凸优化分裂收缩算法的一些新进展

用好变分不等式和邻近点算法两大法宝

中学的数理基础　必要的社会实践
普通的大学数学　一般的优化原理

何 炳 生　　南京大学数学系

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

华东师大　上海大学　复旦大学　南京应用数学中心

西工大　西交大　西电大学　兰州理工　兰州大学

新疆大学　清华大学　北京工业大学

连续优化中一些代表性数学模型

1. 鞍点问题 $\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\{\Phi(x,y)=\theta_1(x)-y^TAx-\theta_2(y)\}$
2. 线性约束的凸优化问题 $\min\{\theta(x)|Ax=b\,(\text{or}\geq b),\; x\in\mathcal{X}\}$
3. 结构型凸优化 $\min\{\theta_1(x)+\theta_2(y)|Ax+By=b,\; x\in\mathcal{X},\; y\in\mathcal{Y}\}$
4. 多块可分离凸优化 $\min\{\sum_{i=1}^p\theta_i(x_i)|\sum_{i=1}^p A_ix_i=b, x_i\in\mathcal{X}_i\}$

变分不等式(VI) 是瞎子爬山的数学表达形式

邻近点算法(PPA) 是步步为营 稳扎稳打的求解方法.

变分不等式和邻近点算法是分析和设计凸优化方法的两大法宝.

分裂是指迭代中子问题都通过分拆求解. 收缩算法有别于可行方向法,
又有别于下降算法, 它的迭代点离优化问题的拉格朗日函数的鞍点越来越近.

先解释上述问题如何化为一个单调变分不等式 并介绍什么是变分不等式的邻近点算法

1 Optimization problem and VI

1.1 Differential convex optimization in Form of VI

Let $\Omega \subset \Re^n$, we consider the convex minimization problem

$$\min\{f(x) \mid x \in \Omega\}. \tag{1.1}$$

What is the first-order optimal condition ?

$x^* \in \Omega^* \quad \Leftrightarrow \quad x^* \in \Omega$ and any feasible direction is not a descent one.

Optimal condition in variational inequality form

- $S_d(x^*) = \{s \in \Re^n \mid s^T \nabla f(x^*) < 0\} \;=\;$ Set of the descent directions.
- $S_f(x^*) = \{s \in \Re^n \mid s = x - x^*,\; x \in \Omega\} \;=$ Set of feasible directions.

$x^* \in \Omega^* \quad \Leftrightarrow \quad x^* \in \Omega$ **and** $S_f(x^*) \cap S_d(x^*) = \emptyset$.

瞎子爬山判定山顶的准则是: 所有可行方向都不再是上升方向

The optimal condition can be presented in a variational inequality (VI) form:

$$x^* \in \Omega, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \Omega. \tag{1.2}$$

Substituting $\nabla f(x)$ with an operator F (from $\Re^n$ into itself), we get a classical VI.

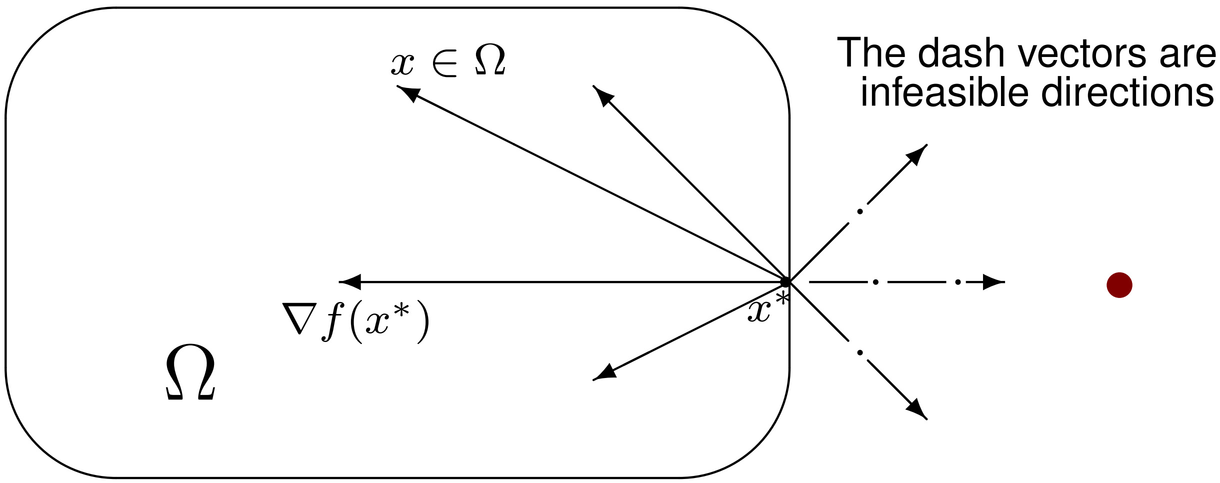


Fig. 1.1 Differential Convex Optimization and VI

Since $f(x)$ is a convex function, we have

$$\begin{array}{l} f(y) \geq f(x) + \nabla f(x)^T (y - x) \\ f(x) \geq f(y) + \nabla f(y)^T (x - y) \end{array} \quad \text{thus} \quad (x - y)^T (\nabla f(x) - \nabla f(y)) \geq 0.$$

We say the gradient ∇f of the convex function f is a monotone operator.

通篇我们需要用到的大学数学　主要是基于微积分学的一个引理

$$x^* \in \operatorname{argmin}\{\theta(x)|x \in \mathcal{X}\} \Leftrightarrow \quad x^* \in \mathcal{X}, \qquad \theta(x) - \theta(x^*) \ge 0, \qquad \forall\, x \in \mathcal{X};$$

$$x^* \in \operatorname{argmin}\{f(x)|x \in \mathcal{X}\} \Leftrightarrow \quad x^* \in \mathcal{X}, \quad (x - x^*)^T \nabla f(x^*) \ge 0, \quad \forall\, x \in \mathcal{X}.$$

上面的凸优化最优性条件是最基本的, 看起来合在一起就是下面的引理:

定理 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,*

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{1.3a}$$

if and only if 　　凸优化最优性条件定理

$$x^* \in \mathcal{X}, \;\; \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \ge 0, \;\; \forall\, x \in \mathcal{X}. \tag{1.3b}$$

定理 1 把优化问题 (1.3a) 转换成了变分不等式 (1.3b).

1.2 Linear constrained convex optimization and VI

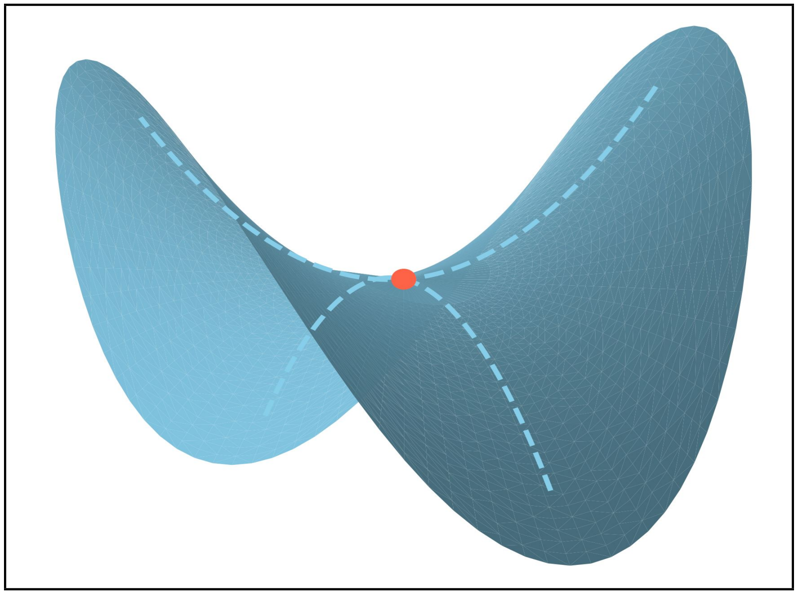
We consider the linearly constrained convex optimization problem

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{1.4}$$

The Lagrangian function of the problem (1.4) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \tag{1.5}$$

which is defined on $\mathcal{U} \times \Re^m$.



A pair of (u^*,λ^*) is called a saddle point of the Lagrange function (1.5), if $(u^*,\lambda^*) \in \mathcal{U} \times \Re^m$, and

$$L(u,\lambda^*) \ge L(u^*,\lambda^*) \ge L(u^*,\lambda), \ \ \forall (u,\lambda) \in \mathcal{U} \times \Re^m.$$

The above inequalities can be written as

$$\left\{\begin{array}{ll} u^* \in \mathcal{U}, & L(u,\lambda^*) - L(u^*,\lambda^*) \ge 0, \quad \forall\, u \in \mathcal{U}, \qquad \text{(1.6a)} \\ \lambda^* \in \Re^m, & L(u^*,\lambda^*) - L(u^*,\lambda) \ge 0, \quad \forall\, \lambda \in \Re^m. \qquad \text{(1.6b)} \end{array}\right.$$

According to the definition of $L(u,\lambda)$ (see(1.5)), it follows from (1.6a) that

$$u^* \in \mathcal{U}, \;\; \theta(u) - \theta(u^*) + (u - u^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, \;\; \forall\, u \in \mathcal{U}. \qquad (1.7)$$

Similarly, for (1.6b), we have

$$\lambda^* \in \Re^m, \;\; (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \ge 0, \;\; \forall\, \lambda \in \Re^m. \qquad (1.8)$$

Writing (1.7) and (1.8) together, we get the following variational inequality:

$$\left\{\begin{array}{ll} u^* \in \mathcal{U}, & \theta(u) - \theta(u^*) + (u - u^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, \quad \forall\, u \in \mathcal{U}, \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \ge 0, \quad \forall\, \lambda \in \Re^m. \end{array}\right.$$

Using a more compact form, the saddle-point can be characterized as the solution of the following VI:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \qquad \text{(1.9a)}$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \text{ and } \quad \Omega = \mathcal{U} \times \Re^m. \qquad \text{(1.9b)}$$

Setting $w = (u, \lambda^*)$ and $w = (u^*, \lambda)$ in (1.9), we get (1.7) and (1.8), respectively. Because F is an affine operator and

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

The matrix is skew-symmetric, we have

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

线性约束的凸优化问题 (1.4), 转换成了混合变分不等式 (1.9).

Two block separable convex optimization

We consider the following structured separable convex optimization

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{1.10}$$

This is a special problem of (1.4) with

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathcal{U} = \mathcal{X} \times \mathcal{Y}, \quad \mathcal{A} = (A, B).$$

The Lagrangian function of the problem (5.5) is

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

The same analysis tells us that the saddle point is a solution of the following VI:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{1.11}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \tag{1.12a}$$

$$F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{1.12b}$$

The affine operator $F(w)$ has the form

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

Again, due to the skew-symmetry, we have $(w - \tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0.$

可分离线性约束凸优化问题 (5.5), 转换成了变分不等式 (1.11)–(1.12).

Convex optimization problem with three separable functions

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

is a special problem of (1.4) with three blocks. The Lagrangian function is

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

The same analysis tells us that the saddle point is a solution of the following VI:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega.$$

where $\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z)$,

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

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

求线性约束凸优化拉格朗日函数的鞍点, 都转换成了相应的变分不等式.

2 Proximal point algorithms and its Beyond

引理 1 *Let the vectors* $a, b \in \Re^n$, $H \in \Re^{n\times n}$ *be a positive definite matrix. If* $b^T H(a-b) \geq 0$, *then we have*

$\|x\|^2 = x^T x, \quad \|x\|_H^2 = x^T H x.$

$$\|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2. \tag{2.1}$$

The assertion follows from $\|a\|_H^2 = \|b+(a-b)\|_H^2 \geq \|b\|_H^2 + \|a-b\|_H^2.$

2.1 Preliminaries of PPA for Variational Inequalities

The optimal condition of the linearly constrained convex optimization is characterized as a mixed monotone variational inequality:

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

$\boldsymbol{u}$ 往往指自变量，向量 $\boldsymbol{w} = (\boldsymbol{u}, \boldsymbol{\lambda})$ 包含自变量 $\boldsymbol{u}$ 和对偶变量 $\boldsymbol{\lambda}$.

Definition: PPA for VI (2.2) in H-norm

For given w^k and $H \succ 0$, find w^{k+1} such that

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \qquad \text{邻近点算法}$$
$$\geq (w - w^{k+1})^T H(w^k - w^{k+1}), \quad \forall\, w \in \Omega, \qquad (2.3)$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.2).

✠ w^{k+1} **is the solution of** (2.2) **if and only if** $w^k = w^{k+1}$ ✠

(2.3) 是求解 VI (2.2) 的 PPA 算法的定义. H 可以是数量矩阵, 也可以是适当的分块矩阵, 当然 H 首先要是对称矩阵. 后面会有大量的例子说明这容易用构造性的方法实现。

Setting $w = w^*$ in (2.3), we obtain

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

Note that (see the structure of $F(w)$ in (1.9b))

$$(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*),$$

and consequently (by using (2.2)) we obtain

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

Thus, we have

$$(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \geq 0. \tag{2.4}$$

By setting $\boldsymbol{a = w^k - w^*}$ and $\boldsymbol{b = w^{k+1} - w^*}$,

the inequality (2.4) means that $\boldsymbol{b^T H(a - b) \geq 0}$.

By using Lemma 1, we obtain

$$\|\boldsymbol{w^{k+1} - w^*}\|_{\boldsymbol{H}}^{\boldsymbol{2}} \leq \|\boldsymbol{w^k - w^*}\|_{\boldsymbol{H}}^{\boldsymbol{2}} - \|\boldsymbol{w^k - w^{k+1}}\|_{\boldsymbol{H}}^{\boldsymbol{2}}. \tag{2.5}$$

We get the nice convergence property of Proximal Point Algorithm.

2.2 Variants of PPA for Variational Inequalities

Let v be a sub-vector of w which contains a part of the elements of the vector w. In some algorithms, the k-th iteration only needs v^k to start.

v 核心变量

PPA for VI (2.2) in H-norm For given v^k and $H \succ 0$, find w^{k+1},

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (v - v^{k+1})^T H(v^k - v^{k+1}), \quad \forall\, w \in \Omega, \tag{2.6}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.2).

✠ w^{k+1} **is the solution of (2.2) if and only if** $v^k = v^{k+1}$ ✠

In this case, v is called the essential variables of w. In addition, we define

$$\mathcal{V}^* = \{v^* \text{ is a subvector of } w^* \mid w^* \in \Omega^*\}.$$

Setting $w = w^*$ in (2.6), we obtain

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

Note that (see the structure of $F(w)$ in (1.9b))

$$(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*),$$

and consequently (by using (2.2)) we obtain

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

Thus, we have

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0. \tag{2.7}$$

By using Lemma 1, we obtain

$$\|\boldsymbol{v^{k+1} - v^*}\|^2_{\boldsymbol{H}} \leq \|\boldsymbol{v^k - v^*}\|^2_{\boldsymbol{H}} - \|\boldsymbol{v^k - v^{k+1}}\|^2_{\boldsymbol{H}}. \tag{2.8}$$

We get the nice convergence property of Proximal Point Algorithm.

The residue sequence $\{\|v^k - v^{k+1}\|_H\}$ is also monotonically no-increasing.

Sequence $\{\|v^k - v^{k+1}\|_H\}$ is non-icreasing. $\|v^k - v^{k+1}\|^2_H \leq \|v^{k-1} - v^k\|^2_H$.

2.3 The relaxed PPA （延伸的邻近点算法）

We shall maintain our focus on the monotone variational inequality (2.2), namely,

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

The PPA form (2.6) reads as

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

Set the output of the above VI as $\tilde{w}^k$, we have

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \quad \forall\, w \in \Omega. & \end{aligned} \tag{2.1}$$

Setting $w = w^*$ in (2.1), we obtain

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \tag{2.2}$$

Applying (see (1.9b)) the identity

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

to (2.2), we obtain

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

Because RHS of the above inequality is , we have

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

We write it as

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

and thus

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.3}$$

The inequality (2.3) means that $(v^k - \tilde{v}^k)$ is the ascent direction of the unknown distance function $\frac{1}{2}\|v - v^*\|_H^2$ at the point v^k.

$$\Big\langle \nabla\big(\tfrac{1}{2}\|v - v^*\|_H^2\big)\Big|_{v=v^k},\ (v^k - \tilde{v}^k)\Big\rangle \geq \|v^k - \tilde{v}^k\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*.$$

The task of the algorithm is to produce a decreasing sequence $\{\|v^k - v^*\|_H^2\}$.

Set

$$v^{k+1}(\alpha) = v^k - \alpha(v^k - \tilde{v}^k) \tag{2.4}$$

which is an α dependent new iterate. It is clear we want to maximize

$$\vartheta(\alpha) = \|v^k - v^*\|_H^2 - \|v^{k+1}(\alpha) - v^*\|_H^2. \tag{2.5}$$

Note that

$$\begin{aligned} \vartheta(\alpha) &= \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \alpha(v^k - \tilde{v}^k)\|_H^2 \\ &= 2\alpha(v^k - v^*)^T H(v^k - \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \end{aligned} \tag{2.6}$$

is a quadratic function of α.

We can not directly maximize $\vartheta(\alpha)$ in (2.6) because the coefficient of the linear term $2(v^k - v^*)^T H(v^k - \tilde{v}^k)$ contains the unknown solution v^*.

Using (2.3), from (2.6) we get

$$\vartheta(\alpha) \geq 2\alpha\|v^k - \tilde{v}^k\|_H^2 - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \tag{2.7}$$

Set

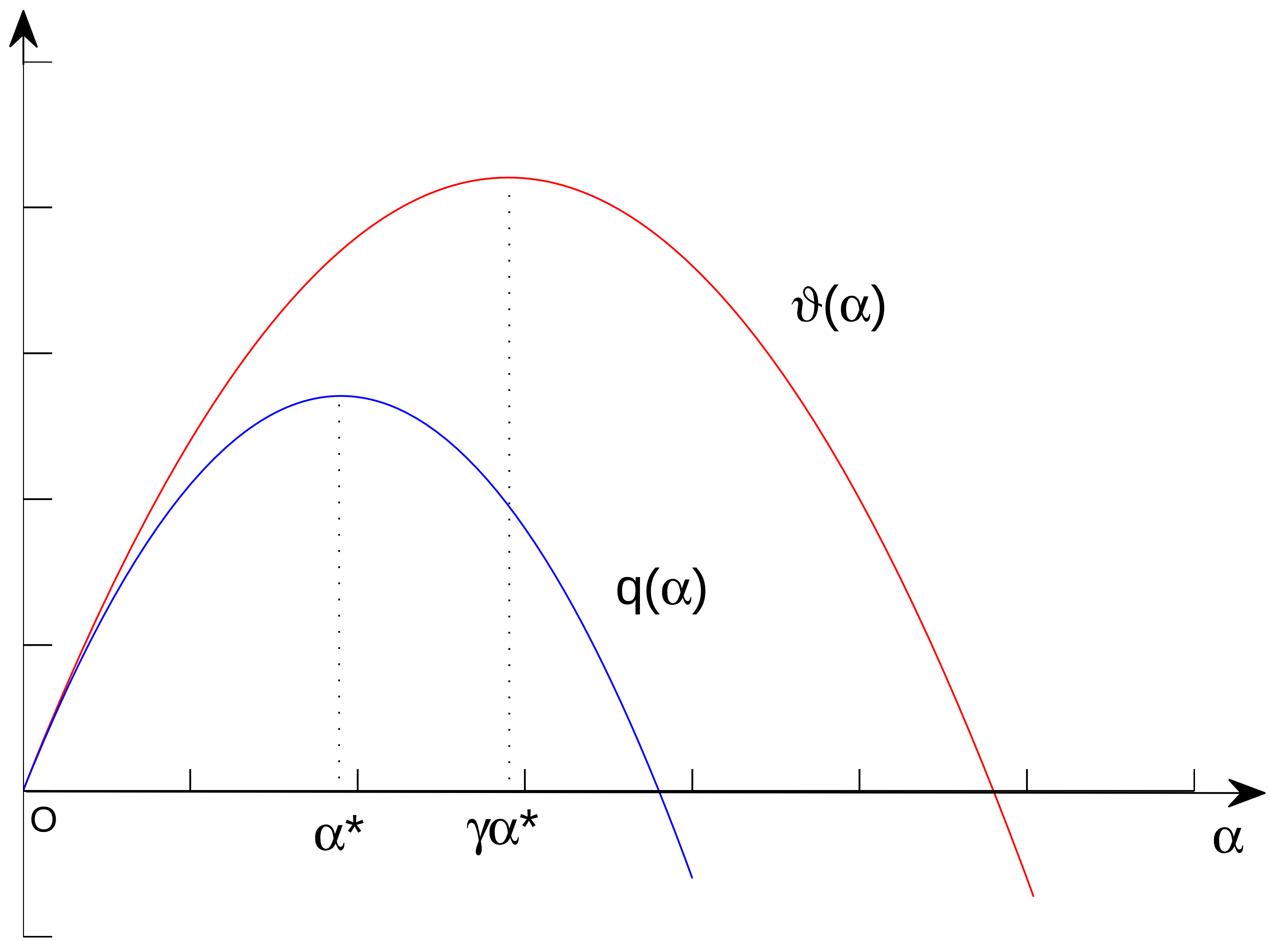
$$q(\alpha) = (2\alpha - \alpha^2)\|v^k - \tilde{v}^k\|_H^2, \tag{2.8}$$

which is a quadratic lower-bound function of $\vartheta(\alpha)$. The quadratic function $q(\alpha)$ reaches its maximum at $\alpha^* \equiv 1$.

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k), \quad \gamma \in (0, 2) \tag{2.9}$$

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

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_H^2. \tag{2.10}$$



取 $\gamma \in [1, 2)$ 的示意图

以上的预备知识. 要求读者理解（或者是先承认）优化问题拉格朗日函数的鞍点和变分不等式 (VI) 解点的等价的关系, 以及 PPA 算法的定义及收缩性质.

3 从原始-对偶混合梯度法到按需定制的邻近点算法

We consider the $\min - \max$ problem (*e. g.* 图像处理中的 ROF Model [4, 30])

$$\min_x \max_y \{\Phi(x,y) = \theta_1(x) - y^T Ax - \theta_2(y) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{3.1}$$

Let (x^*, y^*) be the solution of (6.4), then we have

$$\left\{ \begin{array}{ll} x^* \in \mathcal{X}, \;\; \Phi(x,y^*) - \Phi(x^*,y^*) \geq 0, \;\; \forall\, x \in \mathcal{X}, & \text{(3.2a)} \\ y^* \in \mathcal{Y}, \;\; \Phi(x^*,y^*) - \Phi(x^*,y) \geq 0, \;\; \forall\, y \in \mathcal{Y}. & \text{(3.2b)} \end{array} \right.$$

Using the notation of $\Phi(x,y)$, it can be written as

$$\left\{ \begin{array}{lll} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T y^*) \geq 0, & \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) \;+\; (y - y^*)^T(\,Ax^*) \;\; \geq 0, & \forall\, y \in \mathcal{Y}. \end{array} \right.$$

Furthermore, it can be written as a variational inequality in the compact form:

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T F(u^*) \geq 0, \; \forall\, u \in \Omega, \tag{3.3}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad F(u) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}, \quad \Omega = \mathcal{X} \times \mathcal{Y}.$$

Since $F(u) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix} = \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$, we have

$$(u - v)^T (F(u) - F(v)) \equiv 0.$$

3.1 求解鞍点问题的 原始-对偶混合梯度法 PDHG [32]

For given (x^k, y^k), PDHG [32] produces a pair of (x^{k+1}, y^{k+1}). First,

$$x^{k+1} = \text{argmin}\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \tag{3.4a}$$

and then we obtain y^{k+1} via

$$y^{k+1} = \text{argmax}\{\Phi(x^{k+1}, y) - \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. \tag{3.4b}$$

Ignoring the constant term in the objective function, the subproblems (6.6) are reduced to

$$\begin{cases} x^{k+1} = \text{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(3.5a)} \\ y^{k+1} = \text{argmin}\{\theta_2(y) + y^T A x^{k+1} + \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. & \text{(3.5b)} \end{cases}$$

According to Lemma 1, the optimality condition of (3.5a) is $x^{k+1} \in \mathcal{X}$ and

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T y^k + r(x^{k+1} - x^k)\} \geq 0, \ \forall\, x \in \mathcal{X}. \quad \text{(3.6)}$$

Similarly, from (3.5b) we get $y \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{Ax^{k+1} + s(y^{k+1} - y^k)\} \geq 0, \ \forall\, y \in \mathcal{Y}. \quad \text{(3.7)}$$

Combining (3.6) and (3.7), we have $u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}\Omega$,

$$\begin{aligned} u^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (u - u^{k+1})^T F(u^{k+1}) \\ &\geq (u - u^{k+1})^T Q(u^k - u^{k+1}), \ \ \forall\, u \in \Omega. \end{aligned} \quad \text{(3.8)}$$

where

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \quad \text{is not symmetric.}$$

It does not be the PPA form (2.3), and we can not expect its convergence.

Note that its Lagrange function is

$$L(x, y) = c^T x - y^T (Ax - b) \tag{3.9}$$

which defined on $\Re^2_+ \times \Re$. $x^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $y^* = 1$. is the unique saddle point of the Lagrange function.

For solving the $\min$-$\max$ problem (3.9), by using (6.6), the iterative formula is

$$\begin{cases} x^{k+1} = \arg\min\{c^T x - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 | x \geq 0\} \\ \qquad\quad = \arg\min\{\frac{r}{2}\|x - [x^k + \frac{1}{r}(A^T y^k - c)]\|^2 | x \geq 0\} \\ \qquad\quad = P_{\Re^n_+}[x^k + \frac{1}{r}(A^T y^k - c)] \\ \qquad\quad = \max\{[x^k + \frac{1}{r}(A^T y^k - c)], 0\}, \\ y^{k+1} = y^k - \frac{1}{s}(Ax^{k+1} - b). \end{cases}$$

We use $(x_1^0, x_2^0; y^0) = (0, 0; 0)$ as the start point. For this example, the method is not convergent.

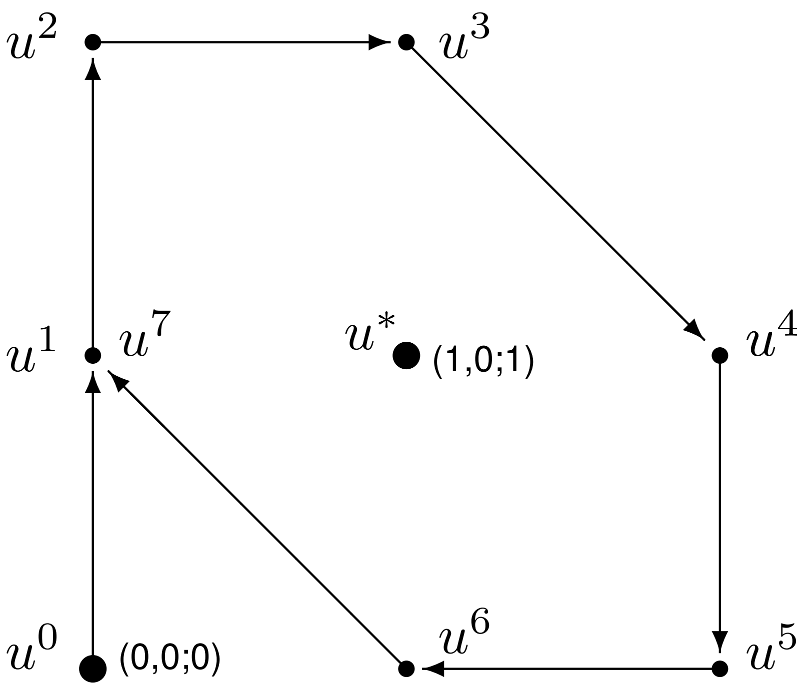
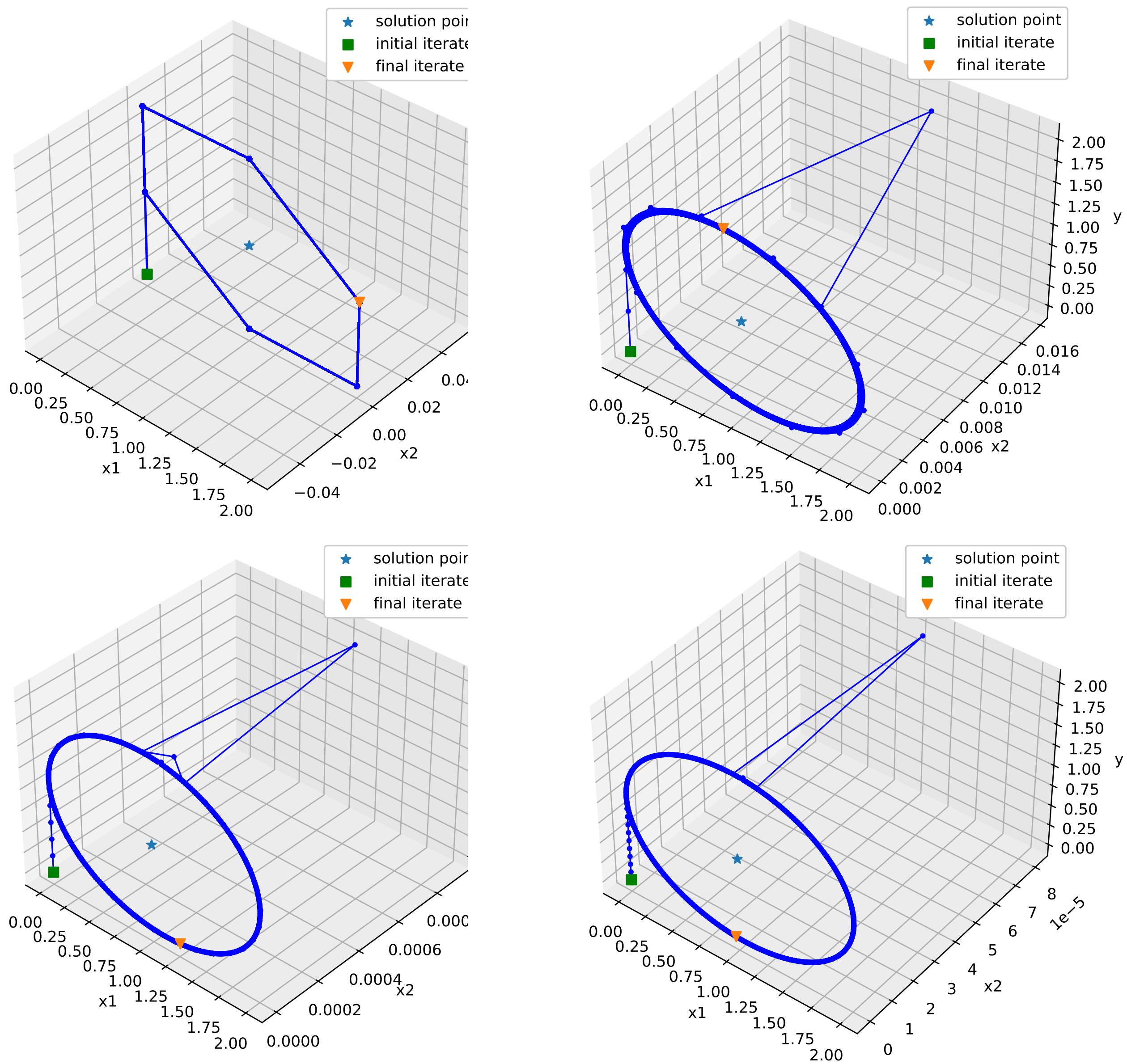


Fig. 2.1 The sequence generated by PDHG Method with $r = s = 1$

$$u^0 = (0, 0; 0)$$
$$u^1 = (0, 0; 1)$$
$$u^2 = (0, 0; 2)$$
$$u^3 = (1, 0; 2)$$
$$u^4 = (2, 0; 1)$$
$$u^5 = (2, 0; 0)$$
$$u^6 = (1, 0; 0)$$
$$u^7 = (0, 0; 1)$$

$$\boldsymbol{u^{k+6} = u^k}$$



For $r = s = 1, 2, 5, 10$, PDHG methods are not convergent

3.2 Customized Proximal Point Algorithm-Classical Version

通常, 我们把这种凑成的邻近点算法称为“按需定制的邻近点算法”.

If we change the non-symmetric matrix Q to a symmetric matrix H such that

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \quad \Rightarrow \quad H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix},$$

then the variational inequality (6.7) will become the following desirable form:

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

For this purpose, we need only to change (3.7) in PDHG, namely,

$$\theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{Ax^{k+1}+s(y^{k+1}-y^k)\} \ge 0, \ \forall\, y \in \mathcal{Y}.$$

to

$$\theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{Ax^{k+1}+A(x^{k+1}-x^k) \\ + s(y^{k+1}-y^k)\} \ge 0, \ \forall\, y \in \mathcal{Y}.$$

$$\theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{A[2x^{k+1}-x^k]+s(y^{k+1}-y^k)\} \ge 0. \quad (3.10)$$

Thus, for given (x^k, y^k), producing a proximal point (x^{k+1}, y^{k+1}) via (6.6a) and (3.10) can be summarized as:

$$x^{k+1} = \text{argmin}\big\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \,\big|\, x \in \mathcal{X}\big\}. \tag{3.11a}$$

$$y^{k+1} = \text{argmax}\big\{\Phi\big([2x^{k+1} - x^k], y\big) - \frac{s}{2}\|y - y^k\|^2\big\} \tag{3.11b}$$

By ignoring the constant term in the objective function, getting x^{k+1} from (3.11a) is equivalent to obtaining x^{k+1} from

$$x^{k+1} = \text{argmin}\big\{\theta_1(x) + \frac{r}{2}\big\|x - \big[x^k + \frac{1}{r}A^T y^k\big]\big\|^2 \,\big|\, x \in \mathcal{X}\big\}.$$

The solution of (3.11b) is given by

$$y^{k+1} = \text{argmin}\big\{\theta_2(y) + \frac{s}{2}\big\|y - \big[y^k + \frac{1}{s}A(2x^{k+1} - x^k)\big]\big\|^2 \,\big|\, y \in \mathcal{Y}\big\}.$$

According to the assumption, there is no difficulty to solve (3.11a)-(3.11b).

In the case that $rs > \|A^T A\|$, the matrix

$$H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \quad \text{is positive definite.}$$

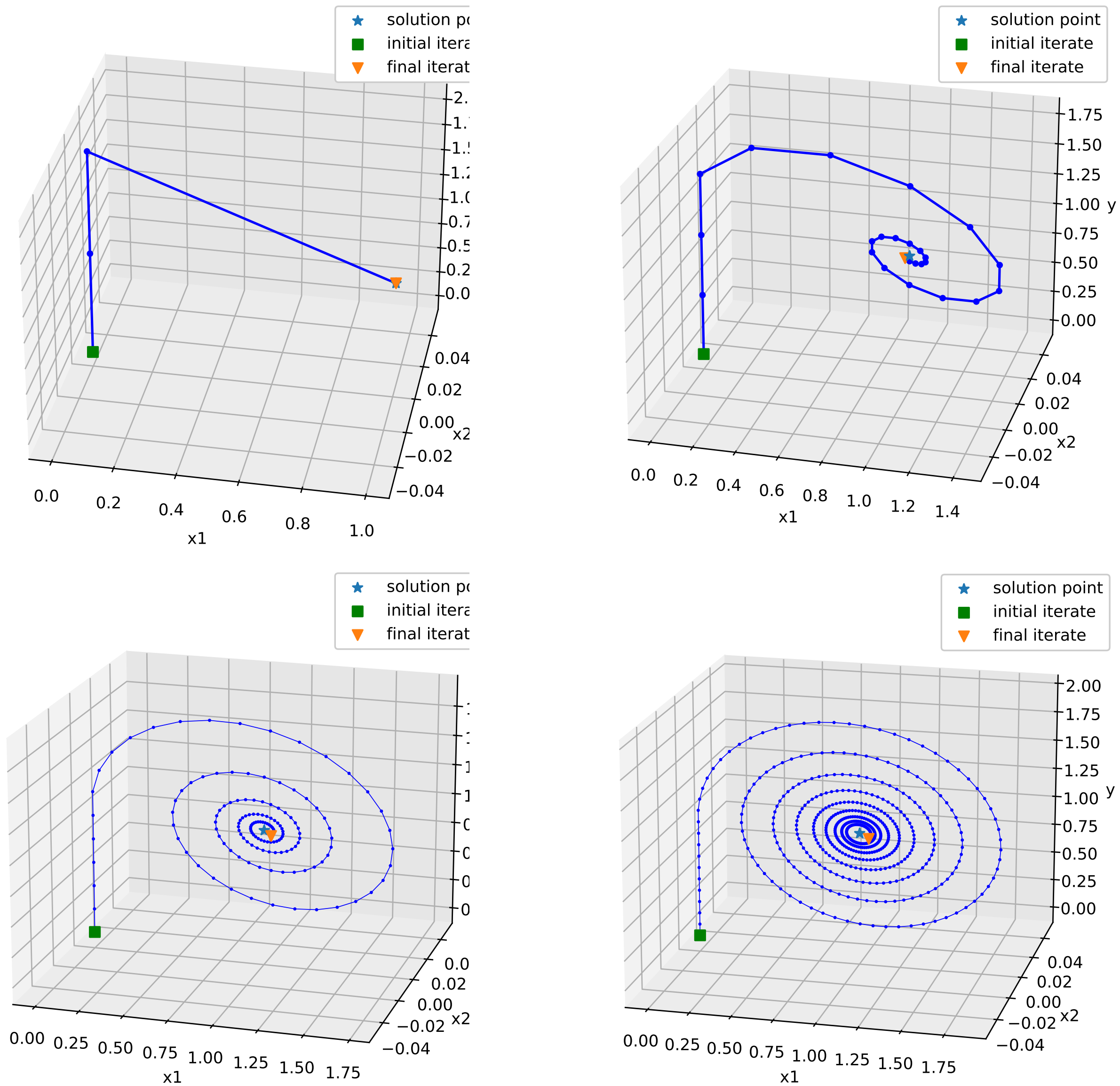
定理 2 *The sequence* $\{u^k = (x^k, y^k)\}$ *generated by the customized PPA* (3.11) *satisfies*

$$\|u^{k+1} - u^*\|_H^2 \le \|u^k - u^*\|_H^2 - \|u^k - u^{k+1}\|_H^2. \tag{3.12}$$

For the minimization problem $\min\{\theta(x) \mid Ax = b, x \in \mathcal{X}\}$,
the iterative scheme is

$$x^{k+1} = \text{argmin}\Big\{\theta(x) + \frac{r}{2}\Big\|x - \big[x^k + \frac{1}{r}A^T y^k\big]\Big\|^2 \,\Big|\, x \in \mathcal{X}\Big\}. \tag{3.13a}$$

$$y^{k+1} = y^k - \frac{1}{s}\big[A(2x^{k+1} - x^k) - b\big]. \tag{3.13b}$$



For $r = s = 1, 2, 5, 10$, C-PPA are convergent. 参数越大, 步子越保守, 收敛越慢

3.3 Simplicity recognition

Frame of VI is recognized by some Researcher in Image Science

Diagonal preconditioning for first order primal-dual algorithms in convex optimization*

Thomas Pock
Institute for Computer Graphics and Vision
Graz University of Technology
pock@icg.tugraz.at

Antonin Chambolle
CMAP & CNRS
École Polytechnique
antonin.chambolle@cmap.polytechnique.fr

- T. Pock and A. Chambolle, IEEE ICCV, 1762-1769, 2011
- 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.

preconditioned algorithm. In very recent work [10], it has been shown that the iterates (2) can be written in form of a proximal point algorithm [14], which greatly simplifies the convergence analysis.

From the optimality conditions of the iterates (4) and the convexity of G and F^* it follows that for any $(x, y) \in X \times Y$ the iterates x^{k+1} and y^{k+1} satisfy

$$\left\langle \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}, F\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} + M\begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\rangle \geq 0\,, \tag{5}$$

where

$$F\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} \partial G(x^{k+1}) + K^T y^{k+1} \\ \partial F^*(y^{k+1}) - Kx^{k+1} \end{pmatrix}$$

and

$$M = \begin{bmatrix} \mathrm{T}^{-1} & -K^T \\ -\theta K & \Sigma^{-1} \end{bmatrix}. \tag{6}$$

It is easy to check, that the variational inequality (5) now takes the form of a proximal point algorithm [10, 14, 16].

作者 C-P 说到我们的 PPA 解释极大地简化了收敛性分析.

我们依然认为, 只有当左边 (6) 式的矩阵 M 对称正定, 才是收敛的 PPA 方法.

否则, 就像我们前面给出的例子, 方法是不一定收敛的.

由 CP 方法演译得来的矩阵 M (相当于我们的 H), 当 $\theta = 0$, 方法不能保证收敛.

对 $\theta \in (0, 1)$, 收敛性没有证明, 至今还是一个 Open Problem.

[9] L. Ford and D. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, New Jersey, 1962.

[10] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for total variation image restoration. Technical report, Nanjing University, China, 2010.

Later, the Reference [10] is published in SIAM J. Imaging Science [19].

Math. Program., Ser. A
DOI 10.1007/s10107-015-0957-3

CrossMark

FULL LENGTH PAPER

On the ergodic convergence rates of a first-order primal–dual algorithm

Antonin Chambolle[1] · Thomas Pock[2,3]

The paper published by Chambolle and Pock in Math. Progr. uses the VI framework

1 Introduction

In this work we revisit a first-order primal–dual algorithm which was introduced in [15, 26] and its accelerated variants which were studied in [5]. We derive new estimates for the rate of convergence. In particular, exploiting a proximal-point interpretation due to [16], we are able to give a very elementary proof of an ergodic $O(1/N)$ rate of convergence (where N is the number of iterations), which also generalizes to non-

Algorithm 1: $O(1/N)$ Non-linear primal–dual algorithm

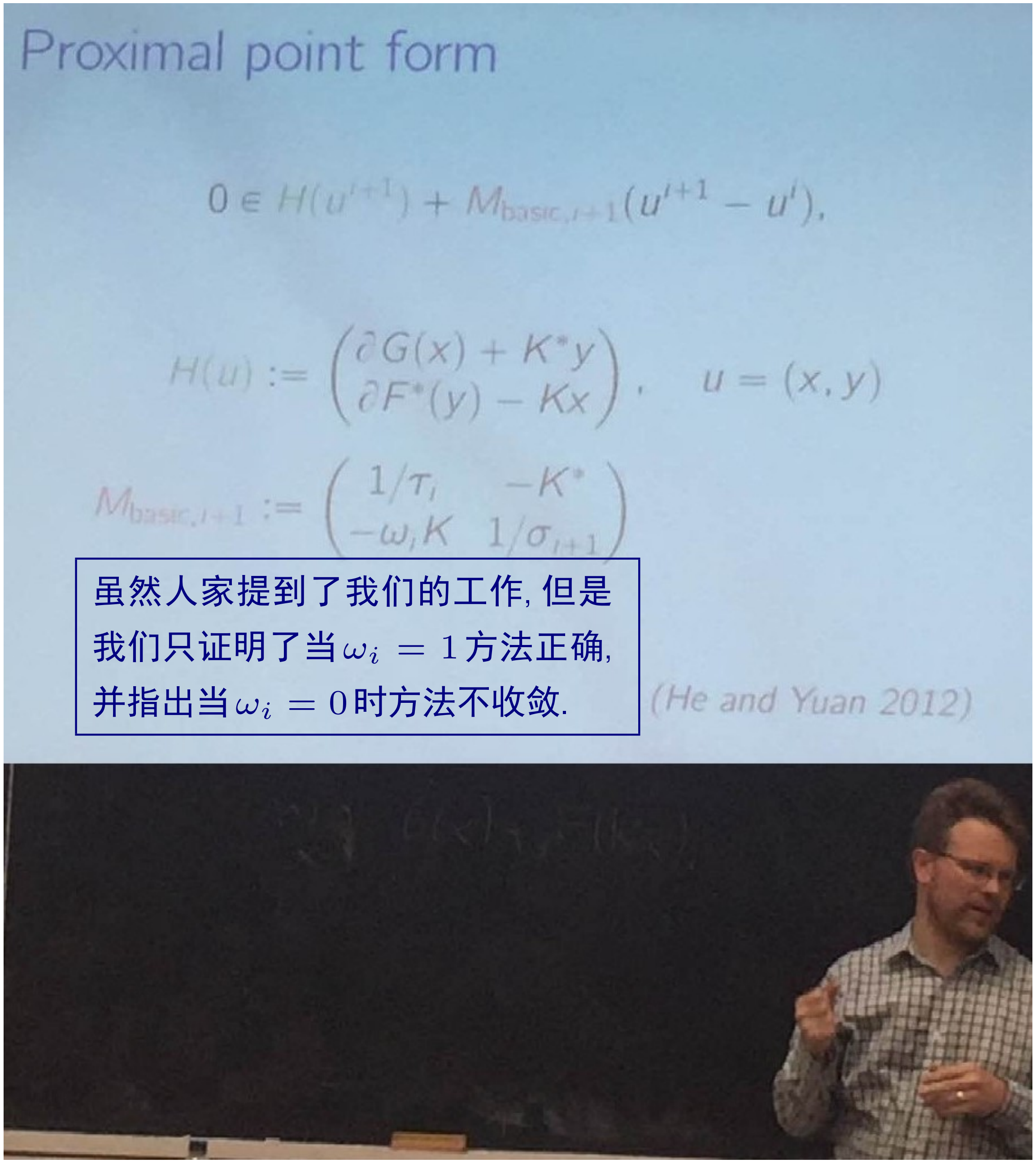
- Input: Operator norm $L := \|K\|$, Lipschitz constant L_f of ∇f, and Bregman distance functions D_x and D_y.
- Initialization: Choose $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$, $\tau, \sigma > 0$
- Iterations: For each $n \geq 0$ let

$$(x^{n+1}, y^{n+1}) = \mathcal{PD}_{\tau,\sigma}(x^n, y^n, 2x^{n+1} - x^n, y^n) \tag{11}$$

The elegant interpretation in [16] shows that by writing the algorithm in this form

♣ 该文的文献 [16] 是我们发表在 SIAM J. Imaging Science 上的文章.

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.



2017 年 7 月, 南方科技大学数学系的一位副主任去英国访问. 在他参加的一个学术会议上, 首位报告人讲: 用 He and Yuan 提出的邻近点形式 (PPF), 处理图像问题。

见到一幅幻灯片介绍我们的工作, 我的同事抢拍了一张照片发给我。

这也说明, 只有简单的思想才容易得到传播, 被人接受。

University of Colorado **Boulder** Technical Report, Department of Applied Mathematics

The Chen-Teboulle algorithm is the proximal point algorithm

Stephen Becker *

November 22, 2011; posted August 13, 2019

Abstract

We revisit the
on the step-size p

Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods.

1 Background

Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods. The idea consists simply of working with a different norm in the *product* Hilbert space. We fix an inner product $\langle x, y\rangle$ on $\mathcal{H} \times \mathcal{H}^*$. Instead of defining the norm to be the induced norm, we define the primal norm as follows (and this induces the dual norm)

$$\|x\|_V = \sqrt{\langle Vx, x\rangle} = \sqrt{\langle x, x\rangle_V}, \quad \|y\|_V^* = \|y\|_{V^{-1}} = \sqrt{\langle y, V^{-1}y\rangle} = \sqrt{\langle y, y\rangle_{V^{-1}}}$$

for any Hermitian positive definite $V \in \mathcal{B}(\mathcal{H}, \mathcal{H})$; we write this condition as $V \succ 0$. For finite dimensional spaces $\mathcal{H}$, this means that V is a positive definite matrix.

4 单块的问题按需设计邻近点算法

根据预设正定矩阵 构造 PPA 算法. 方法可以在 [12] 中查到.

The convex optimization problem,

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

is translated to the equivalent variational inequality :

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

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Re^m. \tag{4.1b}$$

4.1 PPA in Primal-Dual Order

Relaxed PPA for the variational inequality (4.1) : Find $\tilde{w}^k \in \Omega$, such that

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \quad (4.2a)$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_n & A^T \\ A & \frac{1}{\beta} I_m \end{pmatrix}. \quad (4.2b)$$

The concrete formula of (4.2) is

有下划线的部分就是 $F(\tilde{w}^k)$

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

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

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_n)}(\tilde{x}^k - x^k) + \boldsymbol{A^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ \qquad (\underline{A\tilde{x}^k - b}) \quad + \boldsymbol{A}(\tilde{x}^k - x^k) \quad + \quad \boldsymbol{(1/\beta)}(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases} \quad (4.3)$$

把 (4.3) 中的“微小型”变分不等式整理一下, 便是:

$$\begin{cases} \theta(x)-\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T\{-A^T\lambda^k+\ (\beta A^TA+\delta I_n)(\tilde{x}^k-x^k)\}\ge 0, \\ \qquad\qquad\qquad\qquad \big(A[2\tilde{x}^k-x^k]-b\big) \quad + \quad (1/\beta)(\tilde{\lambda}^k-\lambda^k)=0. \end{cases}$$

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (4.3), we need only use the following procedure: (Primal-Dual)

$$\begin{cases} \tilde{x}^k = \text{Argmin}\left\{ \begin{array}{c} \theta(x)-x^TA^T\lambda^k \\ +\frac{1}{2}\beta\|A(x-x^k)\|^2+\frac{1}{2}\delta\|x-x^k\|^2 \end{array} \middle| x\in\mathcal{X} \right\}, \\ \tilde{\lambda}^k=\lambda^k-\beta\big(A[2\tilde{x}^k-x^k]-b\big). \end{cases} \tag{4.4}$$

Then, we use the form

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

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

在 (4.4) 的 x 子问题的目标函数中, 既有非线性函数 $\theta(x)$, 又有非平凡的二次函数, 有时会给求解带来不小的困难!

4.2 均困(均摊困难)的ALM (Balanced ALM) [22]

什么叫均困的ALM, 那是让(4.4)中的x子问题的目标函数只有非线性函数$\theta(x)$和平凡的二次函数$\frac{r}{2}\|x-x^k\|^2$. 把部分困难转移到变量λ的校正.

PPA for the variational inequality (4.1) : Find $\tilde{w}^k \in \Omega$, such that

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (w-\tilde{w}^k)^T H(w^k-\tilde{w}^k), \ \forall w \in \Omega, \tag{4.5a}$$

where

$$H=\begin{pmatrix} rI_n & A^T \\ A & \frac{1}{r}AA^T+\delta I_m \end{pmatrix} \text{ is positive definite.} \tag{4.5b}$$

Then, we use the form

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

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

有下划线的部分就是 $F(\tilde{w}^k)$

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

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

The concrete form of (4.5) is

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{rI_n}(\tilde{x}^k - x^k) + \boldsymbol{A^T}(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \\ \qquad (\underline{A\tilde{x}^k - b}) + \boldsymbol{A}(\tilde{x}^k - x^k) + \boldsymbol{(\frac{1}{r}AA^T + \delta I_m)}\,(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

It can written as

$$\begin{cases} \tilde{x}^k \in \mathcal{X}, \quad \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + r(\tilde{x}^k - x^k)\} \ge 0, \\ \qquad A[(2\tilde{x}^k - x^k) - b] + (\frac{1}{r}AA^T + \delta I_m)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

Thus, the predictor $\tilde{w}^k$ in balanced ALM (4.5) is implemented by

$$\begin{cases} \tilde{x}^k = \arg\min\{\theta(x) - x^TA^T\lambda^k + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(4.6a)} \\ \tilde{\lambda}^k = \arg\min\Big\{\lambda^T\big(A[2\tilde{x}^k - x^k] - b\big) + \frac{1}{2}\|\lambda - \lambda^k\|^2_{(\frac{1}{r}AA^T + \delta I_m)}\Big\}. & \text{(4.6b)} \end{cases}$$

Remark. $\tilde{\lambda}^k$ in (4.6b) is the solution of the following system of linear equations:

$$H_0(\lambda - \lambda^k) + \big(A[2\tilde{x}^k - x^k] - b\big) = 0, \tag{4.7}$$

where

$$H_0 = \frac{1}{r}AA^T + \delta I_m. \tag{4.8}$$

Because the matrix H_0 is positive definite, there are efficient algorithms in literature for solving such a systems of linear equations.

- 均困的增广拉格朗日乘子法, x-子问题 (4.6a) 中的二次项式平凡的, 降低了问题求解的难度.
- λ-子问题 (4.6b) 要求解一个系数矩阵正定的线性方程组. 注意到, 在整个迭代过程中, 我们只要对矩阵 H_0 (see (4.8)) 做一次 Cholesky 分解.
- 请读者比较一下 (4.2b) 和 (4.5b) 中正定矩阵 H 的不同构造形式. 有兴趣的读者可以探索一下其他构造矩阵 H 的方法.

5 两块可分离问题的 ADMM 和 PPA 算法

我们对变分不等式问题

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{5.1}$$

定义了 PPA 算法, 设 H 为对称正定矩阵, H-模下的 PPA 算法的第 k 步从已知的 w^k 出发, 求得的新迭代点 w^{k+1} 使得

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &\geq \; (w - w^{k+1})^T H(w^k - w^{k+1}), \;\; \forall\, w \in \Omega. \end{aligned} \tag{5.2}$$

w^{k+1} 是变分不等式问题 (5.1) 的解的充分必要条件是 (5.2) 中的 $w^k = w^{k+1}$. PPA 算法产生的迭代序列 $\{w^k\}$ 满足

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2, \quad \forall w^* \in \Omega^*. \tag{5.3}$$

并有

$$\|w^k - w^{k+1}\|_H^2 \leq \|w^{k-1} - w^k\|_H^2. \tag{5.4}$$

不等式 (5.3) 和 (5.4) 是 PPA 算法的两条重要而又漂亮的性质.

5.1 ADMM 算法的主要性质

ADMM 是用来求解两块可分离凸优化问题

$$\min\{\theta_1(x) + \theta_2(y) | Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\} \tag{5.5}$$

的有效算法. 我们把问题 (5.5) 转换成变分不等式 (5.1), 其中

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y),$$

$$F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix} \quad \text{和} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m.$$

问题 (5.5) 的增广拉格朗日函数是

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

增广拉格朗日乘子法的第 k 次迭代从给定的 λ^k 开始, 通过

$$\text{(ALM)}\quad \begin{cases} (x^{k+1}, y^{k+1}) = \arg\min\{\mathcal{L}_\beta^{[2]}(x, y, \lambda^k) \mid x \in \mathcal{X}, y \in \mathcal{Y}\}, & \text{(5.6a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & \text{(5.6b)} \end{cases}$$

求得 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$.

用 ALM 方法处理问题 (5.5), 我们把 $v = \lambda$ 叫做核心变量, (x, y) 叫做中间变量.

ADMM 是松弛了的 ALM, k 次迭代从给定的 $v^k = (y^k, \lambda^k)$ 开始, 通过

$$\text{(ADMM)}\quad \begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta^{[2]}(x, y^{k}, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(5.7a)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, & \text{(5.7b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & \text{(5.7c)} \end{cases}$$

求得 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$.

用 ADMM 方法处理问题 (5.5), 我们把 $v = (y, \lambda)$ 是核心变量, x 为做中间变量.

显然, ADMM 就是松弛了的 ALM.

由于改变优化问题中目标函数的常数项对问题的解没有影响, ADMM (5.7) 实际上是通过

$$\begin{cases} x^{k+1} \in \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b) \end{cases} \tag{5.8}$$

求得 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$.

Analysis 根据最优性定理, (5.8) 的 x 和 y 子问题分别满足

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \{-A^T\lambda^k + \beta A^T\big(Ax^{k+1} + By^k - b\big)\} \ge 0, \ \forall\, x \in \mathcal{X} \tag{5.9a}$$

和

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T\lambda^k + \beta B^T\big(Ax^{k+1} + By^{k+1} - b\big)\} \ge 0, \ \forall\, y \in \mathcal{Y}. \tag{5.9b}$$

以 $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$ 代入 (5.9) （消去其中的 λ^k），我们分别得到

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad & \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ & \{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\} \geq 0, \ \forall x \in \mathcal{X}, \end{aligned} \tag{5.10a}$$

和

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \geq 0, \ \forall y \in \mathcal{Y}. \tag{5.10b}$$

将 (5.10) 写成紧凑的形式：$u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$,

$$\begin{aligned} \theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T & \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} \right. \\ & \left. + \beta \begin{pmatrix} A^T B \\ 0 \end{pmatrix} (y^k - y^{k+1}) \right\} \geq 0, \ \forall (x,y) \in \mathcal{X} \times \mathcal{Y}. \end{aligned} \tag{5.11}$$

再把上式改写成

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

$$\left. + \begin{pmatrix} 0 & 0 \\ 0 & \beta B^TB \end{pmatrix} \begin{pmatrix} x^{k+1}-x^k \\ y^{k+1}-y^k \end{pmatrix} \right\} \geq 0, \ \forall (x,y) \in \mathcal{X} \times \mathcal{Y}. \quad (5.12)$$

然后, 我们有如下的引理:

引理 2 对给定的 (y^k,λ^k), 设 $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\Omega$ 是由交替方向法 (5.8) 生成的. 我们有

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

$$\left. + \begin{pmatrix} 0 & 0 \\ \beta B^TB & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} \right\} \geq 0, \ \forall w \in \Omega. \quad (5.13)$$

证明 等式 $(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k) = 0$ 可以改写成

$$\lambda^{k+1} \in \Re^m, \quad (\lambda - \lambda^{k+1})^T \{(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0, \quad \forall \lambda \in \Re^m.$$

将上式加到 (5.13), 就得到引理之结论. □

为了方便, 我们定义

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

得到下面的引理:

引理 3 对给定的 (y^k, λ^k), 设 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ 是由交替方向法 (5.8) 生成的. 我们有

$$(v^{k+1} - v^*)^T H (v^k - v^{k+1}) \geq (y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}), \ \forall\, w^* \in \Omega^*, \quad (5.14)$$

其中

$$H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \quad (5.15)$$

Proof. Setting $w = w^*$ in (5.13), we get

$$
\begin{aligned}
&(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \\
\geq\ & \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} \beta B(y^k - y^{k+1}) \\
& + \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}), \ \forall\, w^* \in \Omega^*. \quad (5.16)
\end{aligned}
$$

Observe the first part of the right hand side of (5.16),

$$
\begin{aligned}
& \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} \beta B(y^k - y^{k+1}) \\
=\ & (y^k - y^{k+1})^T B^T \beta \big(A, B\big) \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix} \\
=\ & (y^k - y^{k+1})^T B^T \beta \big(Ax^{k+1} + By^{k+1} - (Ax^* + By^*)\big) \\
=\ & (y^k - y^{k+1}) B^T \underline{\beta \big(Ax^{k+1} + By^{k+1} - b\big)} \\
=\ & (y^k - y^{k+1}) B^T \underline{(\lambda^k - \lambda^{k+1})}. \quad (5.17)
\end{aligned}
$$

To the second part, since $(w^{k+1}-w^*)^T F(w^{k+1}) = (w^{k+1}-w^*)^T F(w^*)$ and w^* is the optimal solution, it follows that

$$\begin{aligned} &\theta(u^{k+1}) - \theta(u^*) + (w^{k+1}-w^*)^T F(w^{k+1}) \\ = &\quad \theta(u^{k+1}) - \theta(u^*) + (w^{k+1}-w^*)^T F(w^*) \geq 0. \end{aligned} \tag{5.18}$$

The assertion (5.16) immediately. □

引理 4 对给定的 (y^k, λ^k), 设 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ 是由交替方向法 (5.8) 生成的. 我们有

$$(y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}) \geq 0. \tag{5.19}$$

Proof. Because (5.10b) is true for the k-th iteration and the previous iteration, we have

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{5.20}$$

and

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T \{-B^T \lambda^k\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{5.21}$$

Setting $y = y^k$ in (5.20) and $y = y^{k+1}$ in (5.21), respectively, and then adding the two resulting inequalities, we get the assertion (5.19) immediately. □

将 (5.19) 代入 (5.14), 我们得到

$$(v^{k+1}-v^*)^T H(v^k-v^{k+1}) \geq 0, \quad \forall\, v^* \in \mathcal{V}^*. \tag{5.22}$$

在上一讲中我们已经说明：$b^T H(a-b) \geq 0 \quad \Rightarrow \|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2$.
设 $a = v^k - v^*$, $b = v^{k+1} - v^*$, 就有下面的定理.

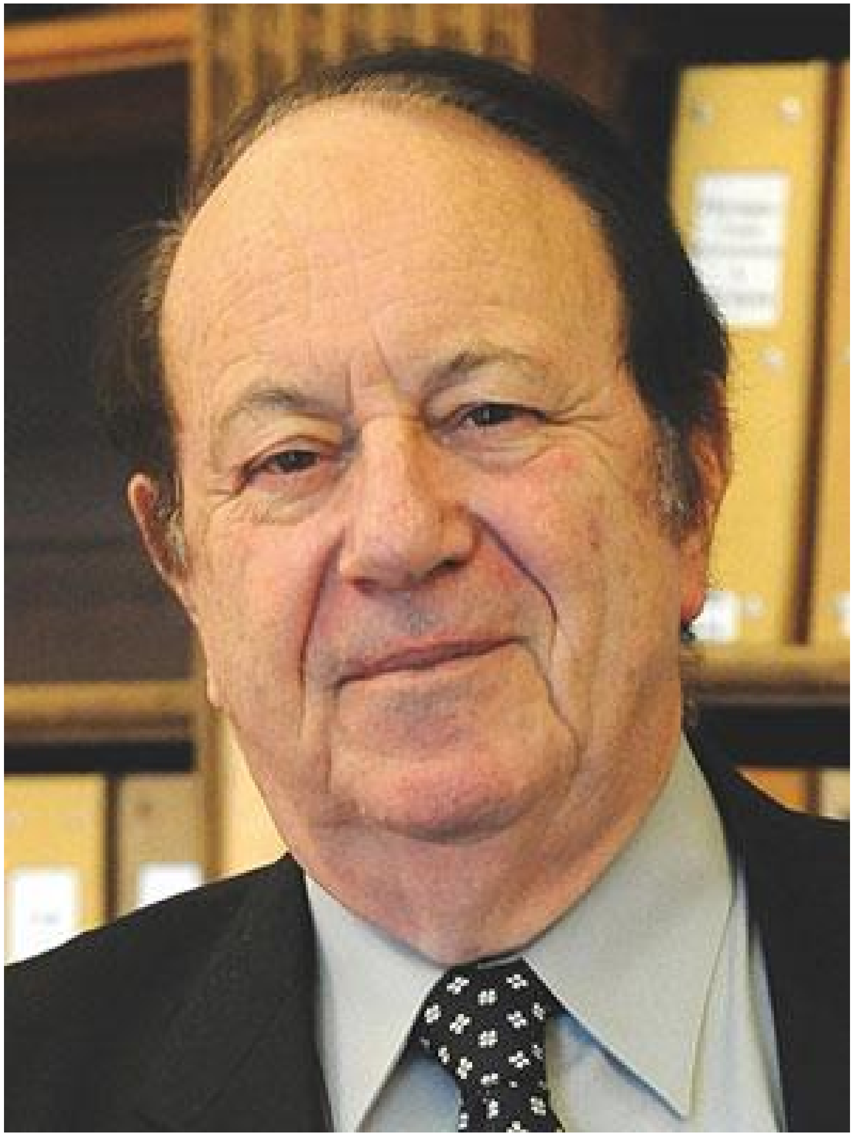
定理 3 对给定的 (y^k, λ^k), 设 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ 是由交替方向法 (5.8) 生成的. 我们有

$$\|\boldsymbol{v}^{k+1}-\boldsymbol{v}^*\|_H^2 \leq \|\boldsymbol{v}^k-\boldsymbol{v}^*\|_H^2 - \|\boldsymbol{v}^k-\boldsymbol{v}^{k+1}\|_H^2, \ \forall\, \boldsymbol{v}^* \in \boldsymbol{\mathcal{V}}^*. \tag{5.23}$$

除此之外, 我们在 [20] 中证明了 ADMM 的迭代序列 $\{v^k\}$ 具备性质

$$\|\boldsymbol{v}^{k+1}-\boldsymbol{v}^{k+2}\|_H^2 \leq \|\boldsymbol{v}^k-\boldsymbol{v}^{k+1}\|_H^2. \tag{5.24}$$

不等式 (5.23) 和 (5.24) 展示了 ADMM 很好的性质. 在一些快速 ADMM 的研究 [8] 中, 都用到了 (5.24) 这条性质.



Roland Glowinski

(1937-2022)

上世纪70 年代中期, 提出了乘子交替方向法, 当时叫做 Algorithm 2. 用来求解下面这类微分方程问题

$$\min\{f(x)+g(Mx)\}$$

他们把它转换成问题

$$\begin{cases} \min & f(x)+g(y), \\ \text{s.t} & Mx-y=0. \end{cases}$$

用交替方向法求解。

我们和Glowinski 从 1998 年开始就有交往 [25]

PERGAMON

Applied Mathematics Letters 13 (2000) 123–130

Applied Mathematics Letters

www.elsevier.nl/locate/aml

A Modified Alternating Direction Method for Convex Minimization Problems

BINGSHENG HE AND JING ZHOU
Department of Mathematics, Nanjing University
Nanjing, 210093, P.R. China

(Received and accepted September 1998)

Communicated by R. Glowinski

我保留的这篇文章的审稿意见一下找不到. 但是 Glowinski 的简短评价和接受意见我还清楚记得.

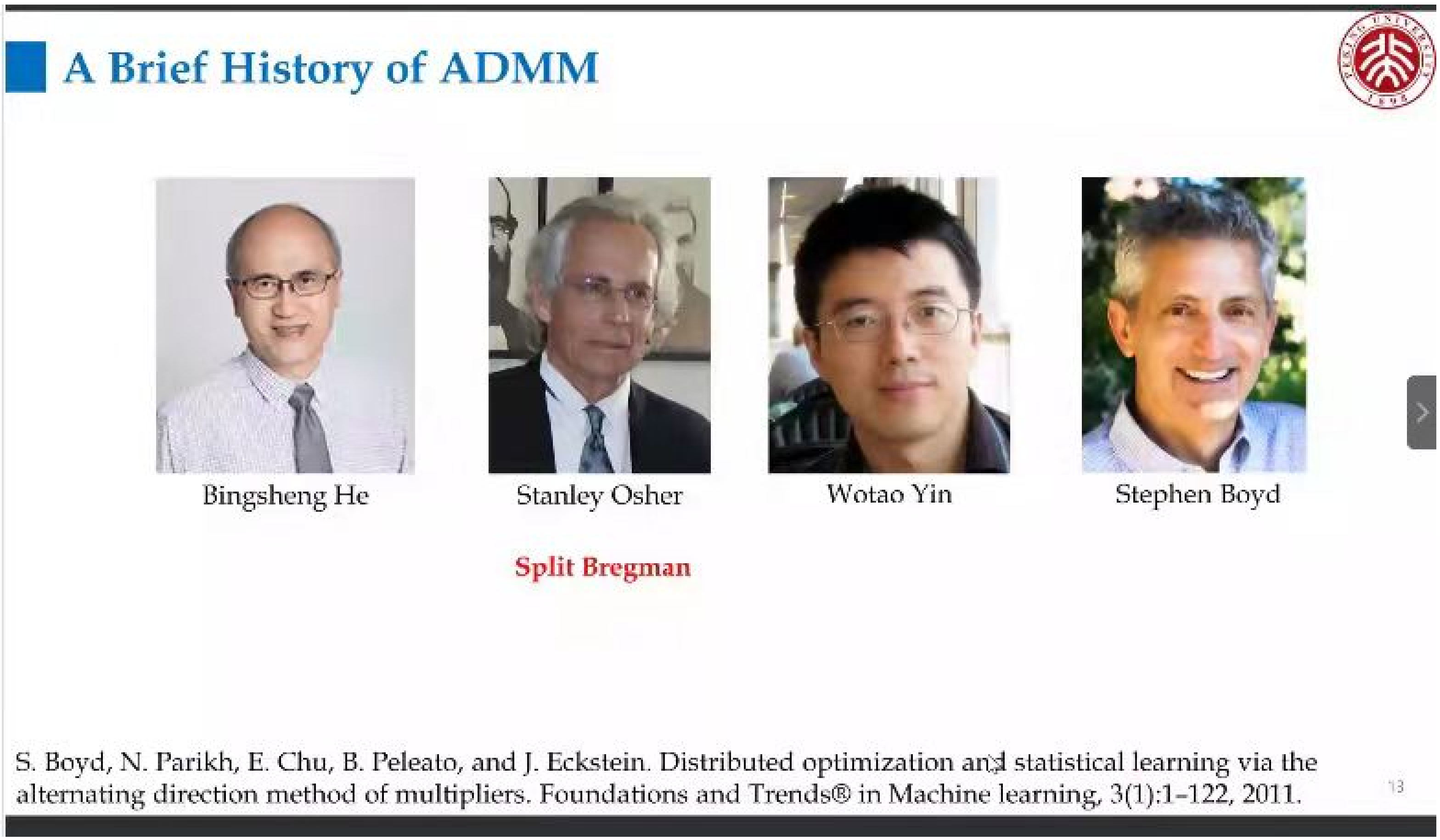
过去十多年, ADMM 方法得到了非常广泛的重视!

由于投入早, 我们做了不少工作, 有理论的, 也有算法的.

我们关于ADMM的工作, 在国内也已经得到越来越多的认可

2021 年7 月, 第三届全国大数据与人工智能科学大会在成都召开.

北京大学智能学院林宙辰教授在大会邀请报告中提到了我的工作。



南方科技大学 2015 级的一位我教过的学生, 那时已经在西安交通大学读硕士, 看到照片上有我, 赶快照了一张照片传给了我。

ADMM 可以求解两块的问题, 三块的问题就必须另想办法了!

5.2 平行求解子问题的 PPA 算法

求解两个可分离块问题 (5.5) 相应的变分不等式 (1.11)-(1.12).
根据 PPA 算法的要求, 设计的右端矩阵为 H 对称正定. $v = w$

假如 PPA 设计成

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \tag{5.1a}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & A^T \\ 0 & \beta B^T B + \delta I_{n_2} & B^T \\ A & B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{5.1b}$$

The both matrices

$$\begin{pmatrix} \beta A^T A + \delta I_{n_1} & A^T \\ A & \frac{1}{\beta} I_m \end{pmatrix} \succ 0, \qquad \begin{pmatrix} \beta B^T B + \delta I_{n_2} & B^T \\ B & \frac{1}{\beta} I_m \end{pmatrix} \succ 0.$$

Because $F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}$, the concrete form of (5.1) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_{n_1})}(\tilde{x}^k - x^k) \boldsymbol{+ A^T}(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{(\beta B^T B + \delta I_{n_2})}(\tilde{y}^k - y^k) \boldsymbol{+ B^T}(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b}) \boldsymbol{+ A}(\tilde{x}^k - x^k) \boldsymbol{+ B}(\tilde{y}^k - y^k) + \boldsymbol{(2/\beta)}\,(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

After simple organization, we obtain

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + (\beta A^T A + \delta I_{n_1})(\tilde{x}^k - x^k)\} \ge 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\lambda^k + (\beta B^T B + \delta I_{n_2})(\tilde{y}^k - y^k)\} \ge 0, \\ [2(A\tilde{x}^k + B\tilde{y}^k - b) - (Ax^k + By^k - b)] + (2/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

In fact, the prediction can be arranged by

$$\left\{\begin{array}{ll} \tilde{x}^k = \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T \lambda^k \\ +\frac{1}{2}\beta\|A(x-x^k)\|^2 + \frac{1}{2}\delta\|x-x^k\|^2 \end{array} \middle| x \in \mathcal{X} \right\} & \text{(5.2a)} \\ \tilde{y}^k = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k \\ +\frac{1}{2}\beta\|B(y-y^k)\|^2 + \frac{1}{2}\delta\|y-y^k\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} & \text{(5.2b)} \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{2}\beta\big[2(A\tilde{x}^k + B\tilde{y}^k - b) - (Ax^k + By^k - b)\big] & \text{(5.2c)} \end{array}\right.$$

$$\left\{\begin{array}{l} \tilde{x}^k = \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}(x-x^k)^T(\beta A^T A + \delta I_{n_1})(x-x^k) | x \in \mathcal{X}\} \\ \tilde{y}^k = \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}(y-y^k)^T(\beta B^T B + \delta I_{n_2})(y-y^k) | y \in \mathcal{Y}\} \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{2}\beta\big[2(A\tilde{x}^k + B\tilde{y}^k - b) - (Ax^k + By^k - b)\big] \end{array}\right.$$

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

利用变分不等式 (VI) 和邻近点算法 (PPA), 更自由地设计 ADMM 类分裂收缩算法

6 凸优化分裂收缩算法的统一框架

我们总是用变分不等式 (VI) 指导算法设计, 把线性约束的凸优化问题归结为下面的变分不等式:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{6.1}$$

Algorithms in a unified framework

A unified Algorithmic Framework for (6.1)

统一框架由预测-校正两部分组成

[Prediction Step.] 从给定的 v^k 出发, 求得预测点 $\tilde{w}^k \in \Omega$ 使其满足

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k),\ \forall w \in \Omega, \tag{6.2a}$$

其中 Q 不一定对称, 但是 $Q^T + Q$ 正定.

[Correction Step.] 给一个合适的非奇异矩阵 M, 由下式确定新的迭代点

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{6.2b}$$

Q 和 M 分别叫做预测矩阵和校正矩阵

Convergence Conditions

对算法框架 (6.2) 中的预测矩阵 Q 和校正矩阵 M, 存在正定矩阵 H, 使得

$$HM = Q, \tag{6.3a}$$

并且

$$G = Q^T + Q - M^T HM \succ 0. \tag{6.3b}$$

预测-校正方法的例子

We consider the $\min - \max$ problem

$$\min_x \max_y \{\Phi(x,y) = \theta_1(x) - y^T Ax - \theta_2(y) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{6.4}$$

Furthermore, it can be written as a variational inequality in the compact form:

$$u \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T F(u^*) \geq 0, \ \forall\, u \in \Omega, \tag{6.5}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad F(u) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}, \quad \Omega = \mathcal{X} \times \mathcal{Y}.$$

The output of Original PDHG algorithm [32] as predictor

For given (x^k, y^k), PDHG [32] produces a pair of $(\tilde{x}^k, \tilde{y}^k)$. First,

$$\tilde{x}^k = \arg\min\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, \tag{6.6a}$$

and then we obtain $\tilde{y}^k$ via

$$\tilde{y}^k = \arg\max\{\Phi(\tilde{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \tag{6.6b}$$

The output $\tilde{u}^k \in \Omega$,

$$\theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T\{F(\tilde{u}^k) + Q(\tilde{u}^k - u^k)\} \geq 0, \;\forall u \in \Omega, \tag{6.7a}$$

where

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}. \tag{6.7b}$$

对于这样的预测, 我们考虑比较简单的校正

$$u^{k+1} = u^k - M(u^k - \tilde{u}^k) \tag{6.8}$$

校正. 其中 M 为单位上三角矩阵或单位下三角矩阵. 收敛性条件 (6.3)

- $H \succ 0$ and $HM = Q$.
- $G = Q^T + Q - M^T HM \succ 0$.

可以改写成等价的

(i) $H \succ 0$ and $H = QM^{-1}$.

(ii) $G = Q^T + Q - Q^T M \succ 0$.

一. 校正矩阵 M 为单位下三角矩阵

其中的 K 是待定的.

$$M = \begin{pmatrix} I_n & 0 \\ K & I_m \end{pmatrix} \quad \text{则} \quad M^{-1} = \begin{pmatrix} I_n & 0 \\ -K & I_m \end{pmatrix}.$$

对条件 (i), 我们在统一框架下指导下求出这个 K 的具体形式. 由于 $H = QM^{-1}$ 正定, 首先必须是对称的. 由

$$H = QM^{-1} = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -K & I_m \end{pmatrix} = \begin{pmatrix} rI_n - A^T K & A^T \\ -sK & sI_m \end{pmatrix}$$

必须对称, 推得

$$-sK = A, \quad \Rightarrow \quad K = -\frac{1}{s}A.$$

因此,

$$M = \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix}, \quad H = \begin{pmatrix} rI_n + \frac{1}{s}A^TA & A^T \\ A & sI_m \end{pmatrix}.$$

对任意的 $r, s > 0$, 矩阵 H 是正定的.

对条件 (ii),

$$\begin{aligned} G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix} \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}. \end{aligned}$$

要矩阵 G 正定, 必须有 $rs > \|A^TA\|$.

采用 PDHG 预测, 前述单位下三角矩阵校正, 当 $rs > \|A^TA\|$, 收敛性条件满足.

二. 校正矩阵 M 为单位上三角矩阵

同样, 其中的 K 是待定的.

$$M = \begin{pmatrix} I_n & K \\ 0 & I_m \end{pmatrix} \quad \text{则} \quad M^{-1} = \begin{pmatrix} I_n & -K \\ 0 & I_m \end{pmatrix}.$$

对条件 (i), 我们在统一框架下指导下求出这个 K 的具体形式. 由于 $H = QM^{-1}$ 正定, 首先必须是对称的. 由

$$H = QM^{-1} = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & -K \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} rI_n & -rK + A^T \\ 0 & sI_m \end{pmatrix}$$

必须对称, 推得

$$rK = A^T, \qquad \Rightarrow \qquad K = \frac{1}{r} A^T.$$

因此,

$$M = \begin{pmatrix} I_n & \frac{1}{r} A^T \\ 0 & I_m \end{pmatrix}, \qquad H = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix}.$$

对任意的 $r, s > 0$, 矩阵 H 是正定的.

而对条件 (ii),

$$
\begin{aligned}
G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\
&= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} \\
&= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \\
&= \begin{pmatrix} rI_n & 0 \\ 0 & sI_m - \frac{1}{r}AA^T \end{pmatrix}.
\end{aligned}
$$

要矩阵 G 正定, 必须有 $rs > \|A^T A\|$.

采用 PDHG 预测, 前述单位上三角矩阵校正, 当 $rs > \|A^T A\|$, 收敛性条件满足.

预测-校正的统一框架, 把不能保证收敛的 PDHG 方法改造成了收敛的方法.

马上我们就会看到, 满足收条件 (6.3) 的预测-校正算法 (6.2) 的收敛性证明非常容易. 如果故意将有些确定的算法 [13, 15, 16] 转换成统一框架 (6.2) 中的预测-校正算法, 然后用 (6.3) 中的两条收敛性条件去验证, 证明就特别简单统一. 对此, 我们在中文综述文章 [10] 中举了大量的例子.

6.1 Convergence proof in the unified framework

In this stage, assuming the conditions (6.3) in the unified framework are satisfied, we prove some convergence properties.

定理 1 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (6.1) *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*

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\\ \geq\quad&\frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{1}{2}\|v^k-\tilde{v}^k\|_G^2,\quad\forall w\in\Omega.\end{aligned} \tag{6.9}$$

Proof. Using $Q=HM$ (see (6.3a)) and the relation (6.2b), the right hand side of (6.3a) can be written as $(v-\tilde{v}^k)^TH(v^k-v^{k+1})$ and hence

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq(v-\tilde{v}^k)^TH(v^k-v^{k+1}),\ \forall w\in\Omega. \tag{6.10}$$

Applying the identity

$$Q(v^k-\tilde{v}^k)=HM(v^k-\tilde{v}^k)=H(v^k-v^{k+1}).$$

$$(a-b)^TH(c-d)=\frac{1}{2}\{\|a-d\|_H^2-\|a-c\|_H^2\}+\frac{1}{2}\{\|c-b\|_H^2-\|d-b\|_H^2\}, \tag{6.11}$$

to the right hand side of (6.10) with

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

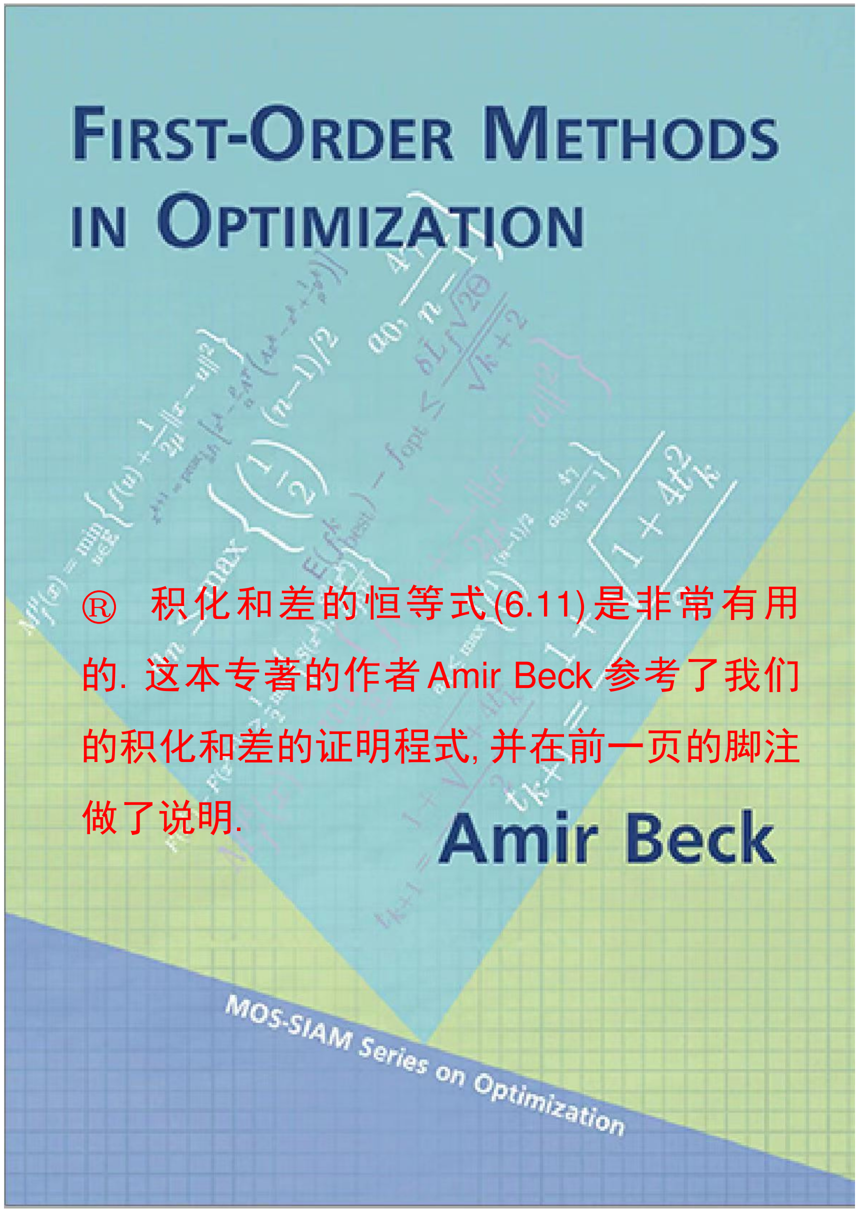
we thus obtain

$$\begin{aligned} & 2(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ & \quad = \big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + (\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \end{aligned} \tag{6.12}$$

For the last term of (6.12), using $HM = Q$ and $2v^T Qv = v^T(Q^T + Q)v$, we have

$$\begin{aligned} & \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ & \quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ & \quad \overset{(6.3a)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ & \quad = \quad 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ & \quad = \quad (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\ & \quad \overset{(6.3b)}{=} \quad \|v^k - \tilde{v}^k\|_G^2. \end{aligned} \tag{6.13}$$

Substituting (6.12), (6.13) in (6.10), the assertion of this theorem is proved. $\square$



® 积化和差的恒等式(6.11)是非常有用的. 这本专著的作者 Amir Beck 参考了我们的积化和差的证明程式, 并在前一页的脚注做了说明.

We will use the following notation:

$$\tilde{\mathbf{x}}^k = \mathbf{x}^{k+1},\quad \tilde{\mathbf{z}}^k = \mathbf{z}^{k+1},\quad \tilde{\mathbf{y}}^k = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k - \mathbf{c}).$$

Using (15.15), (15.16), the subgradient inequality, and the above notation, we obtain that for any $\mathbf{x} \in \mathrm{dom}(h_1)$ and $\mathbf{z} \in \mathrm{dom}(h_2)$,

$$h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \left\langle \rho\mathbf{A}^T\left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right) + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \right\rangle \geq 0,$$
$$h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \left\langle \rho\mathbf{B}^T\left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right) + \mathbf{Q}(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \right\rangle \geq 0.$$

Using the definition of $\tilde{\mathbf{y}}^k$, the above two inequalities can be rewritten as

$$h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \langle \mathbf{A}^T\tilde{\mathbf{y}}^k + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \rangle \geq 0,$$
$$h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \langle \mathbf{B}^T\tilde{\mathbf{y}}^k + (\rho\mathbf{B}^T\mathbf{B} + \mathbf{Q})(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \rangle \geq 0.$$

Adding the above two inequalities and using the identity

$$\mathbf{y}^{k+1} - \mathbf{y}^k = \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c}),$$

we can conclude that for any $\mathbf{x} \in \mathrm{dom}(h_1)$, $\mathbf{z} \in \mathrm{dom}(h_2)$, and $\mathbf{y} \in \mathbb{R}^m$,

$$H(\mathbf{x},\mathbf{z}) - H(\tilde{\mathbf{x}}^k, \tilde{\mathbf{z}}^k) + \left\langle \begin{pmatrix} \mathbf{x} - \tilde{\mathbf{x}}^k \\ \mathbf{z} - \tilde{\mathbf{z}}^k \\ \mathbf{y} - \tilde{\mathbf{y}}^k \end{pmatrix}, \begin{pmatrix} \mathbf{A}^T\tilde{\mathbf{y}}^k \\ \mathbf{B}^T\tilde{\mathbf{y}}^k \\ -\mathbf{A}\tilde{\mathbf{x}}^k - \mathbf{B}\tilde{\mathbf{z}}^k + \mathbf{c} \end{pmatrix} - \begin{pmatrix} \mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) \\ \mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) \\ \frac{1}{\rho}(\mathbf{y}^k - \mathbf{y}^{k+1}) \end{pmatrix} \right\rangle \geq 0, \tag{15.17}$$

where $\mathbf{C} = \rho\mathbf{B}^T\mathbf{B} + \mathbf{Q}$. We will use the following identity that holds for any positive semidefinite matrix $\mathbf{P}$:

$$(\mathbf{a} - \mathbf{b})^T\mathbf{P}(\mathbf{c} - \mathbf{d}) = \frac{1}{2}\left(\|\mathbf{a} - \mathbf{d}\|_{\mathbf{P}}^2 - \|\mathbf{a} - \mathbf{c}\|_{\mathbf{P}}^2 + \|\mathbf{b} - \mathbf{c}\|_{\mathbf{P}}^2 - \|\mathbf{b} - \mathbf{d}\|_{\mathbf{P}}^2\right).$$

Using the above identity, we can conclude that

$$\begin{aligned}(\mathbf{x} - \tilde{\mathbf{x}}^k)^T\mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) &= \frac{1}{2}\left(\|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2 + \|\tilde{\mathbf{x}}^k - \mathbf{x}^k\|_{\mathbf{G}}^2\right)\\ &\geq \frac{1}{2}\|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2, \end{aligned} \tag{15.18}$$

as well as

$$(\mathbf{z} - \tilde{\mathbf{z}}^k)^T\mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) = \frac{1}{2}\|\mathbf{z} - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 - \frac{1}{2}\|\mathbf{z} - \mathbf{z}^k\|_{\mathbf{C}}^2 + \frac{1}{2}\|\mathbf{z}^k - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 \tag{15.19}$$

and

$$\begin{aligned}&2(\mathbf{y} - \tilde{\mathbf{y}}^k)^T(\mathbf{y}^k - \mathbf{y}^{k+1})\\ &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \|\tilde{\mathbf{y}}^k - \mathbf{y}^k\|^2 - \|\tilde{\mathbf{y}}^k - \mathbf{y}^{k+1}\|^2\\ &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2\\ &-\|\mathbf{y}^k + \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}) - \mathbf{y}^k - \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c})\|^2\\ &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2 - \rho^2\|\mathbf{B}(\mathbf{z}^k - \tilde{\mathbf{z}}^k)\|^2.\end{aligned}$$

Strict contraction $\{\|v^k - v^*\|_H^2\}$ 严格下降

定理 2 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (6.1) *and* $\tilde{w}^k$ *is obtained in the* k*-th iteration. If* v^k, v^{k+1} *and* $\tilde{w}^k$ *satisfy the conditions in the unified framework, then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{6.14}$$

Proof. Setting $w = w^*$ in (6.9), we get

$$\begin{aligned} & \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ \ge & \quad \|v^k - \tilde{v}^k\|_G^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \end{aligned} \tag{6.15}$$

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

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \|v^k - \tilde{v}^k\|_G^2. \tag{6.16}$$

The assertion (6.14) follows directly. $\square$

Monotonicity of the series $\{\|v^k - v^{k+1}\|_H^2\}$ 单调不增

定理 3 *For solving the variational inequality* (6.1), *let* $\{w^k\}, \{\tilde{w}^k\}$ *be the sequence generated by* (6.2). *If the conditions* (6.3) *are satisfied, then we have*

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{6.17}$$

Proof Note that for any $w \in \Omega$, we have

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k),$$

and

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

Set the vector w in the above two inequalities by $\tilde{w}^{k+1}$ and $\tilde{w}^k$, respectively, we get

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \ge (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k)$$

and

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

Adding the above two inequalities, it follows that

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

Then, adding $\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}^TQ\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}$ to the both sides of the inequality $(\tilde{v}^k-\tilde{v}^{k+1})^TQ\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}\ge 0$, we get

$$(v^k-v^{k+1})^TQ\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}\ge \frac{1}{2}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)},$$

and thus

利用 $Q=HM$, 然后 $M(v^k-\tilde{v}^k)=(v^k-v^{k+1})$.

$$(v^k-v^{k+1})^TH\{(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\}\ge \frac{1}{2}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \tag{6.18}$$

Finally, by using $\|a\|_H^2-\|b\|_H^2=2a^TH(a-b)-\|a-b\|_H^2$ and (6.18), we get

$$\begin{aligned}
&\|v^k-v^{k+1}\|_H^2-\|v^{k+1}-v^{k+2}\|_H^2\\
&= 2(v^k-v^{k+1})^TH\{(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\}\\
&\quad -\|(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\|_H^2\\
&\ge \|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}-\|(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\|_H^2\\
&= \|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q-M^THM)}\\
&= \|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_G.
\end{aligned}$$

This is the equivalent form of (6.17) and the proof is complete. □

6.2 How to construct the correction matrix M

只要统一框架的预测 (6.2a) 中的预测矩阵 Q 满足

$$Q^T + Q \succ 0,$$

我们总可以取

$$0 \prec G \prec Q^T + Q.$$

然后记

$$D = (Q^T + Q) - G,$$

则 $D \succ 0$. 令

$$M^T H M = D.$$

由矩阵方程组解得

$$\left\{ \begin{array}{r} HM = Q, \\ M^T HM = D. \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{l} HM = Q, \\ Q^T M = D. \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{l} H = QD^{-1}Q^T, \\ M = Q^{-T}D. \end{array} \right. \tag{6.19}$$

就得到满足收敛条件的校正矩阵 M.

实际计算中, 我们只要校正矩阵 M.

H 和 G 只是用来验证收敛条件的.

换句话说, 只要

$$Q^T + Q \succ 0.$$

我们就可以选两个正定矩阵 $D \succ 0$ 和 $G \succ 0$, 使得

$$D \succ 0, \quad G \succ 0, \qquad \text{并且有} \qquad D + G = Q^T + Q.$$

这里可以有无穷多的选择! 将 (6.2b) 中的校正矩阵 M 取成

$$M = Q^{-T} D$$

条件 (6.3) 自然满足.

校正公式 (6.2b) 就是

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

可以通过

$$Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k) \qquad \text{来实现.}$$

6.3 预测-校正的广义 PPA 算法

求解变分不等式 (5.1) 采用单位步长校正的时候, 如果预测公式

$$\tilde{w}^k \in \Omega, \quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k), \quad \forall\, w \in \Omega, \tag{6.20}$$

中的预测矩阵 Q 满足 $Q^T+Q \succ 0$, 若将 Q^T+Q 分拆成

$$D \succ 0, \qquad G \succ 0 \qquad \text{和} \qquad D+G=Q^T+Q, \tag{6.21}$$

再令

$$M = Q^{-T}D \qquad \text{和} \qquad H = QD^{-1}Q^T. \tag{6.22}$$

则由单位步长校正

$$v^{k+1} = v^k - M(v^k-\tilde{v}^k) \tag{6.23}$$

产生的新的迭代序列 $\{v^k\}$ 满足

$$\|v^{k+1}-v^*\|_H^2 \leq \|v^k-v^*\|_H^2 - \|v^k-\tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{6.24}$$

如果我们采用一对特殊的 D 和 G, 使得

$$D = G = \frac{1}{2}(Q^T + Q),$$

那么, (6.24) 就变成了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_D^2, \quad \forall v^* \in \mathcal{V}^*. \tag{6.25}$$

对选定的 D, 根据 (6.22), 总有

$$M^T H M = D,$$

因此, (6.25) 就成了

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

再利用 $M(v^k - \tilde{v}^k) = v^k - v^{k+1}$ (见 (6.23)), 上式就了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{6.26}$$

此外, 关于统一框架中所有固定步长的方法 (见这一讲的定理 3) 都证明了

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{6.27}$$

我们把上述分析结果写成下面的定理.

定理 4 用预测校正方法求解变分不等式 (5.1), 设预测 (6.20) 中的预测矩阵 Q 满足 $Q^T + Q \succ 0$. 若令

$$D = \frac{1}{2}(Q^T + Q), \qquad \text{和} \qquad M = Q^{-T} D$$

则由单位步长校正公式

$$v^{k+1} = v^k - Q^{-T} D(v^k - \tilde{v}^k) \tag{6.28}$$

产生的新的迭代点具有性质 (6.26) 和 (6.27), 其中

$$H = Q[\tfrac{1}{2}(Q^T + Q)]^{-1} Q^T.$$

求解变分不等式 (5.1), 我们把迭代序列具有性质 (6.26) 和 (6.27) 的方法, 称为广义 PPA 方法. 在实际计算中, 我们并不要求显式写出 H 的表达式.

7 p-块可分离凸优化问题和统一框架算法求解

p-块可分离凸优化问题

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\ \Big|\ \sum_{i=1}^{p}A_ix_i=b\ (\text{or}\ \ge b),\ \ x_i\in\mathcal{X}_i\Big\}. \tag{7.1}$$

The Lagrangian function is

$$L(x_1,\dots,x_p,\lambda)=\sum_{i=1}^{p}\theta_i(x_i)-\lambda^T\Big(\sum_{i=1}^{p}A_ix_i-b\Big),$$

which is defined on $\Omega=\prod_{i=1}^{p}\mathcal{X}_i\times\Lambda$, where

$$\Lambda=\begin{cases}\Re^m, & \text{if } \sum_{i=1}^{p}A_ix_i=b,\\[2ex] \Re^m_+, & \text{if } \sum_{i=1}^{p}A_ix_i\ge b.\end{cases}$$

Let $(x_1^*, \ldots, x_p^*, \lambda^*) \in \Omega$ be a saddle point of the Lagrangian function, then

$$L_{\lambda\in\Lambda}(x_1^*, \ldots, x_p^*, \lambda) \le L(x_1^*, \ldots, x_p^*, \lambda^*) \le L_{x_i\in\mathcal{X}_i}(x_1, \ldots, x_p, \lambda^*).$$

The optimality condition of (7.1) can be written as the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{7.2a}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \tag{7.2b}$$

and

$$\theta(x) = \sum_{i=1}^p \theta_i(x_i), \qquad \Omega = \prod_{i=1}^p \mathcal{X}_i \times \Lambda.$$

Again, we denote by Ω^* the solution set of the VI (7.2).

多块问题 (7.2) 的 PRIMAL-DUAL 递推式串行预测 Prediction

从给定的 $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 到预测点 $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_p^k, \tilde{\lambda}^k)$:

Prediction Step. With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$:

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\
\qquad\vdots \\
\tilde{x}_i^k \in \arg\min_{x_i \in \mathcal{X}_i}\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\
\qquad\vdots \\
\tilde{x}_p^k \in \arg\min_{x_p \in \mathcal{X}_p}\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p} A_j\tilde{x}_j^k - b\big)\big].
\end{array}\right. \tag{7.3}$$

预测先原始再对偶. 对可分离的原始变量子问题逐一按序求解.

7.1 采用 Primal-Dual 预测的预测矩阵

Analysis for the P-D Prediction 我们先看 (7.3) 中 x 子问题

$$\tilde{x}_i^k \in \arg\min\Big\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 \,|\, x_i \in \mathcal{X}_i\Big\}.$$

根据最优性引理, 最优性条件是 $\tilde{x}_i^k \in \mathcal{X}_i$ 和

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\Big\{-A_i^T\lambda^k + \beta A_i^T\Big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\Big)\Big\} \ge 0, \quad \forall x_i \in \mathcal{X}_i.$$

它可以改写成 $\tilde{x}_i^k \in \mathcal{X}_i$ 和对所有的 $x_i \in \mathcal{X}_i$ 都有

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T\Big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\Big) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0. \tag{7.4a}$$

预测的对偶部分 $\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^p A_j\tilde{x}_j^k - b\big)\big]$, 等价形式

$$\tilde{\lambda}^k = \arg\min\Big\{\big\|\lambda - \big[\lambda^k - \beta\big(\textstyle\sum_{j=1}^p A_j\tilde{x}_j^k - b\big)\big]\big\|^2 \,\big|\, \lambda \in \Lambda\Big\}.$$

最优性条件是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \{ (\underline{\textstyle\sum_{j=1}^p A_j \tilde{x}_j^k - b}) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \} \geq 0, \quad \forall \lambda \in \Lambda. \tag{7.4b}$$

Summating (7.4a) and (7.4b), for the predictor $\tilde{w}^k$ generated by (7.3), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \geq (w - \tilde{w}^k)^T Q (w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \tag{7.5a}$$

where

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{7.5b}$$

7.2 变量代换下的预测矩阵

The optimization problem (7.1) has been translated to VI (7.2), namely,

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

For the easy analysis, we need to denote the following notations:

$$P = \begin{pmatrix} \sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta})I_m \end{pmatrix}, \quad z = Pw = \begin{pmatrix} \sqrt{\beta}A_1 x_1 \\ \sqrt{\beta}A_2 x_2 \\ \vdots \\ \sqrt{\beta}A_p x_p \\ (1/\sqrt{\beta})\lambda \end{pmatrix}. \tag{7.6}$$

Accordingly, we define

$$\mathcal{Z} = \left\{ z \mid z = Pw,\ w \in \Omega \right\},$$

and

$$\mathcal{Z}^* = \left\{ z^* \mid z^* = Pw^*,\ w^* \in \Omega^* \right\}.$$

Using the notation P in (7.6), for the matrix Q in (7.5b), we have

$$Q = P^T \mathcal{Q} P, \qquad \text{where} \qquad \mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{7.7}$$

Thus, for the right hand side of (7.5a), we have

$$\begin{aligned}(w - \tilde{w}^k)^T Q (w^k - \tilde{w}^k) &= (w - \tilde{w}^k)^T P^T \mathcal{Q} P (w^k - \tilde{w}^k) \\ &= (z - \tilde{z}^k)^T \mathcal{Q} (z^k - \tilde{z}^k).\end{aligned}$$

Then, it follows from (7.5) that we have the following VI for the P-D prediction:

$$\begin{aligned}\tilde{w}^k \in \Omega, \ \ \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \geq (z - \tilde{z}^k)^T \mathcal{Q} (z^k - \tilde{z}^k), \ \ \forall\, w \in \Omega.& \end{aligned} \tag{7.8}$$

where $\mathcal{Q}$ is given in (7.7).

7.3 变量替换下的广义 PPA 算法

仍然考虑线性约束的多块可分离凸优化问题. 这些方法的第 k-步迭代从给定的 $(A_1x_1^k,\ldots,A_px_p^k,\lambda^k)$ 出发, 生成预测点 $\tilde{w}^k$ 满足

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq(w-\tilde{w}^k)^TQ(w^k-\tilde{w}^k),\quad\forall\, w\in\Omega. \tag{7.9}$$

作为合格的预测, 其中的矩阵 Q^T+Q 往往只是本质上正定的. 利用上一讲的变换, 把预测 (7.9) 改写成 $\tilde{w}^k\in\Omega$,

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq(z-\tilde{z}^k)^T\mathcal{Q}(z^k-\tilde{z}^k),\quad\forall\, w\in\Omega, \tag{7.10}$$

其中 $Q=P^T\mathcal{Q}P$,

$$\mathcal{Q}^T+\mathcal{Q}\succ 0 \tag{7.11}$$

是正定矩阵. 在 $\mathcal{Q}$ 非对称但 (7.11) 满足的时候, 必须采用必要的校正. 我们总可以选两个矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 使得

$$\mathcal{D}\succ 0,\quad \mathcal{G}\succ 0,\quad 和\quad \mathcal{D}+\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}. \tag{7.12}$$

根据前一节的分析, 我们有如下的定理.

定理 5 设预测点 $\tilde{z}^k$ 满足条件 (7.10), 其中 $\mathcal{Q}^T+\mathcal{Q}$ 是正定矩阵. 如果由两个正

定矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 使得 (7.12) 成立.

$$\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D} \tag{7.13}$$

那么, 利用矩阵 (7.13) 校正

$$z^{k+1} = z^k - \mathcal{M}(z^k - \tilde{z}^k), \tag{7.14}$$

产生的 z^{k+1} 满足

$$\|z^{k+1} - z^*\|_{\mathcal{H}}^2 \le \|z^k - z^*\|_{\mathcal{H}}^2 - \|z^k - \tilde{z}^k\|_{\mathcal{G}}^2, \quad \forall\, z^* \in \Xi^*, \tag{7.15}$$

其中矩阵 $\mathcal{H} = \mathcal{Q}\mathcal{D}^{-1}\mathcal{Q}^T$.

如果选

$$\mathcal{D} = \mathcal{G} = \frac{1}{2}(\mathcal{Q}^T + \mathcal{Q}) \tag{7.16}$$

那么, (7.15) 就变成了

$$\|z^{k+1} - z^*\|_{\mathcal{H}}^2 \le \|z^k - z^*\|_{\mathcal{H}}^2 - \|z^k - \tilde{z}^k\|_{\mathcal{D}}^2, \quad \forall z^* \in \mathcal{Z}^*.$$

对选定的 $\mathcal{D}$, 根据 $\mathcal{D} = \mathcal{M}^T\mathcal{H}\mathcal{M}$, 并利用 (7.14), 上式就成了

$$\|z^{k+1} - z^*\|_{\mathcal{H}}^2 \le \|z^k - z^*\|_{\mathcal{H}}^2 - \|z^k - z^{k+1}\|_{\mathcal{H}}^2, \quad \forall z^* \in \mathcal{Z}^*. \tag{7.17}$$

下面证明收敛性的另一条重要性质: 序列 $\{\|z^k - z^{k+1}\|_{\mathcal{H}}\}$ 是单调不增的.

定理 6 如果预测点 $\tilde{z}^k$ 满足条件 (7.10), 那么, 由校正 (7.14) 产生的新的迭代点 z^{k+1} 满足

$$\|z^{k+1} - z^{k+2}\|_{\mathcal{H}}^2 \le \|z^k - z^{k+1}\|_{\mathcal{H}}^2. \tag{7.18}$$

不等式 (7.17) 和 (7.18) 说明, 变量替换下的广义 PPA 算法同样具备和 PPA 算法的性质 (5.3) 和 (5.4).

在广义邻近点算法(Generalized PPA)中, 校正矩阵 $\mathcal{M}$ 是由 (7.10) 中的预测矩阵 $\mathcal{Q}$ 唯一确定的. 如果 (7.10) 中的 $\mathcal{Q}$ 是对称的, 根据相关的定义, 校正矩阵为

$$M = \tfrac{1}{2}(I + Q^{-T}Q) \qquad \text{或} \qquad \mathcal{M} = \tfrac{1}{2}\big(\mathcal{I} + \mathcal{Q}^{-T}\mathcal{Q}\big), \tag{7.19}$$

就是单位矩阵. 我们将校正矩阵并非单位矩阵, 迭代序列又具备 (5.3)-(5.4) 这类性质的算法, 称为广义邻近点算法.

7.4 基于秩一校正的广义 PPA 算法

前一节介绍的方法, 预测产生的 $\mathcal{Q}$ 矩阵是一个容易求逆的矩阵与一个广义秩

一矩阵的和. 我们对这样的串型预测, 给出广义邻近点算法的校正公式.

设预测是由 Primal-Dual 预测给出的, 我们得到形如 (7.10) 的变分不等式, 其中

$$\mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{7.20}$$

利用记号

$$\mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}, \qquad \mathcal{E} = \begin{pmatrix} I_m \\ I_m \\ \vdots \\ I_m \end{pmatrix}$$

那么有

$$\mathcal{Q}=\begin{pmatrix}\mathcal{L} & \mathcal{E}\\ 0 & I_m\end{pmatrix}. \tag{7.21}$$

由于

$$\mathcal{Q}^T=\begin{pmatrix}\mathcal{L}^T & 0\\ \mathcal{E}^T & I_m\end{pmatrix} \quad \text{和} \quad \mathcal{Q}^{-T}=\begin{pmatrix}\mathcal{L}^{-T} & 0\\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m\end{pmatrix}.$$

对根据 (7.12) 选择的 $\mathcal{D}$, 校正矩阵 $\mathcal{M}=\mathcal{Q}^{-T}\mathcal{D}$ 都是容易计算的. 特别地, 当

$$\mathcal{D}=\frac{1}{2}\big(\mathcal{Q}^T+\mathcal{Q}\big),$$

$$\mathcal{M}_{PPA}=\frac{1}{2}\mathcal{Q}^{-T}\big(\mathcal{Q}^T+\mathcal{Q}\big)$$

最后得到

$$\mathcal{M}_{PPA} = \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & I_m \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{7.22}$$

利用前一节的变换, 采用 (7.22) 中的矩阵 $\mathcal{M}_{PPA}$ 的校正 (7.14) 可以写成等价的

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & \frac{1}{\beta}I_m \\ -\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix}\begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \tag{7.23}$$

把预测-校正方法写在一起：先根据提供的$(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$进行**预测**:

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\big\{\theta_1(x_1) - x_1^TA_1^T\lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\big\};\\[2mm]
\tilde{x}_2^k \in \arg\min\big\{\theta_2(x_2) - x_2^TA_2^T\lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\big\};\\[2mm]
\qquad\quad\vdots\\[2mm]
\tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\big\{\theta_i(x_i) - x_i^TA_i^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1}A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\big\};\\[2mm]
\qquad\quad\vdots\\[2mm]
\tilde{x}_p^k \in \arg\min_{x_p\in\mathcal{X}_p}\big\{\theta_p(x_p) - x_p^TA_p^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\big\};\\[2mm]
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p}A_j\tilde{x}_j^k - b\big)\big].
\end{array}\right.$$

再为下一次迭代开始给出新的$(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$而进行**校正**:

$$\begin{pmatrix} A_1x_1^{k+1}\\ A_2x_2^{k+1}\\ \vdots\\ A_px_p^{k+1}\\ \lambda^{k+1}\end{pmatrix} = \begin{pmatrix} A_1x_1^{k}\\ A_2x_2^{k}\\ \vdots\\ A_px_p^{k}\\ \lambda^{k}\end{pmatrix} - \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & I_m & -I_m & 0\\ I_m & \cdots & I_m & 2I_m & \frac{1}{\beta}I_m\\ -\beta I_m & 0 & \cdots & 0 & I_m\end{pmatrix}\begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k\\ A_2x_2^k - A_2\tilde{x}_2^k\\ \vdots\\ A_px_p^k - A_p\tilde{x}_p^k\\ \lambda^k - \tilde{\lambda}^k\end{pmatrix}$$

8 Conclusions

为什么说我们在数值优化方面做出了一些颇有特色又自成系统的工作呢？

首先, 变分不等式和邻近点算法是我们的主要工具. 任何一本关于数值优化的书, 都没有专门提及变分不等式 (VI), 也不会刻意介绍邻近点算法 (PPA), 尽管线性约束的凸优化问题的增广拉格朗日乘子法 (ALM) 是乘子 λ 的 PPA 算法.

- 我们把线性约束的凸优化问题转换成一个等价的结构型单调变分不等式，然后说明什么是变分不等式的 PPA 算法, 讨论了 PPA 算法的收敛性质.
- 变分不等式的 PPA 算法迭代的每一步, 都利用其可分离结构, 分解成一些简单的变分不等式, 求解这些小微变分不等式, 又可以通过求解相应的凸优化问题实现.
- 后来我们又有了基于 VI 的预测-校正方法的统一框架, 既可以用它来验证算法的收敛性, 又可以用它“按需设计”求解可分离凸优化问题的算法, 这就是我们与众不同的逻辑.
- 我们又应该保持清醒的头脑, 即使是 ADMM, 它也是松弛了的 ALM, 是关于乘子 λ 的 PPA 算法. 同时也可以强调, 求解线性约束凸优化问题, ALM 是个有竞争力的好方法.

希望各位以质疑的态度审视我的观点, 对的就相信, 不对的请批评指正.

References

[1] S. Becker, The Chen-Teboulle algorithm is the proximal point algorithm, manuscript, 2011, arXiv: 1908.03633[math.OC].

[2] 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.

[3] 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.

[4] 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.

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

[6] R. Fletcher, Practical mathods of Optimization, Second Edition, JOHN WILEY & SONS, 1987.

[7] G.Y. Gu, B.S. He and X.M. Yuan, Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach, Comput. Optim. Appl., 59(2014), 135-161.

[8] T. Goldstein, B. O' Donoghue, S. Setzer and R. Baraniuk, Fast Alternating Direction Optimization Methods, SIAM J. Imaging Science, Vol. 7, No. 3, pp. 1588 – 1623, 2014.

[9] 何炳生, 从变分不等式的投影收缩算法到凸优化的分裂收缩算法, 高等学校计算数学学报 38 (2016), 74-96.

[10] 何炳生, 我和乘子交替方向法20年, 运筹学学报 22 (2018), 1-31.

[11] B. S. He, 何炳生, 凸优化和单调变分不等式收缩算法的统一框架, 中国科学 48 (2018), 255-272, doi: 10.1360/N012017-00034

[12] 何炳生, 利用统一框架设计凸优化的分裂收缩算法, 高等学校计算数学学报, 2022, 44: 1-35.

[13] B. S. He, H. Liu, Z. R. Wang and X. M. Yuan, A strictly Peaceman-Rachford splitting method for convex programming, SIAM J. Optim. 24 (2014), 1011-1040.

[14] 何炳生, 申远, 求解凸规划及鞍点问题定制的PPA 算法及其收敛速率, 中国科学 2012, 42(5): 515 – 525, doi: 10.1360/012011-1049

[15] 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.

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

[17] B. S. He, S. J. Xu and X. M. Yuan, Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints, arXiv:2107.01897v2[math.OC].

[18] 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.

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

[20] 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.

[21] B. S. He and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming, Comput. Optim. Appl. 70 (2018), 791-826.

[22] B. S. He and X. M. Yuan, Balanced Augmented Lagrangian Method for Convex Programming, arXiv 2108.08554 [math.OC]

[23] B. S. He and X. M. Yuan, On construction of splitting contraction algorithms in a prediction- correction framework for separable convex optimization, arXiv 2204.11522 [math.OC]

[24] B.S. He, X.M. Yuan and W.X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, Comput. Optim. Appl., 56(2013), 559-572.

[25] B.S. He and J. Zhou A Modified Alternating Direction Methodfor Convex Minimization Problems, Applied Mathematics Letters 13 (2000) 123-130.

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

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

[28] 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.

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

[30] L. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D, 60 (1992), pp. 227 − 238

[31] Shengjie Xu, A dual-primal balanced augmented Lagrangian method for linearly constrained convex programming, manusript, 2021. arXiv:2109.02106 [math.OC]

[32] M. Zhu and T. F. Chan, An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration, CAM Report 08-34, UCLA, Los Angeles, CA, 2008.