{R}^{m \times n}$ so that $Z + S = D$ for a given D.

Convex approximation

Nuclear norm $\|W\|_*$: the summation of singular values of W

- Matrix completion:
  - Nuclear-norm relaxation:

$$\min_{W \in \mathbb{R}^{m \times n}} \|W\|_*, \text{ s.t. } W_{ij} = M_{ij}, \ \forall (i,j) \in \Omega,$$

  - Nuclear-norm regularized linear least square:

$$\min_{W \in \mathbb{R}^{m \times n}} \mu \|W\|_* + \frac{1}{2} \|\mathcal{P}_\Omega(W - M)\|_F^2,$$

- Matrix separation:

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

Singular value decomposition can be very expensive!

Low-rank factorization model for matrix completion

- Finding a low-rank matrix W so that $\|\mathcal{P}_\Omega(W - M)\|_F^2$ or the distance between W and $\{Z \in \mathbb{R}^{m\times n}, Z_{ij} = M_{ij}, \forall (i,j) \in \Omega\}$ is minimized.
- Any matrix $W \in \mathbb{R}^{m\times n}$ with $\mathrm{rank}(W) \le K$ can be expressed as $W = XY$ where $X \in \mathbb{R}^{m\times K}$ and $Y \in \mathbb{R}^{K\times n}$.

New model

$$\min_{X,Y,Z} \frac{1}{2}\|XY - Z\|_F^2 \ \text{ s.t. } \ Z_{ij} = M_{ij}, \forall (i,j) \in \Omega$$

- Advantage: SVD is no longer needed!
- Related work: the solver `OptSpace` based on optimization on manifold

Low-rank factorization model for matrix separation

- Consider the model

$$\min_{Z,S} \|S\|_1 \text{ s.t. } Z + S = D, \ \mathrm{rank}(L) \leq K$$

- Low-rank factorization: $Z = UV$

$$\min_{U,V,Z} \|Z - D\|_1 \text{ s.t. } UV - Z = 0$$

- Only the entries D_{ij}, $(i,j) \in \Omega$, are given. $\mathcal{P}_\Omega(D)$ is the projection of D onto Ω.

New model

$$\min_{U,V,Z} \quad \|\mathcal{P}_\Omega(Z - D)\|_1 \quad \text{s.t.} \quad UV - Z = 0$$

- Advantage: SVD is no longer needed!

Nonlinear Gauss-Seideal scheme

Consider:

$$\min_{X,Y,Z} \frac{1}{2}\|XY - Z\|_F^2 \text{ s.t. } Z_{ij} = M_{ij}, \forall (i,j) \in \Omega$$

Alternating minimization:

$$X_+ = ZY^\top (YY^\top)^\dagger = \operatorname*{argmin}_{X \in \mathbb{R}^{m\times K}} \frac{1}{2}\|XY - Z\|_F^2,$$

Nonlinear Gauss-Seideal scheme

Consider:

$$\min_{X,Y,Z} \frac{1}{2}\|XY - Z\|_F^2 \text{ s.t. } Z_{ij} = M_{ij}, \forall (i,j) \in \Omega$$

First variant of alternating minimization:

$$\begin{aligned} X_+ &\leftarrow ZY^\dagger \equiv ZY^\top (YY^\top)^\dagger, \\ Y_+ &\leftarrow (X_+)^\dagger Z \equiv (X_+^\top X_+)^\dagger (X_+^\top Z), \\ Z_+ &\leftarrow X_+ Y_+ + \mathcal{P}_\Omega(M - X_+ Y_+). \end{aligned}$$

Let $\mathcal{P}_A$ be the orthogonal projection onto the range space $\mathcal{R}(A)$

- $X_+ Y_+ = \left(X_+ (X_+^\top X_+)^\dagger X_+^\top\right) Z = \mathcal{P}_{X_+} Z$
- One can verify that $\mathcal{R}(X_+) = \mathcal{R}(ZY^\top)$.
- $X_+ Y_+ = \mathcal{P}_{ZY^\top} Z = ZY^\top (YZ^\top ZY^\top)^\dagger (YZ^\top) Z.$
- idea: modify X_+ or Y_+ to obtain the same product $X_+ Y_+$

Nonlinear Gauss-Seideal scheme

Consider:

$$\min_{X,Y,Z} \frac{1}{2}\|XY - Z\|_F^2 \text{ s.t. } Z_{ij} = M_{ij}, \forall (i,j) \in \Omega$$

Second variant of alternating minimization:

$$\begin{aligned} X_+ &\leftarrow ZY^\top, \\ Y_+ &\leftarrow (X_+)^\dagger Z \equiv (X_+^\top X_+)^\dagger (X_+^\top Z), \\ Z_+ &\leftarrow X_+ Y_+ + \mathcal{P}_\Omega(M - X_+ Y_+). \end{aligned}$$

Nonlinear Gauss-Seideal scheme

Consider:

$$\min_{X,Y,Z} \frac{1}{2}\|XY - Z\|_F^2 \text{ s.t. } Z_{ij} = M_{ij}, \forall (i,j) \in \Omega$$

Second variant of alternating minimization:

$$\begin{aligned} X_+ &\leftarrow ZY^\top, \\ Y_+ &\leftarrow (X_+)^\dagger Z \equiv (X_+^\top X_+)^\dagger (X_+^\top Z), \\ Z_+ &\leftarrow X_+ Y_+ + \mathcal{P}_\Omega(M - X_+ Y_+). \end{aligned}$$

Third variant of alternating minimization: $V = \texttt{orth}(ZY^\top)$

$$\begin{aligned} X_+ &\leftarrow V, \\ Y_+ &\leftarrow V^\top Z, \\ Z_+ &\leftarrow X_+ Y_+ + \mathcal{P}_\Omega(M - X_+ Y_+). \end{aligned}$$

Nonlinear SOR

- The nonlinear GS scheme can be slow
- Linear SOR: applying extrapolation to the GS method to achieve faster convergence

The first implementation:

$$\begin{aligned}
X_+ &\leftarrow ZY^\top (YY^\top)^\dagger, \\
X_+(\omega) &\leftarrow \omega X_+ + (1-\omega)X, \\
Y_+ &\leftarrow (X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z), \\
Y_+(\omega) &\leftarrow \omega Y_+ + (1-\omega)Y, \\
Z_+(\omega) &\leftarrow X_+(\omega)Y_+(\omega) + \mathcal{P}_\Omega(M - X_+(\omega)Y_+(\omega)),
\end{aligned}$$

Nonlinear SOR

- Let $S = \mathcal{P}_\Omega(M - XY)$. Then $Z = XY + S$
- Let $Z_\omega \triangleq XY + \omega S = \omega Z + (1-\omega)XY$
- Assume Y has full row rank, then

$$\begin{aligned} Z_\omega Y^\top (YY^\top)^\dagger &= \omega Z Y^\top (YY^\top)^\dagger + (1-\omega) XYY^\top (YY^\top)^\dagger \\ &= \omega X_+ + (1-\omega)X, \end{aligned}$$

Second implementation of our nonlinear SOR:

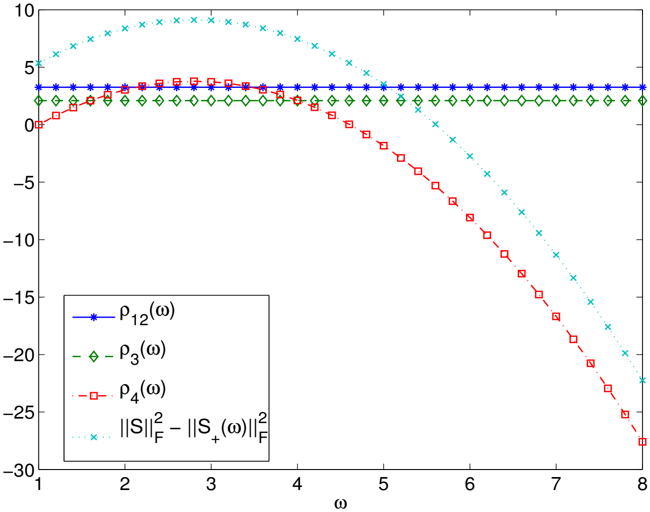
$$\begin{aligned} X_+(\omega) &\leftarrow Z_\omega Y^\top \text{ or } Z_\omega Y^\top (YY^\top)^\dagger, \\ Y_+(\omega) &\leftarrow (X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z_\omega), \\ \mathcal{P}_{\Omega^c}(Z_+(\omega)) &\leftarrow \mathcal{P}_{\Omega^c}(X_+(\omega) Y_+(\omega)), \\ \mathcal{P}_\Omega(Z_+(\omega)) &\leftarrow \mathcal{P}_\Omega(M). \end{aligned}$$

Reduction of the residual $\|S\|_F^2 - \|S_+(\omega)\|_F^2$

Assume that $\mathrm{rank}(Z_\omega) = \mathrm{rank}(Z), \forall\, \omega \in [1, \omega_1]$ for some $\omega_1 \geq 1$. Then there exists some $\omega_2 \geq 1$ such that

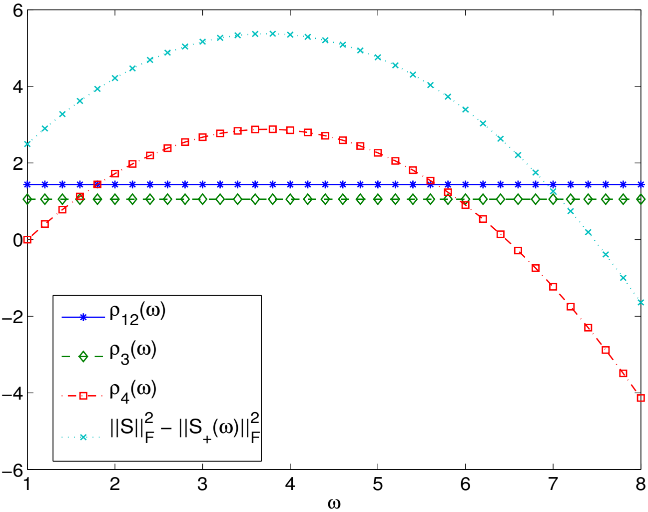
$$\|S\|_F^2 - \|S_+(\omega)\|_F^2 = \rho_{12}(\omega) + \rho_3(\omega) + \rho_4(\omega) > 0, \quad \forall\, \omega \in [1, \omega_2].$$

- $\rho_{12}(\omega) \triangleq \|SP\|_F^2 + \|Q(\omega)S(I-P)\|_F^2 \geq 0$
- $\rho_3(\omega) \triangleq \|\mathcal{P}_{\Omega^c}(SP + Q(\omega)S(I-P))\|_F^2 \geq 0$
- $\rho_4(\omega) \triangleq \frac{1}{\omega^2}\|S_+(\omega) + (\omega-1)S\|_F^2 - \|S_+(\omega)\|_F^2$
- Whenever $\rho_3(1) > 0$ ($\mathcal{P}_{\Omega^c}(X_+(1)Y_+(1) - XY) \neq 0$) and $\omega_1 > 1$, then $\omega_2 > 1$ can be chosen so that $\rho_4(\omega) > 0, \forall\, \omega \in (1, \omega_2]$.


10
5
0
−5
−10
−15
−20
−25
−30
1
2
3
4
5
6
7
8
ω
$\rho_{12}(\omega)$
$\rho_3(\omega)$
$\rho_4(\omega)$
$\|S\|_F^2 - \|S_+(\omega)\|_F^2$

(a)

Reduction of the residual $\|S\|_F^2 - \|S_+(\omega)\|_F^2$



(b)

Nonlinear SOR: convergence guarantee

Problem: how can we select a proper weight ω to ensure convergence for a nonlinear model?

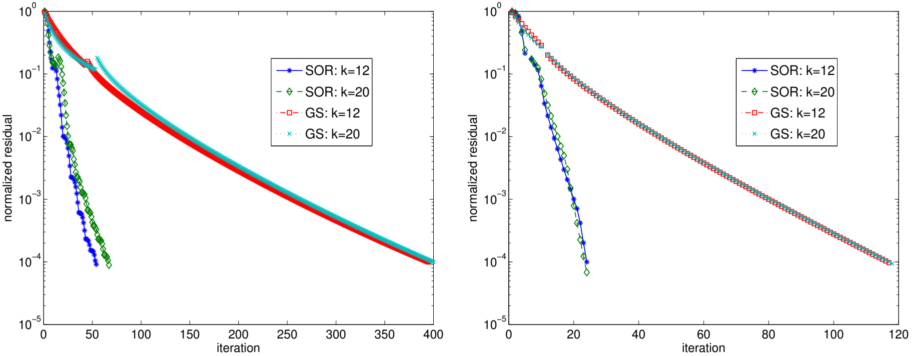
Strategy: Adjust ω dynamically according to the change of the objective function values.

- Calculate the residual ratio $\gamma(\omega) = \frac{\|S_+(\omega)\|_F}{\|S\|_F}$
- A small $\gamma(\omega)$ indicates that the current weight value ω works well so far.
- If $\gamma(\omega) < 1$, accept the new point; otherwise, ω is reset to 1 and this procedure is repeated.
- ω is increased only if the calculated point is acceptable but the residual ratio $\gamma(\omega)$ is considered "too large"; that is, $\gamma(\omega) \in [\gamma_1, 1)$ for some $\gamma_1 \in (0, 1)$.

Nonlinear SOR: complete algorithm

Algorithm 1: A low-rank matrix fitting algorithm (`LMaFit`)

1 Input index set Ω, data $\mathcal{P}_\Omega(M)$ and a rank overestimate $K \geq r$.
2 Set Y^0, Z^0, $\omega = 1$, $\tilde{\omega} > 1$, $\delta > 0$, $\gamma_1 \in (0, 1)$ and $k = 0$.
3 **while** *not convergent* **do**
4   Compute $(X_+(\omega), Y_+(\omega), Z_+(\omega))$.
5   Compute the residual ratio $\gamma(\omega)$.
6   **if** $\gamma(\omega) \geq 1$ **then** set $\omega = 1$ and go to step 4.
7   Update $(X^{k+1}, Y^{k+1}, Z^{k+1})$ and increment k.
8   **if** $\gamma(\omega) \geq \gamma_1$ **then**
9   set $\delta = \max(\delta, 0.25(\omega - 1))$ and $\omega = \min(\omega + \delta, \tilde{\omega})$.



(a) n=1000, r=10, SR = 0.08   (b) n=1000, r=10, SR=0.15

Extension to problems with general linear constraints

- Consider problem:

$$\min_{W \in \mathbb{R}^{m\times n}} \ \mathrm{rank}(W), \ \text{s.t. } \mathcal{A}(W) = b$$

- Nonconvex relaxation:

$$\min_{X,Y,Z} \ \frac{1}{2}\|XY - Z\|_F^2 \ \text{ s.t. } \ \mathcal{A}(Z) = b$$

- Let $S := \mathcal{A}^\top(\mathcal{A}\mathcal{A}^\top)^\dagger(b - \mathcal{A}(XY))$ and $Z_\omega = XY + \omega S$

Nonlinear SOR scheme:

$$\begin{aligned} X_+(\omega) &\leftarrow Z_\omega Y^\top \text{ or } \ Z_\omega Y^\top (YY^\top)^\dagger, \\ Y_+(\omega) &\leftarrow (X_+(\omega)^\top X_+(\omega))^\dagger (X_+(\omega)^\top Z_\omega), \\ Z_+(\omega) &\leftarrow X_+(\omega)Y_+(\omega) + \mathcal{A}^\top(\mathcal{A}\mathcal{A}^\top)^\dagger(b - \mathcal{A}(X_+(\omega)Y_+(\omega))). \end{aligned}$$

Nonlinear SOR: convergence analysis

Main result: Let $\{(X^k, Y^k, Z^k)\}$ be generated by the nonlinear SOR method and $\{\mathcal{P}_{\Omega^c}(X^k Y^k)\}$ be bounded. Then there exists at least a subsequence of $\{(X^k, Y^k, Z^k)\}$ that satisfies the first-order optimality conditions in the limit.

Technical lemmas:

- $\omega S \bullet (X_+(\omega) Y_+(\omega) - XY) = \|X_+(\omega) Y_+(\omega) - XY\|_F^2$.
- Residual reduction from $\|S\|_F^2$ after the first two steps

$$\frac{1}{\omega^2}\|X_+(\omega)Y_+(\omega) - Z_\omega\|_F^2 = \|(I - Q(\omega))S(I - P)\|_F^2 = \|S\|_F^2 - \rho_{12}(\omega)$$

- $\lim_{\omega \to 1^+} \frac{\rho_4(\omega)}{\omega - 1} = 2\|\mathcal{P}_{\Omega^c}(X_+(1) Y_+(1) - XY)\|_F^2 \geq 0$

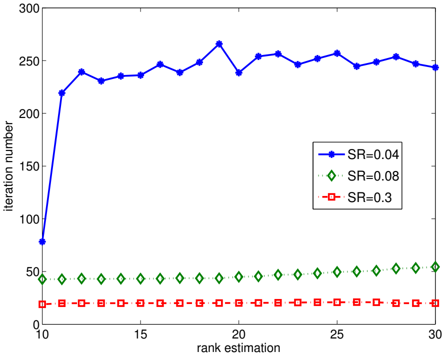
Practical issues

- Variants to obtain $X_+(\omega)$ and $Y_+(\omega)$: "linsolve" .vs. QR
- Storage: full Z or partial $S = \mathcal{P}_\Omega(M - XY)$
- Stopping criteria:

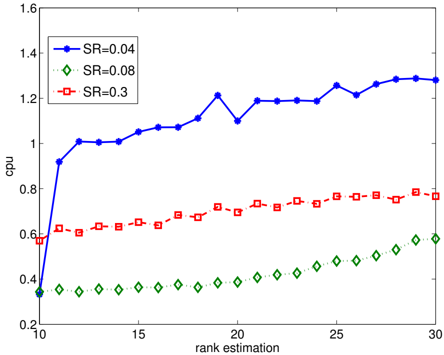
$$\texttt{relres} = \frac{\|\mathcal{P}_\Omega(M - X^k Y^k)\|_F}{\|\mathcal{P}_\Omega(M)\|_F} \leq \texttt{tol} \text{ and } \texttt{reschg} = \left|1 - \frac{\|\mathcal{P}_\Omega(M - X^k Y^k)\|_F}{\|\mathcal{P}_\Omega(M - X^{k-1} Y^{k-1})\|_F}\right| \leq \texttt{tol}/2,$$

- Rank estimation:
  - Start from $K \geq r$ then decrease K aggressively
    - $QR = XE$, E is permutation, $d := |\mathrm{diag}(R)|$ nonincreasing
    - compute the sequnce $\tilde{d}_i = d_i / d_{i+1}$, $i = 1, \cdots, K-1$,
    - examine the ratio $\tau = \frac{(K-1)\tilde{d}(p)}{\sum_{i \neq p} \tilde{d}_i}$, $\tilde{d}(p)$ is the maximal element of $\{\tilde{d}_i\}$ and p is the corresponding index.
    - reset K to p once $\tau > 10$
  - Start from a small K and increase K to $\min(K + \kappa, \texttt{rank\_max})$ when the alg. stagnates

The sensitivity with respect to the rank estimation K

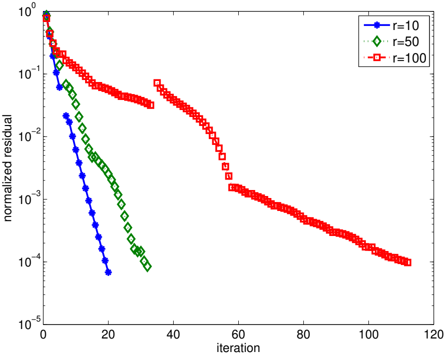


(a) Iteration number

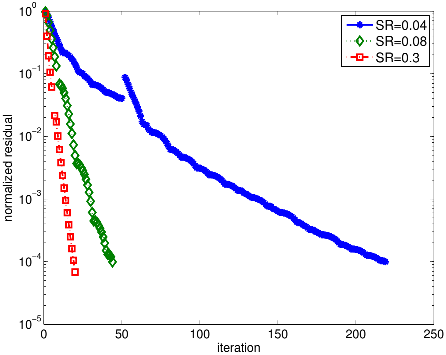


(b) CPU time in seconds

Convergence behavior of the residual

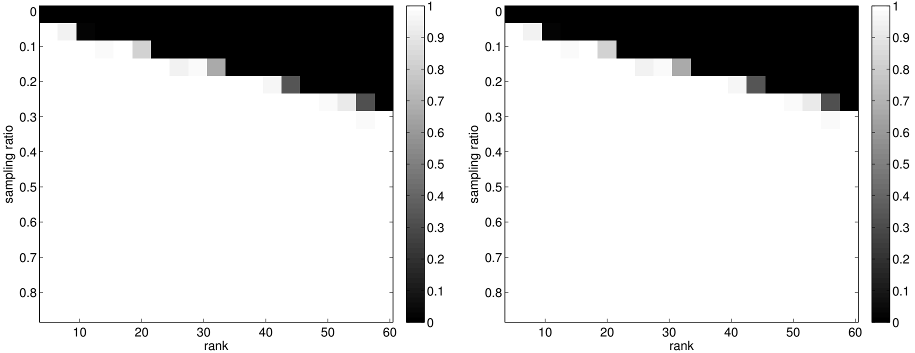


(a) $m = n = 1000$, sr$= 0.3$



(b) $m = n = 1000$, $r = 10$

Phase diagrams for matrix completion recoverability



(a) Model solved by `FPCA` (b) Model solved by `LMaFit`

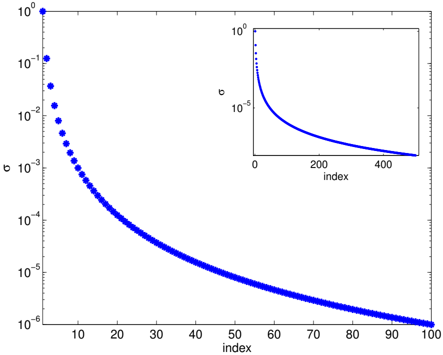
Small random problems: $m = n = 1000$

Problem			APGL				FPCA			LMaFit			
										$K = \lfloor 1.25r \rfloor$		$K = \lfloor 1.5r \rfloor$	
r	SR	FR	μ	time	rel.err	tsvd	time	rel.err	tsvd	time	rel.err	time	rel.err
10	0.04	0.50	5.76e-03	3.89	4.04e-03	82%	32.62	8.21e-01	12%	0.98	4.72e-04	1.00	4.35e-04
10	0.08	0.25	1.02e-02	2.25	6.80e-04	71%	13.24	7.30e-04	19%	0.35	2.27e-04	0.40	2.19e-04
10	0.15	0.13	1.78e-02	2.44	2.14e-04	66%	7.76	4.21e-04	42%	0.39	1.16e-04	0.41	1.48e-04
10	0.30	0.07	3.42e-02	4.11	1.40e-04	58%	17.54	1.97e-04	72%	0.59	8.99e-05	0.62	9.91e-05
50	0.20	0.49	2.94e-02	123.90	2.98e-03	93%	71.43	4.64e-04	56%	3.96	3.03e-04	4.96	2.63e-04
50	0.25	0.39	3.59e-02	23.80	8.17e-04	87%	101.47	3.24e-04	67%	2.98	1.89e-04	3.20	2.11e-04
50	0.30	0.33	4.21e-02	18.64	6.21e-04	85%	146.24	2.64e-04	75%	2.56	1.78e-04	2.78	1.91e-04
50	0.40	0.24	5.53e-02	19.17	3.69e-04	82%	42.28	2.16e-04	77%	2.28	1.11e-04	2.69	1.65e-04
100	0.35	0.54	5.70e-02	73.48	1.24e-03	92%	259.37	5.41e-04	77%	13.07	3.01e-04	17.40	3.09e-04
100	0.40	0.47	6.37e-02	63.08	8.19e-04	91%	302.82	4.11e-04	79%	9.74	2.56e-04	11.39	2.41e-04
100	0.50	0.38	7.71e-02	61.44	4.91e-04	90%	359.66	3.10e-04	82%	7.30	1.55e-04	7.37	1.92e-04
100	0.55	0.35	8.40e-02	50.78	4.12e-04	89%	360.28	2.89e-04	81%	6.23	1.14e-04	7.18	9.99e-05

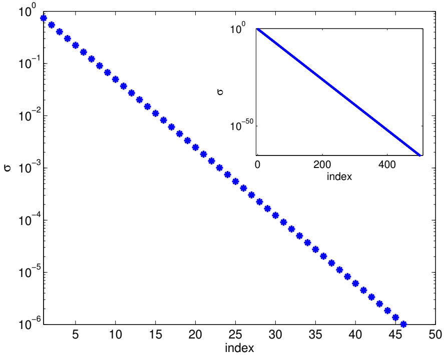
Large random problems

Problem				APGL					LMaFit ($K = \lfloor 1.25r \rfloor$)				LMaFit ($K = \lfloor 1.5r \rfloor$)			
n	r	SR	FR	μ	iter	#sv	time	rel.err	iter	#sv	time	rel.err	iter	#sv	time	rel.err
1000	10	0.119	0.167	1.44e-2	39	10	2.47	3.04e-4	23	10	0.29	1.67e-4	23	10	0.28	1.73e-4
1000	50	0.390	0.250	5.36e-2	40	50	14.48	3.08e-4	18	50	1.88	6.58e-5	18	50	2.04	7.62e-5
1000	100	0.570	0.334	8.58e-2	53	100	49.67	3.98e-4	20	100	5.47	1.86e-4	21	100	5.99	1.42e-4
5000	10	0.024	0.166	1.37e-2	52	10	12.48	2.17e-4	29	10	1.99	1.71e-4	29	10	2.17	1.77e-4
5000	50	0.099	0.200	6.14e-2	76	50	161.82	1.26e-3	20	50	15.87	2.72e-5	20	50	16.49	3.86e-5
5000	100	0.158	0.250	1.02e-1	60	100	316.02	3.74e-4	26	100	57.85	1.57e-4	27	100	60.69	1.47e-4
10000	10	0.012	0.166	1.37e-2	53	10	23.45	3.61e-4	34	10	5.08	1.54e-4	34	10	5.56	1.66e-4
10000	50	0.050	0.200	5.97e-2	56	50	225.21	2.77e-4	23	50	44.80	4.76e-5	23	50	48.85	5.70e-5
10000	100	0.080	0.250	9.94e-2	71	100	941.38	2.87e-4	30	100	168.44	1.63e-4	30	100	176.45	1.70e-4
20000	10	0.006	0.167	1.35e-2	57	10	60.62	2.37e-4	38	10	12.46	1.44e-4	38	10	13.60	1.57e-4
30000	10	0.004	0.167	1.35e-2	59	10	95.50	1.96e-4	39	10	20.55	1.71e-4	39	10	23.48	1.73e-4
50000	10	0.002	0.167	1.35e-2	66	10	192.28	1.58e-4	42	10	43.43	1.81e-4	42	10	49.49	1.84e-4
100000	10	0.001	0.167	1.34e-2	92	10	676.11	2.10e-4	46	10	126.59	1.33e-4	46	10	140.32	1.30e-4

Random low-rank approximation problems



(a) power-low: $\sigma_i = i^{-3}$



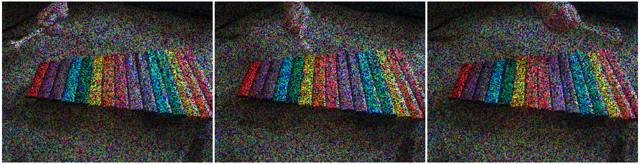
(b) exponentially: $\sigma_i = e^{-0.3i}$

Random low-rank approximation problems

Problem		APGL				FPCA			LMaFit(est_rank=1)			LMaFit(est_rank=2)		
SR	FR	μ	#sv	time	rel.err	#sv	time	rel.err	#sv	time	rel.err	#sv	time	rel.err
power-low decaying														
0.04	0.99	1.00e-04	90	16.30	6.48e-01	1	40.49	1.39e-01	5	0.70	3.68e-01	11	0.31	8.96e-03
0.08	0.49	1.00e-04	85	19.95	2.00e-01	2	45.81	4.38e-02	5	1.59	2.20e-01	20	0.60	1.13e-03
0.15	0.26	1.00e-04	7	1.73	4.05e-03	4	14.46	1.78e-02	5	1.47	1.52e-01	20	0.75	4.57e-04
0.30	0.13	1.00e-04	11	1.85	1.86e-03	4	31.48	1.04e-02	5	3.20	8.12e-02	22	1.33	2.36e-04
exponentially decaying														
0.04	0.99	1.00e-04	100	15.03	7.50e-01	14	35.79	5.05e-01	5	0.48	3.92e-01	16	0.86	4.08e-01
0.08	0.49	1.00e-04	100	21.60	3.31e-01	8	39.82	1.24e-01	5	0.44	2.66e-01	26	1.84	1.98e-02
0.15	0.26	1.00e-04	100	17.43	4.71e-02	13	12.31	2.76e-02	5	0.63	2.39e-01	28	1.62	7.26e-04
0.30	0.13	1.00e-04	42	9.50	3.31e-03	14	29.13	1.71e-02	6	1.03	1.71e-01	30	2.01	2.38e-04

Video Denoising

original video



50% masked original video

Video Denoising

recovered video by `LMaFit`

recovered video by `APGL`

	APGL					LMaFit			
m/n	μ	iter	#sv	time	rel.err	iter	#sv	time	rel.err
76800/423	3.44e+01	34	80	516.22	4.58e-02	64	80	92.47	4.93e-02

Contribution

- Explore matrix completion by using the low-rank facotrization model
- Reduce the cost of the nonlinear Gauss-Seideal scheme by eliminating an unecessary least square problem
- Propose a nonlinear successive over-relaxation (SOR) algorithm with convergence guarantee
- Adjust the relaxation weight of SOR dynamically
- Excellent computational performance on a wide range of test problems

Matrix separation

Consider:

$$\min_{U,V,Z} \quad \|\mathcal{P}_\Omega(Z - D)\|_1 \quad \text{s.t.} \quad UV - Z = 0$$

Introduce the augmented Lagrangian function

$$\mathcal{L}_\beta(U, V, Z, \Lambda) = \|\mathcal{P}_\Omega(Z - D)\|_1 + \langle \Lambda, UV - Z \rangle + \frac{\beta}{2}\|UV - Z\|_F^2,$$

Alternating direction augmented Lagrangian framework:

$$\begin{aligned}
U^{j+1} &:= \arg\min_{U \in \mathbb{R}^{m\times k}} \mathcal{L}_\beta(U,\ V^j,\ Z^j,\ \Lambda^j),\\
V^{j+1} &:= \arg\min_{V \in \mathbb{R}^{k\times n}} \mathcal{L}_\beta(U^{j+1},\ V,\ Z^j,\ \Lambda^j),\\
Z^{j+1} &:= \arg\min_{Z \in \mathbb{R}^{m\times n}} \mathcal{L}_\beta(U^{j+1},\ V^{j+1},\ Z,\ \Lambda^j),\\
\Lambda^{j+1} &:= \Lambda^j + \gamma\beta(U^{j+1}V^{j+1} - Z^{j+1}).
\end{aligned}$$

ADM subproblems

- Let $B = Z - \Lambda/\beta$, then

$$U_+ = BV^\top (VV^\top)^\dagger \text{ and } V_+ = (U_+^\top U_+)^\dagger U_+^\top B$$

Since $U_+ V_+ = U_+ (U_+^\top U_+)^\dagger U_+^\top B = \mathcal{P}_{U_+} B$, then:

$$Q := \mathtt{orth}(BV^\top), \quad U_+ = Q \text{ and } V_+ = Q^\top B$$

- Variable Z:

$$\begin{aligned} \mathcal{P}_\Omega(Z_+) &= \mathcal{P}_\Omega\left(\mathcal{S}\left(U_+V_+ - D + \frac{\Lambda}{\beta}, \frac{1}{\beta}\right) + D\right) \\ \mathcal{P}_{\Omega^c}(Z_+) &= \mathcal{P}_{\Omega^c}\left(U_+V_+ + \frac{\Lambda}{\beta}\right) \end{aligned}$$

Theoretical results

- Let $X \triangleq (U, V, Z, \Lambda)$. The KKT conditions are

$$\Lambda V^\top = 0, \;\; U^\top \Lambda = 0, \;\; \partial_Z(\|Z - D\|_1) = \Lambda, \;\; UV - Z = 0.$$

The equation $\partial_Z(\|Z - D\|_1) = \Lambda$ is equivalent to

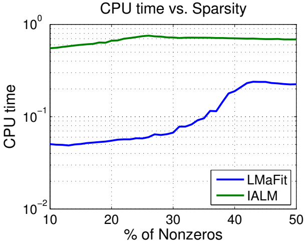
$$Z - D = \mathcal{S}(Z - D + \Lambda/\beta, 1/\beta) = \mathcal{S}(UV - D + \Lambda/\beta, 1/\beta),$$

- Let $\{X^j\}_{p=1}^{\infty}$ be an uniformly bounded sequence generated by ADM with

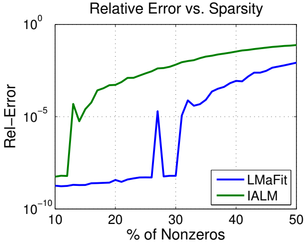
$$\lim_{p\to\infty}(X^{j+1} - X^j) = 0.$$

Then any accumulation point of $\{X^j\}_{p=1}^{\infty}$ satisfies the KKT conditions.

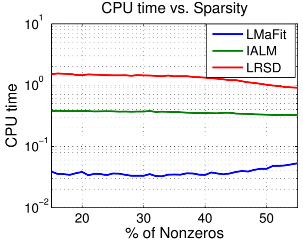
CPU and rel. err vs. sparsity: m = 100 and $k^* = 5$



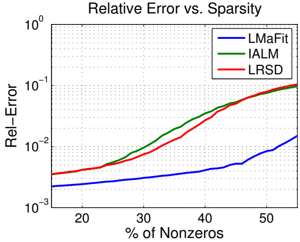
(a) CPU seconds



(b) Relative Error

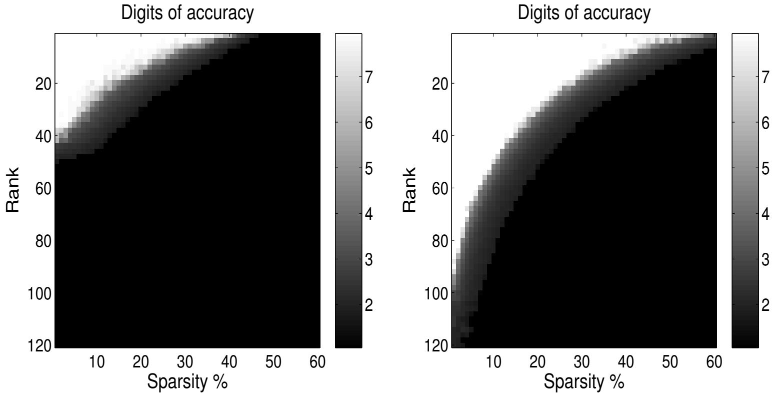


(c) CPU seconds (with noise)



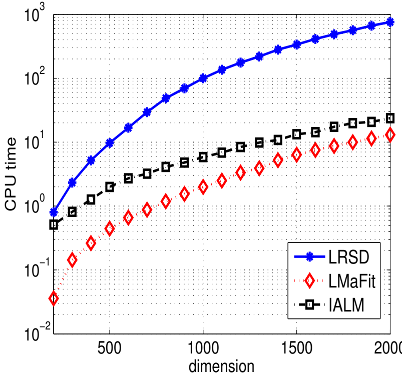
(d) Relative Error (with noise)

Recoverability test: phase plots

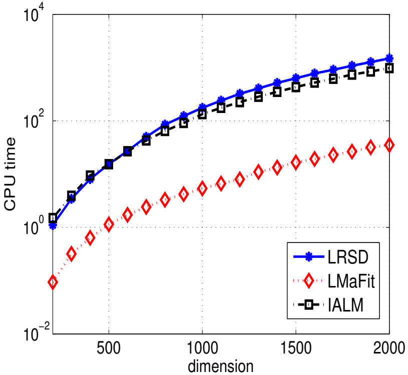


(a) IALM (b) LMaFit

Performance with respect to size



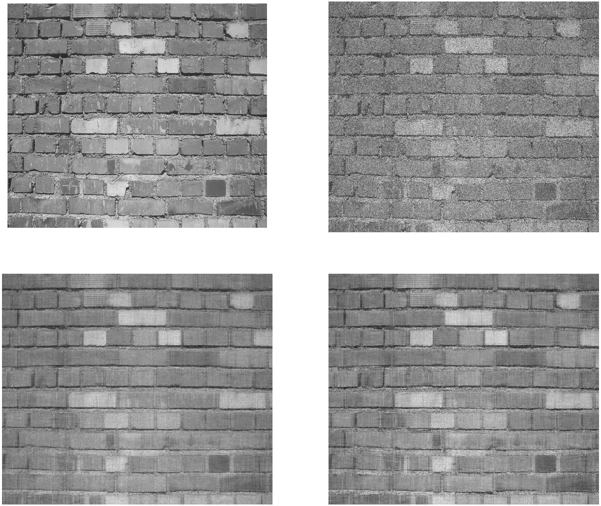
(a) k^*=0.05m and $\|S^*\|_0$=0.05m^2



(b) k^*=0.15m and $\|S^*\|_0$=0.15m^2

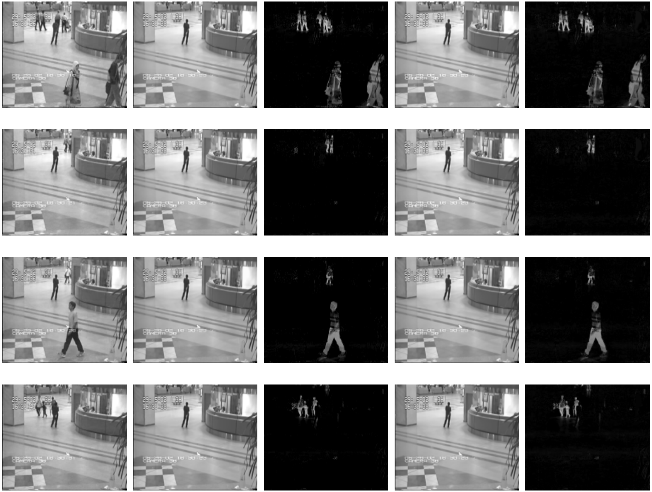
Deterministic low-rank matrices

Figure: Recovered results of “brickwall". Upper left: original image; upper right: corrupted image; lower left: IALM; lower right: `LMaFit`.



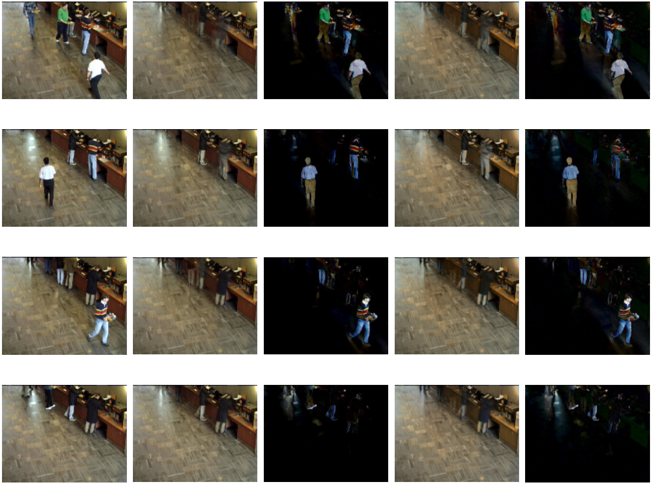
Video separation

Figure: original, separated results by IALM and `LMaFit` .



Video separation

Figure: original, separated results by IALM and LMaFit .



Upcoming Talks: Oct. 1?th

Wotao, Yin, **Fast Curvilinear Search Algorithms for Optimization with Constraints** $\|x\|_2 = 1$ **or** $X^\top X = I$

- p-harmonic flow into spheres
- polynomial optimization with normalized constraints
- maxcut SDP relaxation
- low-rank nearest correlation matrix estimation
- linear and nonlinear eigenvalue problems
- quadratic assignment problem