A non-monotone alternating updating method for a class of matrix factorization problems

Ting Kei Pong
Department of Applied Mathematics
The Hong Kong Polytechnic University
Hong Kong

SIAM-ALA 2018
May 2018

(Joint work with Xiaojun Chen and Lei Yang)

Slides largely adapted from Lei Yang's previous talk

Motivating applications

Factorization-related optimization problems:

- Non-negative Matrix Factorization (NMF) (Paatero, Tapper, '94; Lee, Seung, '99)

$$\begin{aligned} \min_{X,Y} \quad & \tfrac{1}{2}\|XY^T - M\|_F^2 \\ \text{s.t.} \quad & X \geq 0, \quad Y \geq 0, \end{aligned}$$

  where $M \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{m\times r}$ and $Y \in \mathbb{R}^{n\times r}$, $r \leq \min\{m, n\}$.

- Matrix Completion (MC) (Shang, Liu, Cheng, '16)

$$\min_{X,Y} \quad \eta\|X\|_* + \eta\|Y\|_* + \|\mathcal{P}_\Omega(XY^\top - M)\|_F^2,$$

  where $\|\cdot\|_*$ is the nuclear norm.

- Other matrix completion models (Wen, Yin, Zhang, '12; Sun, Luo, '16)

General model

$$\min_{X,Y} \ \mathcal{F}(X,Y) := \Psi(X) + \Phi(Y) + \frac{1}{2}\|\mathcal{A}(XY^T) - \boldsymbol{b}\|^2,$$

where:

- $X \in \mathbb{R}^{m\times r}$ and $Y \in \mathbb{R}^{n\times r}$ with $r \leq \min\{m,n\}$, $\boldsymbol{b} \in \mathbb{R}^q$;
- Ψ, $\Phi : \mathbb{R}^{m\times n} \to \mathbb{R}_+ \cup \{\infty\}$ are proper closed and level-bounded, and are continuous on their domain;
- $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^q$ is a linear map with $q \leq mn$ and $\mathcal{A}\mathcal{A}^* = \mathcal{I}$.

General model

$$\min_{X,Y} \;\; \mathcal{F}(X, Y) := \Psi(X) + \Phi(Y) + \frac{1}{2}\|\mathcal{A}(XY^T) - \boldsymbol{b}\|^2,$$

where:

- $X \in \mathbb{R}^{m\times r}$ and $Y \in \mathbb{R}^{n\times r}$ with $r \le \min\{m, n\}$, $\boldsymbol{b} \in \mathbb{R}^q$;
- Ψ, $\Phi : \mathbb{R}^{m\times n} \to \mathbb{R}_+ \cup \{\infty\}$ are proper closed and level-bounded, and are continuous on their domain;
- $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^q$ is a linear map with $q \le mn$ and $\mathcal{A}\mathcal{A}^* = \mathcal{I}$.

Idea: Adapt techniques in NMF to deal with XY^T?

General model

$$\min_{X,Y} \; \mathcal{F}(X,Y) := \Psi(X) + \Phi(Y) + \frac{1}{2}\|\mathcal{A}(XY^T) - \boldsymbol{b}\|^2,$$

where:

- $X \in \mathbb{R}^{m\times r}$ and $Y \in \mathbb{R}^{n\times r}$ with $r \leq \min\{m,n\}$, $\boldsymbol{b} \in \mathbb{R}^q$;
- Ψ, $\Phi : \mathbb{R}^{m\times n} \to \mathbb{R}_+ \cup \{\infty\}$ are proper closed and level-bounded, and are continuous on their domain;
- $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^q$ is a linear map with $q \leq mn$ and $\mathcal{A}\mathcal{A}^* = \mathcal{I}$.

Idea: Adapt techniques in NMF to deal with XY^T?

Not trivial! The product XY^T is coupled with $\mathcal{A}$... Can we "split" them?

Key ideas I

Perform "lifting": Consider

$$\Theta_{\alpha,\beta}(X, Y, Z) = \Psi(X) + \Phi(Y) + \frac{\alpha}{2}\|XY^\top - Z\|_F^2 + \frac{\beta}{2}\|\mathcal{A}(Z) - \boldsymbol{b}\|^2 .$$

Key ideas I

Perform "lifting": Consider

$$\Theta_{\alpha,\beta}(X, Y, Z) = \Psi(X) + \Phi(Y) + \frac{\alpha}{2}\|XY^\top - Z\|_F^2 + \frac{\beta}{2}\|\mathcal{A}(Z) - \boldsymbol{b}\|^2 .$$

Fact 1: (Yang, P., Chen '17)
If $\alpha\mathcal{I} + \beta\mathcal{A}^*\mathcal{A} \succ 0$ and $\frac{1}{\alpha} + \frac{1}{\beta} = 1$, then

$$\min_{X,Y} \{\mathcal{F}(X, Y)\} \text{ is equivalent to } \min_{X,Y,Z} \{\Theta_{\alpha,\beta}(X, Y, Z)\}.$$

Key ideas I

Perform "lifting": Consider

$$\Theta_{\alpha,\beta}(X,Y,Z) = \Psi(X) + \Phi(Y) + \frac{\alpha}{2}\|XY^\top - Z\|_F^2 + \frac{\beta}{2}\|\mathcal{A}(Z) - \boldsymbol{b}\|^2.$$

Fact 1: (Yang, P., Chen '17)
If $\alpha\mathcal{I} + \beta\mathcal{A}^*\mathcal{A} \succ 0$ and $\frac{1}{\alpha} + \frac{1}{\beta} = 1$, then

$$\min_{X,Y}\{\mathcal{F}(X,Y)\} \text{ is equivalent to } \min_{X,Y,Z}\{\Theta_{\alpha,\beta}(X,Y,Z)\}.$$

Conceptual algorithm 1: In the k-th iteration,

- Approx. minimize Θ w.r.t Z to get Z^k.
- Approx. minimize Θ w.r.t X to get X^{k+1}.
- Approx. minimize Θ w.r.t Y to get Y^{k+1}.

Globalize using suitable line-search techniques.

Key ideas II

In our tests, requiring $\alpha\mathcal{I} + \beta\mathcal{A}^*\mathcal{A} \succ 0$ leads to slow convergence. ☹

Key ideas II

In our tests, requiring $\alpha\mathcal{I} + \beta\mathcal{A}^*\mathcal{A} \succ 0$ leads to slow convergence. ☹

Fact 2: (Yang, P., Chen '17)
Suppose that $\frac{1}{\alpha} + \frac{1}{\beta} = 1$.

- If $0 \in \partial\Theta_{\alpha,\beta}(X^*, Y^*, Z^*)$, then $0 \in \partial\mathcal{F}(X^*, Y^*)$;
- If $0 \in \partial\mathcal{F}(X^*, Y^*)$, then $0 \in \partial\Theta_{\alpha,\beta}(X^*, Y^*, Z^*)$, where

$$Z^* = \left(\mathcal{I} - \tfrac{\beta}{\alpha+\beta}\mathcal{A}^*\mathcal{A}\right)(X^*(Y^*)^T) + \tfrac{\beta}{\alpha+\beta}\mathcal{A}^*\boldsymbol{b}.$$

Key ideas II

In our tests, requiring $\alpha\mathcal{I} + \beta\mathcal{A}^*\mathcal{A} \succ 0$ leads to slow convergence. ☹

Fact 2: (Yang, P., Chen '17)
Suppose that $\frac{1}{\alpha} + \frac{1}{\beta} = 1$.

- If $0 \in \partial\Theta_{\alpha,\beta}(X^*, Y^*, Z^*)$, then $0 \in \partial\mathcal{F}(X^*, Y^*)$;
- If $0 \in \partial\mathcal{F}(X^*, Y^*)$, then $0 \in \partial\Theta_{\alpha,\beta}(X^*, Y^*, Z^*)$, where

$$Z^* = \left(\mathcal{I} - \frac{\beta}{\alpha+\beta}\mathcal{A}^*\mathcal{A}\right)(X^*(Y^*)^T) + \frac{\beta}{\alpha+\beta}\mathcal{A}^*\boldsymbol{b}.$$

Conceptual algorithm 2: In the k-th iteration,

- Obtain Z^k by solving $\nabla_Z\Theta_{\alpha,\beta}(X^k, Y^k, Z) = 0$.
- Approx. minimize Θ w.r.t X to get X^{k+1}.
- Approx. minimize Θ w.r.t Y to get Y^{k+1}.

Globalize using suitable line-search techniques.

Key ideas II

In our tests, requiring $\alpha\mathcal{I} + \beta\mathcal{A}^*\mathcal{A} \succ 0$ leads to slow convergence. ☹

Fact 2: (Yang, P., Chen '17)
Suppose that $\frac{1}{\alpha} + \frac{1}{\beta} = 1$.

- If $0 \in \partial\Theta_{\alpha,\beta}(X^*, Y^*, Z^*)$, then $0 \in \partial\mathcal{F}(X^*, Y^*)$;
- If $0 \in \partial\mathcal{F}(X^*, Y^*)$, then $0 \in \partial\Theta_{\alpha,\beta}(X^*, Y^*, Z^*)$, where

$$Z^* = \left(\mathcal{I} - \frac{\beta}{\alpha+\beta}\mathcal{A}^*\mathcal{A}\right)(X^*(Y^*)^T) + \frac{\beta}{\alpha+\beta}\mathcal{A}^*\boldsymbol{b}.$$

Conceptual algorithm 2: In the k-th iteration,

- Obtain Z^k by solving $\nabla_Z\Theta_{\alpha,\beta}(X^k, Y^k, Z) = 0$. **Not a descent step!**
- Approx. minimize Θ w.r.t X to get X^{k+1}.
- Approx. minimize Θ w.r.t Y to get Y^{k+1}.

Globalize using suitable line-search techniques.

General Framework

Non-monotone Alternating Updating Method (NAUM)

Step 1. Compute Z^k by

$$Z^k = \left(\mathcal{I} - \frac{\beta}{\alpha+\beta}\mathcal{A}^*\mathcal{A}\right)\left(X^k(Y^k)^T\right) + \frac{\beta}{\alpha+\beta}\mathcal{A}^*\boldsymbol{b}.$$

Step 2. Perform non-monotone line search:

- **(2a)** Compute U by "**Prox**", "**Prox-linear**" or "**Hi-prox**".
- **(2b)** Compute V by "**Prox**", "**Prox-linear**" or "**Hi-prox**".
- **(2c)** If a certain line search criterion is satisfied, go to **Step 3**.
- **(2d)** Update related parameters

Step 3. Set $(X^{k+1}, Y^{k+1}) \leftarrow (U, V)$, $k \leftarrow k+1$ and go to **Step 1**.

The Complete NAUM

Non-monotone Alternating Updating Method (NAUM)

Input. (X^0, Y^0), α, β, γ, ρ, $\tau > 1$, $c > 0$, $\mu^{\min} > 0$, $\sigma^{\max} > \sigma^{\min} > 0$, $N \geq 0$, where $\frac{1}{\alpha} + \frac{1}{\beta} = 1$, $\rho = \|\mathcal{I} - \frac{\beta}{\alpha+\beta}\mathcal{A}^*\mathcal{A}\|^2$ and $(\alpha + \gamma)\mathcal{I} + \beta\mathcal{A}^*\mathcal{A} \succeq 0$.

Step 1. Compute Z^k by

$$Z^k = \left(\mathcal{I} - \frac{\beta}{\alpha + \beta}\mathcal{A}^*\mathcal{A}\right)\left(X^k(Y^k)^\top\right) + \frac{\beta}{\alpha + \beta}\mathcal{A}^*\boldsymbol{b}.$$

Step 2. Choose $\mu_k^0 \geq \mu^{\min}$ and $\sigma_k^0 \in [\sigma^{\min}, \sigma^{\max}]$ arbitrarily. Set $\tilde{\mu}_k = \mu_k^0$, $\sigma_k = \sigma_k^0$ and $\mu_k^{\max} = (\alpha + 2\gamma\rho)\|Y^k\|^2 + c$.

- **(2a)** Set $\mu_k \leftarrow \min\{\tilde{\mu}_k, \mu_k^{\max}\}$. Compute U by "**Prox**", "**Prox-linear**" or "**Hi-prox**".
- **(2b)** Compute V by "**Prox**", "**Prox-linear**" or "**Hi-prox**".
- **(2c)** If $\mathcal{F}(U, V) - \max_{[k-N]_+ \leq i \leq k} \mathcal{F}(X^i, Y^i) \leq -\frac{c}{2}\left(\|U - X^k\|_F^2 + \|V - Y^k\|_F^2\right)$, is satisfied, go to **Step 3**.
- **(2d)** If $\mu_k = \mu_k^{\max}$, set $\sigma_k^{\max} = (\alpha + 2\gamma\rho)\|U\|^2 + c$, $\sigma_k \leftarrow \min\{\tau\sigma_k, \sigma_k^{\max}\}$ and then, go to step **(2b)**; otherwise, set $\tilde{\mu}_k \leftarrow \tau\mu_k$ and $\sigma_k \leftarrow \tau\sigma_k$ and then, go to step **(2a)**.

Step 3. Set $(X^{k+1}, Y^{k+1}) \leftarrow (U, V)$, $k \leftarrow k + 1$ and go to **Step 1**.

Updating Schemes of U

Let $\mathcal{H}_\alpha(X, Y, Z) := \frac{\alpha}{2}\|XY^\top - Z\|_F^2$. Compute a candidate U of X^{k+1}:

- **Prox**: $U \in \operatorname*{Argmin}_X \; \Psi(X) + \mathcal{H}_\alpha(X, Y^k, Z^k) + \frac{\mu_k}{2}\|X - X^k\|_F^2$.

Updating Schemes of U

Let $\mathcal{H}_\alpha(X,Y,Z) := \frac{\alpha}{2}\|XY^\top - Z\|_F^2$. Compute a candidate U of X^{k+1}:

- **Prox**: $U \in \operatorname*{Argmin}_X \ \Psi(X) + \mathcal{H}_\alpha(X, Y^k, Z^k) + \frac{\mu_k}{2}\|X - X^k\|_F^2$.
- **Prox-linear**:

$$U \in \operatorname*{Argmin}_X \ \Psi(X) + \langle \nabla_X \mathcal{H}_\alpha(X^k, Y^k, Z^k),\ X - X^k\rangle + \frac{\mu_k}{2}\|X - X^k\|_F^2.$$

Updating Schemes of U

Let $\mathcal{H}_\alpha(X, Y, Z) := \frac{\alpha}{2}\|XY^\top - Z\|_F^2$. Compute a candidate U of X^{k+1}:

- **Prox**: $U \in \operatorname*{Argmin}_X \; \Psi(X) + \mathcal{H}_\alpha(X, Y^k, Z^k) + \frac{\mu_k}{2}\|X - X^k\|_F^2$.
- **Prox-linear**:

$$U \in \operatorname*{Argmin}_X \; \Psi(X) + \langle \nabla_X \mathcal{H}_\alpha(X^k, Y^k, Z^k), \, X - X^k \rangle + \frac{\mu_k}{2}\|X - X^k\|_F^2.$$

If $\Psi(X) = \sum_{i=1}^{r} \psi_i(\boldsymbol{x}_i)$ with $X = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r]$, update U columnwise[1]:

- **Hi-prox**:

$$\boldsymbol{u}_i \in \operatorname*{Argmin}_{\boldsymbol{x}_i} \; \psi_i(\boldsymbol{x}_i) + \mathcal{H}_\alpha(\boldsymbol{u}_{j<i}, \boldsymbol{x}_i, \boldsymbol{x}^k_{j>i}, Y^k, Z^k) + \frac{\mu_k}{2}\|\boldsymbol{x}_i - \boldsymbol{x}_i^k\|^2.$$

[1] thanks to $XY^\top = \sum_{i=1}^r \boldsymbol{x}_i \boldsymbol{y}_i^\top$

Updating Schemes of V

After computing U, we compute a candidate V of Y:

- **Prox**: $V \in \operatorname*{Argmin}_{Y} \; \Phi(Y) + \mathcal{H}_\alpha(U, Y, Z^k) + \frac{\sigma_k}{2}\|Y - Y^k\|_F^2$
- **Prox-linear**:

$$V \in \operatorname*{Argmin}_{Y} \; \Phi(Y) + \langle \nabla_Y \mathcal{H}_\alpha(U, Y^k, Z^k),\; Y - Y^k \rangle + \frac{\sigma_k}{2}\|Y - Y^k\|_F^2.$$

If $\Phi(Y) = \sum_{i=1}^{r} \phi_i(\boldsymbol{y}_i)$ with $Y = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_r]$, update V columnwise:

- **Hi-prox**:

$$\boldsymbol{v}_i \in \operatorname*{Argmin}_{\boldsymbol{y}_i} \; \phi_i(\boldsymbol{y}_i) + \mathcal{H}_\alpha(U, \boldsymbol{v}_{j<i}, \boldsymbol{y}_i, \boldsymbol{y}_{j>i}^k, Z^k) + \frac{\sigma_k}{2}\|\boldsymbol{y}_i - \boldsymbol{y}_i^k\|^2.$$

Potential Advantages of 'Hi-prox'

The associated subproblems can be reduced to proximal mapping computations:

$$\begin{cases} \boldsymbol{u}_i \in \operatorname*{Argmin}_{\boldsymbol{x}_i} \left\{ \psi_i(\boldsymbol{x}_i) + \frac{\alpha}{2} \|\boldsymbol{x}_i \boldsymbol{y}_i^\top - P_i^k\|_F^2 + \frac{\mu_k}{2} \|\boldsymbol{x}_i - \boldsymbol{x}_i^k\|^2 \right\}, \\ \boldsymbol{v}_i \in \operatorname*{Argmin}_{\boldsymbol{y}_i} \left\{ \phi_i(\boldsymbol{y}_i) + \frac{\alpha}{2} \|\boldsymbol{u}_i \boldsymbol{y}_i^\top - Q_i^k\|_F^2 + \frac{\sigma_k}{2} \|\boldsymbol{y}_i - \boldsymbol{y}_i^k\|^2 \right\}, \end{cases}$$

where

$$P_i^k := Z^k - \sum_{j=1}^{i-1} \boldsymbol{u}_j (\boldsymbol{y}_j^k)^\top - \sum_{j=i+1}^{r} \boldsymbol{x}_j^k (\boldsymbol{y}_j^k)^\top,$$
$$Q_i^k := Z^k - \sum_{j=1}^{i-1} \boldsymbol{u}_j \boldsymbol{v}_j^\top - \sum_{j=i+1}^{r} \boldsymbol{u}_j (\boldsymbol{y}_j^k)^\top.$$

Potential Advantages of 'Hi-prox'

The associated subproblems can be reduced to proximal mapping computations:

$$\begin{cases} \boldsymbol{u}_i \in \operatorname*{Argmin}_{\boldsymbol{x}_i} \left\{ \psi_i(\boldsymbol{x}_i) + \frac{\alpha}{2} \|\boldsymbol{x}_i \boldsymbol{y}_i^\top - P_i^k\|_F^2 + \frac{\mu_k}{2} \|\boldsymbol{x}_i - \boldsymbol{x}_i^k\|^2 \right\}, \\ \boldsymbol{v}_i \in \operatorname*{Argmin}_{\boldsymbol{y}_i} \left\{ \phi_i(\boldsymbol{y}_i) + \frac{\alpha}{2} \|\boldsymbol{u}_i \boldsymbol{y}_i^\top - Q_i^k\|_F^2 + \frac{\sigma_k}{2} \|\boldsymbol{y}_i - \boldsymbol{y}_i^k\|^2 \right\}, \end{cases}$$

where

$$P_i^k := Z^k - \sum_{j=1}^{i-1} \boldsymbol{u}_j (\boldsymbol{y}_j^k)^\top - \sum_{j=i+1}^{r} \boldsymbol{x}_j^k (\boldsymbol{y}_j^k)^\top,$$
$$Q_i^k := Z^k - \sum_{j=1}^{i-1} \boldsymbol{u}_j \boldsymbol{v}_j^\top - \sum_{j=i+1}^{r} \boldsymbol{u}_j (\boldsymbol{y}_j^k)^\top.$$

No need to form P_i^k and Q_i^k explicitly: only need $P_i^k \boldsymbol{y}_i^k$ and $(Q_i^k)^T \boldsymbol{x}_i^k$.

Convergence Analysis

Theorem 1. (Yang, P., Chen '17)
Let $\frac{1}{\alpha} + \frac{1}{\beta} = 1$. Then

1. The line search criterion is satisfied after finitely many inner iterations.

Convergence Analysis

Theorem 1. (Yang, P., Chen '17)
Let $\frac{1}{\alpha} + \frac{1}{\beta} = 1$. Then

1. The line search criterion is satisfied after finitely many inner iterations.

Let $\{X^k\}$, $\{Y^k\}$ and $\{Z^k\}$ be the sequences generated by NAUM. Then

2. The sequence $\{(X^k, Y^k)\}$ is bounded;
3. $\lim\limits_{k\to\infty} \|X^{k+1} - X^k\|_F + \|Y^{k+1} - Y^k\|_F = 0$;
4. Any cluster point (X^*, Y^*) of $\{(X^k, Y^k)\}$ is a stationary point of $\mathcal{F}$, i.e.,
$$0 \in \partial \mathcal{F}(X^*, Y^*).$$

Non-negative Matrix Factorization

- NMF model:

$$\begin{aligned} \min_{X,Y} \quad & \frac{1}{2}\|XY^T - M\|_F^2 \\ \text{s.t.} \quad & 0 \le X, \qquad\qquad 0 \le Y. \end{aligned}$$

Non-negative Matrix Factorization

- NMF model: Add upper bound

$$\begin{aligned} \min_{X,Y} \quad & \frac{1}{2}\|XY^T - M\|_F^2 \\ \text{s.t.} \quad & 0 \le X \le 10^{16}, 0 \le Y \le 10^{16}. \end{aligned}$$

Non-negative Matrix Factorization

- NMF model: Add upper bound

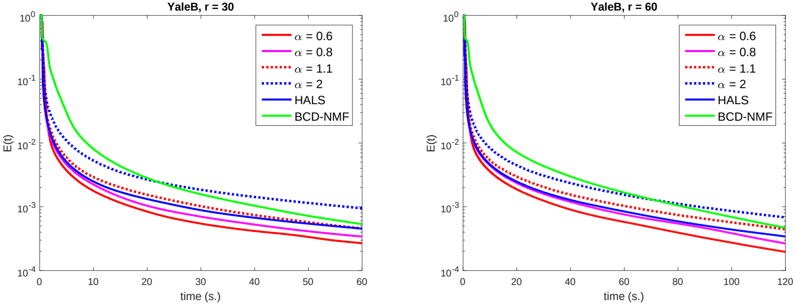
$$\min_{X,Y} \quad \frac{1}{2}\|XY^T - M\|_F^2$$
$$\text{s.t.} \quad 0 \le X \le 10^{16},\ 0 \le Y \le 10^{16}.$$

- NAUM with **Hi-prox**:

$$\begin{cases} Z^k = \dfrac{\alpha}{\alpha+\beta} X^k (Y^k)^\top + \dfrac{\beta}{\alpha+\beta} M, \\ \boldsymbol{u}_i = \max\left\{0,\ \min\left\{10^{16},\ \dfrac{\alpha P_i^k \boldsymbol{y}_i^k + \mu_k \boldsymbol{x}_i^k}{\alpha \|\boldsymbol{y}_i^k\|^2 + \mu_k}\right\}\right\}, \quad i = 1, 2 \cdots, r, \\ \boldsymbol{v}_i = \max\left\{0,\ \min\left\{10^{16},\ \dfrac{\alpha (Q_i^k)^\top \boldsymbol{u}_i + \sigma_k \boldsymbol{y}_i^k}{\alpha \|\boldsymbol{u}_i\|^2 + \sigma_k}\right\}\right\}, \quad i = 1, 2 \cdots, r. \end{cases}$$

Evolution for Face Dataset 'YaleB'

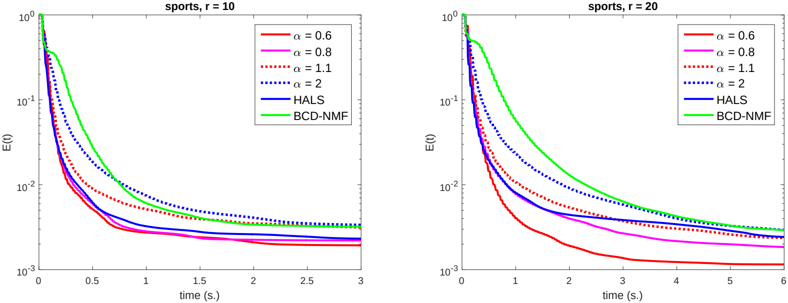
$E(t)$: a normalized measure of function value reduction w.r.t. time.



(a) $T^{max} = 60$ (b) $T^{max} = 120$

Figure: Average $E(t)$ of 20 indep. trials on 'YaleB' (32256×2414, dense).

Evolution for Text Dataset ‘sports’



(a) $T^{max} = 3$ (b) $T^{max} = 6$

Figure: Average $E(t)$ of 20 indep. trials on ‘sports’ (8580×14870, sparse).

Matrix Completion

- An alternative MC model (Shang, Liu, Cheng '16):

$$\min_{X,Y} \ \eta\|X\|_* + \eta\|Y\|_* + \left\|\mathcal{P}_\Omega(XY^T - M)\right\|_F^2.$$

Matrix Completion

- An alternative MC model (Shang, Liu, Cheng '16):

$$\min_{X,Y} \ \eta\|X\|_* + \eta\|Y\|_* + \left\|\mathcal{P}_\Omega(XY^T - M)\right\|_F^2.$$

- NAUM with **Prox-linear**:

$$\begin{cases} Z^k = X^k(Y^k)^\top + \frac{\beta}{\alpha+\beta}\mathcal{P}_\Omega\left(M - X^k(Y^k)^\top\right), \\ U = \mathcal{S}_{\eta/(2\mu_k)}\left(X^k - \mu_k^{-1}\alpha(X^k(Y^k)^\top - Z^k)Y^k\right), \\ V = \mathcal{S}_{\eta/(2\sigma_k)}\left(Y^k - \sigma_k^{-1}\alpha(U(Y^k)^\top - Z^k)^\top U\right). \end{cases}$$

Matrix Completion

- An alternative MC model (Shang, Liu, Cheng '16):

$$\min_{X,Y} \eta\|X\|_* + \eta\|Y\|_* + \left\|\mathcal{P}_\Omega(XY^T - M)\right\|_F^2.$$

- NAUM with **Prox-linear**:

$$\begin{cases} Z^k = X^k(Y^k)^\top + \frac{\beta}{\alpha+\beta}\mathcal{P}_\Omega\left(M - X^k(Y^k)^\top\right), \\ U = \mathcal{S}_{\eta/(2\mu_k)}\left(X^k - \mu_k^{-1}\alpha(X^k(Y^k)^\top - Z^k)Y^k\right), \\ V = \mathcal{S}_{\eta/(2\sigma_k)}\left(Y^k - \sigma_k^{-1}\alpha(U(Y^k)^\top - Z^k)^\top U\right). \end{cases}$$

- Proximal Alternating Linearized Minimization (PALM) (Bolte, et al '14):

$$\begin{cases} X^{k+1} = \mathcal{S}_{\frac{\eta}{2\|Y^k\|^2}}\left(X^k - \|Y^k\|^{-2}\left[\mathcal{P}_\Omega(X^k(Y^k)^\top - M)\right]Y^k\right), \\ Y^{k+1} = \mathcal{S}_{\frac{\eta}{2\|X^{k+1}\|^2}}\left(Y^k - \|X^{k+1}\|^{-2}\left[\mathcal{P}_\Omega(X^{k+1}(Y^k)^\top - M)\right]^\top X^{k+1}\right). \end{cases}$$

Numerical Results for MC on Face Dataset

η	data	sr	r	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 1.1$	PALM	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 1.1$	PALM
				iter				Normalized fval[2]			
5	CBCL	0.5	30	780	1189	3320	3306	1.13e-01	7.50e-02	4.52e-01	1
		0.5	60	921	1218	3850	4654	3.24e-02	5.10e-02	3.85e-01	1
		0.2	30	1174	2366	4767	3573	8.01e-03	2.21e-01	6.87e-01	9.60e-01
		0.2	60	1577	1919	5360	5037	1.03e-02	8.95e-02	8.08e-01	8.86e-01
	ORL	0.5	30	1218	1243	1241	1468	0	2.94e-01	5.06e-01	1
		0.5	60	1049	1051	1051	1327	0	1	4.00e-01	7.73e-01
		0.2	30	2074	325	385	2691	2.59e-03	7.01e-01	1	1.31e-01
		0.2	60	1551	1551	356	2222	0	3.82e-01	1	2.12e-01
				CPU time				RecErr			
5	CBCL	0.5	30	35.56	54.14	151.23	119.05	1.05e-01	1.05e-01	1.06e-01	1.08e-01
		0.5	60	57.66	76.09	240.19	206.47	8.81e-02	9.02e-02	9.04e-02	8.99e-02
		0.2	30	34.04	68.57	137.97	75.56	1.37e-01	1.37e-01	1.38e-01	1.43e-01
		0.2	60	72.01	87.82	245.21	147.08	1.34e-01	1.35e-01	1.35e-01	1.36e-01
	ORL	0.5	30	294.20	300	300	300	1.72e-01	1.84e-01	2.01e-01	2.12e-01
		0.5	60	300	300	300	300	1.66e-01	2.11e-01	2.05e-01	2.11e-01
		0.2	30	300	47.35	55.86	300	2.08e-01	3.04e-01	3.81e-01	2.24e-01
		0.2	60	300	300	69.21	300	2.16e-01	2.35e-01	3.49e-01	2.61e-01

[2] Normalized fval: $(\mathcal{F}(X^*, Y^*) - \mathcal{F}_{\min})/(\mathcal{F}_{\max} - \mathcal{F}_{\min})$

Conclusion

- A potential function is introduced for designing a new algorithm for a class of matrix factorization problems.
- Our approach decouples the linear map from the product XY^T, and hence allows adaptation of efficient techniques in NMF.
- We established convergence of our NAUM, even though the Z-update may not induce a descent in general.

Reference:

- Lei Yang, Ting Kei Pong and Xiaojun Chen.
  A non-monotone alternating updating method for a class of matrix factorization problems
  Available at *https://arxiv.org/abs/1705.06499*.

Thanks for coming! ☺