Alternating direction method of multiplier: a powerful tool for difficult optimization problems

Xin Liu (liuxin@lsec.cc.ac.cn)
State Key Laboratory of Scientific and Engineering Computing
Academy of Mathematics and Systems Science, Chinese Academy of Sciences

Brief Introduction

Philosophy of ADMM

An ancient strategy: **divide** and **conquer**

Mathematical view: **split** and **alternate**

Optimization model description

$$\min_{x\in\Omega} f(x) \ \text{ s.t. } c(x)=0,$$

with a splitting structure:

- $x := (x_1, x_2, ..., x_p)$
- $\{x \mid x\in\Omega\} = \bigcap_{i=1}^{p}\{x \mid x_i \in \Omega_i\}$

* Split variables are connected by equality constraints.

Algorithm Framework

Augmented Lagrangian function (Henstenes 1969, Powell 1969, Rockafellar 1973)

$$\mathcal{L}_\beta(x,\lambda) = f(x) - \lambda^{\mathrm{T}} c(x) + \frac{\beta}{2}\|c(x)\|_2^2.$$

Alternating direction method of multiplier (ADMM) (Glowinski-Marocco 1975, Gabay-Mercier 1976, $p=2$, ...)

$$\begin{cases} x_1^{k+1} \leftarrow \arg\min\limits_{x_1\in\Omega_1} \mathcal{L}_\beta(x_1, x_2^k, ..., x_p^k, \lambda^k); \\ x_2^{k+1} \leftarrow \arg\min\limits_{x_2\in\Omega_2} \mathcal{L}_\beta(x_1^{k+1}, x_2, x_3^k, ..., x_p^k, \lambda^k); \\ \ldots\ldots \\ x_p^{k+1} \leftarrow \arg\min\limits_{x_p\in\Omega_p} \mathcal{L}_\beta(x_1^{k+1}, ..., x_{p-1}^{k+1}, x_p, \lambda^k); \\ \lambda^{k+1} \leftarrow \lambda^k - \tau\,\beta\, c(x_1^{k+1}, ..., x_p^{k+1}). \end{cases}$$

Convergence

Existent results – based on strict conditions

- Two blocks, joint convexity, separability (Gabay-Mercier 1976)
- Multi-blocks, joint convexity, separability
  - variant versions (He-Yuan et al., Goldfarb-Ma, ...)
  - strongly convexity (Luo, 2012)
- Global linear convergence rate
  - linear programming (Eckstein-Bertsekas, 1990)
  - strongly convexity, Lipschitz gradient (Deng-Yin)

Nonconvex and nonseparable case (Yang-L.-Zhang) [1]

- Some pioneering results on the local convergence and linear local convergence rate
- Milder restriction on the optimization model: the second order sufficiency at the solution

References

[1] J. Yang, X. Liu and Y. Zhang.: *A Class of Stationary Iterative Method for Saddle Point Problems: Convergence and Extension*, finished

[2] Z. Wen, C. Yang, X. Liu and S. Marchesini: *Alternating Direction Methods for Classical and Ptychographic Phase Retrieval*, accepted by Inverse Problem

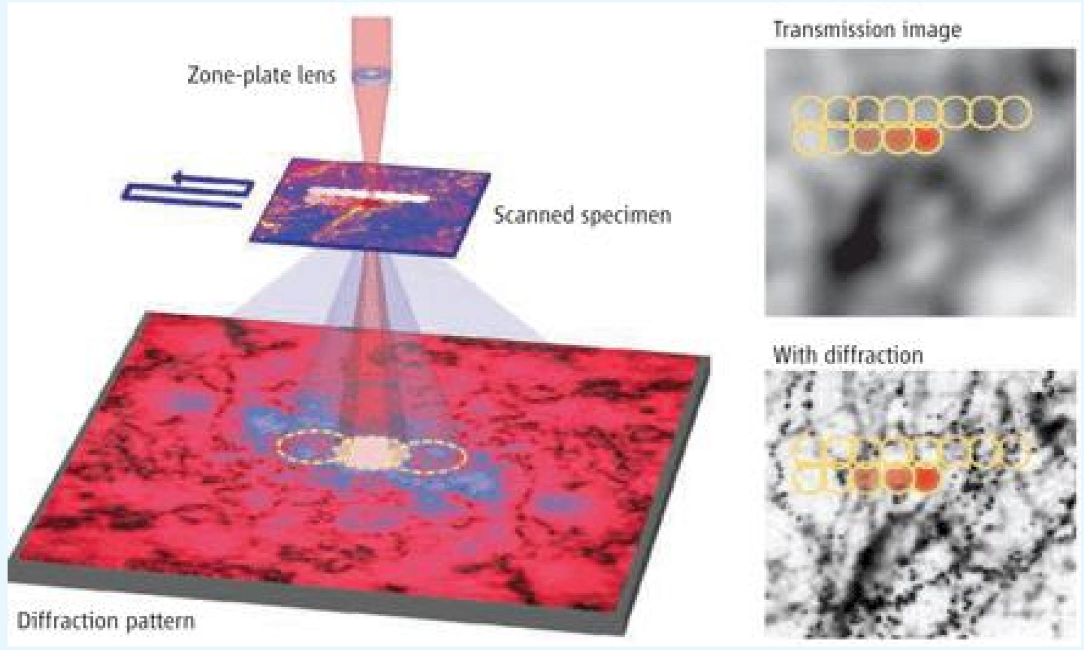
[3] Z. Wen, X. Peng, X. Liu, X. Bai and X. Sun: *Asset Allocation under the Basel Accord Risk Measures*, finished

[4] Y. Zhang: *An Alternating Direction Algorithm for Nonnegative Matrix Factorization* , Rice technical report, 2010.

Acknowledgements

Research supported in part by NSFC grant 11101409 and 10831006, and the National Center for Mathematics and Interdisciplinary Sciences, CAS.

Ptychographic Phase Retrieval



- Details refer to (Wen-Yang-L.-Marchesini, 2012) [2]
- Background: X-ray diffractive imaging, transmission electron microscopy
- Mathematical problem: given $|\mathcal{F}(Q_i\psi)|$ for $i=1,\ldots,k$, can we recover ψ?
- Optimization model: nonconvex, nonsmooth

$$\min_{\hat\psi\in\mathbb{C}^n} \sum_{i=1}^{k}\frac{1}{2}\left\||\mathcal{F}Q_i\hat\psi| - b_i\right\|_2^2.$$

Splitting reformulation: $\min\limits_{\hat\psi\in\mathbb{C}^n, z\in\mathbb{C}^{m\times k}} \sum_{i=1}^k \frac{1}{2}\||z_i| - b_i\|_2^2$ s.t. $z_i = \mathcal{F}Q_i\hat\psi, \quad i=1,...,k.$

Augmented Lagrangian: $\mathcal{L}_\beta(z_i,\psi,y_i) = \sum_{i=1}^k\left(\frac{1}{2}\||z_i| - b_i\|_2^2 + y_i^*(\mathcal{F}Q_i\psi - z_i) + \frac{\beta}{2}\|\mathcal{F}Q_i\psi - z_i\|_2^2\right).$

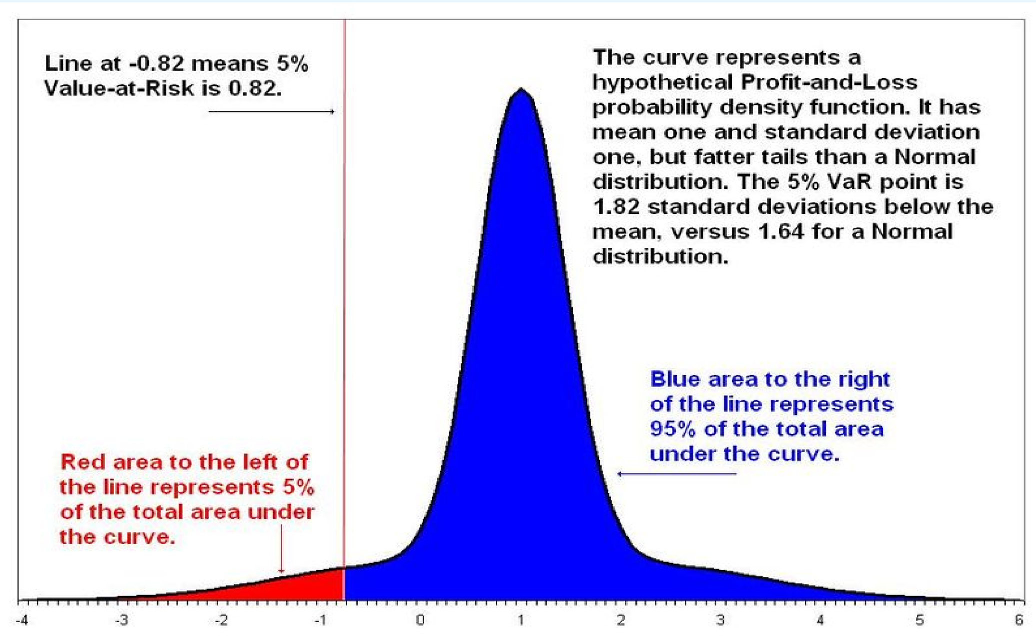
Updating z: $(z_i^+)_{(l)} = \begin{cases} \frac{|(s_i)_{(l)}| + (b_i)_{(l)}}{(1+\beta)|(s_i)_{(l)}|}(s_i)_{(l)}, & \text{if } (s_i)_{(l)} \neq 0 \text{ and } (b_i)_{(l)} > 0; \\ \pm\frac{(b_i)_{(l)}}{1+\beta}, & \text{if } (s_i)_{(l)} = 0 \text{ and } (b_i)_{(l)} > 0; \\ 0, & \text{otherwise}, \end{cases}$

where $s_i = y_i + \beta\mathcal{F}Q_i\psi,\ i=1,...,k.$ (closed-form formula)

Updating ψ: $\psi^+ = \frac{1}{\beta}\left(\sum_{i=1}^k Q_i^*Q_i\right)^{-1}\sum_{i=1}^k Q_i^*\mathcal{F}^*\left(\beta z_i^+ - y_i\right).$ (solving linear system)

Updating Lagrangian multiplier y: $y_i^{j+1} = y_i^j + \tau\,\beta\,(\mathcal{F}Q_i\psi^{j+1} - z_i^{j+1}), \quad i=1,\ldots,k.$

Portfolio Optimization



- Details refer to (Peng-Wen-L.-Bai-Sun) [3]
- Value at risk: $\mathrm{VaR}_\alpha(X) \triangleq -\inf\limits_{x\in\mathbb{R}} x$ s.t. $\mathrm{P}(X > x) \leq 1-\alpha = -\inf\limits_{x\in\mathbb{R}} x$ s.t. $\mathrm{F}_X(x) > \alpha$.
- Optimization model: combinatorial objective

$$\min_{u\in\mathcal{U}_{r_0}} (-\tilde{R}u)_{(p)},$$

where $\mathcal{U}_{r_0} = \{u\in\mathbb{R}^d \mid \mu^{\mathrm{T}}u \geq r_0, \mathbf{1}^{\mathrm{T}}u = 1, u\geq 0\}$; $(\cdot)_{(p)}$ refers to the p-th smallest component of a vector.

Splitting reformulation: $\min_{u\in\mathcal{U}_{r_0},\, x\in\mathbb{R}^n} x_{(p)}$ s.t. $x + \tilde{R}u = 0.$

Augmented Lagrangian: $\mathcal{L}_\beta(x,u,\lambda) := x_{(p)} - \lambda^{\mathrm{T}}(x+\tilde{R}u) + \frac{\beta}{2}\|x+\tilde{R}u\|^2.$

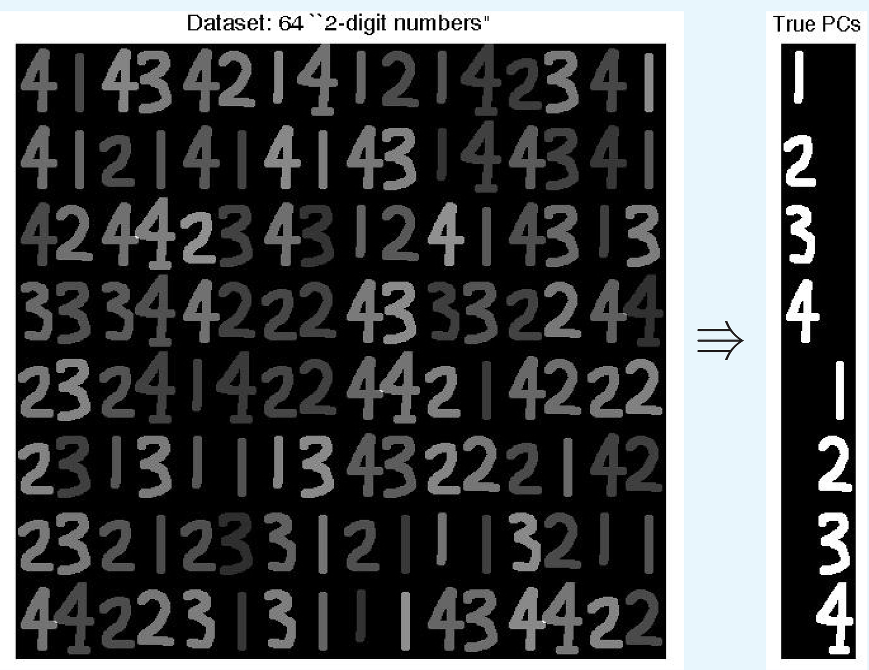
Updating x: $x_{(i)} = \begin{cases} \gamma_{i^*}, & \text{if } i^* \leq i \leq p; \\ v_i, & \text{otherwise}, \end{cases}$ where $v^{(j)} = -\left(\tilde{R}u^{(j)} + \frac{1}{\beta}\lambda^{(j)}\right)$, $\gamma_i = \frac{\beta\sum_{j=i}^p v_j - 1}{\beta(p-i+1)}$

$i^* := \max\{i \mid i\leq p,\ v_{i-1} < \gamma_i\}$ after sorting $v_1 \leq v_2 \leq \ldots \leq v_n$. (closed-form formula)

Updating u: $u^{(j+1)} = \arg\min_{u\in\mathcal{U}_{r_0}} \frac{1}{2}u^\top\tilde{R}^\top\tilde{R}u + b^\top u$, where $b = \tilde{R}^\top(\frac{1}{\beta}\lambda^{(j)} + x^{(j+1)})$ (solving QP)

Updating Lagrangian multiplier λ: $\lambda^{(j+1)} = \lambda^{(j)} + \beta(x^{(j+1)} + \tilde{R}u^{(j+1)}).$

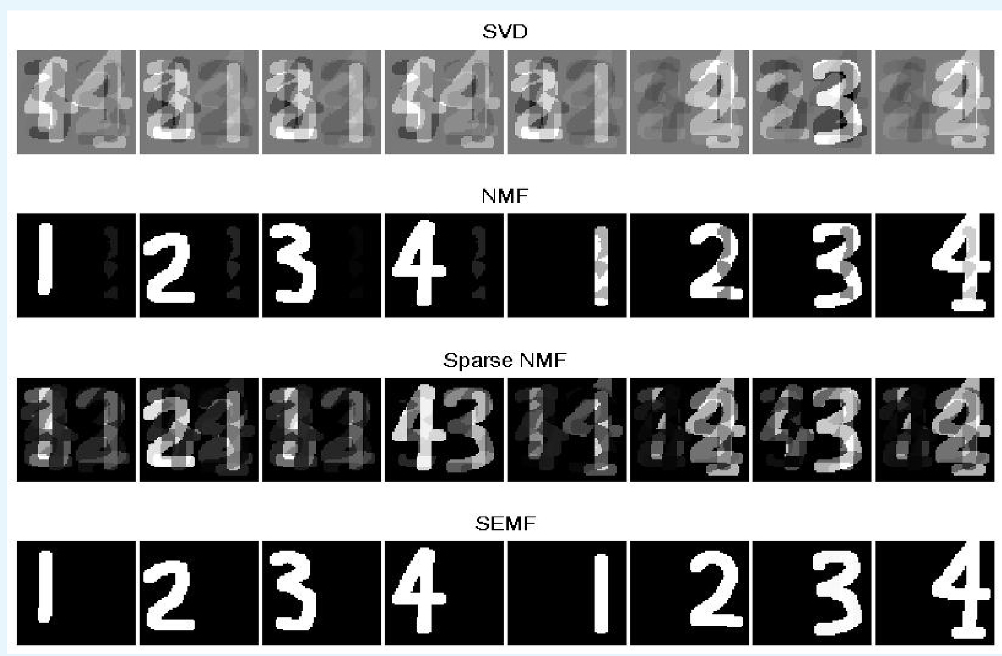
Structure Enforcing Matrix Factorization



- Principal component analysis (PCA) with structures: given dataset $A\in\mathbb{R}^{m\times n}$, find W, H with $k \ll n$ columns so that $A\approx WH^{\mathrm{T}} \Leftrightarrow \mathbf{a}_j \approx \mathbf{w}_1h_{j1} + \mathbf{w}_2h_{j2} + \cdots + \mathbf{w}_kh_{jk}$. with prior information on decomposition pattern (W,H)
- Optimization model: nonconvex, combinatorial constraints

$$\min_{W\in\mathbb{R}^{m\times k},\ H\in\mathbb{R}^{n\times k}} \|A - WH^{\mathrm{T}}\|_{\mathrm{F}}^2 \ \text{ s.t. } W\in\mathbb{T}_1,\ H\in\mathbb{T}_2,$$

where $\mathbb{T}_1$, $\mathbb{T}_2$ can be $\{X \mid X^{\mathrm{T}}X = I\}$, or $\{X \mid X\geq 0\}$ (nonnegative matrix factorization, (Zhang, 2010) [4]), or any other matrix sets allowing 'easy projection'.



- Splitting reformulation: $\min_{W,\ H,\ S_1,\ S_2} \|A - WH^{\mathrm{T}}\|_{\mathrm{F}}^2$ s.t. $W = S_1\in\mathbb{T}_1,\ H = S_2 \in\mathbb{T}_2.$
- Augmented Lagrangian: $\mathcal{L}_{(\beta_1,\ \beta_2)}(W,H,S_1,S_2,\Lambda) = \|A - WH^{\mathrm{T}}\|_{\mathrm{F}}^2 - \Lambda_1\bullet(W - S_1) - \Lambda_2\bullet(H - S_2) + \frac{\beta_1}{2}\cdot\|W - S_1\|_{\mathrm{F}}^2 + \frac{\beta_2}{2}\cdot\|H - S_2\|_{\mathrm{F}}^2.$
- $\begin{cases} W^{k+1} \leftarrow \arg\min_W \mathcal{L}_{(\beta_1,\ \beta_2)}(W, H^k, S_1^k, S_2^k, \Lambda^k); \\ H^{k+1} \leftarrow \arg\min_H \mathcal{L}_{(\beta_1,\ \beta_2)}(W^{k+1}, H, S_1^k, S_2^k, \Lambda^k); \\ S_i^{k+1} \leftarrow \mathrm{Proj}_{\mathbb{T}_i}(V_i^{k+1} - \Lambda_i^k/\beta_i); \quad \text{Here, } i = 1,2; \\ \Lambda_i^{k+1} \leftarrow \Lambda_i^k - \beta_i(V_i^{k+1} - S_i^{k+1}). \quad V_1 = W, V_2 = H. \end{cases}$