图像处理中的一些基本优化问题
和求解的分裂收缩算法

三. 多个可分离函数的分裂收缩算法和按需的算法设计

何 炳 生

南方科技大学数学系　　南京大学数学系

Homepage: maths.nju.edu.cn/~hebma

感谢电子科技大学 张文星博士 提供图像素材

《医学图像处理与分析》暑期学校

西安交大　　2019 年 7 月

1 三个可分离目标函数的凸优化问题

$$
\begin{aligned}
\min\ & \theta_1(x) + \theta_2(y) + \theta_3(z) \\
\text{s.t}\ & Ax + By + Cz = b \\
& x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}
\end{aligned}
\tag{1.1}
$$

Background extraction of surveillance video (II)

The original surveillance video has missing information and additive noise

$$P_\Omega(D) = P_\Omega(X+Y)\text{+noise}$$

P_Ω — indicating missing data, Z — noise/outliers

Model

$$\min\left\{\|X\|_* + \tau\|Y\|_1 + \|P_\Omega(Z)\|_F^2 \;\mid\; X+Y-Z=D\right\}$$

observed video | foreground | background

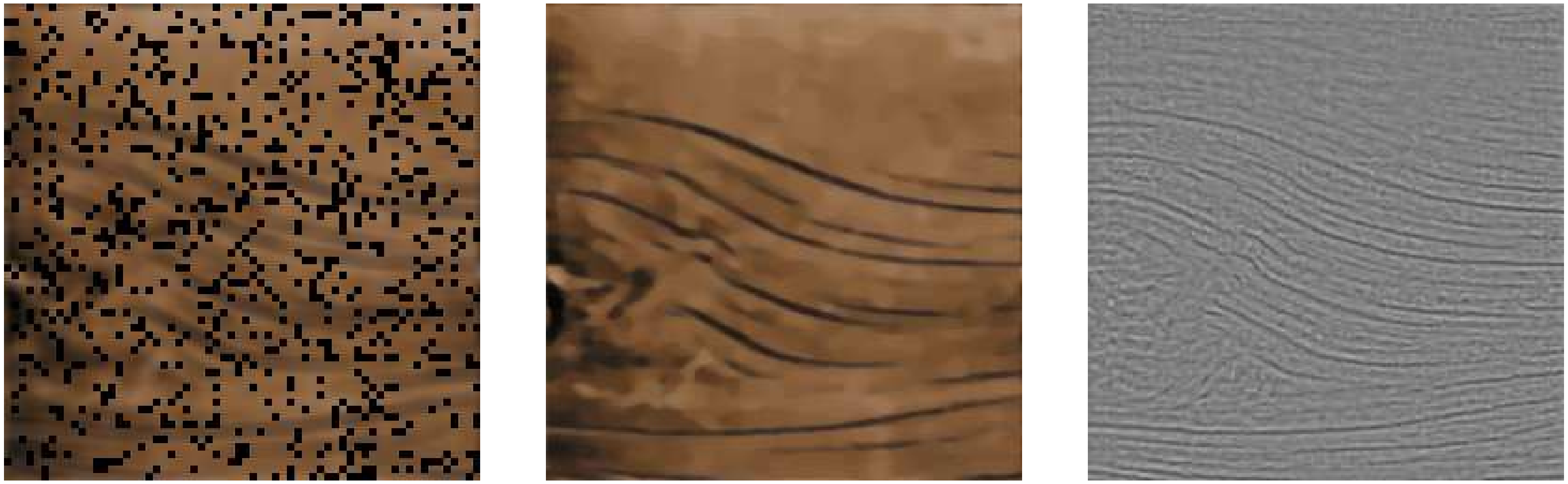
Image decomposition with degradations

The target image for decomposition contains degradations, e.g., blur, missing pixels, $\cdots$

$$\mathbf{f} = K(\mathbf{u} + \text{div}\,\mathbf{v}) + \mathbf{z}, \quad K \text{ — degradation operator}, \quad \mathbf{z} \text{ — noise/outlier}$$

Model

$$\min \left\{ \|\nabla \mathbf{u}\|_1 + \tau \|\mathbf{v}\|_\infty + \|\mathbf{z}\|_2^2 \;\middle|\; K(\mathbf{u} + \text{div}\,\mathbf{v}) + \mathbf{z} = \mathbf{f} \right\}$$



target image cartoon texture

2 Mathematical Background

两大基本概念：变分不等式　和　邻近点 (PPA) 算法

Lemma 1 *Let* $\mathcal{X} \subset \Re^n$ *be a closed convex set,* $\theta(x)$ *and* $f(x)$ *be convex functions and* $f(x)$ *is differentiable. Assume that the solution set of the minimization problem* $\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}$ *is nonempty. Then,*

$$x^* \in \arg\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\} \tag{2.1a}$$

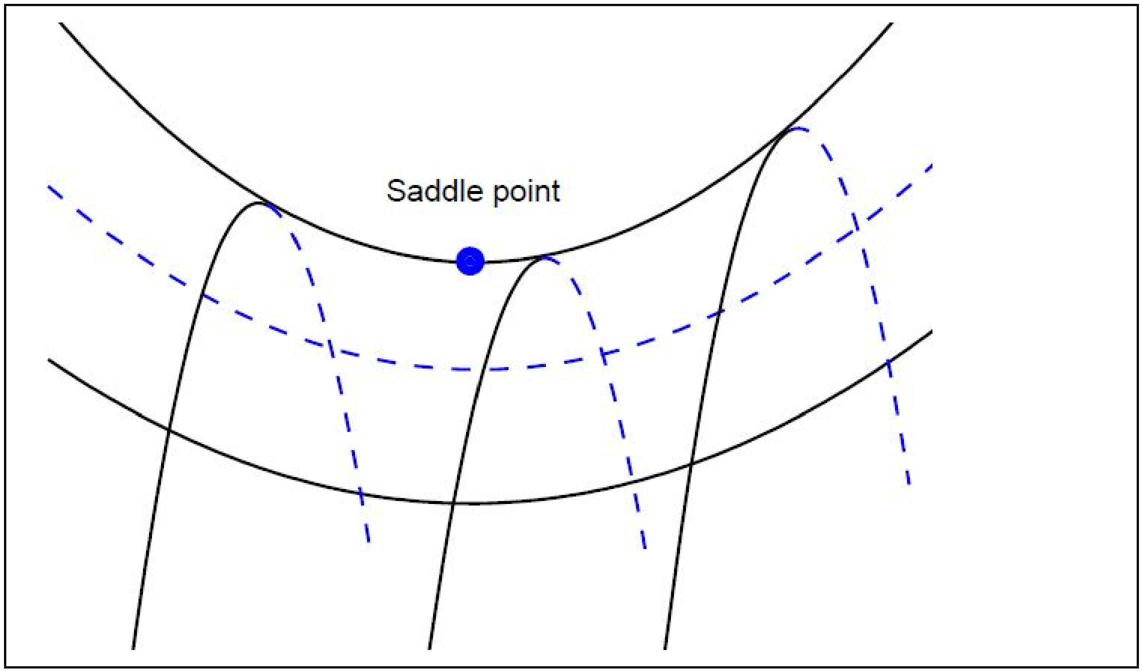
if and only if

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.1b}$$

2.1 Linearly constrained convex optimization and VI

The Lagrangian function of the problem (1.1) is

$$L^3(x, y, z, \lambda) = \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax + By + Cz - b).$$



The saddle point $(x^*, y^*, z^*, \lambda^*) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m$ of $L^3(x, y, z, \lambda)$

satisfies

$$L^3_{\lambda\in\Re^m}\big(x^*,y^*,z^*,\lambda\big) \le L^3\big(x^*,y^*,z^*,\lambda^*\big) \le L^3_{x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}}\big(x,y,z,\lambda^*\big).$$

In other words, for any saddle point (x^*,λ^*), we have

$$\left\{\begin{array}{l} x^* \in \text{argmin}\{L^3(x,y^*,z^*,\lambda^*)|x\in\mathcal{X}\},\\ y^* \in \text{argmin}\{L^2(x^*,y,z^*,\lambda^*)|y\in\mathcal{Y}\},\\ z^* \in \text{argmin}\{L^2(x^*,y^*,z,\lambda^*)|y\in\mathcal{Z}\},\\ \lambda^* \in \text{argmax}\{L(x^*,y^*,z^*,\lambda)|\lambda\in\Re^m\}.\end{array}\right.$$

According to Lemma 1, the saddle point is a solution of the following VI:

$$\left\{\begin{array}{lrl} x^*\in\mathcal{X}, & \theta_1(x)-\theta_1(x^*)+(x-x^*)^T(-A^T\lambda^*)\ge 0, & \forall\, x\in\mathcal{X},\\ y^*\in\mathcal{X}, & \theta_2(y)-\theta_2(y^*)+(y-y^*)^T(-B^T\lambda^*)\ge 0, & \forall\, y\in\mathcal{Y},\\ z^*\in\mathcal{Z}, & \theta_3(z)-\theta_3(z^*)+(z-z^*)^T(-C^T\lambda^*)\ge 0, & \forall\, x\in\mathcal{Z},\\ \lambda^*\in\Re^m, & (\lambda-\lambda^*)^T(Ax^*+By^*+Cz^*-b)\ge 0, & \forall\, \lambda\in\Re^m.\end{array}\right.$$

Its compact form is the following variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{2.2}$$

where

$$w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix},$$

and

$$\theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z), \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m.$$

Note that the operator F is monotone, because

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \geq 0, \text{ Here } (w - \tilde{w})^T (F(w) - F(\tilde{w})) = 0. \tag{2.3}$$

2.2 Preliminaries of PPA for Variational Inequalities

The optimal condition of the problem (1.1) is characterized as a mixed monotone variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{2.4}$$

PPA for monotone mixed VI in H-norm

For given w^k, find the proximal point w^{k+1} in H-norm which satisfies

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T \\ &\{F(w^{k+1}) + H(w^{k+1} - w^k)\} \geq 0, \ \forall\, w \in \Omega, \end{aligned} \tag{2.5}$$

where H is a symmetric positive definite matrix.

Convergence Property of Proximal Point Algorithm in H-norm

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \tag{2.6}$$

2.3 Splitting Methods in a Unified Framework

We study the algorithms using the guidance of variational inequality.

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{2.7}$$

Algorithms in a unified framework

[Prediction Step.] With given v^k, find a vector $\tilde{w}^k \in \Omega$ such that

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k), \ \forall w \in \Omega, \tag{2.8a}$$

where the matrix Q is not necessary symmetric, but $Q^T + Q$ is positive definite.

[Correction Step.] The new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \tag{2.8b}$$

Convergence Conditions

For the matrices Q and M, there is a positive definite matrix H such that

$$HM = Q. \tag{2.9a}$$

Moreover, the matrix

$$G = Q^T + Q - \alpha M^T HM \tag{2.9b}$$

is positive semi-definite.

Convergence using the unified framework

Theorem 1 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (3.1) *and* $\tilde{w}^k$ *is obtained in the* k*-th iteration. If* v^k, v^{k+1} *and* $\tilde{w}^k$ *satisfy the conditions in the unified framework, then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.10}$$

定理 1 的主要结论

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \alpha\|v^k-\tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*.$$

是跟 PPA 类似的收缩不等式, 所以说这类方法是 PPA Like 方法.

关于统一框架下算法及其收敛性证明可以参考下面的文章:

- B.S. He, and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming. Comput. Optim. Appl. 70 (2018), 791 – 826.
- 何炳生, 我和乘子交替方向法 20 年,《运筹学学报》22 卷第1期, pp. 1-31, 2018.

PPA 类算法步步为营, 稳扎稳打; 缺点是思想保守, 影响速度与精度.

3 Two special prediction-correction methods

We study the optimization algorithms using the guidance of variational inequality.

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{3.1}$$

3.1 Algorithms I $Q = H$, H is positive definite

[Prediction Step.] With given v^k, find a vector $\tilde{w}^k \in \Omega$ such that

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{3.2a}$$

where the matrix H is symmetric and positive definite.

[Correction Step.] The new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0, 2) \tag{3.2b}$$

H is a symmetric positive definite matrix. 预测往往对参数有要求

The sequence $\{v^k\}$ generated by the prediction-correction method (3.2) satisfies

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha(2-\alpha)\|v^k - \tilde{v}^k\|_H^2. \quad \forall v^* \in \mathcal{V}^*.$$

The above inequality is the Key for convergence analysis !

上式是跟 (**??**) 类似的不等式, 方法具有 PPA Like 收敛性质.

Set $\alpha = 1$ in (3.2b), the prediction (3.2a) becomes: $w^{k+1} \in \Omega$ such that

$$\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \ge (v - v^{k+1})^T H(v^k - v^{k+1}), \ \forall w \in \Omega.$$

The generated sequence $\{v^k\}$ satisfies

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2. \quad \forall v^* \in \mathcal{V}^*.$$

上式是跟 (**??**) 类似的不等式, 是关于核心变量 v 的 PPA 方法.

3.2 Algorithms II $\quad Q$ is the sum of two matrices

[Prediction Step.] With given v^k, find a vector $\tilde{w}^k \in \Omega$ such that

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k), \ \forall w \in \Omega, \quad \text{(3.3a)}$$

where

$$Q = D + K, \quad \text{(3.3b)}$$

D is a block diagonal positive definite matrix

K is skew-symmetric (反对称) $\quad Q^T + Q = 2D$

[Correction Step.] For the positive matrix D, the new iterate v^{k+1} is given by

$$v^{k+1} = v^k - \gamma \alpha_k^* M(v^k - \tilde{v}^k), \quad \text{(3.4a)}$$

where $\quad M = D^{-1}Q, \quad \gamma \in (0,2)$, and the optimal step size is given by

$$\alpha_k^* = \frac{\|v^k - \tilde{v}^k\|_D^2}{\|M(v^k - \tilde{v}^k)\|_D^2}. \quad \text{(3.4b)}$$

Since $M^T DM = M^T Q$, we have

$$\|M(v^k - \tilde{v}^k)\|_D^2 = \left[M(v^k - \tilde{v}^k)\right]^T \left[Q(v^k - \tilde{v}^k)\right]$$

and thus

$$\alpha_k^* = \frac{\|v^k - \tilde{v}^k\|_D^2}{\left[M(v^k - \tilde{v}^k)\right]^T \left[Q(v^k - \tilde{v}^k)\right]}.$$

步长计算很容易实现

The sequence $\{v^k\}$ generated by the prediction-correction Algorithm II satisfies

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_D^2 - \gamma(2-\gamma)\alpha_k^* \|v^k - \tilde{v}^k\|_D^2. \quad \forall v^* \in \mathcal{V}^*.$$

上式是跟 (**??**) 类似的不等式, 预测-校正方法都具有 PPA Like 收敛性质.

所以, 这个报告中所说的方法, 都是邻近点类 (PPA Like) 算法.

Convergence of the prediction-correction method II

Lemma 2 *For given* v^k*, let the predictor* $\tilde{w}^k$ *be generated by* (3.3a)*, then we have*

$$(v^k - v^*)^T Q(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_D^2, \tag{3.5}$$

where Q *is given in the right hand side of* (3.3a) *and* D *is given in* (3.3b)*.*

Proof. Set $w = w^*$ in (3.3a), we get

$$(\tilde{v}^k - v^*)^T Q(v^k - \tilde{v}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \tag{3.6}$$

Because

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) = (\tilde{w}^k - w^*)^T F(w^*)$$

and

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \geq 0,$$

the right hand side of (3.6) is non-negative. Thus, we have

$$\{(v^k - v^*) - (v^k - \tilde{v}^k)\}^T Q(v^k - \tilde{v}^k) \geq 0$$

and

$$(v^k - v^*)^T Q(v^k - \tilde{v}^k) \geq (v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{3.7}$$

For the right hand side of the above inequality, by using $Q = D + K$ and the skew-symmetry of K, we obtain

$$\begin{aligned}(v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) &= (v^k - \tilde{v}^k)^T (D + K)(v^k - \tilde{v}^k) \\ &= \|v^k - \tilde{v}^k\|_D^2.\end{aligned}$$

The lemma is proved. □

Theorem 2 *For given* v^k*, let the predictor* $\tilde{w}^k$ *be generated by* (3.3a)*. If the new iterate* v^{k+1} *is given by*

$$v^{k+1}(\alpha) = v^k - \alpha M(v^k - \tilde{v}^k), \quad \gamma \in (0, 2), \tag{3.8}$$

then we have

$$\|v^{k+1} - v^*\|_D^2 \leq \|v^k - v^*\|_D^2 - q_k^{II}(\alpha), \quad \forall v^* \in \mathcal{V}^*, \tag{3.9}$$

where

$$q_k^{II}(\alpha) = 2\alpha\|w^k - \tilde{w}^k\|_D^2 - \alpha^2\|M(w^k - \tilde{w}^k)\|_D^2. \quad (3.10)$$

Proof. First, we define the profit function by

$$\vartheta_k^{II}(\alpha) = \|v^k - v^*\|_D^2 - \|v^{k+1}(\alpha) - v^*\|_D^2. \quad (3.11)$$

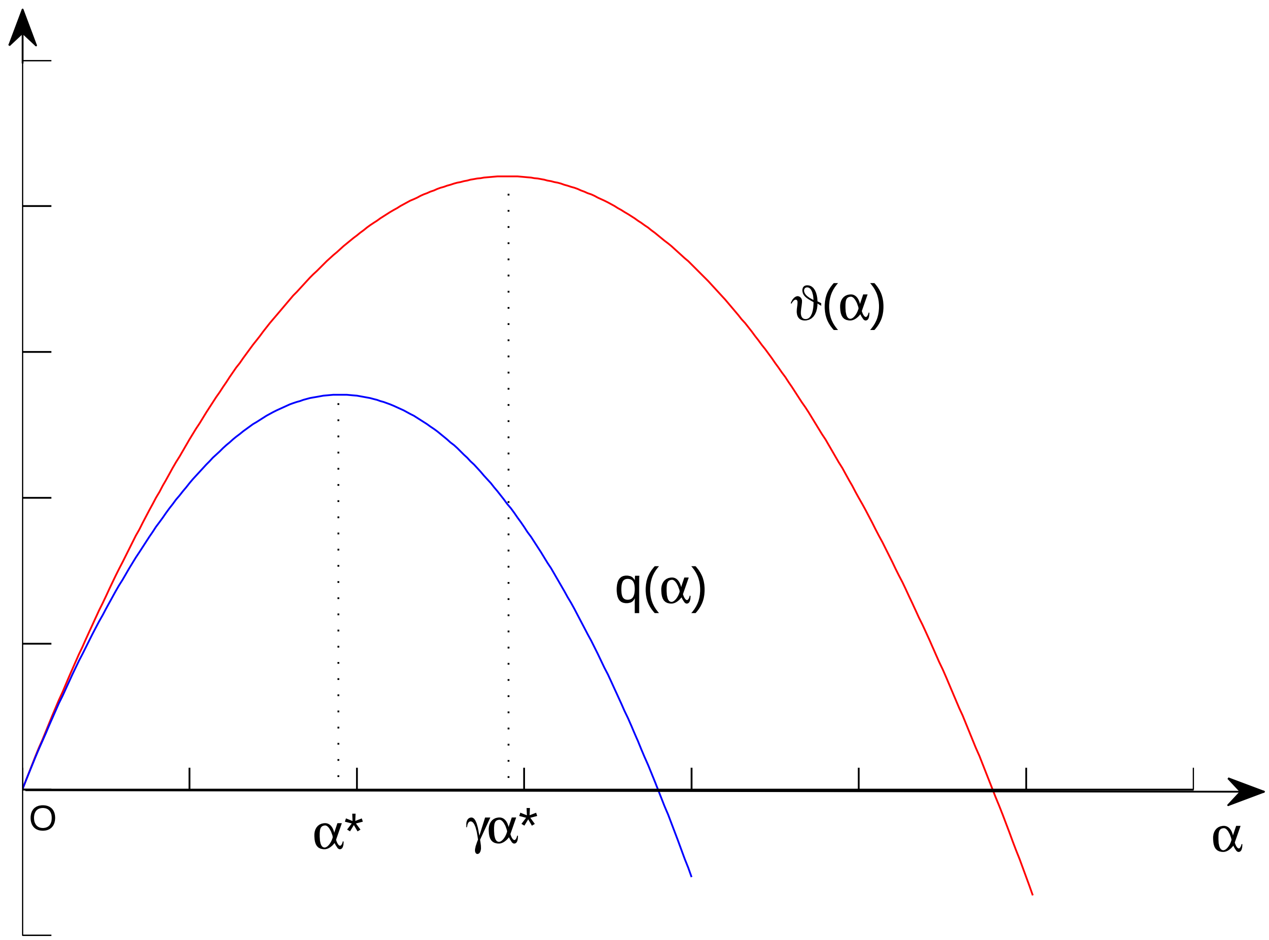
Thus, it follows from (3.8) that

$$\begin{aligned}\vartheta_k^{II}(\alpha) &= \|v^k - v^*\|_D^2 - \|(v^k - v^*) - \alpha M(v^k - \tilde{v}^k)\|_D^2 \\ &= 2\alpha(v^k - v^*)^T DM(v^k - \tilde{v}^k) - \alpha^2\|M(v^k - \tilde{v}^k)\|_D^2.\end{aligned}$$

By using $DM = Q$ and (3.5), we get

$$\vartheta_k^{II}(\alpha) \geq 2\alpha\|v^k - \tilde{v}^k\|_D^2 - \alpha^2\|M(v^k - \tilde{v}^k)\|_D^2 = q_k^{II}(\alpha). \quad \square$$

$q_k^{II}(\alpha)$ reaches its maximum at α_k^* which is given by $(3.4b)$.



取 $\gamma \in [1,2)$ 的示意图

Since we take $\alpha = \gamma\alpha_k^*$, it follows from (3.10) that

$$q_k^{II}(\alpha) = 2\gamma\alpha_k^*\|v^k - \tilde{v}^k\|_D^2 - \gamma^2(\alpha_k^*)^2\|M(v^k - \tilde{v}^k)\|_D^2. \tag{3.12}$$

By using (3.4b), we get

$$\begin{aligned} &(\alpha_k^*)^2\|M(v^k - \tilde{v}^k)\|_D^2 \\ &= \quad \alpha_k^* \frac{\|v^k - \tilde{v}^k\|_D^2}{\|M(v^k - \tilde{v}^k)\|_D^2}\|M(v^k - \tilde{v}^k)\|_D^2 \\ &= \quad \alpha_k^*\|v^k - \tilde{v}^k\|_D^2. \end{aligned}$$

Substituting it in (3.12) we get $\quad q_k^{II}(\alpha) \geq \gamma(2-\gamma)\alpha_k^*\|v^k - \tilde{v}^k\|_D^2.$

$$\|v^{k+1} - v^*\|_D^2 \leq \|v^k - v^*\|_D^2 - \gamma(2-\gamma)\alpha_k^*\|v^k - \tilde{v}^k\|_D^2. \quad \forall v^* \in \mathcal{V}^*.$$

4 Applications for separable problems

This section presents various applications of the proposed algorithms for the separable convex optimization problem

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b, \, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{4.1}$$

Its VI-form is

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{4.2}$$

where

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \tag{4.3a}$$

and

$$\theta(u) = \theta_1(x) + \theta_2(y), \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{4.3b}$$

The augmented Lagrangian Function of the problem (4.1) is

$$\mathcal{L}_\beta(x,y,\lambda) = \theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)+\frac{\beta}{2}\|Ax+By-b\|^2. \quad (4.4)$$

Solving the problem (4.1) by using ADMM, the k-th iteration begins with given (y^k, λ^k), it offers the new iterate (y^{k+1}, λ^{k+1}) via

$$\text{(ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(4.5a)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, & \text{(4.5b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(4.5c)} \end{array} \right.$$

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad v = \begin{pmatrix} y \\ \lambda \end{pmatrix} \quad \text{and} \quad \mathcal{V}^* = \{(y^*, \lambda^*) \mid (x^*, y^*, \lambda^*) \in \Omega^*\}.$$

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \|v^k-v^{k+1}\|_H^2, \quad H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}$$

根据算法 I 的要求 设计预测公式.

4.1 ADMM in PPA-sense

In order to solve the separable convex optimization problem (4.1), we construct a method whose prediction-step is

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{4.6a}$$

where

$$H = \begin{pmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0\text{, say } \delta = 0.05\text{)}. \tag{4.6b}$$

Since H is positive definite, we can use the update form of Algorithm I to produce the new iterate $v^{k+1} = (y^{k+1}, \lambda^{k+1})$. (In the algorithm [2], we took $\delta = 0$).

The concrete form of (4.6) is

$$\left\{\begin{array}{l} \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad\qquad \{\underline{-A^T\tilde{\lambda}^k}\}\ge 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T \\ \qquad\qquad \{\underline{-B^T\tilde{\lambda}^k}+\boldsymbol{(1+\delta)\beta B^T B}(\tilde{y}^k-y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k-\lambda^k)\}\ge 0, \\ (\underline{A\tilde{x}^k+B\tilde{y}^k-b}) \qquad \boldsymbol{-B}(\tilde{y}^k-y^k) \quad + \quad \boldsymbol{(1/\beta)}\,(\tilde{\lambda}^k-\lambda^k)=0. \end{array}\right.$$

The underline part is $F(\tilde{w}^k)$:

$$\boldsymbol{F(w)}=\begin{pmatrix} \boldsymbol{-A^T\lambda} \\ \boldsymbol{-B^T\lambda} \\ \boldsymbol{Ax+By-b} \end{pmatrix}$$

In fact, the prediction can be arranged by

$$\left\{\begin{array}{ll} \tilde{x}^k=\text{Argmin}\{\mathcal{L}_\beta(x,y^k,\lambda^k)\,|\,x\in\mathcal{X}\}, & (4.7a) \\ \tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b), & (4.7b) \\ \tilde{y}^k=\text{Argmin}\left\{\begin{array}{c}\theta_2(y)-y^TB^T\boldsymbol{[2\tilde{\lambda}^k-\lambda^k]} \\ +\frac{1+\delta}{2}\beta\|B(y-y^k)\|^2\end{array}\,\middle|\,y\in\mathcal{Y}\right\}. & (4.7c) \end{array}\right.$$

这个预测与经典的交替方向法 (6.10) 相当, 采用(3.2b) 校正, 会加快速度.

4.2 Linearized ADMM-Like Method

当子问题 (4.7c) 求解有困难时, 用 $\frac{s}{2}\|y - y^k\|^2$ 代替 $\frac{1+\delta}{2}\beta\|B(y - y^k)\|^2$.

By using the linearized version of (4.7), the prediction step becomes

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T H(v^k-\tilde{v}^k), \ \forall w \in \Omega, \quad (4.8)$$

where

$$H = \begin{bmatrix} sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}, \text{ 代替 (4.6) 中的 } \begin{bmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}. \quad (4.9)$$

The concrete formula of (4.8) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + \ \boldsymbol{s}(\tilde{y}^k - y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b})\boldsymbol{-B}(\tilde{y}^k - y^k) + \boldsymbol{(1/\beta)}(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases} \quad (4.10)$$

The underline part is $F(\tilde{w}^k)$:

$$\boldsymbol{F(w)} = \begin{pmatrix} -\boldsymbol{A^T\lambda} \\ -\boldsymbol{B^T\lambda} \\ \boldsymbol{Ax + By - b} \end{pmatrix}$$

Then, we use the form

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0,2)$$

to update the new iterate v^{k+1}.

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (4.10),

we need only use the following procedure:

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\theta_2(y) - y^T B^T [\mathbf{2\tilde{\lambda}^k - \lambda^k}] + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \end{cases}$$

用 $\frac{s}{2}\|y - y^k\|^2$ 代替 $\frac{1+\delta}{2}\beta\|B(y - y^k)\|^2$, 为保证收敛, 需要 $s > \beta\|B^T B\|$.

对给定的 $\beta > 0$，要求 $s > \beta\|B^T B\|$，太大的 s 会影响收敛速度

4.3 Method without $s > \beta\|B^TB\|$

当矩阵 B^TB 的条件不好, 又必须线性化, 就采取以下的方法

For solving the same problem, we give the following prediction:

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{4.11a}$$

where

$$Q = \begin{pmatrix} sI & B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} = D + K. \tag{4.11b}$$

Because

$$D = \begin{pmatrix} sI & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \quad \text{and} \quad K = \begin{pmatrix} 0 & B^T \\ -B & 0 \end{pmatrix},$$

根据这样的预测, 可以用算法 II 的校正公式 (3.4) 产生新的迭代点.

How to implement the prediction?

The concrete formula of (4.11) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{s}(\tilde{y}^k - y^k) \ \boldsymbol{+B^T} \ (\tilde{\lambda}^k - \lambda^k)\} \ge 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b})\boldsymbol{-B}(\tilde{y}^k - y^k)\boldsymbol{+(1/\beta)}(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

The underline part is $F(\tilde{w}^k)$:

$$\boldsymbol{F(w)} = \begin{pmatrix} \boldsymbol{-A^T\lambda} \\ \boldsymbol{-B^T\lambda} \\ \boldsymbol{Ax + By - b} \end{pmatrix}$$

This can be implemented by

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\theta_2(y) - y^T B^T \boldsymbol{\lambda^k} + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \end{cases}$$

The y-subproblem is easy. 对给定的 $\beta > 0$, 可以取任意的 $s > 0$.

对可分离目标函数的优化问题, 我们在 §4 中提出三种预测-校正方法

- 如果子问题中求解过程中, 二次项不带来任何困难的时候, 建议采用 §4.1 中的方法.
- 如果子问题中求解中, 必须对一个子问题中的二次项线性化, 并且矩阵条件好的时候, 建议采用 §4.2 中的方法.
- 如果必须线性化, 矩阵条件又不好的时候, 建议分别采用 §**??**.3 和 §4.3 中的方法.

希望这些框架能为针对实际问题设计算法提供帮助.

5 求解三个可分离目标函数的凸优化问题

这个问题的 Lagrange 函数是

$$L(x, y, z, \lambda) = \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax + By + Cz - b).$$

增广 Lagrange 函数是

$$\mathcal{L}_\beta^3(x, y, z, \lambda) \quad = \quad L(x, y, z, \lambda) + \frac{\beta}{2}\|Ax + By + Cz - b\|^2.$$

直接推广的交替方向法

$$\left\{\begin{array}{lcl} x^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \,\middle|\, x \in \mathcal{X}\right\}, \\ y^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) \,\middle|\, y \in \mathcal{Y}\right\}, \\ z^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y^{k+1}, z, \lambda^k) \,\middle|\, z \in \mathcal{Z}\right\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{array}\right. \tag{5.12}$$

对 $m \geq 3$, 一般形式的问题直接推广的交替方向法不能保证收敛 [4].

♣ 感谢堵丁柱教授注意到我们的有关工作.

直接推广 ADMM： 我们发表在 2016 Math.Progr. 的三个算子问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)\,|\,Ax+By+Cz=b,\ x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}\}$$

的第一个例子中, $\theta_1(x)=\theta_2(y)=\theta_3(z)=0,\mathcal{X}=\mathcal{Y}=\mathcal{Z}=\Re,$

$\mathcal{A}=[A,B,C]\in\Re^{3\times 3}$ 是个非奇异矩阵, $b=0\in\Re^3.$

还有一些据此延伸的例子, 证明了直接推广的 ADMM 并不收敛.
这些例子更多的是在理论方面的意义.

值得继续研究的问题：三个算子的实际问题中, 线性约束矩阵

$\mathcal{A}=[A,B,C]$ 往往至少有一个是单位矩阵, 即, $\mathcal{A}=[A,B,I].$

直接推广的 ADMM 处理这种更贴近实际的三个算子的问题,
既没有证明收敛，也没有举出反例，至今我们于心不甘！！

举个简单的例子来说：

- 乘子交替方向法 (ADMM) 处理问题

$$\min\{\theta_1(x)+\theta_2(y)|Ax+By=b,\ x\in\mathcal{X},y\in\mathcal{Y}\}$$ 是收敛的.

- 将等式约束换成不等式约束, 问题就变成

$$\min\{\theta_1(x)+\theta_2(y)|Ax+By\le b,\ x\in\mathcal{X},y\in\mathcal{Y}\}.$$

- 再化成三个算子的等式约束问题

$$\min\{\theta_1(x)+\theta_2(y)+0\,|\,Ax+By+z=b,\ x\in\mathcal{X},y\in\mathcal{Y},z\ge 0\}$$

- 直接推广的 ADMM 处理上面这种问题, 不少人做过尝试, 但是至今既没有证明收敛性, 也没有举出反例！

基于上述认知, 我们对三个算子的问题提出了一些修正算法. 注意：
我们的方法对问题不加任何条件！ 对 β 不加限制, 只对方法动手术！

5.1 带高斯回代的 ADMM 方法

以 (5.12) 提供的 (y^{k+1}, z^{k+1}) 为预测, 取 $\alpha \in (0,1)$, 校正公式为

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} := \begin{pmatrix} y^{k} \\ z^{k} \end{pmatrix} - \alpha \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}. \quad (5.13)$$

由于为下一步迭代只要准备 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$, 我们只要做

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} := \begin{pmatrix} By^{k} \\ Cz^{k} \end{pmatrix} - \alpha \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} B(y^k - y^{k+1}) \\ C(z^k - z^{k+1}) \end{pmatrix}.$$

- B. S. He, M. Tao and X.M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, *SIAM Journal on Optimization* **22**(2012), 313-340.

对 y 和 z, 有先后, 不公平, 那就要做找补, 调整

5.2 ADMM + Prox-Parallel Splitting ALM

[简单地强制 y 和 z 平等不能保证方法收敛]

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases}$$

y, z 子问题平行, 如果不想做后处理, 就给它们俩预先都加个正则项

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \qquad (\tau > 1) \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\beta\|B(y - y^k)\|^2 \big| y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\beta\|C(z - z^k)\|^2 \big| z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases}$$

上述做法相当于：

$$\begin{cases} x^{k+1} = \text{Argmin}\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b - \frac{1}{\beta}\lambda^k\|^2 \,|\, x \in \mathcal{X}\}, \\ \qquad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\ y^{k+1} = \text{Argmin}\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, \\ z^{k+1} = \text{Argmin}\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{cases} \tag{5.14}$$

其中 $\mu > 2$. 例如, 可以取 $\mu = 2.01$.

- B. He, M. Tao and X. Yuan, A splitting method for separable convex programming. IMA J. Numerical Analysis, 31(2015), 394-426.

太自由, 又不校正, 就加正则项, 不忘自己昨天的承诺.

This method is accepted by Osher's research group

- E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 21, NO. 7, JULY 2012 3239

A Convex Model for Nonnegative Matrix Factorization and Dimensionality Reduction on Physical Space

Ernie Esser, Michael Möller, Stanley Osher, Guillermo Sapiro, *Senior Member, IEEE*, and Jack Xin

$$\min_{T\geq 0, V_j \in D_j, e \in E} \zeta \sum_i \max_j (T_{i,j}) + \langle R_w \sigma C_w, T \rangle$$

$$\text{such that} \quad YT - X_s = V - X_s \text{diag}(e). \quad (15)$$

Since the convex functional for the extended model (15) is slightly more complicated, it is convenient to use a variant of ADMM that allows the functional to be split into more than two parts. The method proposed by He *et al.* in [34] is appropriate for this application. Again, introduce a new variable Z

Using the ADMM-like method in [34], a saddle point of the augmented Lagrangian can be found by iteratively solving the subproblems with parameters $\delta > 0$ and $\mu > 2$, shown in the tion refinement step. Due to the different algorithm used to solve the extended model, there is an additional numerical parameter μ, which for this application must be greater than two according to [34]. We set μ equal to 2.01. There are also model parame-

[33] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," 2009 [Online]. Available: http://arxiv.org/PS cache/arxiv/pdf/0912/0912.3599v1.pdf

[34] B. He, M. Tao, and X. Yuan, "A splitting method for separate convex programming with linking linear constraints," Tech. Rep., 2011 [Online]. Available: http://www.optimization-online.org/DB_FILE/2010/06/ 2665.pdf

6 最优参数

Some linearly constrained convex optimization problems

1. Linearly constrained convex optimization $\min\{\theta(x) | Ax = b, \; x \in \mathcal{X}\}$

2. Convex optimization problem with separable objective function

$$\min\{\theta_1(x) + \theta_2(y) | Ax + By = b, \; x \in \mathcal{X}, \; y \in \mathcal{Y}\}$$

3. Convex optimization problem with 3 separable objective functions

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) | Ax+By+Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}$$

There are some crucial parameters：

- Crucial parameter in the so called linearized ALM for the first problem,
- Crucial parameter in the so called linearized ADMM for the second problem,
- Crucial proximal parameter in the Proximal Parallel ADMM-like Method for the convex optimization problem with 3 separable objective functions.

6.1 Linearized Augmented Lagrangian Method

Consider the following convex optimization problem:

$$\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}. \tag{6.1}$$

The augmented Lagrangian function of the problem (6.1) is

$$\mathcal{L}_\beta(x,\lambda) = \theta(x) - \lambda^T(Ax-b) + \frac{\beta}{2}\|Ax-b\|^2.$$

Starting with a given λ^k, the k-th iteration of the Augmented Lagrangian Method [15, 19] produces the new iterate $w^{k+1} = (x^{k+1}, \lambda^{k+1})$ via

$$\text{(ALM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x,\lambda^k) \;\big|\; x \in \mathcal{X}\big\}, & \text{(6.2a)} \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} - b), \quad \gamma \in (0,2) & \text{(6.2b)} \end{array} \right.$$

In the classical ALM, the optimization subproblem (6.2a) is

$$\min\{\theta(x) + \frac{\beta}{2}\|Ax - (b + \frac{1}{\beta}\lambda^k)\|^2 | x \in \mathcal{X}\}.$$

Sometimes, because of the structure of the matrix A, we should simplify the subproblem (6.2a). Notice that

- Ignore the constant term in the objective function of $\mathcal{L}_\beta(x,\lambda^k)$, we have

$$\begin{array}{rl} & \arg\min\{\mathcal{L}_\beta(x,\lambda^k) \mid x\in\mathcal{X}\} \\ = & \arg\min\{\theta(x)-(\lambda^k)^T(Ax-b)+\frac{\beta}{2}\|Ax-b\|^2 \mid x\in\mathcal{X}\} \\ = & \arg\min\left\{ \begin{array}{c} \theta(x)-(\lambda^k)^T(Ax-b)+ \\ \frac{\beta}{2}\|(Ax^k-b)+A(x-x^k)\|^2 \end{array} \middle| x\in\mathcal{X}\right\} \\ = & \arg\min\left\{ \begin{array}{c} \theta(x)-x^TA^T[\lambda^k-\beta(Ax^k-b)] \\ +\frac{\beta}{2}\|A(x-x^k)\|^2 \end{array} \middle| x\in\mathcal{X}\right\}. \end{array} \tag{6.3}$$

- In the so called **Linearized ALM**, the term $\frac{\beta}{2}\|A(x-x^k)\|^2$ is replaced with $\frac{r}{2}\|x-x^k\|^2$. In this way, the x-subproblem becomes

$$x^{k+1}=\arg\min\{\theta(x)-x^TA^T[\lambda^k-\beta(Ax^k-b)]+\frac{r}{2}\|x-x^k\|^2 \mid x\in\mathcal{X}\}. \tag{6.4}$$

In fact, the linearized ALM simplifies the quadratic term $\frac{\beta}{2}\|A(x-x^k)\|^2$.

In comparison with (6.3), the simplified x-subproblem (6.4) is equivalent to

$$x^{k+1} = \arg\min\{\mathcal{L}_\beta(x,\lambda^k) + \tfrac{1}{2}\|x - x^k\|_{D_A}^2 \mid x \in \mathcal{X}\}, \tag{6.5}$$

where

$$D_A = rI - \beta A^T A. \tag{6.6}$$

In order to ensure the convergence, it **was** required that $\boldsymbol{r > \beta\|A^T A\|}$.

Thus, the mathematical form of the Linearized ALM can be written as

$$\left\{\begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x,\lambda^k) + \tfrac{1}{2}\|x - x^k\|_{D_A}^2 \;\big|\; x \in \mathcal{X}\}, & \text{(6.7a)} \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} - b), \qquad \gamma \in (0,2). & \text{(6.7b)} \end{array}\right.$$

where D_A is defined by (6.6).

Large parameter r in (6.6) **will lead a slow convergence !**

Recent Advance. Bingsheng He, Feng Ma, Xiaoming Yuan:
Optimal proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems, IMA Journal of Numerical Analysis. 39(2019).

Our new result in the above paper:

For the matrix D_A in (6.7a) with the form (6.6)

- if $r > \frac{2+\gamma}{4}\beta\|A^TA\|$ is used in the method (6.7), it is still convergent;
- if $r < \frac{2+\gamma}{4}\beta\|A^TA\|$ is used in the method (6.7), there is divergent example.

Especially, when $\gamma = 1$,

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x,\lambda^k) + \frac{1}{2}\|x - x^k\|_{D_A}^2 \mid x \in \mathcal{X}\}, & \text{(6.8a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b). & \text{(6.8b)} \end{cases}$$

According to our new result: For the matrix D_A in in (6.7a) with the form (6.6),

- if $r > \frac{3}{4}\beta\|A^TA\|$ is taken in the method (6.8), it is still convergent;
- if $r < \frac{3}{4}\beta\|A^TA\|$ is taken in the method (6.8), there is divergent example.

$r = 0.75$ is the threshold factor in the matrix D_A for linearized ALM (6.8) !

6.2 Linearized ADMM

Consider the convex optimization problem with separable objective function:

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{6.9}$$

The augmented Lagrangian function of the problem (6.9) is

$$\mathcal{L}^2_\beta(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2.$$

Starting with a given (y^k, λ^k), the k-th iteration of the classical ADMM [**?**, 7] generates the new iterate $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ via

$$\text{(ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x, y^k, \lambda^k) \ \big|\ x \in \mathcal{X}\big\}, & \text{(6.10a)} \\ y^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \ \big|\ y \in \mathcal{Y}\big\}, & \text{(6.10b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(6.10c)} \end{array} \right.$$

In (6.10a) and (6.10a), the optimization subproblems are

$$\min\{\theta_1(x) + \tfrac{\beta}{2}\|Ax - p^k\|^2 | x \in \mathcal{X}\} \quad \text{and} \quad \min\{\theta_2(y) + \tfrac{\beta}{2}\|By - q^k\|^2 | y \in \mathcal{Y}\},$$

respectively. We assume that one of the minimization subproblems (without loss of the generality, say, (6.10b)) should be simplified. Notice that

- Using the notation $\mathcal{L}_\beta(x^{k+1}, y, \lambda^k)$ and ignoring the constant term in the objective function, we have

$$
\begin{aligned}
&\arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) \\ +\frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T By + \\ \frac{\beta}{2}\|(Ax^{k+1} + By^k - b) + B(y - y^k)\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T[\lambda^k - \beta(Ax^{k+1} + By^k - b)] \\ +\frac{\beta}{2}\|B(y - y^k)\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\}
\end{aligned} \tag{6.11}
$$

- In the so called **Linearized ADMM**, the term $\frac{\beta}{2}\|B(y - y^k)\|^2$ is replaced with $\frac{s}{2}\|y - y^k\|^2$. Thus, the y-subproblem becomes

$$y^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T[\lambda^k - \beta(Ax^{k+1} + By^k - b)] \\ + \frac{s}{2}\|y - y^k\|^2 \end{array} \,\middle|\, y \in \mathcal{Y} \right\}. \tag{6.12}$$

In fact, the linearized ADMM simplifies the quadratic term $\frac{\beta}{2}\|B(y - y^k)\|^2$.

In comparison with (6.11), the simplified y-subproblem (6.12) is equivalent to

$$y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \tfrac{1}{2}\|y - y^k\|_{D_B}^2 \mid y \in \mathcal{Y}\}, \tag{6.13}$$

where

$$D_B = sI - \beta B^T B. \tag{6.14}$$

In order to ensure the convergence, it **was** required that $\boldsymbol{s > \beta\|B^T B\|}$.

Thus, the mathematical form of the Linearized ADMM can be written as

$$\left\{ \begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (6.15a) \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \tfrac{1}{2}\|y - y^k\|_{D_B}^2 \mid y \in \mathcal{Y}\}, & (6.15b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & (6.15c) \end{array} \right.$$

where D_B is defined by (6.14).

A large parameter s will lead a slow convergence of the linearized ADMM.

最新进展：最优线性化因子的选择– OO6228 的结论

Recent Advance. Bingsheng He, Feng Ma, Xiaoming Yuan:
Optimal Linearized Alternating Direction Method of Multipliers for Convex Programming. http://www.optimization-online.org/DB_HTML/2017/09/6228.html

Our new result in the above paper: For the matrix D_B in (6.15b) with the form (6.14)

- if $s > \frac{3}{4}\beta\|B^TB\|$ is taken in the method (6.15), it is still convergent;
- if $s < \frac{3}{4}\beta\|B^TB\|$ is taken in the method (6.15), there is divergent example.

$s = 0.75$ is the threshold factor in the matrix D_B for linearized ADMM (6.15) !

Notice that the matrix D_B defined in (6.14) is indefinite for $s \in (0.75, 1)$!

6.3 Parameters improvements in the method for problem with 3 separable objective functions

For the problem with three separable objective functions

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\}, \quad (6.16)$$

the augmented Lagrangian function is

$$\begin{aligned}\mathcal{L}^3_\beta(x,y,z,\lambda) \quad = \quad & \theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b) \\ & +\frac{\beta}{2}\|Ax+By+Cz-b\|^2.\end{aligned}$$

Using the **direct extension of ADMM** to solve the problem (6.16), the formula is

$$\begin{cases} x^{k+1} = \text{Argmin}\{\mathcal{L}^3_\beta(x,y^k,z^k,\lambda^k)\,|\,x\in\mathcal{X}\}, \\ y^{k+1} = \text{Argmin}\{\mathcal{L}^3_\beta(x^{k+1},y,z^k,\lambda^k)\,|\,y\in\mathcal{Y}\}, \\ z^{k+1} = \text{Argmin}\{\mathcal{L}^3_\beta(x^{k+1},y^{k+1},z,\lambda^k)\,|\,z\in\mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b). \end{cases} \quad (6.17)$$

Unfortunately, the direct extension (6.17) **is not necessarily convergent [4] !**

ADMM + Parallel Splitting ALM

强制 y, z 平等

$$\begin{cases} x^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \mid x\in\mathcal{X}\right\}, \\ y^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) \mid y\in\mathcal{Y}\right\}, \\ z^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) \mid z\in\mathcal{Z}\right\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases}$$

平行处理 y, z 子问题, 各自为政, 不能保证方法收敛!

ADMM + Parallel-Prox Splitting ALM

各自为政, 过分自由. 给它们加个适当的正则项($\tau > 1$), 方法就能保证收敛.

$$x^{k+1} = \arg\min\{\mathcal{L}(x, y^k, z^k, \lambda^k) \mid x\in\mathcal{X}\}, \quad (6.18a)$$

$$\begin{cases} y^{k+1} = \arg\min\{\mathcal{L}(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\|B(y-y^k)\|^2 \mid y\in\mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\|C(z-z^k)\|^2 \mid z\in\mathcal{Z}\}, \end{cases} \quad (6.18b)$$

$$\lambda^{k+1} = \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \quad (6.18c)$$

Notice that (6.18b) can be written as

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} = \arg\min\left\{ \mathcal{L}(x^{k+1}, y, z, \lambda^k) + \frac{1}{2}\left\| \begin{array}{c} y - y^k \\ z - z^k \end{array} \right\|^2_{D_{BC}} \,\middle|\, \begin{array}{l} y \in \mathcal{Y} \\ z \in \mathcal{Z} \end{array} \right\},$$

where

$$D_{BC} = \begin{pmatrix} \tau B^T B & -B^T C \\ -C^T B & \tau C^T C \end{pmatrix}. \tag{6.19}$$

D_{BC} is positive semidefinite when $\tau \geq 1$.

However, the matrix D_{BC} is indefinite for $\tau \in (0, 1)$.

In other words, the scheme (6.18) can be rewritten as

$$\left\{ \begin{array}{rcl} x^{k+1} & = & \arg\min\{\mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ \begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} & = & \arg\min\left\{ \mathcal{L}(x^{k+1}, y, z, \lambda^k) + \frac{1}{2}\left\| \begin{array}{c} y - y^k \\ z - z^k \end{array} \right\|^2_{D_{BC}} \,\middle|\, \begin{array}{l} y \in \mathcal{Y} \\ z \in \mathcal{Z} \end{array} \right\}, \\ \lambda^{k+1} & = & \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{array} \right.$$

The algorithm (6.18) can be rewritten in an equivalent form: $(\mu = \tau + 1 > 2)$.

$$\left\{\begin{array}{l} x^{k+1} = \arg\min\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b - \frac{1}{\beta}\lambda^k\|^2 \,|\, x \in \mathcal{X}\}, \\ \qquad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\ y^{k+1} = \arg\min\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{array}\right. \tag{6.20}$$

The related publication:

- B. He, M. Tao and X. Yuan, A splitting method for separable convex programming. IMA J. Numerical Analysis, 31(2015), 394-426.

In the above paper, in order to ensure the convergence, it **was** required

$\tau > 1$ (in (6.18)) which is equivalent to $\mu > 2$ (in (6.20)).

This method is accepted by Osher's research group

- E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

> tion refinement step. Due to the different algorithm used to solve the extended model, there is an additional numerical parameter μ, which for this application must be greater than two according to [34]. We set μ equal to 2.01. There are also model parame-

Thus, Osher's research group utilize the iterative formula (6.20), according to our previous paper, they set

$$\mu = 2.01, \qquad \text{it is only a pity larger than 2.}$$

Large parameter μ (or τ) will lead a slow convergence.

最新进展：最优正则化因子的选择– OO6235 的结论

Recent Advance in : Bingsheng He, Xiaoming Yuan: On the Optimal Proximal Parameter of an ADMM-like Splitting Method for Separable Convex Programming http://www.optimization-online.org/DB_HTML/2017/ 10/6235.html

Our new assertion: In (6.18)

- if $\tau > 0.5$, the method is still convergent;
- if $\tau < 0.5$, there is divergent example.

Equivalently in (6.20) :

- if $\mu > 1.5$, the method is still convergent;
- if $\mu < 1.5$, there is divergent example.

For convex optimization problem (6.16) with three separable objective functions, the parameters in the equivalent methods (6.18) and (6.20) :

- **0.5** is the threshold factor of the parameter τ in (6.18) !
- **1.5** is the threshold factor of the parameter μ in (6.20) !

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Foundations and Trends in Machine Learning Vol. 3, No. 1 (2010) 1 – 122.

[2] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, Science China Mathematics, 56 (2013), 2179-2186.

[3] A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

[4] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, *Mathematical Programming, Series A* 2016.

[5] E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

[6] D. Gabay, Applications of the method of multipliers to variational inequalities, *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, edited by M. Fortin and R. Glowinski, North Holland, Amsterdam, The Netherlands, 1983, pp. 299–331.

[7] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[8] D. Hallac, Ch. Wong, S. Diamond, A. Sharang, R. Sosič, S. Boyd and J. Leskovec, SnapVX: A Network-Based Convex Optimization Solver, Journal of Machine Learning Research 18 (2017) 1-5.

[9] B. S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for

convex programming, *SIAM Journal on Optimization* **24**(2014), 1011-1040.

[10] B. S. He, M. Tao and X.M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, *SIAM Journal on Optimization* **22**(2012), 313-340.

[11] B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming, *IMA J. Numerical Analysis* **31**(2015), 394-426.

[12] B.S. He, H. Yang, and S.L. Wang, Alternating directions method with self-adaptive penalty parameters for monotone variational inequalities, *JOTA* **23**(2000), 349–368.

[13] B. S. He and X. M. Yuan, On the $O(1/t)$ convergence rate of the alternating direction method, *SIAM J. Numerical Analysis* **50**(2012), 700-709.

[14] B.S. He and X.M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, *SIAM J. Imag. Science* **5**(2012), 119-149.

[15] M. R. Hestenes, Multiplier and gradient methods, *JOTA* **4**, 303-320, 1969.

[16] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.*, **4**, 154-159, 1970.

[17] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15** (2004), 229–251.

[18] J. Nocedal and S. J. Wright, Numerical Optimization, Springer Verlag, New York, 1999.

[19] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in Optimization, R. Fletcher, ed., Academic Press, New York, NY, pp. 283-298, 1969.

Thank you very much for your attention ！

Thank you very much for reading ！