# 凸优化的一些典型问题及其求解方法

## 二. 交替方向法和分裂收缩算法的统一框架


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


统计与数据科学前沿理论及应用 教育部重点实验室

华东师范大学计算机科学与技术学院

$2020$ 年 $11$ 月 $28-29$ 日

# 1 两个可分离目标函数的凸优化问题

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

## 1.1 图像中的问题 **Image decomposition**

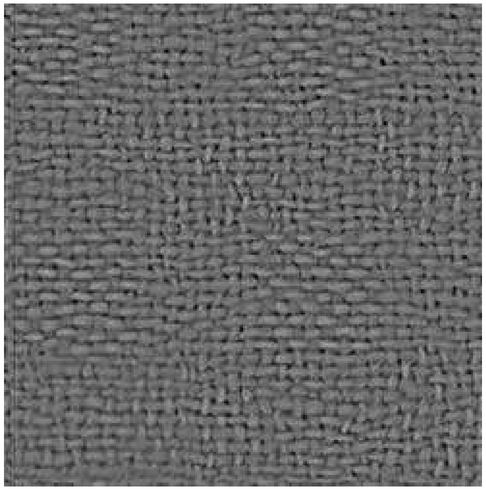
Separate the sketch (cartoon) and oscillating component (texture) of image

$$\mathbf{f} = \mathbf{u} + \mathbf{v}, \quad \mathbf{u} \text{ — cartoon part, } \mathbf{v} \text{ — texture part}$$

**Model** $$\min \ \{\|\nabla \mathbf{u}\|_1 + \tau \|\mathbf{v}\|_{-1,\infty} \ | \ \mathbf{u} + \mathbf{v} = \mathbf{f}\}$$

original image

cartoon part



texture part

## Background extraction of surveillance video (I)

Considering the foreground object detection in complex environments and extract the background in surveillance video

$$D = X + Y, \quad D \text{ — original video, } X \text{ — background, } Y \text{ — foreground}$$

**Model** $\min \ \{\|X\|_* + \tau\|Y\|_1 \ | \ X + Y = D\}$

original video

foreground

background

## Image denoising

Pixels are perturbed by a whole range of external and unwanted disturbances

$$\mathbf{g} = \mathbf{f} + noise$$

## Model

$$\min \ \left\{ \|\nabla \mathbf{f}\|_1 + \tfrac{1}{2}\|\mathbf{f} - \mathbf{g}\|_2^2 \right\} \Leftrightarrow \min \ \left\{ \|\mathbf{y}\|_1 + \tfrac{1}{2}\|\mathbf{f} - \mathbf{g}\|_2^2 \mid \nabla \mathbf{f} - \mathbf{y} = 0 \right\},$$

original image

noised image

restored image

## 1.2 矩阵优化问题

**Best matrix approximation under some conditions**

$$\min_{X} \{\frac{1}{2}\|X - C\|_F^2 \,|\, X \in S_\Lambda^n \cap S_B\},$$

where

$$S_\Lambda^n = \{H \in \mathcal{S}^n \,|\, \lambda_{\min} I \preceq H \preceq \lambda_{\max} I\}$$

and

$$S_B = \{H \in \mathcal{S}^n \,|\, H_L \le H \le H_U\}.$$

转换成等价的结构型优化问题:

$$\begin{aligned} \min_{X,Y} \quad & \tfrac{1}{2}\|X - C\|^2 + \tfrac{1}{2}\|Y - C\|^2 \\ \text{s.t} \quad & X - Y = 0, \\ & X \in S_+^n,\ Y \in S_B. \end{aligned} \tag{1.2}$$

The problem (1.2) is a concrete problem of type (1.1).

## Best matrix approximation under some conditions

Matrix completion is to recover an unknown matrix from a sampling of its entries.

For an $m \times n$ matrix $M$, $\Omega$ denotes the indices subset of the matrix

$$\Omega = \{(ij) \,|\, i \in \{1, 2, \ldots, m\},\ j \in \{1, 2, \ldots, n\}\}.$$

Only some entries of the matrix $M_{ij}$, $(ij) \in \Omega$ are known.

- Obviously, without additional information, one could not complete the matrix in general. 要由部分信息获取全部信息
- Fortunately, in many cases, the matrix that we want to recover is low-rank.

A common model to complete the low-rank matrix:

$$\min\{\text{rank}(X) \,:\, X_{ij} = M_{ij},\ (ij) \in \Omega\}.$$

However, this problem is NP-hard

根据 Candés, Recht, Tao 的工作

- E. J. Candés and B. Recht, Exact Matrix Completion via Convex Optimization, 2008.
- E. J. Candés and T. Tao, The Power of Convex Relaxation: Near-Optimial Matrix Completion, 2009.

在适当(实际问题具备的)条件下, 大多数不完整信息的低秩矩阵可以通过求解松弛问题

$$\min\{\|X\|_* \mid X_{ij} = M_{ij},\ (ij) \in \Omega\}$$

或者

$$(\spadesuit) \qquad \min\{\|X\|_* \mid \|X_\Omega - M_\Omega\|_F \le \Delta\}$$

得到精确恢复. 其中 $\|X\|_*$ 表示矩阵 $X$ 的奇异值的和. 通常称为矩阵 $X$ 的核模—- Nuclear Norm. 问题 ($\spadesuit$) 的等价形式是

$$\begin{aligned} \min_{X,Y} \quad & \|X\|_* + 0 \cdot Y \\ \text{s. t} \quad & X - Y = 0 \\ & X \in \Re^{m\times n}, Y \in \odot \end{aligned} \tag{1.3}$$

where $\odot = \{Y \mid \sum_{(ij)\in\Omega}(Y_{ij} - M_{ij})^2 \le \Delta^2\}$

## Smooth Optimization Approach for Covariance Selection — Statistics

$$\min_{X} \{\textbf{Tr}(CX) - \log(\det(X)) + \rho e^T|X|e \;\Big|\; X \in S^n_+\}$$

where $C$ is a given symmetric matrix, $e^T|X|e = \sum_{i=1}^n \sum_{j=1}^n |X_{ij}|$. Its equivalent optimization problem is

$$(\clubsuit) \qquad \begin{array}{rl} \min_{X,Y} & \textbf{Tr}(CX) - \log(\det(X)) + \rho e^T|Y|e \\ \text{s.t} & X - Y = 0, \\ & X \in S^n_+,\ Y \in R^{n\times n}. \end{array}$$

## Low rank and sparse optimization problem in statistics

$$\begin{array}{rl} \min_{X,Y} & \|X\|_* + \rho e^T|Y|e \\ \text{s.t} & X + Y = H \\ & X,\ Y \in R^{n\times n}. \end{array} \tag{1.4}$$

这些矩阵优化的数学模型是本身就是一个形如 (1.1) 的结构型优化问题.

# 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* $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ *is nonempty. Then,*

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

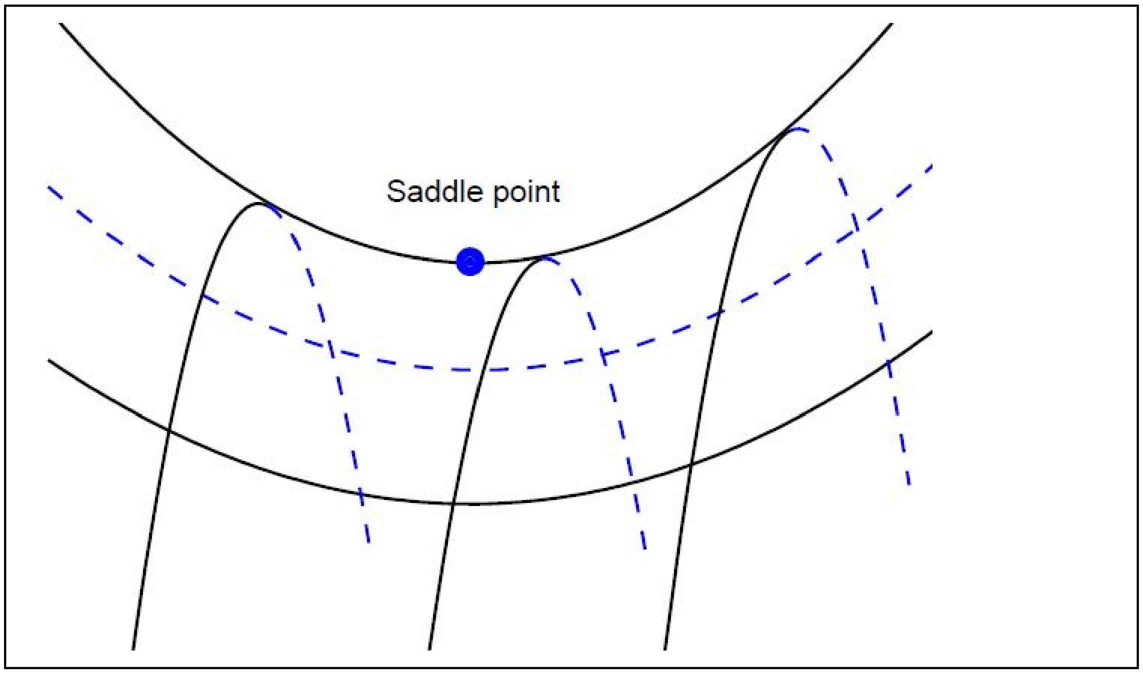
*if and only if*

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

## 2.1 Linearly constrained convex optimization and VI

The Lagrangian function of the problem (1.1) is

$$L^2(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T (Ax + By - b).$$




The saddle point $(x^*, y^*, \lambda^*) \in \mathcal{X} \times \mathcal{Y} \times \Re^m$ of $L^2(x, y, \lambda)$ satisfies

$$L^2_{\lambda \in \Re^m}\big(x^*, y^*, \lambda\big) \le L^2\big(x^*, y^*, \lambda^*\big) \le L^2_{x \in \mathcal{X}, y \in \mathcal{Y}}\big(x, y, \lambda^*\big).$$

In other words, for any saddle point $(x^*, y^*, \lambda^*)$, we have

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

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

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

Its compact form is the following variational inequality:

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

where

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

and

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

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

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \{F(w^{k+1}) + H(w^{k+1} - w^k)\} \ge 0, \ \forall\, w \in \Omega, \qquad (2.5)$$

where $H$ is a symmetric positive definite matrix.

✠ Again, $w^k$ is the solution of (2.4) if and only if $w^k = w^{k+1}$ ✠

**Convergence Property of Proximal Point Algorithm in $H$-norm**

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \qquad (2.6)$$

The sequence $\{w^k\}$ is Fejér monotone in $H$-norm. In customized PPA, via choosing a proper positive definite matrix $H$, the solution of the subproblem (2.5) has a closed form. An iterative algorithm is called the contraction method, if its generated sequence $\{w^k\}$ satisfies $\|w^{k+1} - w^*\|_H^2 < \|w^k - w^*\|_H^2$.

# 3 ADMM 求解两个可分离目标函数凸优化问题

**问题**

$$\min\{\theta_1(x)+\theta_2(y)\,|\,Ax+By=b,\ x\in\mathcal{X},\ y\in\mathcal{Y}\}. \quad (1.1)$$

**罚函数–Penalty Function**

$$P(x,y,\beta_k)=\theta_1(x)+\theta_2(y)+\frac{\beta_k}{2}\|Ax+By-b\|^2.$$

**Lagrange 函数**

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

**增广 Lagrange 函数**

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

**求解问题 (1.1) 的罚函数方法** 理论上需要数列 $\{\beta_k\} \nearrow$ 趋于无穷

$$(x^{k+1}, y^{k+1}) = \text{Argmin}\Big\{\theta_1(x) + \theta_2(y) + \frac{\beta_k}{2}\|Ax + By - b\|^2 \Big| x \in \mathcal{X},\ y \in \mathcal{Y}\Big\}$$

**求解问题 (1.1) 的增广 Lagrange 乘子法** 从给定的 $\lambda^k$ 开始

$$(x^{k+1}, y^{k+1}) = \text{Argmin}\left\{ \begin{array}{l|l} \theta_1(x) + \theta_2(y) - (\lambda^k)^T(Ax + By - b) & x \in \mathcal{X} \\ \quad + \frac{\beta}{2}\|Ax + By - b\|^2 & y \in \mathcal{Y} \end{array} \right\}$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b).$$ 原则上 $\beta$ 可以固定.

子问题难能度一样, 增广 Lagrange 乘子法(ALM)优于罚函数方法 [15]

原因：迭代犹如博弈双方讨价还价, (ALM) 同时考虑了对方的感受。

**共同的缺点** 没有利用 $x$ 和 $y$ 的可分离结构！求解会无从着手。

## 求解问题 (1.1) 的松弛的罚函数方法 — 交替极小化方法(AMA)

从给定的 $y^k$ 开始

$$x^{k+1} = \text{Argmin}\left\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,\middle|\, x \in \mathcal{X}\right\},$$

$$y^{k+1} = \text{Argmin}\left\{\theta_2(y) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,\middle|\, y \in \mathcal{Y}\right\}.$$

## 求解问题 (1.1) 的松弛的增广 Lagrange 乘子法 — ADMM

从给定的 $(y^k, \lambda^k)$ 开始

$$x^{k+1} = \text{Argmin}\left\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,\middle|\, x \in \mathcal{X}\right\},$$

$$y^{k+1} = \text{Argmin}\left\{\theta_2(y) - (\lambda^k)^T By + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,\middle|\, y \in \mathcal{Y}\right\},$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b).$$

**都松弛, 乘子交替方向法 (ADMM) 应该优于交替极小化方法 (AMA)**

## 3.1 两个可分离目标函数问题的 ADMM 方法

**Applied ADMM to the structured COP:** $(y^k, \lambda^k) \Rightarrow (y^{k+1}, \lambda^{k+1})$

First, for given $(y^k, \lambda^k)$, $x^{k+1}$ is the solution of the following problem

$$x^{k+1} = \text{Argmin}\left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ +\frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \qquad \text{(3.1a)}$$

Use $\lambda^k$ and the obtained $x^{k+1}$, $y^{k+1}$ is the solution of the following problem

$$y^{k+1} = \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) \\ +\frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \qquad \text{(3.1b)}$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \qquad \text{(3.1c)}$$

**Advantages** The $x$ and $y$ sub-problems are separately solved one by one.

**Remark** Ignoring the constant term in the objective function, the sub-problems (3.1a) and (3.1b) is equivalent to

$$x^{k+1} = \text{Argmin}\big\{\theta_1(x) + \tfrac{\beta}{2}\|(Ax + By^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 \big| x \in \mathcal{X}\big\} \tag{3.2a}$$

and

$$y^{k+1} = \text{Argmin}\big\{\theta_2(y) + \tfrac{\beta}{2}\|(Ax^{k+1} + By - b) - \tfrac{1}{\beta}\lambda^k\|^2 \big| y \in \mathcal{Y}\big\} \tag{3.2b}$$

respectively. Note that the equation (3.1c) can be written as

$$(\lambda - \lambda^{k+1})\{(Ax^{k+1} + By^{k+1} - b) + \tfrac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0, \ \forall \lambda \in \Re^m. \tag{3.2c}$$

Notice that the sub-problems (3.2a) and (3.2b) are the type of

$$x^{k+1} = \text{Argmin}\big\{\theta_1(x) + \tfrac{\beta}{2}\|Ax - p^k\|^2 \big| x \in \mathcal{X}\big\}$$

and

$$y^{k+1} = \text{Argmin}\big\{\theta_2(y) + \tfrac{\beta}{2}\|By - q^k\|^2 \big| y \in \mathcal{Y}\big\},$$

respectively.

**Analysis** According to Lemma 1, the solution of (3.1a) and (3.1b) satisfies

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \big\{-A^T\lambda^k + \beta A^T\big(Ax^{k+1} + By^k - b\big)\big\} \geq 0, \ \forall\, x \in \mathcal{X} \tag{3.3a}$$

and

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\ \big\{-B^T\lambda^k + \beta B^T\big(Ax^{k+1} + By^{k+1} - b\big)\big\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.3b}$$

respectively. Substituting $\lambda^{k+1}$ (see (3.1c)) in (3.3) (eliminating $\lambda^k$ in (3.3)), we get

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \big\{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\big\} \geq 0, \ \forall\, x \in \mathcal{X}, \tag{3.4a}$$

and

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\big\{-B^T\lambda^{k+1}\big\} \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{3.4b}$$

The compact form of (3.4) is $u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$ and

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T B \\ 0 \end{pmatrix} (y^k - y^{k+1}) \right\} \geq 0, \tag{3.5}$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

By adding and subtracting the term $\beta B^T B(y^k - y^{k+1})$, we rewrite the about variational inequality in our desirable form

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T B \\ B^T B \end{pmatrix} (y^k - y^{k+1}) \right.$$
$$\left. + \begin{pmatrix} 0 & 0 \\ 0 & \beta B^T B \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\} \geq 0, \ \forall\, (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

Combining the last inequality with (3.2c), we have $w^{k+1} \in \Omega$ and

$$\theta(u)-\theta(u^{k+1})+\begin{pmatrix} x-x^{k+1} \\ y-y^{k+1} \\ \lambda-\lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \\ Ax^{k+1}+By^{k+1}-b \end{pmatrix} +\beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k-y^{k+1}) \right.$$
$$\left. + \begin{pmatrix} 0 & 0 \\ \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} \right\} \geq 0, \ \forall\, w \in \Omega. \quad (3.6)$$

For convenience we use the notations

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

Then, we get the following lemma:

**Lemma 2** *Let the sequence* $\{w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})\} \in \Omega$ *be generated by*

(3.1). *Then, we have*

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}), \ \forall\, w^* \in \Omega^*, \quad (3.7)$$

*where*

$$H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \quad (3.8)$$

*and*

$$\eta(y^k, y^{k+1}) = \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}). \quad (3.9)$$

**Proof**. Setting $w = w^*$ in (3.26), and using $H$ and $\eta(y^k, y^{k+1})$, we get

$$\begin{aligned} &(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \\ \geq \quad & (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) \\ & + \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}), \ \forall\, w^* \in \Omega^*. \quad (3.10) \end{aligned}$$

Since $F$ is monotone and $w^*$ is the optimal solution, it follows that

$$\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}) \geq 0.$$

Using the above inequality, the assertion (3.7) follows from (3.10) immediately. $\square$

**Lemma 3** *Let the sequence* $\{w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})\} \in \Omega$ *be generated by* (3.1)*. Then, we have*

$$(w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) \geq 0, \ \forall\, w^* \in \Omega^*, \tag{3.11}$$

*where* $\eta(y^k, y^{k+1})$ *is defined in* (3.9)*.*

**Proof**. By using $\eta(y^k, y^{k+1})$ (see (3.9)), $Ax^* + By^* = b$ and (3.1c), we have

$$\begin{aligned} & (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) \\ = \quad & \beta\{(Ax^{k+1} + By^{k+1}) - (Ax^* + By^*)\}^T B(y^k - y^{k+1}) \\ = \quad & (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}), \ \forall\, w^* \in \Omega^*. \end{aligned} \tag{3.12}$$

Because (3.4b) is true for the $k$-th iteration and the previous iteration, we have

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.13}$$

and

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T\{-B^T\lambda^k\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.14}$$

Setting $y = y^k$ in (3.13) and $y = y^{k+1}$ in (3.14), respectively, and then adding the two resulting inequalities, we get

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0. \tag{3.15}$$

Substituting (3.15) in (3.12), the assertion (3.11) follows immediately. $\square$

Substituting (3.11) in (3.7), we get

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0, \ \ \forall\, v^* \in \mathcal{V}^*. \tag{3.16}$$

Using the above inequality, as in the last lecture, we have the following theorem, which is the key for the proof of the convergence of ADMM.

**Theorem 1** *Let the sequence* $\{w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})\} \in \Omega$ *be generated by* (3.1)*. Then, we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \tag{3.17}$$

**How to choose the parameter $\beta$.** The efficiency of ADMM is heavily dependent on the parameter $\beta$ in (3.1). We discuss how to choose a suitable $\beta$ in the practical computation.

Note that if $\boldsymbol{\beta A^T B(y^k - y^{k+1}) = 0}$, then it follows from (3.5)

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} \geq 0, \ \forall (x,y) \in \mathcal{X} \times \mathcal{Y}. \quad (3.18)$$

In this case, if additionally $\boldsymbol{Ax^{k+1} + By^{k+1} - b = 0}$, then we have

$$\left\{ \begin{array}{rl} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T(-A^T\lambda^{k+1}) \geq 0, & \forall x \in \mathcal{X} \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T(-B^T\lambda^{k+1}) \geq 0, & \forall y \in \mathcal{Y} \\ (\lambda - \lambda^{k+1})^T(Ax^{k+1} + By^{k+1} - b) \geq 0, & \forall \lambda \in \Re^m \end{array} \right.$$

and consequently $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is a solution of the variational inequality (2.2).

In other words, $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is not a solution of (2.2) because

$$\beta A^T B(y^k - y^{k+1}) \neq 0 \qquad \text{and/or} \qquad Ax^{k+1} + By^{k+1} - b \neq 0.$$

We call

$$\boldsymbol{\|\beta A^T B(y^k - y^{k+1})\|} \qquad \text{and} \qquad \boldsymbol{\|Ax^{k+1} + By^{k+1} - b\|}$$

the primal-residual and the dual-residual, respectively. It seems that we should balance the primal and the dual residuals dynamically. If

$$\mu\|\beta A^T B(y^k - y^{k+1})\| < \|Ax^{k+1} + By^{k+1} - b\| \quad \text{with a} \quad \mu > 1,$$

it means that the dual residual is too large and thus we should enlarge the parameter $\beta$ in the augmented Lagrangian function. Otherwise, we should reduce the parameter $\beta$. A simple scheme that often works well is (see, e.g., [10]):

$$\beta_{k+1} = \begin{cases} \beta_k * \tau, & \text{if } \mu\|\beta A^T B(y^k - y^{k+1})\| < \|Ax^{k+1} + By^{k+1} - b\|; \\ \beta_k/\tau, & \text{if } \|\beta A^T B(y^k - y^{k+1})\| > \mu\|Ax^{k+1} + By^{k+1} - b\|; \\ \beta_k, & \text{otherwise.} \end{cases}$$

where $\mu > 1, \tau > 1$ are parameters. Typical choices might be $\mu = 10$ and $\tau = 2$. The idea behind this penalty parameter update is to try to keep the primal and dual residual norms within a factor of $\mu$ of one another as they both converge to zero. This self adaptive adjusting rule has been used by S. Boyd's group [1] and in their Optimization Solver [6].

## 3.2 Linearized ADMM

The augmented Lagrangian Function of the problem (1.1) is

$$\mathcal{L}_\beta(x,y,\lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax+By-b) + \frac{\beta}{2}\|Ax+By-b\|^2. \quad (3.19)$$

Solving the problem (1.1) by using ADMM, the $k$-th iteration begins with given $(y^k, \lambda^k)$, it offers the new iterate $(y^{k+1}, \lambda^{k+1})$ via

$$\text{(ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x, y^k, \lambda^k) \;\big|\; x \in \mathcal{X}\big\}, & \text{(3.20a)} \\ y^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \;\big|\; y \in \mathcal{Y}\big\}, & \text{(3.20b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(3.20c)} \end{array} \right.$$

In optimization problem, the solution is invariant by changing the constant term in the objective function. Thus, by using the augmented Lagrangian function,

$$\begin{aligned} y^{k+1} &= \arg\min\big\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \;\big|\; y \in \mathcal{Y}\big\} \\ &= \arg\min\big\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \;\big|\; y \in \mathcal{Y}\big\}. \end{aligned}$$

Thus, by denoting $q^k = b - Ax^{k+1} + \frac{1}{\beta}\lambda^k$, the solution of (3.20b) is given by

$$\min\{\theta_2(y) + \frac{\beta}{2}\|By - q^k\|^2 \mid y \in \mathcal{Y}\}. \tag{3.21}$$

In some practical applications, because of the structure of the matrix $B$, the subproblem (3.21) are not so easy to be solved. In this case, it is necessary to use the linearized version of the ADMM.

Notice that the Taylor expansion of the quadratic term of (3.20b), namely,

$$\begin{aligned}
&\frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \\
&= \quad \frac{\beta}{2}\|(Ax^{k+1} + By^k - b) + B(y - y^k)\|^2 \\
&= \quad \beta(y - y^k)^T B^T (Ax^{k+1} + By^k - b) + \frac{\beta}{2}\|B(y - y^k)\|^2 \\
&\qquad + \frac{\beta}{2}\|Ax^{k+1} + By^k - b\|^2 \\
&= \quad \beta y^T B^T (Ax^{k+1} + By^k - b) + \frac{\beta}{2}\|B(y - y^k)\|^2 + \text{constant.}
\end{aligned}$$

Changing the constant term in the objective function of (3.1b) accordingly, we have

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

So-called linearized version of ADMM, we remove the unfavorable term $\frac{\beta}{2}\|B(y - y^k)\|^2$ in the objective function, and add the term $\frac{s}{2}\|y - y^k\|^2$.

Strictly speaking, it should be a ”linearization” plus ”regularization” method. It can also be interpreted as:

$$
\text{The term} \quad \frac{\beta}{2}\|B(y - y^k)\|^2 \quad \text{is approximated by} \quad \frac{s}{2}\|y - y^k\|^2.
$$

In other words, it is equivalent to adding the term

$$
\frac{1}{2}\|y - y^k\|_D^2 \qquad (\text{with } \; D = sI_{n_2} - \beta B^T B) \tag{3.22}
$$

to the objective function of (3.20b), we get

$$\begin{aligned} y^{k+1} &= \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_D^2 \mid y \in \mathcal{Y}\} \\ &= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T\big[\lambda^k - \beta(Ax^{k+1} + By^k - b)\big] \\ +\dfrac{s}{2}\ \|y - y^k\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} \\ &= \arg\min\big\{\theta_2(y) + \frac{s}{2}\big\|y - d^k\big\|^2 \mid y \in \mathcal{Y}\big\}, \end{aligned} \quad (3.23)$$

where

$$d^k = y^k + \frac{1}{s}B^T\big[\lambda^k - \beta(Ax^{k+1} + By^k - b)\big].$$

By using such strategy, the sub-problems of ADMM is simplified. The linearized version of ADMM are applied successfully in scientific computing. The following analysis is based on the fact that the sub-problem

$$\min\{\theta_2(y) + \frac{s}{2}\|y - d^k\|^2 \mid y \in \mathcal{Y}\}$$

are easy to be solved.

**Linearized ADMM.** For solving the problem (1.1), the $k$-th iteration of the linearized ADMM begins with given $w^k = (x^k, y^k, \lambda^k)$, produces the

$w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ via the following procedure:

$$\left\{\begin{array}{l} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \qquad (3.24a)\\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_D^2 \mid y \in \mathcal{Y}\}, \qquad (3.24b)\\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \qquad (3.24c)\end{array}\right.$$

where $D$ is defined by (3.22).

First, using the optimality of the sub-problems of (3.24), we prove the following lemma as the base of convergence.

**Lemma 4** *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (3.24) *for the problem* (1.1)*. Then, we have* $w^{k+1} \in \Omega$ *and*

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \\ Ax^{k+1} + By^{k+1} - b \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}) + \begin{pmatrix} 0 & 0 \\ \beta B^T B + D & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} \right\} \geq 0, \ \forall\, w \in \Omega. \qquad (3.25)$$

**Proof.** By using the classical ADMM (3.20), we have (see (3.26)) $w^{k+1} \in \Omega$ and

$$\theta(u)-\theta(u^{k+1})+\begin{pmatrix} x-x^{k+1} \\ y-y^{k+1} \\ \lambda-\lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \\ Ax^{k+1}+By^{k+1}-b \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k-y^{k+1}) \right.$$

$$\left. + \begin{pmatrix} 0 & 0 \\ \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} \right\} \geq 0, \ \forall\, w \in \Omega.$$

In comparison (3.24b) with (3.20b), the assertion (3.25) is proved. $\square$

This lemma is the base for the convergence analysis of the linearized ADMM.

By using the notation $F(w)$ and $(w^{k+1}-w)^T F(w^{k+1}) = (w^{k+1}-w)^T F(w)$, it follows from (3.25) that

$$\begin{aligned} w^{k+1} \in \Omega, \ \ & \theta(u)-\theta(u^{k+1}) + (w-w^{k+1})^T F(w) \\ & + \beta \begin{pmatrix} x-x^{k+1} \\ y-y^{k+1} \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k-y^{k+1}) \\ & \geq (v-v^{k+1})^T G(v^k-v^{k+1}), \ \ \forall w \in \Omega, \end{aligned} \tag{3.26a}$$

where $G$ is given by

$$G = \begin{pmatrix} D+\beta B^T B & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}. \tag{3.26b}$$

## The contractive property of the sequence $\{w^k\}$ by Linearized ADMM (3.24)

In the following we will prove, for any $w^* \in \Omega^*$, the sequence

$$\{\|v^{k+1} - v^*\|_G + \|y^k - y^{k+1}\|_D^2\}$$

is monotonically decreasing. For this purpose， we prove some lemmas.

**Lemma 5** *Let $\{w^k\}$ be the sequence generated by Linearized ADMM* (3.24) *for the problem* (1.1). *Then, we have*

$$(v^{k+1}-v^*)^T G(v^k-v^{k+1}) \geq (\lambda^k-\lambda^{k+1})^T B(y^k-y^{k+1}), \quad \forall w^* \in \Omega^*. \tag{3.27}$$

**Proof**. Setting the $w \in \Omega$ in (3.26a) by any $w^* \in \Omega^*$, we obtain

$$
\begin{aligned}
&(v^{k+1} - v^*)^T G(v^k - v^{k+1}) \\
&\quad \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \\
&\qquad + \beta \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}).
\end{aligned} \tag{3.28}
$$

According to the optimality, a part of the terms in the right hand side of the above inequality,

$$\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0.$$

Using $Ax^* + By^* = b$ and $\lambda^k - \lambda^{k+1} = \beta(Ax^{k+1} + By^{k+1} - b)$ (see (3.24c)) to deal the last term in the right hand side of (3.28) , it follows that

$$
\begin{aligned}
&\beta \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \\
&= \beta[(Ax^{k+1} - Ax^*) + (By^{k+1} - By^*)]^T B(y^k - y^{k+1}) \\
&= (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}).
\end{aligned}
$$

The lemma is proved. $\square$

**Lemma 6** *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (3.24) *for the problem* (1.1)*. Then, we have*

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq \frac{1}{2}\|y^k - y^{k+1}\|_D^2 - \frac{1}{2}\|y^{k-1} - y^k\|_D^2. \qquad (3.29)$$

**Proof**. First, for the classical ADMM we have (see (3.4b))

$$y^{k+1} \in \mathcal{Y}, \ \ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}.$$

In comparison (3.24b) with (3.20b), for linearized ADMM, we have

$$\begin{aligned} y^{k+1} \in \mathcal{Y}, \ \ \theta_2(y) - \theta_2(y^{k+1}) &+ (y - y^{k+1})^T \\ &\{-B^T\lambda^{k+1} + D(y^{k+1} - y^k)\} \geq 0, \ \forall\, y \in \mathcal{Y}. \end{aligned} \qquad (3.30)$$

Setting $k$ in (3.30) by $k-1$, we have

$$\begin{aligned} y^k \in \mathcal{Y}, \ \ \theta_2(y) - \theta_2(y^k) &+ (y - y^k)^T \\ &\{-B^T\lambda^k + D(y^k - y^{k-1})\} \geq 0, \ \ \forall\, y \in \mathcal{Y}. \end{aligned} \qquad (3.31)$$

Setting the $y$ in (3.30) and (3.31) by $y^k$ and $y^{k+1}$, respectively, and adding them, we get

$$(y^k - y^{k+1})^T \{ B^T(\lambda^k - \lambda^{k+1}) + D[(y^{k+1} - y^k) - (y^k - y^{k-1})] \} \geq 0.$$

From the above inequality we get

$$(y^k - y^{k+1})^T B^T(\lambda^k - \lambda^{k+1}) \geq (y^k - y^{k+1})^T D[(y^k - y^{k+1}) - (y^{k-1} - y^k)].$$

Using the Cauchy-Schwarz inequality for the right hand side term of the above inequality, we get

$$\begin{aligned} &(y^k - y^{k+1})^T D[(y^k - y^{k+1}) - (y^{k-1} - y^k)] \\ &\qquad \geq \|y^k - y^{k+1}\|_D^2 - (y^k - y^{k+1})^T D(y^{k-1} - y^k) \\ &\qquad \geq \frac{1}{2}\|y^k - y^{k+1}\|^2 - \frac{1}{2}\|y^{k-1} - y^k\|_D^2, \end{aligned}$$

The assertion (3.29) of this lemma follows directly from the above two inequalities. □

By using Lemma 5 and Lemma 6, we can prove the following convergence theorem.

**Theorem 2** *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (3.24) *for the problem* (1.1). *Then, we have*

$$\begin{aligned}&\big(\|v^{k+1}-v^*\|_G^2+\|y^k-y^{k+1}\|_D^2\big)\\ &\quad\le\big(\|v^k-v^*\|_G^2+\|y^{k-1}-y^k\|_D^2\big)-\|v^k-v^{k+1}\|_G^2,\ \forall w^*\in\Omega^*,\end{aligned}\tag{3.32}$$

*where* $G$ *is given by* (3.26b).

**Proof**. From Lemma 5 and Lemma 6, it follows that

$$(v^{k+1}-v^*)^TG(v^k-v^{k+1})\ge\frac{1}{2}\|y^k-y^{k+1}\|_D^2-\frac{1}{2}\|y^{k-1}-y^k\|_D^2,\ \ \forall w^*\in\Omega^*.$$

Using the above inequality, for any $w^*\in\Omega^*$, we get

$$\begin{aligned}\|v^k-v^*\|_G^2&=\|(v^{k+1}-v^*)+(v^k-v^{k+1})\|_G^2\\ &\ge\|v^{k+1}-v^*\|_G^2+\|v^k-v^{k+1}\|_G^2+2(v^{k+1}-v^*)^TG(v^k-v^{k+1})\\ &\ge\|v^{k+1}-v^*\|_G^2+\|v^k-v^{k+1}\|_G^2\\ &\qquad+\|y^k-y^{k+1}\|_D^2-\|y^{k-1}-y^k\|_D^2.\end{aligned}$$

The assertion of the Theorem 2 is proved from the last inequality. $\square$

# 4 Splitting Methods in a Unified Framework

We study the algorithms using the guidance of variational inequality. The optimal condition of the linearly constrained convex optimization is resulted in a variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{4.1}$$

## 4.1 Algorithms in a unified framework

**A Prototype Algorithm for (4.1)**

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{4.2a}$$

where the matrix $Q$ has the property: $Q^T + Q$ is positive definite.

[Correction Step.] Determine a nonsingular matrix $M$ and a scalar $\alpha > 0$, let

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \tag{4.2b}$$

**Convergence Conditions**

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

$$HM = Q. \tag{4.3a}$$

For the given $H$, $M$ and $Q$ satisfied the condition (4.3a), and the step size $\alpha$ determined in (4.2), the matrix

$$G = Q^T + Q - \alpha M^T HM \succ 0. \tag{4.3b}$$

There are many possibilities, the principle is simplicity and efficiency. See an example:

- The simplest case is $H = I$ and $M = Q$.
- To ensure the symmetry and positivity of $H = QM^{-1}$, we take $H = QD^{-1}Q^T$, where $D$ is a symmetric investable block diagonal matrix.
  Set $M^{-1} = D^{-1}Q^T$, Thus $M = Q^{-T}D$ satisfies the condition (4.3a).
- After choosing the matrix $M$, let

$$\alpha_{\max} = \arg\max\{\alpha \,|\, Q^T + Q - \alpha M^T HM \succ 0\},$$

  the condition (4.3b) is satisfied for any $\alpha \in (0, \alpha_{\max})$.

## 4.2 Methods for Linearly Constrained Problems

We consider the convex optimization, namely

$$\min\{\theta(u) \mid Au = b,\ u \in \mathcal{U}\}. \tag{4.4}$$

The related variational inequality of the saddle point of the Lagrangian function is

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

where

$$w = \left(\begin{array}{c} u \\ \lambda \end{array}\right), \quad F(w) = \left(\begin{array}{c} -A^T\lambda \\ Au - b \end{array}\right) \ \text{ and } \quad \Omega = \mathcal{U} \times \Re^m.$$

### Augmented Lagrangian Method

Its augmented Lagrangian function is

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

The $k$-th iteration of the **Augmented Lagrangian Method** [**?**, 16] begins with a given $\lambda^k$, obtain $w^{k+1} = (x^{k+1}, \lambda^{k+1})$ via

$$\text{(ALM)} \quad \begin{cases} \tilde{u}^k = \arg\min\{\mathcal{L}_\beta(u, \lambda^k) \mid u \in \mathcal{U}\}, & \text{(4.5a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{u}^k - b). & \text{(4.5b)} \end{cases}$$

In (4.5), $\tilde{u}^k$ is only a computational result of (4.5a) from given $\lambda^k$, it is called the intermediate variable. In order to start the $k$-th iteration of ALM, we need only to have $\lambda^k$ and thus we call it as the essential variable.

The subproblem (4.5a) is a problem of mathematical form

$$\min\{\theta(u) + \tfrac{\beta}{2}\|Au - p^k\|^2 \mid u \in \mathcal{U}\} \tag{4.6}$$

where $\beta > 0$ is a given scalar and $p^k = b + \frac{1}{\beta}\lambda^k$.

**Assumption**: The solution of problem (4.6) has closed-form solution or can be efficiently computed with a high precision.

The optimal condition can be written as $\tilde{w}^k \in \Omega$ and

$$\begin{cases} \theta(u) - \theta(\tilde{u}^k) \; + \; (u - \tilde{u}^k)^T \{-A^T\lambda^k + \beta A^T(A\tilde{u}^k - b)\} \geq 0, \;\; \forall u \in \mathcal{U}, \\ \qquad\qquad\qquad\qquad\quad (\lambda - \tilde{\lambda}^k)^T\{(A\tilde{u}^k - b) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \;\; \forall \lambda \in \Re^m. \end{cases}$$

The above relations can be written as

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} u - \tilde{u}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{u}^k - b \end{pmatrix} \geq (\lambda - \tilde{\lambda}^k)^T \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k), \; \forall \, w \in \Omega. \quad (4.7)$$

Setting $v = \lambda$ in (4.7), it can be written as (4.2a),

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

with $Q = \frac{1}{\beta} I$.

**Correction** $\qquad \lambda^{k+1} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k), \quad \alpha \in (0, 2).$

Indeed, because $M = I, \; H = QM^{-1} = Q \succ 0,$

$$G = Q^T + Q - \alpha M^T H M = \frac{2-\alpha}{\beta} I \succ 0.$$

## Customized PPA

Recall the convex optimization problem (4.4), namely,

$$\min\{\theta(u) \mid Au = b,\ u \in \mathcal{U}\}.$$

The related variational inequality of the saddle point of the Lagrangian function is

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

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Au - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{U} \times \Re^m.$$

For given $v^k = w^k = (u^k, \lambda^k)$, the predictor is given by

$$\text{(CPPA)} \quad \begin{cases} \tilde{u}^k = \arg\min\big\{L(u, \lambda^k) + \frac{r}{2}\|u - u^k\|^2 \;\big|\; u \in \mathcal{U}\big\}, & \text{(4.8a)} \\ \tilde{\lambda}^k = \arg\max\big\{L\big([2\tilde{u}^k - u^k], \lambda\big) - \frac{s}{2}\|\lambda - \lambda^k\|^2\big\} & \text{(4.8b)} \end{cases}$$

The output $\tilde{w}^k \in \Omega$ of the iteration (4.8) satisfies

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

It is a form of (4.2a) where

$$Q = \begin{pmatrix} rI & A^T \\ A & sI \end{pmatrix}$$ is symmetric

The subproblem (4.8a) is a problem of mathematical form

$$\min\{\theta(u) + \frac{r}{2}\|u - a^k\|^2 \,|\, u \in \mathcal{U}\} \tag{4.9}$$

where $r > 0$ is a given scalar and $a^k = x^k + \frac{1}{r} A^T \lambda^k$

**Assumption**:

- The solution of problem (4.9) has has closed-form solution.
- To ensure the positiveness of the matrix $Q$, we have to set $rs > \|A^T A\|$.

We take $M = I$ in the correction (4.2b) and the new iterate is updated by

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

Then, we have and

$$H = QM^{-1} = Q \succ 0 \quad \text{and} \quad G = Q^T + Q - \alpha M^T HM = (2-\alpha)H \succ 0.$$

The convergence conditions (4.3) are satisfied. More about customized PPA, please see

♣ G.Y. Gu, B.S. He and X.M. Yuan, Customized Proximal point algorithms for linearly constrained convex minimization and saddle-point problem: a unified Approach, Comput. Optim. Appl., 59(2014), 135-161.

# 5 Convergence proof in the unified framework

In this section, assuming the conditions (4.3) in the unified framework are satisfied, we prove some convergence properties.

**Theorem 1** *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (4.1) *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*

$$\begin{aligned}&\alpha\big(\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k)\big)\\ \geq\quad& \frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{\alpha}{2}\|v^k-\tilde{v}^k\|_G^2,\quad \forall w\in\Omega. \qquad (5.1)\end{aligned}$$

**Proof**. Using $Q=HM$ (see (4.3a)) and the relation (4.2b), the right hand side of (4.3a) can be written as $(v-\tilde{v}^k)^T\frac{1}{\alpha}H(v^k-v^{k+1})$ and hence

$$\alpha\{\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k)\}\geq(v-\tilde{v}^k)^T H(v^k-v^{k+1}),\ \forall w\in\Omega. \qquad (5.2)$$

Applying the identity

$$(a-b)^T H(c-d)=\frac{1}{2}\{\|a-d\|_H^2-\|a-c\|_H^2\}+\frac{1}{2}\{\|c-b\|_H^2-\|d-b\|_H^2\},$$

to the right hand side of (5.2) with

$$a = v, \quad b = \tilde{v}^k, \quad c = v^k, \quad \text{and} \quad d = v^{k+1},$$

we thus obtain

$$\begin{aligned} & 2(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ & \quad = \big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + (\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \end{aligned} \tag{5.3}$$

For the last term of (5.3), we have

$$\begin{aligned} & \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ & \quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ & \quad \overset{(4.3a)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - \alpha M(v^k - \tilde{v}^k)\|_H^2 \\ & \quad = \quad 2\alpha(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - \alpha^2(v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ & \quad = \quad \alpha(v^k - \tilde{v}^k)^T(Q^T + Q - \alpha M^T HM)(v^k - \tilde{v}^k) \\ & \quad \overset{(4.3b)}{=} \quad \alpha\|v^k - \tilde{v}^k\|_G^2. \end{aligned} \tag{5.4}$$

Substituting (5.3), (5.4) in (5.2), the assertion of this theorem is proved. $\square$

## 5.1 Convergence in a strictly contraction sense

**Theorem 2** *Let $\{v^k\}$ be the sequence generated by a method for the problem* (4.1) *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 ($G \succ 0$), then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.5}$$

**Proof**. Setting $w = w^*$ in (5.1), we get

$$\begin{aligned} &\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &\quad \ge \quad \alpha\|v^k - \tilde{v}^k\|_G^2 + 2\alpha\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \end{aligned} \tag{5.6}$$

By using the optimality of $w^*$ and the monotonicity of $F(w)$, we have

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

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \alpha\|v^k - \tilde{v}^k\|_G^2. \tag{5.7}$$

The assertion (5.5) follows directly. $\square$

## 5.2 Convergence rate in an ergodic sense

**Equivalent Characterization of the Solution Set of VI**

For the convergence rate analysis, we need another characterization of the solution set of VI (4.1). It can be described the following theorem.

**Theorem 3** *The solution set of VI*$(\Omega, F, \theta)$ *is convex and it can be characterized as*

$$\Omega^* = \{\tilde{w} \in \Omega \,|\, \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0, \;\; \forall\, w \in \Omega.\} \tag{5.8}$$

**Proof**. According to the definition of the solution of VI, for any $\tilde{w} \in \Omega^*$, we have

$$\tilde{w} \in \Omega, \;\; \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(\tilde{w}) \geq 0, \;\; \forall\, w \in \Omega.$$

Since $(w - \tilde{w})^T F(\tilde{w}) = (w - \tilde{w})^T F(w)$,

$$\tilde{w} \in \Omega, \;\; \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0, \;\; \forall\, w \in \Omega.$$

The theorem is proved. $\square$

We use (5.8) to define the approximate solution of VI (4.1). Namely, for given $\epsilon > 0$, $\tilde{w} \in \Omega$ is called an $\epsilon$-approximate solution of VI$(\Omega, F, \theta)$, if it satisfies

$$\tilde{w} \in \Omega, \quad \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq -\epsilon, \quad \forall\, w \in \mathcal{D}_{(\tilde{w})},$$

where

$$\mathcal{D}_{(\tilde{w})} = \{w \in \Omega \,|\, \|w - \tilde{w}\| \leq 1\}.$$

We need to show that for given $\epsilon > 0$, after $t$ iterations, it can offer a $\tilde{w} \in \mathcal{W}$, such that

$$\tilde{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \Big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\Big\} \leq \epsilon. \tag{5.9}$$

Theorem 1 is also the base for the convergence rate proof. Substituting

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

in (5.1), we obtain

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2\alpha}\|v - v^k\|_H^2 \geq \frac{1}{2\alpha}\|v - v^{k+1}\|_H^2, \ \forall w \in \Omega. \tag{5.10}$$

Note that the above assertion is hold for $G \succeq 0$.

**Theorem 4** *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (4.1) *and* $\tilde{w}^k$ *is obtained in the* $k$*-th iteration. Assume that* $v^k$, $v^{k+1}$ *and* $\tilde{w}^k$ *satisfy the conditions in the unified framework and let* $\tilde{w}_t$ *be defined by*

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{w}^k. \tag{5.11}$$

*Then, for any integer number* $t > 0$, $\tilde{w}_t \in \Omega$ *and*

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2\alpha(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{5.12}$$

**Proof**. First, it holds that $\tilde{w}^k \in \Omega$ for all $k \ge 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (5.11) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (5.10) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t}\theta(\tilde{u}^k) + \Big((t+1)w - \sum_{k=0}^{t}\tilde{w}^k\Big)^T F(w) + \frac{1}{2\alpha}\|v - v^0\|_H^2 \ge 0, \quad \forall w \in \Omega.$$

Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k)-\theta(u)+(\tilde{w}_t-w)^TF(w)\leq\frac{1}{2\alpha(t+1)}\|v-v^0\|_H^2,\quad\forall w\in\Omega. \tag{5.13}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t=\frac{1}{t+1}\sum_{k=0}^{t}\tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_t)\leq\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k).$$

Substituting it in (5.13), the assertion of this theorem follows directly. $\square$

Recall (5.9). The conclusion (5.12) thus indicates obviously that the method is able to generate an approximate solution (i.e., $\tilde{w}_t$) with the accuracy $O(1/t)$ after $t$ iterations. That is, in the case $G\succeq 0$, the convergence rate $O(1/t)$ of the method is established.


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

## 5.3 Convergence rate in a pointwise iteration-complexity

In this subsection, we show that if the matrix $G$ defined in (4.3b) is positive definite, a worst-case $O(1/t)$ convergence rate in a pointwise iteration-complexity can also be established for the prototype algorithm (4.2). Note in general a pointwise convergence rate is stronger than the ergodic convergence rate.

Recall that we have proved

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

In fact, $\|v^k - \tilde{v}^k\|$ can be used to measure how much $w^k$ fails to be a solution point. In the following, we will prove that

$$\|v^{k+1} - \tilde{v}^{k+1}\|_{M^T HM} \le \|v^k - \tilde{v}^k\|_{M^T HM}$$

By using the correction form $v^k - v^{k+1} = \alpha(v^k - \tilde{v}^k)$, consequently, we get

$$\|v^{k+1} - v^{k+2}\|_H \le \|v^k - v^{k+1}\|_H.$$

We first need to prove the following lemma.

**Lemma 7** *For the sequence generated by the prototype algorithm (4.2) where the Convergence Condition is satisfied, we have*

$$\begin{aligned} &(v^k - \tilde{v}^k)^T M^T H M \{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\ \geq \quad & \frac{1}{2\alpha} \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \end{aligned} \tag{5.14}$$

**Proof**. First, set $w = \tilde{w}^{k+1}$ in (4.2a), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{5.15}$$

Note that (4.2a) is also true for $k := k+1$ and thus we have

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

Set $w = \tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}). \tag{5.16}$$

Combining (5.15) and (5.16) and using the monotonicity of $F$, we get

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq 0. \tag{5.17}$$

Adding the term

$$\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}$$

to the both sides of (5.17), and using $v^T Q v = \frac{1}{2} v^T (Q^T + Q) v$, we obtain

$$(v^k - v^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \frac{1}{2} \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}.$$

Substituting $(v^k - v^{k+1}) = \alpha M(v^k - \tilde{v}^k)$ in the left-hand side of the last inequality and using $Q = HM$, we obtain (5.14) and the lemma is proved. $\square$

Now, we are ready to prove (5.18), the key inequality in this section.

**Theorem 5** *For the sequence generated by the prototype algorithm (4.2) where the Convergence Condition is satisfied, we have*

$$\|M(v^{k+1} - \tilde{v}^{k+1})\|_H \leq \|M(v^k - \tilde{v}^k)\|_H, \quad \forall\, k > 0. \tag{5.18}$$

**Proof**. Setting $a = M(v^k - \tilde{v}^k)$ and $b = M(v^{k+1} - \tilde{v}^{k+1})$ in the identity

$$\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a-b) - \|a-b\|_H^2,$$

we obtain

$$\begin{aligned} & \|M(v^k - \tilde{v}^k)\|_H^2 - \|M(v^{k+1} - \tilde{v}^{k+1})\|_H^2 \\ = \quad & 2(v^k - \tilde{v}^k)^T M^T HM[(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})] \\ & \quad -\|M[(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})]\|_H^2. \end{aligned}$$

Inserting (5.14) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned} & \|M(v^k - \tilde{v}^k)\|_H^2 - \|M(v^{k+1} - \tilde{v}^{k+1})\|_H^2 \\ \geq \quad & \frac{1}{\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2 - \|M[(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})]\|_H^2 \\ = \quad & \frac{1}{\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2 \geq 0, \end{aligned}$$

where the last inequality is because of the positive definiteness of the matrix $(Q^T + Q) - \alpha M^T HM \succeq 0$. The assertion (5.18) follows immediately. $\square$

Note that it follows from $G \succ 0$ and Theorem 2 there is a constant $c_0 > 0$ such that

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - c_0\|M(v^k - \tilde{v}^k)\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.19}$$

Now, with (5.19) and (5.18), we can establish the worst-case $O(1/t)$ convergence rate in a nonergodic sense for the prototype algorithm (4.2).

**Theorem 6** *Let $\{v^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prototype algorithm (4.2) under the Convergence Condition. For any integer $t > 0$, we have*

$$\|M(v^t - \tilde{v}^t)\|_H^2 \le \frac{1}{(t+1)c_0}\|v^0 - v^*\|_H^2. \tag{5.20}$$

**Proof**. First, it follows from (5.19) that

$$\textstyle\sum_{k=0}^{\infty} c_0\|M(v^k - \tilde{v}^k)\|_H^2 \le \|v^0 - v^*\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{5.21}$$

According to Theorem 5, the sequence $\{\|M(v^k - \tilde{v}^k)\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(t+1)\|M(v^t - \tilde{v}^t)\|_H^2 \le \textstyle\sum_{k=0}^{t}\|M(v^k - \tilde{v}^k)\|_H^2. \tag{5.22}$$

The assertion (5.20) follows from (5.21) and (5.22) immediately. $\square$

# 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] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming, *IMA J. Numerical*

*Analysis* **31**(2015), 394-426.

[10] 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.

[11] 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.

[12] 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.

[13] B.S. He and X.M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, Numerische Mathematik, 130 (2015) 567-577.

[14] 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.

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

[16] 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.