均困平衡的增广拉格朗日乘子法

Balanced ALM for convex optimization

B.S. He and X.M. Yuan, Balanced Augmented Lagrangian Method for Convex Optimization. arXiv:2108.08554

何 炳 生

南京大学数学系

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

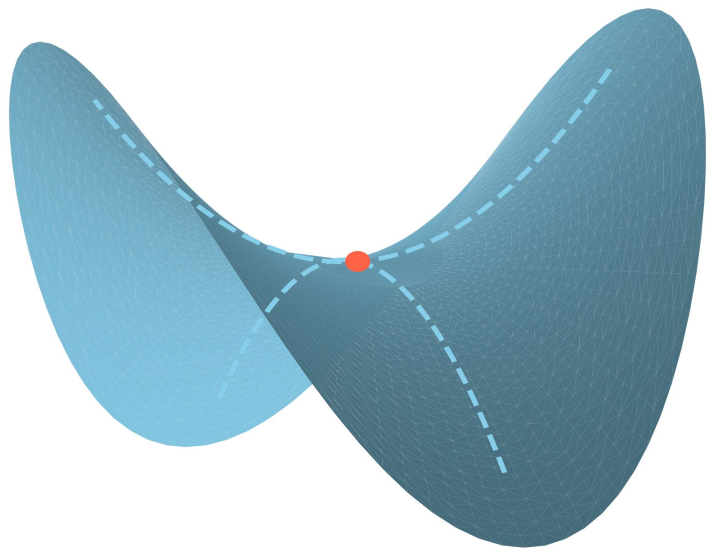
1 线性约束的凸优化问题及其求解方法

我们以线性等式约束的凸优化问题

$$\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\} \tag{1.1}$$

为例介绍均困平衡的增广拉格朗日乘子法. 问题 (1.1) 的 Lagrange 函数是

$$L(x,\lambda) = \theta(x) - \lambda^T(Ax - b), \qquad (x,\lambda) \in \mathcal{X} \times \Re^m. \tag{1.2}$$



Saddle Point of the Lagrange Function

2 增广拉格朗日乘子法 (ALM)

我们从增广拉格朗日乘子法 [6, 9, 10] (ALM) 谈起。许多有效算法（例如交替方向法ADMM）起源于ALM。用 ALM 求解问题 (1.1), 迭代公式是

$$\begin{cases} x^{k+1} \in \arg\min\{L(x,\lambda^k) + \frac{r}{2}\|Ax-b\|^2 \mid x \in \mathcal{X}\}, & \text{(2.1a)} \\ \lambda^{k+1} = \operatorname{argmax}\{L(x^{k+1},\lambda) - \frac{1}{2r}\|\lambda-\lambda^k\|^2 \mid \lambda \in \Re^m\}. & \text{(2.1b)} \end{cases}$$

ALM 的 x-子问题 (2.1a) 中的 $r>0$ 是等式约束 $Ax-b=0$ 的罚参数。把子问题目标函数里的常数项去掉, ALM 可以通过

$$\begin{cases} x^{k+1} \in \operatorname{argmin}\{\theta(x) + \frac{r}{2}\|Ax-[b+\frac{1}{r}\lambda^k]\|^2 \mid x \in \mathcal{X}\}, & \text{(2.2a)} \\ \lambda^{k+1} = \lambda^k - r(Ax^{k+1}-b), & \text{(2.2b)} \end{cases}$$

实现。ALM 的 x-子问题 (2.2a) 中的形如 $\frac{r}{2}\|Ax-p^k\|^2$ 的二次项和原问题的目标函数加在一起, 会给求解带来一定的困难。ALM 通过 (2.2b) 更新对偶变量 λ^{k+1} 却非常简单！

这就是我们考虑均困平衡 ALM 的出发点.

3 按需定制的邻近点算法 (CP-PPA)

求解 (1.1) 的原始–对偶迭代方法中, 简化 x-子问题的方法有按需定制的 PPA 方法 [1, 2, 4], 我们称它为 CP-PPA 算法。这个方法先是由 Chambolle and Pock 在文献 [1] 中提出, 他们在（也只是在）遍历意义下证明了收敛性。我们在文献 [4] 中用变分不等式邻近点算法的观点证明了该方法点列意义下的收敛性。Chambolle and Pock 在他们后来发表在 MP 的文章 [2] 中说到, 他们的后继研究是充分利用了我们在文献 [4] 中提出的 PPA 解释。CP-PPA 的迭代公式是:

$$\left\{\begin{array}{ll} x^{k+1} = \operatorname{argmin}\{L(x,\lambda^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(3.1a)} \\ \lambda^{k+1} = \operatorname{argmax}\{L([2x^{k+1} - x^k],\lambda) - \frac{s}{2}\|\lambda - \lambda^k\|^2\}. & \text{(3.1b)} \end{array}\right.$$

这里的 $r, s > 0$, 是迫使新的 x^{k+1} 和 λ^{k+1} 分别离原来的 x^k 和 λ^k 不要太远. 忽略常数项, 迭代公式可以简化成

$$\left\{\begin{array}{ll} x^{k+1} = \operatorname{argmin}\big\{\theta(x) + \frac{r}{2}\big\|x - \big[x^k + \frac{1}{r}A^T\lambda^k\big]\big\|^2 \,\big|\, x \in \mathcal{X}\big\}, & \text{(3.2a)} \\ \lambda^{k+1} = \lambda^k - \frac{1}{s}\big(A[2x^{k+1} - x^k] - b\big). & \text{(3.2b)} \end{array}\right.$$

CP-PPA 方法中 x-子问题 (3.2a) 的二次项为 $\frac{r}{2}\|x-q^k\|^2$, 比起 (2.2a) 中的二次项 $\frac{r}{2}\|Ax-p^k\|^2$ 简单，因此 x-子问题的求解也比较容易。然而, 为了保证收敛, 要求参数

$$rs > \|A^T A\|. \tag{3.3}$$

从 (3.1) 可以看出, 过大的 r 和 s, 分别迫使 x^{k+1} 离 x^k, λ^{k+1} 离 λ^k 很近, 相当于迭代中步长受限, 影响整体收敛速度。CP-PPA 方法在图像处理中的确取得了比较好的效果, 是得益于其中用到的全变差矩阵的 $\|A^T A\| \le 8$。

从迭代公式 (3.1) 可以看出, CP-PPA 也属于原始–对偶混合梯度法 PDHG。

4 均困平衡的增广拉格朗日乘子法 (Balanced ALM)

均困平衡的增广拉格朗日乘子法 — (Balanced ALM) 是把 ALM 的 x-子问题 (2.2a) 简化成 CP-PPA 的 x-子问题 (3.2a) 同样的形式, 而新的对偶变量 λ^{k+1}

的获取要通过求解一个线性方程组实现。Balanced ALM 的迭代公式是

$$\left\{\begin{array}{ll} x^{k+1} = \arg\min\big\{L(x,\lambda^k) + \frac{r}{2}\|x - x^k\|^2 \;\big|\; x \in \mathcal{X}\big\}, & \text{(4.1a)} \\ \lambda^{k+1} = \text{argmax}\big\{L([2x^{k+1} - x^k], \lambda) - \frac{1}{2}\|\lambda - \lambda^k\|^2_{(\frac{1}{r}AA^T + \delta I_m)}\big\}. & \text{(4.1b)} \end{array}\right.$$

这里的 $r > 0$ 和 $\delta > 0$ 是任意的, 其中的 $\delta > 0$ 只是为了保证某种正定性。因此, 就像 ALM 方法一样, Balanced ALM 中只有一个参数 r 需要根据具体问题而选择。 Balanced ALM 方法迭代公式的等价形式是

$$\left\{\begin{array}{ll} x^{k+1} = \text{argmin}\big\{\theta(x) + \frac{r}{2}\big\|x - \big[x^k + \frac{1}{r}A^T\lambda^k\big]\big\|^2 \;\big|\; x \in \mathcal{X}\big\}. & \text{(4.2a)} \\ \lambda^{k+1} = \arg\min\big\{\frac{1}{2}\|\lambda - \lambda^k\|^2_{(\frac{1}{r}AA^T + \delta I_m)} + \lambda^T\big(A[2x^{k+1} - x^k] - b\big)\big\}. & \text{(4.2b)} \end{array}\right.$$

这里的 x-子问题 (4.2a) 和 CP-PPA 的 x-子问题 (3.2a) 形式完全一样。(4.2b) 中的 λ^{k+1} 是线性方程组

$$\Big(\frac{1}{r}AA^T + \delta I_m\Big)(\lambda - \lambda^k) + \big(A[2x^{k+1} - x^k] - b\big) = 0 \qquad \text{(4.3)}$$

的解. 注意到上述线性方程组的系数矩阵是对称正定的, 在整个迭代过程中我们只要做一次正定矩阵

$$\left(\frac{1}{r}AA^T+\delta I_m\right)$$

的 Cholesky 分解。矩阵计算 [3] 里有非常成熟的方法求解这类线性方程组。

线性方程组 (4.3) 告诉我们, 对偶变量的更新采用的是 Levenberg-Marquardt 方法 [7, 8]。 Levenberg-Marquardt 方法的优点做优化计算的都是很清楚的。

从 (4.1b) 和 (3.1b) 的差别可以看出, 由于采用了 Levenberg-Marquardt 方向, (4.1) 已经不再属于原始-对偶混合梯度法。

世界上没有一个算法是对所有的问题最好的.
Balanced ALM 的优越性还是显而易见的.

有兴趣的读者可以参阅我们在 arXiv 上的文章 [5] .
或者看后面 §5 的分析和 §6 的证明, 都相当简短.

5 Mathematical Background for Convergence

5.1 Variational Inequality Formulation

Recall the Lagrangian function of the problem (1.1)

$$L(x,\lambda) = \theta(x) - \lambda^T(Ax-b), \qquad (x,\lambda) \in \mathcal{X} \times \Re^m. \tag{5.1}$$

A pair of $(x^*,\lambda^*) \in \mathcal{X} \times \Re^m$ is called a saddle point of the Lagrangian function, if

$$L_{\lambda\in\Re^m}(x^*,\lambda) \le L(x^*,\lambda^*) \le L_{x\in\mathcal{X}}(x,\lambda^*).$$

The above inequalities can be written as

$$\left\{ \begin{array}{ll} x^* \in \mathcal{X}, \quad L(x,\lambda^*) - L(x^*,\lambda^*) \ge 0, \ \ \forall\, x \in \mathcal{X}, & \text{(5.2a)} \\ \lambda^* \in \Re^m, \ L(x^*,\lambda^*) - L(x^*,\lambda) \ge 0, \ \ \forall\, \lambda \in \Re^m. & \text{(5.2b)} \end{array} \right.$$

According to the definition of $L(x,\lambda)$ (see(5.1)),

$$L(x,\lambda^*) - L(x^*,\lambda^*) = \theta(x) - \theta(x^*) + (x-x^*)^T(-A^T\lambda^*),$$

it follows from (5.2a) that

$$x^* \in \mathcal{X}, \ \ \theta(x) - \theta(x^*) + (x-x^*)^T(-A^T\lambda^*) \ge 0, \ \ \forall\, x \in \mathcal{X}. \tag{5.3}$$

Similarly, for (5.2b), since

$$L(x^*, \lambda^*) - L(x^*, \lambda) = (\lambda - \lambda^*)^T (Ax^* - b),$$

we have

$$\lambda^* \in \Re^m, \quad (\lambda - \lambda^*)^T (Ax^* - b) \geq 0, \quad \forall\, \lambda \in \Re^m. \tag{5.4}$$

Notice that the above expression is equivalent to $Ax^* = b$.

Writing (5.3) and (5.4) together, we get the following variational inequality:

$$\begin{cases} x^* \in \mathcal{X}, & \theta(x) - \theta(x^*) + (x - x^*)^T (-A^T \lambda^*) \geq 0, \quad \forall\, x \in \mathcal{X}, \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T (Ax^* - b) \geq 0, \quad \forall\, \lambda \in \Re^m. \end{cases}$$

A more compact form of the saddle-points is the following variational inequality:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{5.5a}$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T \lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Re^m. \tag{5.5b}$$

Because F is a affine operator and

$$F(w) = \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ b \end{pmatrix},$$

the matrix is skew-symmetric, we have

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0.$$

我们把线性约束的凸优化问题 (1.1), 转换成了混合变分不等式 (5.5).

Lemma 1 *Let $\mathcal{X} \subset \Re^n$ be a closed convex set, $\theta(x)$ and $f(x)$ be convex functions. If f is differentiable and the solution set of the minimization problem*

$$\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$$

is nonempty, then it holds that

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{5.6a}$$

if and only if

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \mathcal{X}. \tag{5.6b}$$

5.2 PPA for Variational Inequalities

The variational inequality:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega. \tag{5.7}$$

PPA for VI (5.7) in H-norm For given w^k and $H \succ 0$, find w^{k+1},

$$w^{k+1} \in \Omega, \quad \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \ge (v - v^{k+1})^T H(v^k - v^{k+1}), \quad \forall\, w \in \Omega. \tag{5.8}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (5.7).

In the right hand side of (5.8), it can be $v = w = (x, \lambda)$, it is also possible that $v = \lambda$.

Theorem 1 *Let w^{k+1} be generated by* (5.8) *with given w^k. Then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall w^* \in \Omega^*, \tag{5.9}$$

Proof. Setting $w = w^*$ in (5.8), we obtain

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \ge \theta(x^{k+1}) - \theta(x^*) + (w^{k+1} - w^*)^T F(w^{k+1}).$$

By using $(w^{k+1}-w^*)^T F(w^{k+1}) = (w^{k+1}-w^*)^T F(w^*)$ and

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

we get

$$(v^{k+1}-v^*)^T H(v^k - v^{k+1}) \geq 0. \tag{5.10}$$

Applying the identity

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

to the left-hand side of (5.10) with $a = w^k - w^*$ and $b = w^{k+1} - w^*$, we thus obtain

$$\begin{aligned} & 2(v^{k+1}-v^*)^T H(v^k - v^{k+1}) \\ = \quad & \left(\|v^k - v^*\|_H^2 - \|v^{k+1}-v^*\|_H^2\right) - \|v^k - v^{k+1}\|_H^2. \end{aligned}$$

Thus, from (5.10) and the above identity, we obtain

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

We get the nice convergence property of Proximal Point Algorithm. □

6 Convergence analysis for different methods

The convergence proof of the various methods is regarded as the PPA for variational inequalities as in Section 5.2.

6.1 Convergence Analysis for ALM

The iterative scheme of the classical ALM is

$$\left\{\begin{array}{l} x^{k+1} \in \text{argmin}\big\{\theta(x) + \frac{r}{2}\big\|Ax - \big[b + \frac{1}{r}\lambda^k\big]\big\|^2 \,\big|\, x \in \mathcal{X}\big\}. \quad (6.1a) \\ \lambda^{k+1} = \lambda^k - r(Ax^{k+1} - b). \quad (6.1b) \end{array}\right.$$

Lemma 2 *For given* λ^k, *let* $w^{k+1} = (x^{k+1}, \lambda^{k+1})$ *be generated by* (6.1), *then we have*

$$w^{k+1} \in \Omega, \ \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \geq (\lambda - \lambda^{k+1})^T \frac{1}{r}(\lambda^k - \lambda^{k+1}), \ \ \forall\, w \in \Omega. \quad (6.2)$$

Proofs. According to Lemma 1, the solution x^{k+1} of the subproblem (6.1a) satisfies

$$x^{k+1} \in \mathcal{X}, \ \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \big\{A^T[r(Ax^{k+1} - b) - \lambda^k]\big\} \geq 0, \ \forall\, x \in \mathcal{X}.$$

By using (6.1b),

$$x^{k+1} \in \mathcal{X}, \ \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \{-A^T \lambda^{k+1}\} \geq 0, \ \forall x \in \mathcal{X}. \quad (6.3)$$

The equation (6.1b) can be written as

$$(\lambda - \lambda^{k+1})^T (Ax^{k+1} - b) \geq (\lambda - \lambda^{k+1})^T \frac{1}{r} (\lambda^k - \lambda^{k+1}), \ \forall \lambda \in \Re^m. \quad (6.4)$$

Combining (6.3) and (6.4), we get

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(x) - \theta(x^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} -A^T \lambda^{k+1} \\ Ax^{k+1} - b \end{pmatrix} \\ \geq \quad (\lambda - \lambda^{k+1})^T \Big(\frac{1}{r}\Big) (\lambda^k - \lambda^{k+1}), \quad \forall w \in \Omega. \end{aligned} \quad (6.5)$$

Using the notation of (5.5), the assertion of this lemma is proved. □

$v = \lambda$. $H = \frac{1}{r} I_m$. Thus, ALM is a PPA method ! Sincr r is a scalar, we have

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \|\lambda^k - \lambda^*\|^2 - \|\lambda^k - \lambda^{k+1}\|^2. \quad (6.6)$$

引理 2 说明增广拉格朗日乘子法 (6.1) 是对偶变量 λ 的 PPA 算法.

6.2 Convergence Analysis for CP-PPA

The iterative scheme of the CP-PPA is

$$\left\{\begin{array}{l} x^{k+1} = \arg\min\{\theta(x) + \frac{r}{2}\|x - [x^k + \frac{1}{r}A^T\lambda^k]\|^2 \mid x \in \mathcal{X}\}, \qquad (6.7a) \\ \lambda^{k+1} = \lambda^k - \frac{1}{s}\big(A[2x^{k+1} - x^k] - b\big). \qquad (6.7b) \end{array}\right.$$

Lemma 3 *For given w^k, let w^{k+1} be generated by* (6.7)*, then we have*

$$w^{k+1} \in \Omega, \;\; \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \geq (w - w^{k+1})^T H(w^k - w^{k+1}), \;\; \forall\, w \in \Omega. \tag{6.8a}$$

where

$$H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}. \tag{6.8b}$$

Proof. According to Lemma 1, the solution x^{k+1} of the subproblem (6.7a) satisfies

$$x^{k+1} \in \mathcal{X}, \;\theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T\{r(x^{k+1} - x^k) - A^T\lambda^k]\} \geq 0, \;\; \forall\, x \in \mathcal{X}.$$

It can be rewritten as

$$x^{k+1} \in \mathcal{X}, \;\; \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T(-A^T\lambda^{k+1}) \\ \geq (x - x^{k+1})^T\{r(x^k - x^{k+1}) + A(\lambda^k - \lambda^{k+1})\}, \;\; \forall x \in \mathcal{X}. \quad (6.9)$$

The equation (6.7b) can be written as

$$\lambda^{k+1} \in \Re^m, \;\; (\lambda - \lambda^{k+1})^T(Ax^{k+1} - b) \\ \geq (\lambda - \lambda^{k+1})^T\{A(x^k - x^{k+1}) + s(\lambda^k - \lambda^{k+1})\}, \; \forall \lambda \in \Re^m. \quad (6.10)$$

Combining (6.9) and (6.10), we get

$$w^{k+1} \in \Omega, \quad \theta(x) - \theta(x^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} -A^T\lambda^{k+1} \\ Ax^{k+1} - b \end{pmatrix} \\ \geq \begin{pmatrix} x - x^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \begin{pmatrix} x^k - x^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix}, \quad \forall w \in \Omega. \quad (6.11)$$

In the case that $rs > \|A^TA\|$, the matrix H is positive definite. □

引理 3 说明 CP-PPA 方法 (6.7) 是变量 w 的 PPA 算法.

6.3 Convergence Analysis for Balanced ALM

The iterative scheme of the Balanced ALM is

$$
\begin{cases}
x^{k+1} = \arg\min\big\{ L(x,\lambda^k) + \frac{r}{2}\|x - x^k\|^2 \;\big|\; x \in \mathcal{X} \big\}, & \text{(6.12a)} \\
\lambda^{k+1} = \operatorname{argmax}\big\{ L([2x^{k+1} - x^k], \lambda) - \frac{1}{2}\|\lambda - \lambda^k\|^2_{(\frac{1}{r}AA^T + \delta I_m)} \big\}. & \text{(6.12b)}
\end{cases}
$$

Lemma 4 *For given $w^k = (x^k, \lambda^k)$, let u^{k+1} be generated by* (6.12)*, then we have*

$$
\begin{aligned}
w^{k+1} \in \Omega,\;\; \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\
\geq (w - w^{k+1})^T H(w^k - w^{k+1}), \;\; \forall\, w \in \Omega. \qquad \text{(6.13a)}
\end{aligned}
$$

where

$$
H = \begin{pmatrix} rI_n & A^T \\ A & \frac{1}{r}AA^T + \delta I_m \end{pmatrix}. \qquad \text{(6.13b)}
$$

Proof. According to Lemma 1, x^{k+1} offered by (6.12a) satisfies

$$
x^{k+1} \in \mathcal{X},\;\; \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \big\{ -A^T\lambda^k + r(x^{k+1} - x^k) \big\} \geq 0, \;\; \forall\, x \in \mathcal{X}.
$$

Then, for any unknown λ^{k+1}, we have

$$x^{k+1} \in \mathcal{X}, \quad \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T(-A^T\lambda^{k+1}) \\ \geq (x - x^{k+1})^T\{r(x^k - x^{k+1}) + A^T(\lambda^k - \lambda^{k+1})\}, \;\; \forall x \in \mathcal{X}. \quad (6.14)$$

Similarly, according to Lemma 1, λ^{k+1} offered by (6.12b) is characterized by the variational inequality

$$(\lambda - \lambda^{k+1})^T\Big\{\Big(A[2x^{k+1} - x^k] - b\Big) + \Big(\frac{1}{r}AA^T + \delta I_m\Big)(\lambda^{k+1} - \lambda^k)\Big\} \geq 0, \quad \forall\, \lambda \in \Lambda.$$

It can be rewritten as

$$\lambda^{k+1} \in \Lambda, \quad (\lambda - \lambda^{k+1})^T(Ax^{k+1} - b) \\ \geq (\lambda - \lambda^{k+1})^T\{(A(x^k - x^{k+1}) + \Big(\frac{1}{r}AA^T + \delta I_m\Big)(\lambda^k - \lambda^{k+1})\}, \quad (6.15)$$

for all $\lambda \in \Re^m$. Combining (6.14) and (6.15), and using the notation in (5.5), we get (6.13). The assertion of Lemma 4 is proved. $\Box$

引理 4 说明 Balanced ALM 方法 (6.12) 是变量 w 的 PPA 算法.

References

[1] 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.

[2] A. Chambolle and T Pock, On the ergodic convergence rates of a first-order primal-dual algorithm, Math. Program., A 159 (2016) 253-287.

[3] G. Golub and C. F. Van Loan, Matrix Computations, The Johns Hopkins University Press, The Fourth Edition, 2013.

[4] 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.

[5] B.S. He and X.M. Yuan, Balanced Augmented Lagrangian Method for Convex Optimization. manusript, 2021. arXiv:2108.08554

[6] M. R. Hestenes, Multiplier and gradient methods, J. Optim. Theory Appli, 4 (1969), pp. 303-320.

[7] K. Levenberg, A method for the solution of certain nonlinear problems in least squares, Quart. Appl. Math., **2** (1944) 164-168.

[8] D. W. Marquardt, An algorithm for least squares estimation of nonlinear parameters, SIAM J. Appl. Math., **11** (1963) 431-441.

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

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

Thank you very much for reading !