凸优化和单调变分不等式的收缩算法

第十三讲: 定制 PPA 意义下的交替方向法及其线性化方法

Alternating direction method of multipliers in sense of customized PPA and its linearized version

南京大学数学系　何炳生

hebma@nju.edu.cn

The context of this lecture is based on the manuscript [2]

1 Structured constrained convex optimization

We consider the following structured constrained convex optimization problem

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

where $\theta_1(x): \Re^{n_1}\rightarrow\Re,\ \theta_2(y): \Re^{n_2}\rightarrow\Re$ are convex functions (but not necessary smooth), $A\in\Re^{m\times n_1}$, $B\in\Re^{m\times n_2}$ and $b\in\Re^m$, $\mathcal{X}\subset\Re^{n_1}$, $\mathcal{Y}\subset\Re^{n_2}$ are given closed convex sets.

The task of solving the problem (1.1) is to find an $(x^*,y^*,\lambda^*)\in\Omega$, such that

$$\left\{\begin{array}{l}\theta_1(x)-\theta_1(x^*)+(x-x^*)^T(-A^T\lambda^*)\ge 0,\\ \theta_2(y)-\theta_2(y^*)+(y-y^*)^T(-B^T\lambda^*)\ge 0, \quad \forall\,(x,y,\lambda)\in\Omega,\\ (\lambda-\lambda^*)^T(Ax^*+By^*-b)\ge 0,\end{array}\right. \quad (1.2)$$

where

$$\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m.$$

By denoting

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

and

$$\theta(u) = \theta_1(x) + \theta_2(y),$$

the first order optimal condition (1.2) can be written in a compact form such as

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \; \forall\, w \in \Omega. \quad (1.3)$$

Note that the mapping F is monotone. We use Ω^* to denote the solution set of the variational inequality (1.3). For convenience we use the notations

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

Applied ADMM to the structure VI $(y^k, \lambda^k) \Rightarrow (y^{k+1}, \lambda^{k+1})$

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

$$\tilde{x}^k = \text{Argmin}\left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ +\frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \quad (1.4a)$$

Use λ^k and the obtained $\tilde{x}^k$, $\tilde{y}^k$ is the solution of the following problem

$$\tilde{y}^k = \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(A\tilde{x}^k + By - b) \\ +\frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \quad (1.4b)$$

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b). \quad (1.4c)$$

The sub-problems (1.4a) and (1.4b) are separately solved.

Classical Alternating Direction Method of Multipliers: $v^{k+1} = \tilde{v}^k$.

2 ADMM based customized PPA

The k-th iteration of the proposed Alternating Direction Method of Multipliers in this section is also from a pair of (y^k, λ^k) to a new pair of (y^{k+1}, λ^{k+1}). In the prediction step, we generate a $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ which satisfies

$$\tilde{w}^k \in \Omega, \;\; \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T \{F(\tilde{w}^k) + Q(\tilde{v}^k - v^k)\} \geq 0, \;\; \forall\, w \in \Omega, \quad (2.1)$$

where Q is a 3×2 block matrix whose first low is zero, and the rest sub-matrix is symmetric and positive semi-definite. In details, the matrices Q and M have the following forms

$$Q = \begin{pmatrix} 0 & 0 \\ \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \qquad (2.2)$$

Note that the matrix H is symmetric and positive semidefinite. If we replace $Q(\tilde{v}^k - v^k)$ by $G(\tilde{w}^k - w^k)$ with a symmetric positive definite matrix G, then (2.1) becomes a sub-problem of the proximal point algorithm. Thus, the method in this lecture is called the ADMM-based customized PPA or Alternating direction method in the sense of customized PPA.

2.1 Motivation

In the classical ADMM, x is the intermediate variable. For given (y^k, λ^k), we denote the the minimizer of the augmented Lagrangian function by $\tilde{x}^k$, *i. e.,*

$$\tilde{x}^k = \text{Argmin}\{\theta_1(x) - (\lambda^k)^T(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}. \tag{2.3}$$

Thus, we have $\tilde{x}^k \in \mathcal{X}$ and

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + \beta A^T\big(A\tilde{x}^k + By^k - b\big)\} \geq 0, \ \forall\, x \in \mathcal{X}.$$

If we write the above variational inequality as

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\} \geq 0, \ \forall\, x \in \mathcal{X}, \tag{2.4}$$

it implies that

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \tag{2.5}$$

According to the above definition, for any $\tilde{y}^k \in \mathcal{Y}$, we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \qquad (2.6)$$

Combining (2.4) and (2.6) together, we get $(\tilde{x}^k, \tilde{\lambda}^k) \in \mathcal{X} \times \Re^m$,

$$\theta_1(x) - \theta_1(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -B & \frac{1}{\beta} \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \right\} \ge 0, \qquad (2.7)$$

for all $(x, \lambda) \in \mathcal{X} \times \Re^m$. In order to get $\tilde{w}^k \in \Omega$, such that

$$\begin{pmatrix} \big(\theta_1(x) - \theta_1(\tilde{x}^k)\big) + \\ \big(\theta_2(y) - \theta_2(\tilde{y}^k)\big) \end{pmatrix} + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \beta B^T B & -B^T \\ -B & \frac{1}{\beta} \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \right\} \ge 0, \ \forall w \in \Omega, \qquad (2.8)$$

we need only to find $\tilde{y}^k \in \mathcal{Y}$, such that

$$\tilde{y}^k \in \mathcal{Y}, \quad \big(\theta_2(y) - \theta_2(\tilde{y}^k)\big) + (y - \tilde{y}^k)^T \\ \big\{-B^T\tilde{\lambda}^k + B^T\big(\beta B(\tilde{y}^k - y^k) - (\tilde{\lambda}^k - \lambda^k)\big)\big\} \geq 0, \ \forall\, y \in \mathcal{Y}. \quad (2.9)$$

By using (2.5), we have

$$\beta B(\tilde{y}^k - y^k) - (\tilde{\lambda}^k - \lambda^k) = \beta\big(A\tilde{x}^k + B\tilde{y}^k - b\big).$$

Thus, the variational inequality (2.9) is

$$\big(\theta_2(y) - \theta_2(\tilde{y}^k)\big) + (y - \tilde{y}^k)^T\big\{-B^T\tilde{\lambda}^k + \beta B^T\big(A\tilde{x}^k + B\tilde{y}^k - b\big)\big\} \geq 0, \ \forall\, y \in \mathcal{Y}.$$

For given $\tilde{x}^k$ and the defined $\tilde{\lambda}^k$ in (2.5), such a $\tilde{y}^k$ can be obtained via solving the following convex optimization problem:

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|A\tilde{x}^k + By - b - \tfrac{1}{\beta}\tilde{\lambda}^k\|^2 \mid y \in \mathcal{Y}\}. \quad (2.10)$$

The above analysis guides us to construct the ADMM based customized PPA.

2.2 The proposed ADMM based customized PPA

From given $v^k = (y^k, \lambda^k)$, the prediction step produces $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$.

The prediction step:

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

$$\tilde{x}^k = \text{Argmin}\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k - b - \frac{1}{\beta}\lambda^k\|^2 \mid x \in \mathcal{X}\} \quad (2.11a)$$

2. Set the multipliers by

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \quad (2.11b)$$

3. Finally, use the obtained $\tilde{x}^k$ and $\tilde{\lambda}^k$, find $\tilde{y}^k$ by

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \frac{\beta}{2}\|A\tilde{x}^k + By - b - \frac{1}{\beta}\tilde{\lambda}^k\|^2 \mid y \in \mathcal{Y}\} \quad (2.11c)$$

In the ADMM view of point, we generate the predictor in the order

$$\tilde{x}^k, \quad \tilde{\lambda}^k \quad \text{and} \quad \tilde{y}^k.$$

As illustrated in the motivation, we get (2.8). This variational inequality can be written in the form of

$$\tilde{w}^k \in \Omega, \ (w - \tilde{w}^k)^T \{F(\tilde{w}^k) + Q(\tilde{v}^k - v^k)\} \geq 0, \ \forall\, w \in \Omega, \qquad (2.12)$$

where Q is just the same matrix defined in (2.2). The above variational inequality is essential in the unified framework of the contraction methods.

The correction step: Update the new iterate v^{k+1} by

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k), \quad \gamma \in (0, 2). \qquad (2.13)$$

To get the new iterate v^{k+1}, this method does not need to calculate the step size.

2.3 Convergence of the ADMM in sense of customized PPA

Based on the analysis in the last subsection, we have the following lemma.

Lemma 2.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (2.11) *from the given*

$v^k = (y^k, \lambda^k)$. *Then, we have*

$$(\tilde{w}^k - w^*)^T Q(v^k - \tilde{v}^k) \geq 0, \ \forall w^* \in \Omega^*, \tag{2.14}$$

where the matrix Q *is defined in* (2.2).

Proof. Setting $(x, y, \lambda) = (x^*, y^*, \lambda^*)$ in (2.8), we obtain

$$(\tilde{w}^k - w^*)^T Q(v^k - \tilde{v}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \tag{2.15}$$

Since F is monotone and $\tilde{w}^k \in \Omega$, it follows that

$$\begin{aligned} & \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \\ \geq \quad & \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \geq 0. \end{aligned}$$

The last inequality is due to $\tilde{w}^k \in \Omega$ and $w^* \in \Omega^*$ (see (1.3)). Therefore, the right hand side of (2.15) is non-negative and the lemma is proved. $\square$

Lemma 2.2 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (2.11) *from the given*

$v^k = (y^k, \lambda^k)$*. Then, we have*

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_H^2, \; \forall \, v^* \in \mathcal{V}^*, \qquad (2.16)$$

where H *is defined in* (2.2)*.*

Proof. Recall the matrices Q and H in (2.2). It follows from (2.14) that

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \geq 0, \; \forall \, v^* \in \mathcal{V}^*.$$

Assertion (2.16) follows from the last inequality directly. □

The matrix H is symmetric and positive semi-definite. We still use $\|v - \tilde{v}\|_H$ to denote that

$$\|v - \tilde{v}\|_H = \sqrt{(v - \tilde{v})^T H(v - \tilde{v})}.$$

If $\|v^k - \tilde{v}^k\|_H^2 = 0$, because H is symmetric and positive semi-definite, we have $H(v^k - \tilde{v}^k) = 0$. In this case, $\tilde{w}^k$ is a solution of the variational inequality (see (1.2) and (2.8)). Thus, we can take $\|v^k - \tilde{v}^k\|_H^2 \leq \epsilon$ as the stopping criterium in the iteration process.

Theorem 2.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (2.11) *from the given* $v^k = (y^k, \lambda^k)$ *and the new iterate* v^{k+1} *be given by* (2.13). *Then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \quad (2.17)$$

Proof. By a simple manipulation, we obtain

$$\begin{aligned}
&\|v^{k+1} - v^*\|_H^2 \\
&\quad \overset{(2.16)}{=} \quad \|(v^k - v^*) - \gamma(v^k - \tilde{v}^k)\|_H^2 \\
&\quad = \quad \|v^k - v^*\|_H^2 - 2\gamma(v^k - v^*)^T H(v^k - \tilde{v}^k) + \gamma^2\|v^k - \tilde{v}^k\|_H^2 \\
&\quad \overset{(2.13)}{\le} \quad \|v^k - v^*\|_H^2 - 2\gamma\|v^k - \tilde{v}^k\|_H^2 + \gamma^2\|v^k - \tilde{v}^k\|_H^2 \\
&\quad = \quad \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_H^2.
\end{aligned}$$

This is true for any $v^* \in \mathcal{V}^*$ and the theorem is proved. $\square$

The inequality (2.17) is essential for the convergence of the proposed alternating direction method. The detailed convergence proof can be found in [2]. For the convergence rate of the customized PPA, the reader are refereed to [6].

2.4 Ensure the matrix H to be positive definite

In the ADMM based customized PPA (2.11), the subproblem (2.11c) can be written as

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|By - p^k\|^2 \,|\, y \in \mathcal{Y}\}, \tag{2.18}$$

where

$$p^k = b + \tfrac{1}{\beta}\tilde{\lambda}^k - A\tilde{x}^k.$$

If we add an additional term $\frac{\delta\beta}{2}\|B(y - y^k)\|^2$ (with any small $\delta > 0$) to the objective function of the subproblem (2.11c), we will get $\tilde{y}^k$ via

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|By - p^k\|^2 + \tfrac{\delta\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}.$$

By a manipulation, the solution point of the above subproblem is obtained via

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{(1+\delta)\beta}{2}\|By - q^k\|^2 \,|\, y \in \mathcal{Y}\}, \tag{2.19}$$

where

$$q^k = \tfrac{1}{1+\delta}(p^k + \delta B y^k).$$

In this way, the matrix Q in (2.12) will be modified to

$$Q = \begin{pmatrix} 0 & 0 \\ (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix},$$

and the related matrix H in (2.2) becomes

$$\begin{aligned} H &= \begin{pmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I_m \end{pmatrix} \begin{pmatrix} (1+\delta) I & -I \\ -I & I_m \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I_m \end{pmatrix}. \end{aligned} \quad (2.20)$$

Thus, for any $\delta > 0$, H is positive definite when B is a full rank matrix. In other words, instead of (2.18), using (2.19) to get $\tilde{y}^k$, it will ensure the positivity of H theoretically. However, in practical computation, it works still well by using $\delta = 0$.

3 Application and Numerical Experiments

3.1 Applications to least-squares problems

We consider the following problem:

$$\min\{\frac{1}{2}\|X - C\|_F^2 \mid X \in S_+^n \cap S_B\}, \tag{3.1}$$

where

$$S_+^n = \{H \in R^{n\times n} \mid H^T = H, \; H \succeq 0\}.$$

and

$$S_B = \{H \in R^{n\times n} \mid H^T = H, \; H_L \le H \le H_U\}.$$

Use the following MATLAB Code to produce the matrices C, H_L and H_U

```
rand('state',0);   C=rand(n,n);      C=(C'+C)-ones(n,n)+eye(n);
%% C is symmetric and C_{ij} is in (-1,1), C_{jj} is in (0,2)
HU=ones(n)*0.1;  HL=-HU; for i=1:n HU(i,i)=1; HL(i,i)=1; end;
```

The problem is converted to the following equivalent one:

$$\begin{array}{rl} \min & \frac{1}{2}\|X-C\|^2 + \frac{1}{2}\|Y-C\|^2 \\ \text{s.t} & X - Y = 0, \\ & X \in S^n_+,\ Y \in S_B. \end{array} \qquad (3.2)$$

The basic sub-problems in the ADMM based customized PPA

- For fixed Y^k and Z^k,

$$\tilde{X}^k = \text{Argmin}\{\frac{1}{2}\|X-C\|_F^2 - \mathbf{Tr}(Z^k X) + \frac{\beta}{2}\|X - Y^k\|_F^2 \mid X \in S^n_+\}$$

- Set $\tilde{Z}^k$ by

$$\tilde{Z}^k = Z^k - \beta(\tilde{X}^k - Y^k).$$

- With fixed $\tilde{X}^k$ and $\tilde{Z}^k$,

$$\tilde{Y}^k = \text{Argmin}\{\frac{1}{2}\|Y-C\|_F^2 + \mathbf{Tr}(\tilde{Z}^k Y) + \frac{\beta}{2}\|\tilde{X}^k - Y\|_F^2 \mid Y \in S_B\}$$

$\tilde{X}^k$ can be directly obtained via

$$\tilde{X}^k = P_{S_+^n}\{\frac{1}{1+\beta}(\beta Y^k + Z^k + C)\}. \tag{3.3}$$

$$P_{S_+^n}(A) = U\Lambda^+U^T, \qquad [U,\Lambda] = \text{eig}(A), \qquad \Lambda^+ = \max(\Lambda, 0).$$

Similarly, $\tilde{Y}^k$ in is given by

$$\tilde{Y}^k = P_{S_B}\{\frac{1}{1+\beta}(\beta \tilde{X}^k - \tilde{Z}^k + C)\}. \tag{3.4}$$

$$S_B = \{H \,|\, H_L \le H \le H_U\}, \qquad P_{S_B}(A) = \min(\max(H_L, A), H_U)$$

The most time consuming calculation is $[U,\Lambda] = \textbf{eig}(A)$, $9n^3$

MATLAB Code – An iteration of the classical ADMM

```
Y0= Y;                          Z0    = Z;             k = k+1;
X = (Y0*beta+Z0+C)/(1+beta);   [V,D] = eig(X);         D = max(0,D);
X = (V*D)*V';
Y = min(max((X*beta-Z0+C)/(1+beta),HL),HU);
Z = Z0-(X-Y)*beta;
```

MATLAB Code – An iteration of the new order ADMM

```
Y0= Y;                          Z0    = Z;             k = k+1;
X = (Y0*beta+Z0+C)/(1+beta);   [V,D] = eig(X);         D = max(0,D);
X = (V*D)*V';                   Z     = Z0-(X-Y0)*beta;;
Y = min(max((X*beta-Z+C)/(1+beta),HL),HU);
```

MATLAB Code – An iteration of the extended ADMM

```
Y0= Y;                          Z0    = Z;             k = k+1;
X = (Y0*beta+Z0+C)/(1+beta);   [V,D] = eig(X);         D = max(0,D);
X = (V*D)*V';                   Z     = Z0-(X-Y0)*beta;;
Y = min(max((X*beta-Z+C)/(1+beta),HL),HU);
Y = Y0-(Y0-Y)*1.5;
Z = Z0-(Z0-Z)*1.5;
```

Numerical results for problem (3.1)

C=rand(n,n); C=(C'+C)-ones(n,n) + eye(n)

H_U=ones(n,n)/10; H_L=-ones(n,n)/10; $H_U(jj) = H_L(jj)$=1.

Table 1. Numerical results

$n \times n$ Matrix	Classical ADMM		Customized PPA		Extended C-PPA		
$n =$	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.	β
100	46	1.39	44	1.37	28	0.94	5
200	50	3.07	50	3.05	31	4.41	10
500	48	25.50	49	24.52	32	16.50	10
800	51	110.18	50	107.29	33	72.12	10
1000	51	208.93	52	212.74	34	140.70	10
2000	55	1578.96	55	1579.68	36	1053.87	10

3.2 Applications to image restoration

The mathematical form of the image restoration problem is

$$\min \||\nabla x|\|_1 + \frac{\mu}{2}\|Kx - f\|^2, \tag{3.5}$$

where $\mu > 0$ is trade-off; K is a blur operator and f is observed image.

The equivalent problem:

$$\begin{aligned} &\min \||y|\|_1 + \frac{\mu}{2}\|Kx - f\|^2 \\ &\text{s. t. } \nabla x = y, \end{aligned} \tag{3.6}$$

This is a problem of form (1.1) where $\mathcal{X}$, $\mathcal{Y}$ are full spaces,

$$\theta_1(x) = \frac{\mu}{2}\|Kx - f\|^2,$$

$$\theta_2(y) = \||y|\|_1,$$

$$A = \nabla, \quad B = -I \quad \text{and} \quad b = 0.$$

The augmented Lagrangian function

$$\mathcal{L}_A(x,y,\lambda) = |||y|||_1 + \frac{\mu}{2}\|Kx - f\|^2 - \lambda^T(\nabla x - y) + \frac{\beta}{2}\|\nabla x - y\|^2,$$

where λ is Lagrange multiplier and β is the penalty parameter.

For given (y^k, λ^k), get $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ as follows:

1. $\tilde{x}^k$ is the solution of the following least square problem

$$\tilde{x}^k = \arg\min_x \left\{ \tfrac{\mu}{2}\|Kx - f\|^2 - (\lambda^k)^T(\nabla x - y^k) + \tfrac{\beta}{2}\|\nabla x - y^k\|^2 \right\}.$$

2. Set $\tilde{\lambda}^k$ by

$$\tilde{\lambda}^k = \lambda^k - \beta(\nabla \tilde{x}^k - y^k).$$

3. Finally, with fixed $(\tilde{x}^k, \tilde{\lambda}^k)$, $\tilde{y}^k$ are solutions of

$$\tilde{y}^k = \arg\min_y \left\{ |||y|||_1 - (\tilde{\lambda}^k)^T(\nabla \tilde{x}^k - y) + \tfrac{\beta}{2}\|\nabla \tilde{x}^k - y\|^2 \right\}.$$

Solving the x subproblem for getting $\tilde{x}^k$:

$$(\beta\nabla^T\nabla + \mu K^T K)\tilde{x}^k = \nabla^T(\beta y^k + \lambda^k) + \mu K^T f.$$

- If ∇ and K satisfy some periodic boundary conditions, they can be factored by Fourier transform as $\nabla = \mathcal{F}^{-1}\Lambda_D\mathcal{F}$ and $K = \mathcal{F}^{-1}\Lambda_K\mathcal{F}$.
- If ∇ and K satisfy some reflective boundary conditions, they can be factored by discrete cosine transform as $\nabla = \mathcal{C}^{-1}\Lambda_D\mathcal{C}$ and $K = \mathcal{C}^{-1}\Lambda_K\mathcal{C}$.

Solving the y subproblem for getting $\tilde{y}^k$:

$$\tilde{y}^k = \text{shrink}_{\frac{1}{\beta}}\left(\nabla\tilde{x}^k - \frac{\tilde{\lambda}^k}{\beta}\right),$$

where

$$\text{shrink}_c(v) = v - \min(c, \|v\|)\frac{v}{\|v\|}.$$

Note that

$$\text{shrink}_c(v) = v - P_{B_2^c}(v) \quad \text{where} \quad B_2^c = \{v \in \Re^n | \|v\|_2 \le c\}.$$

MATLAB Code – An iteration of the classical ADMM

```
%%%%%%%%%%%%%%%%%%%%   step 1  x^{k+1}   %%%%%%%%%%%%%%%%%%%%%%
Temp= PTx(beta*v1+lbd11) + PTy(beta*v2+lbd12) + HTx0;
un  = real(ifft2(fft2(Temp)./MDu));
%%%%%%%%%%%%%%%%%%%%   step 2  y^{k+1}   %%%%%%%%%%%%%%%%%%%%%%
dxun= Px(un);
dyun= Py(un);
sk1 = dxun - lbd11/beta;
sk2 = dyun - lbd12/beta;
nsk = sqrt(sk1.^2 + sk2.^2); nsk(nsk==0)=1;
nsk = max(1-1./(beta*nsk),0);
vn1 = sk1.*nsk;
vn2 = sk2.*nsk;
%%%%%%%%%%%%%%%%%%%% update  \lambda     %%%%%%%%%%%%%%%%%%%%%%
lbdn11 = lbd11 - beta*(dxun - vn1);
lbdn12 = lbd12 - beta*(dyun - vn2);
%%%%%%%%%%%%%%%%%%%% New iterative point %%%%%%%%%%%%%%%%%%%%%%
u = un; v1 = vn1; v2 = vn2; lbd11 = lbdn11; lbd12 = lbdn12;
```

MATLAB Code – An iteration of the new order ADMM

```
%%%%%%%%%%%%%%%%%%%%   step 1  x^{k+1}  %%%%%%%%%%%%%%%%%%%%%%
Temp= PTx(beta*v1+lbd11) + PTy(beta*v2+lbd12) + HTx0;
un  = real(ifft2(fft2(Temp)./MDu));
dxun= Px(un);
dyun= Py(un);
%%%%%%%%%%%%%%%%%%%%   update \lambda   %%%%%%%%%%%%%%%%%%%%%%
lbdn11 = lbd11 - beta*(dxun - v1);
lbdn12 = lbd12 - beta*(dyun - v2);
%%%%%%%%%%%%%%%%%%%%   step 2  y^{k+1}  %%%%%%%%%%%%%%%%%%%%%%
sk1 = dxun - lbdn11/beta;
sk2 = dyun - lbdn12/beta;
nsk = sqrt(sk1.^2 + sk2.^2); nsk(nsk==0)=1;
nsk = max(1-1./(beta*nsk),0);
vn1 = sk1.*nsk;
vn2 = sk2.*nsk;
%%%%%%%%%%%%%%%%%%%% New iterative point %%%%%%%%%%%%%%%%%%%%%%
u = un; v1 = vn1; v2 = vn2; lbd11 = lbdn11; lbd12 = lbdn12;
```

MATLAB Code – An iteration of the extended C-PPA

```
%%%%%%%%%%%%%%%%%%%%   step 1  x^{k+1}  %%%%%%%%%%%%%%%%%%%%%
Temp= PTx(beta*v1+lbd11) + PTy(beta*v2+lbd12) + HTx0;
un  = real(ifft2(fft2(Temp)./MDu));
dxun= Px(un);
dyun= Py(un);
%%%%%%%%%%%%%%%%%%%%  update \lambda    %%%%%%%%%%%%%%%%%%%%%
lbdn11 = lbd11 - beta*(dxun - v1);
lbdn12 = lbd12 - beta*(dyun - v2);
%%%%%%%%%%%%%%%%%%%%  step 2  y^{k+1}  %%%%%%%%%%%%%%%%%%%%%
sk1 = dxun - lbd11/beta;
sk2 = dyun - lbd12/beta;
nsk = sqrt(sk1.^2 + sk2.^2); nsk(nsk==0)=1;
nsk = max(1-1./(beta*nsk),0);
vn1 = sk1.*nsk;
vn2 = sk2.*nsk;
%%%%%%%%%%%%%%%%%% New iterative point %%%%%%%%%%%%%%%%%%%%%
u  = un;
v1 = v1 - gamma*(v1-vn1);
v2 = v2 - gamma*(v2-vn2);
lbd11 = lbd11 - gamma*(lbd11-lbdn11);
lbd12 = lbd12 - gamma*(lbd12-lbdn12);
```

Numerical results for image restoration

```
I = double(imread('chart.tiff'))/255; I = double(imread('house.png'))/255;
h = fspecial('disk',7); x0 = imfilter(I,h,'circular')+0.02*randn(size(I));
```

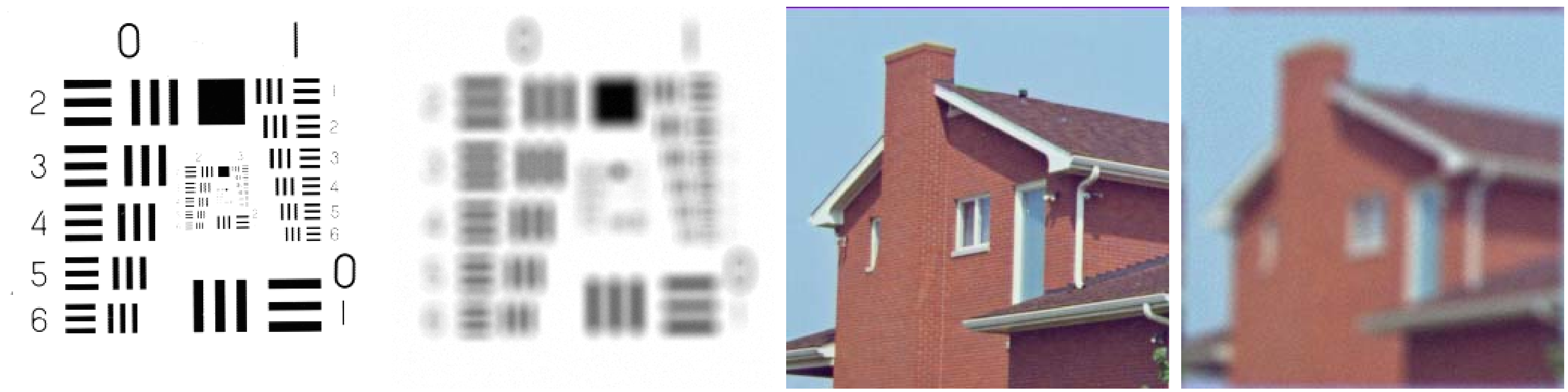

Figure 1: Original and degraded images. Left: Chart. Right: House

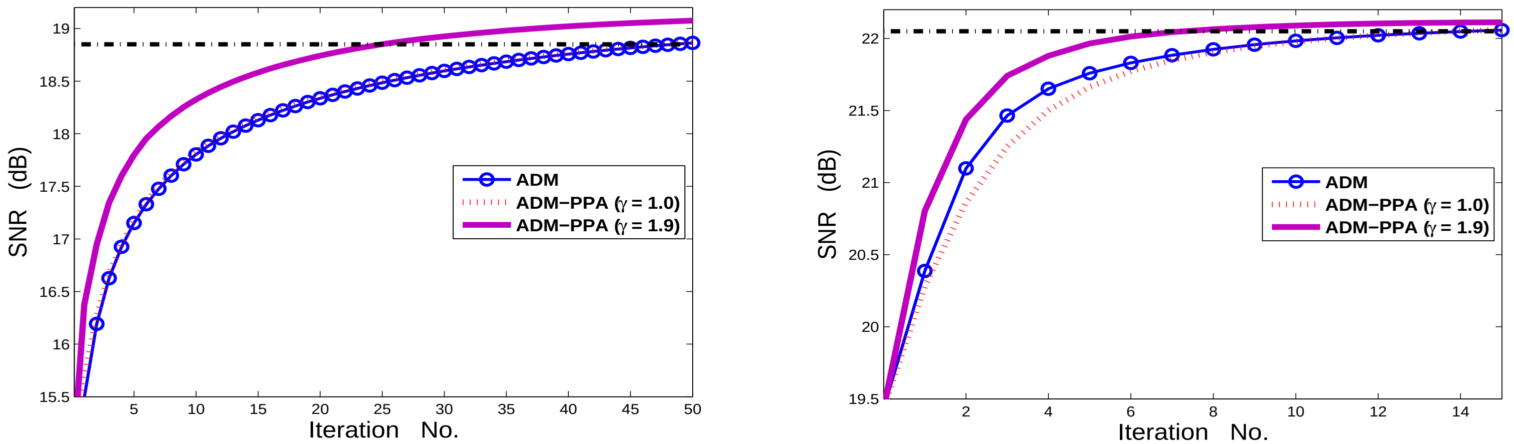

Figure 2: Performances of ADMM and two variants methods on TV-l2. Left: Chart. Right: House.

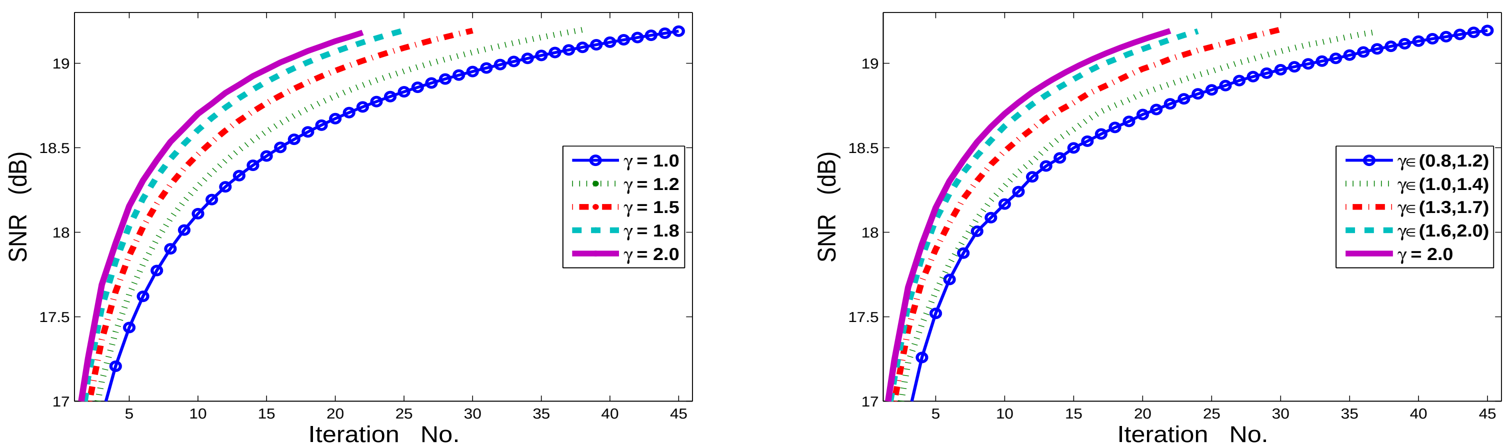

Figure 3: Performances of Algorithm 2 with different values of γ for Chart. Top: fixed γ. Bottom: random generated γ.

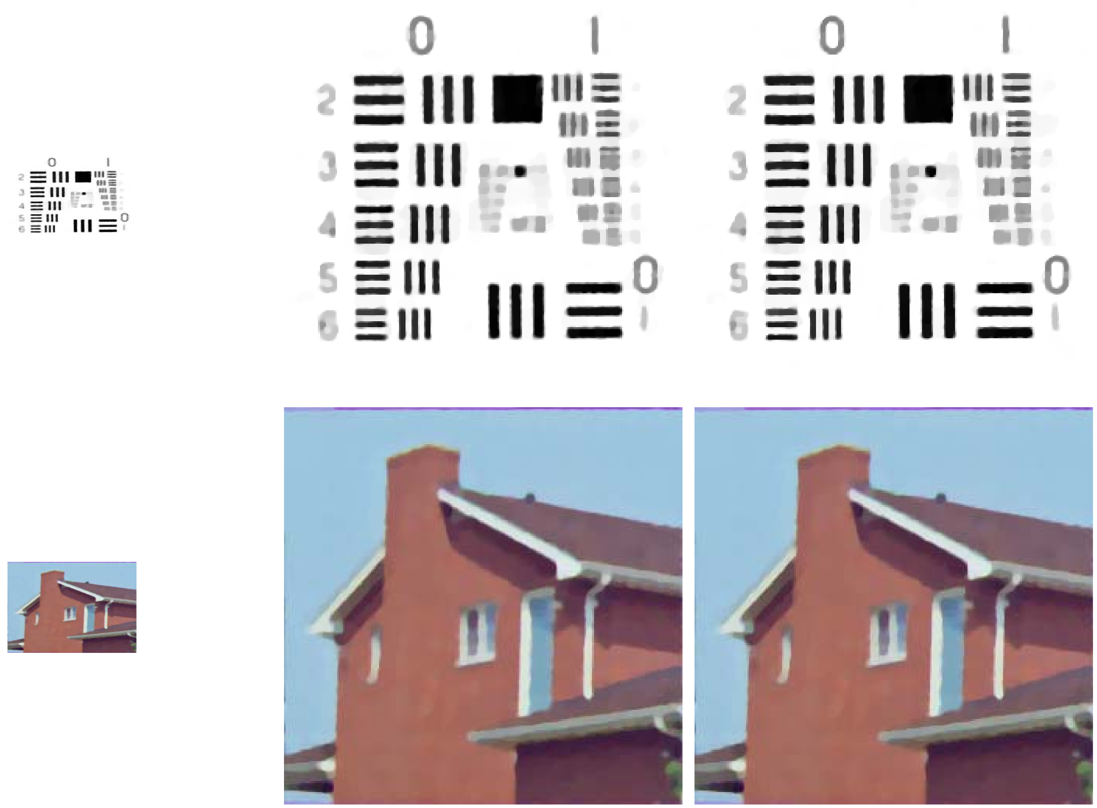

Figure 4: Restorations. From left column to right column: ADMM, new order ADM-M, and the extended new order ADMM.

Table 1: Numerical comparisons of the classical ADMM (ADMM), the customized PPA and the extended customized PPA for TV-l^2 image restoration.

	Chart			House		
	ADMM	C-PPA	E-C-PPA	ADMM	C-PPA	E-C-PPA
		$\gamma = 1$	$\gamma = 1.8$		$\gamma = 1$	$\gamma = 1.8$
Iter	74	74	42	57	56	33
CPU	2.32	2.29	1.29	2.79	2.75	1.63
SNR	19.01	19.01	19.02	22.11	22.11	22.12

SNR $= 20 \log_{10} \frac{\|x\|}{\|x-I\|}$, where x is restoration and I is original image.

It seems that the new order ADMM is so good as the classical one. However, the extended-ADMM converges much faster than the other both ADMMs.

4 Linearized ADMM-based PPA Method

Note that the subproblems (2.11a) and (2.11c) in the last section are equivalent to the problems

$$\tilde{x}^k = \text{Argmin}\big\{\theta_1(x) + \tfrac{\beta}{2}\|(Ax + By^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 \big| x \in \mathcal{X}\big\} \qquad \text{(4.1a)}$$

and

$$\tilde{y}^k = \text{Argmin}\big\{\theta_2(y) + \tfrac{\beta}{2}\|(A\tilde{x}^k + By - b) - \tfrac{1}{\beta}\tilde{\lambda}^k\|^2 \big| y \in \mathcal{Y}\big\} \qquad \text{(4.1b)}$$

respectively. In some structured optimization (1.1), the subproblem (4.1a) is easy because A is usually a scalar matrix. However, to obtain the solution of the subproblem (4.1b) is expensive in the case that B does not have a special form. In this lecture, we suppose that only the solution of the problem

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

has a closed form, and consider to linearize the quadratic function of the subproblem (4.1b) ADMM in sense of the customized PPA.

4.1 Linearized alternating direction method

The prediction step:

1. First, for given (x^k, y^k, λ^k), solving the x subproblem to get $\tilde{x}^k$ by

$$\tilde{x}^k = \text{Argmin}\Big\{\theta_1(x) + \frac{\beta}{2}\|(Ax + By^k - b) - \frac{1}{\beta}\lambda^k\|^2 \Big| x \in \mathcal{X}\Big\} \tag{4.2a}$$

2. Set the new multipliers by

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \tag{4.2b}$$

3. Finally, use the obtained $\tilde{x}^k$ and $\tilde{\lambda}^k$, solving the y subproblem to get $\tilde{y}^k$ by

$$\tilde{y}^k = \text{Argmin}\left(\begin{array}{c} \Big\{\theta_2(y) + \beta y^T B^T (A\tilde{x}^k + By^k - b - \frac{1}{\beta}\tilde{\lambda}^k) \\ +\frac{s}{2}\|y - y^k\|^2 \quad \Big| \; y \in \mathcal{Y}\Big\} \end{array} \right). \tag{4.2c}$$

Request on the parameter s

For given $\beta > 0$, s should satisfy

$$sI - \beta B^T B \succeq 0. \tag{4.3}$$

The correction step: Update the new iterate v^{k+1} by

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k), \quad \gamma \in [1, 2). \tag{4.4}$$

To get the new iterate v^{k+1}, this method does not need to calculate the step size. However, it needs to estimate the max-eigenvalue of $B^T B$, *i. e.*, $\lambda_{\max}(B^T B)$.

4.2 Analysis in the PPA framework

Note that the solution of (4.2a), $\tilde{x}^k$ satisfies

$$\begin{aligned} \tilde{x}^k \in \mathcal{X}, \ \ \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \{-A^T \lambda^k + \beta A^T (A\tilde{x}^k + By^k - b)\} \geq 0, \ \forall\, x \in \mathcal{X}. \end{aligned} \tag{4.5}$$

Substituting $\tilde{\lambda}^k$ (see (4.2b)) in (4.5) (eliminating λ^k), we get

$$\tilde{x}^k \in \mathcal{X}, \ \ \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \big\{ -A^T \tilde{\lambda}^k \big\} \ge 0, \ \forall\, x \in \mathcal{X}. \quad (4.6)$$

The solution of (4.2c), $\tilde{y}^k$ satisfies

$$\begin{aligned} \tilde{y}^k \in \mathcal{Y}, \ \ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \big\{ & -B^T \tilde{\lambda}^k \\ & + \beta B^T \big(A\tilde{x}^k + By^k - b \big) + s(\tilde{y}^k - y^k) \big\} \ge 0, \ \forall\, y \in \mathcal{Y}. \end{aligned} \quad (4.7)$$

Note that $\beta\big(A\tilde{x}^k + By^k - b\big) = (\lambda^k - \tilde{\lambda}^k)$ (see (4.2b)). Substituting it in (4.7), we obtain

$$\begin{aligned} \tilde{y}^k \in \mathcal{Y}, \ \ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \big\{ & -B^T \tilde{\lambda}^k \\ & - B^T(\tilde{\lambda}^k - \lambda^k) + s(\tilde{y}^k - y^k) \big\} \ge 0, \ \forall\, y \in \mathcal{Y}. \end{aligned} \quad (4.8)$$

From (4.2b) we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \quad (4.9)$$

Combining the inequalities (4.6), (4.8) and (4.9), we obtain

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

$$\left. +\begin{pmatrix} 0 & 0 \\ sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k-y^k \\ \tilde{\lambda}^k-\lambda^k \end{pmatrix} \right\} \ge 0, \ \forall\, w\in\Omega. \qquad (4.10)$$

The last variational inequality can be written in form of $\tilde{w}^k\in\Omega$ and

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T\{F(\tilde{w}^k)+Q(\tilde{v}^k-v^k)\}\ge 0, \ \forall\, w\in\Omega, \qquad (4.11)$$

where

$$Q=\begin{pmatrix} 0 & 0 \\ sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix} \qquad (4.12)$$

which is essential in the framework of the PPA contraction methods.

4.3 Convergence of the Linearized ADM-based PPA Method

Based on the analysis in the last subsection, we have the following lemma.

Lemma 4.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (4.2) *from the given* $v^k = (y^k, \lambda^k)$*. Then, we have*

$$(\tilde{v}^k - v^*)^T G(v^k - \tilde{v}^k) \geq 0, \ \forall w^* \in \Omega^*, \tag{4.13}$$

where

$$G = \begin{pmatrix} sI & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{4.14}$$

Proof. Setting $w = w^*$ in (4.10), we get

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

Since F is monotone and $\tilde{w}^k \in \Omega$, it follows that

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

The right hand side of the last inequality is non-negative because $\tilde{w}^k \in \Omega$ and $w^* \in \Omega^*$. the assertion follows directly. $\square$

Lemma 4.2 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (4.2) *from the given* $v^k = (y^k, \lambda^k)$*. Then, we have*

$$(v^k - v^*)^T G(v^k - \tilde{v}^k) \ge \|v^k - \tilde{v}^k\|_G^2, \ \forall\, w^* \in \Omega^*, \qquad (4.15)$$

where G *is defined in* (4.14).

Proof. Assertion (4.15) follows from the last inequality directly. $\square$

Since G is symmetric and positive semi-definite, we have

$$v^k - \tilde{v}^k = 0 \quad \text{or} \quad G(v^k - \tilde{v}^k) = 0,$$

whenever $\|v^k - \tilde{v}^k\|_G^2 = 0$. Therefore, it follows from (4.10) that $\tilde{w}^k$ is a solution of the variational inequality when $\|v^k - \tilde{v}^k\|_G^2 = 0$.

Theorem 4.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (4.2) *from the given* $v^k = (y^k, \lambda^k)$ *and the new iterate* w^{k+1} *be given by* (4.4). *Then we have*

$$\|v^{k+1} - v^*\|_G^2 \le \|v^k - v^*\|_G^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_G^2, \; \forall\, w^* \in \Omega^*. \quad (4.16)$$

Proof. By using (4.4) and (4.15), we obtain

$$\begin{aligned} &\|v^{k+1} - v^*\|_G^2 \\ &\quad \overset{(4.4)}{=} \quad \|(v^k - v^*) - \gamma(v^k - \tilde{v}^k)\|_G^2 \\ &\quad \overset{(4.15)}{\le} \quad \|v^k - v^*\|_G^2 - 2\gamma\|v^k - \tilde{v}^k\|_G^2 + \gamma^2\|v^k - \tilde{v}^k\|_G^2 \\ &\quad = \quad \|v^k - v^*\|_G^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_G^2. \end{aligned}$$

This is true for any $w^* \in \Omega^*$ and the theorem is proved. $\square$

The inequality (4.16) is essential for the convergence of the Linearized alternating direction method. By using (4.4), the result of Theorem 4.1 can be written as

$$\|v^{k+1} - v^*\|_G^2 \le \|v^k - v^*\|_G^2 - \frac{2-\gamma}{\gamma}\|v^k - v^{k+1}\|_G^2, \; \forall\, w^* \in \Omega^*.$$

5 Convergence rate of ADMM-CPPA Method

Lemma 5.1 *Let $\{w^k\}$ be the sequence generated by the customized PPA (4.2) with (4.4). Then, we have*

$$(\tilde{v}^k - \tilde{v}^{k+1})^T G\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \geq 0. \tag{5.1}$$

Proof. Set $w = \tilde{w}^{k+1}$ in (4.11), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T\{F(\tilde{w}^k) + Q(\tilde{v}^k - v^k)\} \geq 0. \tag{5.2}$$

Note that (4.11) is also true for $k := k+1$ and thus we have

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T\{F(\tilde{w}^{k+1}) + Q(\tilde{v}^{k+1} - v^{k+1})\} \geq 0, \ \forall\, w \in \Omega.$$

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

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T\{F(\tilde{w}^{k+1}) + Q(\tilde{v}^{k+1} - v^{k+1})\} \geq 0. \tag{5.3}$$

Adding (5.2) and (5.3) and using the monotonicity of F, we get

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

we obtain (5.1) immediately. □

Lemma 5.2 *Let* $\{w^k\}$ *be the sequence generated by the customized PPA (4.2) with (4.4). Then, we have*

$$\begin{aligned} & (v^k - \tilde{v}^k)^T G\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\ \geq \quad & \frac{1}{\gamma}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2. \end{aligned} \tag{5.4}$$

Proof. Adding the term $\|(v^k - v^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\|_G^2$ to the both sides of (5.1), we obtain

$$\begin{aligned} & (v^k - v^{k+1})^T G\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\ \geq \quad & \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2. \end{aligned} \tag{5.5}$$

Substituting the term $(v^k - v^{k+1})$ in the left hand side of (5.5) by $\gamma(v^k - \tilde{v}^k)$ (see (4.4)), we obtain (5.4) and the lemma is proved. $\square$

Lemma 5.3 *Let* $\{v^k\}$ *be the sequence generated by the customized PPA (4.2) with (4.4). Then, we have*

$$\|v^{k+1} - \tilde{v}^{k+1}\|_G^2 \le \|v^k - \tilde{v}^k\|_G^2. \tag{5.6}$$

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

$$\|a\|_G^2 - \|b\|_G^2 = 2a^T G(a-b) - \|a-b\|_G^2,$$

we obtain

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_G^2 - \|v^{k+1} - \tilde{v}^{k+1}\|_G^2 \\
&\quad = \quad 2(v^k - \tilde{v}^k)^T G\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\
&\qquad\qquad - \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2.
\end{aligned}$$

By using (5.4) to the first term of the right hand side of the last equality, we obtain

$$\|v^k - \tilde{v}^k\|_G^2 - \|v^{k+1} - \tilde{v}^{k+1}\|_G^2 \geq \frac{2-\gamma}{\gamma}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2.$$

The assertion of this lemma is proved. $\square$

Having the assertion (4.16) and Lemma 5.3, we are ready to present the $O(1/t)$ convergence rate of the customized PPA in the residue sense.

Theorem 5.1 *Let $\{w^k\}$ be the sequence generated by the customized PPA (4.2) with (4.4). Then, we have*

$$\|v^k - \tilde{v}^k\|_G^2 \leq \frac{1}{(k+1)\gamma(2-\gamma)}\|v^0 - v^*\|_G^2, \quad \forall w^* \in \Omega^*. \tag{5.7}$$

Proof. First, it follows from (4.16) that

$$\gamma(2-\gamma)\sum_{t=0}^{\infty}\|v^t - \tilde{v}^t\|_G^2 \leq \|v^0 - v^*\|_G^2, \quad \forall\, w^* \in \Omega^*. \tag{5.8}$$

According to Lemma 5.3, the sequence $\{\|v^t - \tilde{v}^t\|_G^2\}$ is non-increasing. Therefore, we have

$$(k+1)\|v^k - \tilde{v}^k\|_G^2 \le \sum_{i=0}^{k} \|v^i - \tilde{v}^i\|_G^2. \qquad (5.9)$$

The assertion of this theorem follows from (5.8) and (5.9) directly. $\square$

The solution set of the variational inequality VI(Ω, F, θ) is convex and closed. Theorem 5.1 indicates that ADMM has $O(1/k)$ iteration convergence rate. Let

$$d = \inf\{\|v^0 - v^*\|_G \,|\, w^* \in \Omega^*\}.$$

For any given $\epsilon > 0$, in order to enforce the error $\|v^k - \tilde{v}^k\|_G^2 \le \epsilon$, according to (5.7), it needs at most $k = \lfloor d^2/\gamma(2-\gamma)\epsilon \rfloor$ iterations.

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, **3**, 1-122, 2010.

[2] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A relaxed customized proximal point algorithm for separable convex programming. http://www.optimization-online.org/DB_HTML/2011/08/3141.html

[3] J. Eckstein and M. C. Ferris, Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control, INFORMS J. Comput., **10**, 218-235, 1998.

[4] D. Gabay, Applications of the Method of Multipliers to Variational Inequalities, in *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, 299-331, 1983.

[5] E. G. Gol'shtein and N. V. Tret'yakov, Modified Lagrangian in convex programming and their generalizations, Math. Program. Studies, **10**, 86-97, 1979.

[6] G.Y. Gu, B.S. He and X.M. Yuan: Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a uniform approach. http://www.optimization-online.org/DB_HTML/ 2011/10/ 3220.html

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

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

[9] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim., **14**, 877-898, 1976.