Fixed Point and Bregman Iterative Methods for Matrix Rank Minimization

Donald Goldfarb

Columbia University

Joint with Shiqian Ma and Lifeng Chen

Compressive Sensing Workshop
Duke University
25-26 February 2009

Matrix Rank Minimization

Affinely Constrained Matrix Rank Minimization (ACMRM) problem

$$\begin{array}{ll} \min & \operatorname{rank}(X) \\ \text{s.t.} & \mathcal{A}(X) = b, \end{array}$$

where $X \in \mathbb{R}^{m \times n}$, $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$, $b \in \mathbb{R}^p$.
Special case: Matrix Completion (MC) problem

$$\begin{array}{ll} \min & \operatorname{rank}(X) \\ \text{s.t.} & X_{ij} = M_{ij}, (i,j) \in \Omega \end{array}$$

Analogy to Compressed Sensing

- If x is square and diagonal, ACMRM becomes CS problem

$$\begin{array}{ll} \min & \|x\|_0 \\ \text{s.t.} & Ax = b, \end{array}$$

  where $x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ and $\|x\|_0 \equiv \text{card}\{x_i \neq 0\}$

- Basis Pursuit (BP):

$$\begin{array}{ll} \min & \|x\|_1 \\ \text{s.t.} & Ax = b. \end{array}$$

Theorem (Candès and Tao 2006, Rudelson and Vershynin 2005) When A is Gaussian random and partial Fourier, with high probability, BP gives the optimal solution of the CS problem for b of a size of $m = O(k \log(n/k))$ and $O(k \log(n)^4)$, respectively.

NNM for Affinely Constrained MRM

Nuclear Norm Minimization (NNM):

$$\begin{array}{ll} \min & \|X\|_* \\ \text{s.t.} & \mathcal{A}(X) = b, \end{array}$$

where $\|X\|_* = \sum_i \sigma_i$ and $\sigma_i = i$th singular value of matrix X.

NNM for Affinely Constrained MRM

Nuclear Norm Minimization (NNM):

$$\begin{array}{ll} \min & \|X\|_* \\ \text{s.t.} & \mathcal{A}(X) = b, \end{array}$$

where $\|X\|_* = \sum_i \sigma_i$ and $\sigma_i = i$th singular value of matrix X.

Theorem (Recht, Fazel and Parrilo, 2007)
Rewrite $\mathcal{A}(X) = b$ as $A\,\mathrm{vec}(X) = b$. If the entries of $A \in \mathbf{R}^{p \times mn}$ are suitably random, e.g., i.i.d. Gaussian, then with very high probability, $m \times n$ matrices of rank r can be recovered by solving the NNM problem whenever

$$p \geq Cr(m+n)\log(mn),$$

where C is a positive constant.

NNM for Matrix Completion

Theorem (Candès and Recht, 2008)
Let $M \in \mathbb{R}^{n_1 \times n_2}$ have rank r with SVD $M = \sum_{k=1}^{r} \sigma_k u_k v_k^\top$, where the families $\{u_k\}_{1 \le k \le r}$ and $\{v_k\}_{1 \le k \le r}$ are selected uniformly at random among all families of r orthonormal vectors. Let $n = \max(n_1, n_2)$. Then $\exists C, c$ s.t. if

$$|\Omega| \equiv p \ge C n^{5/4} r \log n,$$

the minimizer of the problem NNM is unique and equal to M with probability at least $1 - cn^{-3}$. In addition, if $r \le n^{1/5}$, then the recovery is exact with probability at least $1 - cn^{-3}$ provided that

$$p \ge C n^{6/5} r \log n.$$

Dual Problem of NNM

Dual Problem of NNM:

$$\begin{array}{ll} \max & b^\top z \\ \text{s.t.} & \|\mathcal{A}^*(z)\|_2 \leq 1. \end{array}$$

SDP formulation of NNM:

$$\begin{array}{ll} \min\limits_{X,W_1,W_2} & \frac{1}{2}(\mathrm{Tr}(W_1) + \mathrm{Tr}(W_2)) \\ \text{s.t.} & \begin{bmatrix} W_1 & X \\ X^\top & W_2 \end{bmatrix} \succeq 0 \\ & \mathcal{A}(X) = b. \end{array}$$

SDP formulation of Dual of NNM:

$$\begin{array}{ll} \max\limits_{z} & b^\top z \\ \text{s.t.} & \begin{bmatrix} I_m & \mathcal{A}^*(z) \\ \mathcal{A}^*(z)^\top & I_n \end{bmatrix} \succeq 0. \end{array}$$

Optimality Conditions for Unconstrained NNM Problem

- Unconstrained Nuclear Norm Minimization (UNNM):

$$\min \mu\|X\|_* + \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2.$$

- Optimality condition:

$$\mathbf{0} \in \mu\partial\|X^*\|_* + \mathcal{A}^*(\mathcal{A}(X^*) - b).$$

$$\partial\|X\|_* = \{UV^\top + W : U^\top W = 0, WV = 0, \|W\|_2 \leq 1\}.$$

Optimality Conditions for Unconstrained NNM Problem

- Unconstrained Nuclear Norm Minimization (UNNM):

$$\min \mu\|X\|_* + \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2.$$

- Optimality condition:

$$\mathbf{0} \in \mu\partial\|X^*\|_* + \mathcal{A}^*(\mathcal{A}(X^*) - b).$$

$$\partial\|X\|_* = \{UV^\top + W : U^\top W = 0, WV = 0, \|W\|_2 \leq 1\}.$$

Theorem: Let $X \in \mathbb{R}^{m\times n}$ have SVD $X = U\Sigma V^\top$. Then X is optimal for UNNM iff $\exists$ a matrix $W \in \mathbb{R}^{m\times n}$ s.t.

$$\begin{aligned} &\mu(UV^\top + W) + \mathcal{A}^*(\mathcal{A}(X) - b) = 0, \\ &U^\top W = 0, WV = 0, \|W\|_2 \leq 1. \end{aligned}$$

Operator Splitting

$$\mathbf{0} \in \mu \partial \|X^*\|_* + \mathcal{A}^*(\mathcal{A}(X^*) - b),$$

Let

$$Y^* = X^* - \tau \mathcal{A}^*(\mathcal{A}(X^*) - b),$$

then the optimality condition reduces to

$$\mathbf{0} \in \tau \mu \partial \|X^*\|_* + X^* - Y^*,$$

i.e., X^* is the optimal solution to

$$\min_{X \in \mathbb{R}^{m \times n}} \tau \mu \|X\|_* + \frac{1}{2}\|X - Y^*\|_F^2$$

Matrix Shrinkage Operator

Nonnegative Vector Shrinkage Operator. Assume $x \in \mathbb{R}^n_+$. $\forall \nu > 0$,

$$s_\nu(x) := \bar{x}, \text{ with } \bar{x}_i = \begin{cases} x_i - \nu, & \text{if } x_i - \nu > 0 \\ 0, & \text{o.w.} \end{cases}$$

Matrix Shrinkage Operator. Assume $X \in \mathbb{R}^{m \times n}$ and the SVD of X is $X = U\text{Diag}(\sigma)V^\top$, $U \in \mathbb{R}^{m \times r}, \sigma \in \mathbb{R}^r_+, V \in \mathbb{R}^{n \times r}$. $\forall \nu > 0$,

$$S_\nu(X) := U\text{Diag}(\bar{\sigma})V^\top, \qquad \text{with } \bar{\sigma} = s_\nu(\sigma).$$

Matrix Shrinkage Operator (Cont.)

Theorem: Given $Y \in \mathbb{R}^{m\times n}$, $\mathrm{rank}(Y) = t$ and SVD $Y = U_Y \mathrm{Diag}(\gamma) V_Y^\top$, where $U_Y \in \mathbb{R}^{m\times t}, \gamma \in \mathbb{R}^t_+, V_Y \in \mathbb{R}^{n\times t}$, and a scalar $\nu > 0$,

$$X := S_\nu(Y) = U_Y \mathrm{Diag}(s_\nu(\gamma)) V_Y^\top$$

is an optimal solution of the problem

$$\min_{X\in\mathbb{R}^{m\times n}} f(X) := \nu\|X\|_* + \frac{1}{2}\|X - Y\|_F^2.$$

Fixed Point Method for UNNM

Fixed Point Iterative Scheme

$$\begin{cases} Y^k = X^k - \tau \mathcal{A}^*(\mathcal{A}(X^k) - b) \\ X^{k+1} = S_{\tau\mu}(Y^k). \end{cases}$$

Lemma: Matrix shrinkage operator is non-expansive. i.e.,

$$\|S_\nu(Y_1) - S_\nu(Y_2)\|_F \leq \|Y_1 - Y_2\|_F.$$

Theorem: The sequence $\{X^k\}$ generated by the fixed point iterations converges to some $X^* \in \mathcal{X}^*$ (the optimal set of UNNM).

Fixed Point Continuation Algorithm for UNNM

- Initialize: Given X_0, $\bar{\mu} > 0$. Select $\mu_1 > \mu_2 > \cdots > \mu_L = \bar{\mu} > 0$. Set $X = X_0$.
- **for** $\mu = \mu_1, \mu_2, \ldots, \mu_L$, **do**
  - **while** NOT converged, **do**
    - select $\tau > 0$
    - compute $Y = X - \tau \mathcal{A}^*(\mathcal{A}(X) - b)$, and SVD of Y, $Y = U\text{Diag}(\sigma)V^\top$
    - compute $X = U\text{Diag}(s_{\tau\mu}(\sigma))V^\top$
  - **end while**

end for

Bregman Iterative Method

- ℓ_1-regularized problem

$$\min_x J(x) + \frac{1}{2}\|Ax - b\|_2^2, \text{ where } J(x) = \mu\|x\|_1.$$

- Bregman distance:

$$D_J^p(u, v) := J(u) - J(v) - < p, u - v >, \text{ where } p \in \partial J(v).$$

- Bregman iterative regularization procedure

$$x^{k+1} \leftarrow \min_x D_J^{p^k}(x, x^k) + \frac{1}{2}\|Ax - b\|_2^2$$

Bregman Iterative Scheme

Optimality condition: $\mathbf{0} \in \partial J(x^{k+1}) - p^k + A^\top(Ax^{k+1} - b)$, thus

$$p^{k+1} := p^k - A^\top(Ax^{k+1} - b).$$

So the Bregman iterative scheme is

$$\begin{cases} x^{k+1} \leftarrow \min_x D_J^{p^k}(x, x^k) + \frac{1}{2}\|Ax - b\|_2^2 \\ p^{k+1} = p^k - A^\top(Ax^{k+1} - b). \end{cases}$$

or equivalently,

$$\begin{cases} b^{k+1} = b + (b^k - Ax^k) \\ x^{k+1} \leftarrow \min_x J(x) + \frac{1}{2}\|Ax - b^{k+1}\|_2^2. \end{cases}$$

Bregman Iterative Method for NNM

Bregman Iterative Method

- $b^0 \leftarrow \mathbf{0}, X^0 \leftarrow \mathbf{0}$,
- for $k = 0, 1, \ldots$ do
- $b^{k+1} \leftarrow b + (b^k - \mathcal{A}(X^k))$,
- $X^{k+1} \leftarrow \arg\min_X \mu\|X\|_* + \frac{1}{2}\|\mathcal{A}(X) - b^{k+1}\|_2^2$.

Approximate SVD Technique

Monte-Carlo approximate SVD (Drineas et.al.2006)

- **Input:** $A \in \mathbb{R}^{m \times n}$, $1 \leq k \leq c \leq n$.
- **Output:** $U_k \in \mathbb{R}^{m \times k}$ and Σ_k.
  - For $j = 1$ to c,
    - Randomly choose a column $A^{(i)}$ of A
    - Set $C^{(j)} = A^{(i)}/\sqrt{c/n}$.
  - Compute SVD of $C^\top C$: $\sum_{j=1}^{c} \sigma_j^2 y^j {y^j}^\top$.
  - Compute $u^j = Cy^j/\sigma_j$ for $j = 1, \ldots, k$.
  - Return U_k, where $U_k^{(j)} = u^j$, and $\Sigma_k = diag(\sigma_j, j = 1, \cdots, k)$.

Approximate SVD Technique

Monte-Carlo approximate SVD (Drineas et.al.2006)

- **Input:** $A \in \mathbb{R}^{m\times n}$, $1 \le k \le c \le n$.
- **Output:** $U_k \in \mathbb{R}^{m\times k}$ and Σ_k.
  - For $j = 1$ to c,
    - Randomly choose a column $A^{(i)}$ of A
    - Set $C^{(j)} = A^{(i)}/\sqrt{c/n}$.
  - Compute SVD of $C^\top C$: $\sum_{j=1}^{c} \sigma_j^2 y^j {y^j}^\top$.
  - Compute $u^j = Cy^j/\sigma_j$ for $j = 1, \ldots, k$.
  - Return U_k, where $U_k^{(j)} = u^j$, and $\Sigma_k = diag(\sigma_j, j = 1, \cdots, k)$.

Theorem: With high probability, the following estimate holds for both $\xi = 2$ and $\xi = F$:

$$\|A - A_{k_s}\|_\xi^2 \le \min_{D:\text{rank}(D)\le k_s} \|A - D\|_\xi^2 + poly(k_s, 1/c_s)\|A\|_F^2,$$

where $A_k = U_k \Sigma_k V_k^\top$, $V_k = A^\top U_k \Sigma_k^{-1}$.

Numerical Tests: Stopping Rules and Solvers

$$(1) \quad \|U_k V_k^\top + g^k/\mu\|_2 - 1 < gtol, \quad (2) \quad \frac{\|X^{k+1} - X^k\|_F}{\max\{1, \|X^k\|_F\}} < xtol,$$

- FPC1. Exact SVD, stopping rule: (2).
- FPC2. Exact SVD, stopping rule: (1) and (2).
- FPC3. Exact SVD with debiasing, stopping rule: (2).
- FPCA. Approximate SVD, stopping rule: (2).
- Bregman. Bregman iterative method using FPC2 to solve the subproblems.

Numerical Tests Randomly Created MC Problems

- Generation: generate matrices $M_L \in \mathbb{R}^{m\times r}$ and $M_R \in \mathbb{R}^{n\times r}$ with i.i.d. Gaussian entries; set $M = M_L M_R^\top$.
- Sample a subset Ω of p entries of M uniformly at random.

Measures:

- $rel.err. := \frac{\|X_{opt} - M\|_F}{\|M\|_F}$; Claim recovery if $rel.err. < 1e-3$.
- $SR = p/(mn)$ (sampling ratio)
- $FR = r(m+n-r)/p$ (Note if $FR > 1$, it is not possible to recover the matrix)
- $NS =$ the number of problems successfully solved

Comparisons on small problems (m=n=40,p=800,SR=0.5)

r	FR	Solver	NS	avg. secs.	avg. rel.err.
1	0.0988	FPC1	50	1.81	1.67e-9
		FPC2	50	3.61	1.32e-9
		FPC3	50	16.81	1.06e-9
		SDPT3	50	1.81	6.30e-10
2	0.1950	FPC1	42	3.05	1.01e-6
		FPC2	42	17.97	1.01e-6
		FPC3	49	16.86	1.26e-5
		SDPT3	44	1.90	1.50e-9
3	0.2888	FPC1	35	5.50	9.72e-9
		FPC2	35	20.33	2.17e-9
		FPC3	42	16.87	3.58e-5
		SDPT3	37	1.95	2.66e-9
4	0.3800	FPC1	22	9.08	7.91e-5
		FPC2	22	18.43	7.91e-5
		FPC3	29	16.95	3.83e-5
		SDPT3	29	2.09	1.18e-8
5	0.4688	FPC1	1	10.41	2.10e-8
		FPC2	1	17.88	2.70e-9
		FPC3	5	16.70	1.78e-4
		SDPT3	8	2.26	1.83e-7
6	0.5550	FPC1	0	—	—
		FPC2	0	—	—
		FPC3	0	—	—
		SDPT3	1	2.87	6.58e-7

Comparison between FPC and Bregman (m=n=40, p=800, SR = 0.5)

Problem			FPC2	Bregman
r	FR	NIM (NS)	max. rel.err	max. rel.err
1	0.0988	32 (50)	2.22e-9	1.87e-15
2	0.1950	29 (42)	5.01e-9	2.96e-15
3	0.2888	24 (35)	2.77e-9	2.93e-15
4	0.3800	10 (22)	5.51e-9	3.11e-15

Comparison of FPCA and SDPT3 (m=n=40,p=800,SR=0.5)

Problems		FPCA			SDPT3		
r	FR	NS	avg. sec.	avg. rel.err	NS	avg. secs.	avg. rel.err
1	0.0988	50	4.24	6.60e-7	50	1.84	6.30e-1
2	0.1950	50	4.35	1.08e-6	44	1.93	1.50e-9
3	0.2888	50	4.83	1.83e-6	37	1.99	2.66e-9
4	0.3800	50	4.92	2.56e-6	29	2.12	1.18e-8
5	0.4688	50	5.06	3.38e-6	8	2.30	1.83e-7
6	0.5550	50	5.48	3.72e-6	1	2.89	6.58e-7
7	0.6388	50	5.79	4.78e-6	0	—	—
8	0.7200	50	6.03	8.57e-6	0	—	—
9	0.7987	49	6.75	1.27e-5	0	—	—
10	0.8750	32	8.71	7.49e-5	0	—	—
11	0.9487	0	—	—	0	—	—

$FR = r(m + n - r)/p$

Medium sized matrices: (m=n=100,p=2000,SR=0.2)

Problems		FPCA			SDPT3		
r	FR	NS	avg. secs.	avg. rel.err	NS	avg. secs.	avg. rel.err
1	0.0995	50	7.94	6.11e-6	47	15.10	1.55e-9
2	0.1980	50	8.17	6.51e-6	31	16.02	7.95e-9
3	0.2955	50	9.09	7.36e-6	13	19.23	1.05e-4
4	0.3920	50	9.33	1.09e-5	0	—	—
5	0.4875	49	9.91	2.99e-5	0	—	—
6	0.5820	47	10.81	3.99e-5	0	—	—
7	0.6755	44	12.63	8.87e-5	0	—	—
8	0.7680	31	16.30	1.24e-4	0	—	—
9	0.8595	2	17.88	6.19e-4	0	—	—
10	0.9500	0	—	—	0	—	—

$FR = r(m + n - r)/p$

Medium sized matrices: (m=n=100,p=3000,SR=0.3)

Problems		FPCA			SDPT3		
r	FR	NS	avg. secs.	avg. rel.err	NS	avg. secs.	avg. rel.err
1	0.0663	50	8.39	1.83e-6	50	36.68	2.01e-10
2	0.1320	50	8.53	1.86e-6	50	36.50	1.13e-9
3	0.1970	50	9.30	2.11e-6	46	38.50	1.28e-5
4	0.2613	50	9.72	2.88e-6	42	41.28	4.60e-6
5	0.3250	50	9.87	3.60e-6	32	43.92	7.82e-8
6	0.3880	50	9.96	3.93e-6	17	49.60	3.44e-7
7	0.4503	50	10.19	4.27e-6	3	59.18	1.43e-4
8	0.5120	50	10.65	4.38e-6	0	—	—
9	0.5730	50	11.74	5.01e-6	0	—	—
10	0.6333	50	11.76	6.30e-6	0	—	—
11	0.6930	50	12.08	8.29e-6	0	—	—
12	0.7520	50	13.67	2.64e-5	0	—	—
13	0.8103	48	16.00	2.95e-5	0	—	—
14	0.8680	40	20.51	1.35e-4	0	—	—
15	0.9250	0	—	—	0	—	—
16	0.9813	0	—	—	0	—	—

$FR = r(m + n - r)/p$

Large matrices: (m=n=1000,p=2e+5,SR=0.2)

Problems		FPCA		
r	FR	NS	avg. secs.	avg. rel.err
50	0.4875	10	1500.7	2.73e-6
51	0.4970	10	1510.2	2.75e-6
52	0.5065	10	1515.0	2.80e-6
53	0.5160	10	1520.6	2.79e-6
54	0.5254	10	1535.9	2.77e-6
55	0.5349	10	1543.6	2.80e-6
56	0.5443	10	1556.3	2.78e-6
57	0.5538	10	1567.3	2.74e-6
58	0.5632	10	1586.4	2.69e-6
59	0.5726	10	1576.1	2.66e-6
60	0.5820	10	1602.0	2.55e-6

Real Data Set: Jester Joke Set (Goldberg, 2001)

- Hold out 2 ratings for each user.
- Mean Absolute Error(MAE)

$$MAE = \frac{1}{2N} \sum_{i=1}^{N} |\hat{r}_{i_1}^i - r_{i_1}^i| + |\hat{r}_{i_2}^i - r_{i_2}^i|.$$

- Normalized Mean Absolute Error(NMAE) $NMAE = \frac{MAE}{r_{max} - r_{min}}$

Numerical Results

Table: Numerical results of FPC1 for Jester joke data set

num.user	num.samp	samp.ratio	rank	$\sigma_{\max}$	$\sigma_{\min}$	NMAE	Time
100	7172	0.7172	79	285.6520	3.4916e-004	0.1727	34.3
1000	71152	0.7115	100	786.3651	38.4326	0.1667	304.8125
2000	140691	0.7035	100	1.1242e+003	65.0607	0.1582	661.6563

Table: Numerical results of FPCA for Jester joke data set

num.user	num.samp	samp.ratio	ϵ_{k_s}	c_s	rank	$\sigma_{\max}$	$\sigma_{\min}$	NMAE	Time
100	7172	0.7172	1e-2	25	20	295.1449	32.6798	0.1627	26.7344
1000	71152	0.7115	1e-2	100	85	859.2710	48.0393	0.2008	808.5156
1000	71152	0.7115	1e-4	100	90	859.4588	44.6220	0.2101	778.5625
2000	140691	0.7035	1e-4	200	100	1.1518e+003	63.5244	0.1564	1.1345e+003