凸优化分裂收缩算法统一框架的最新进展

从好不容易凑出一个算法
到并不费劲构造一簇算法

何 炳 生

南京大学数学系

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

西电最优化前沿论坛 2022 年 11 月 26 – 27 日

数学之美，不是纯数学的专利。为应用服务的最优化方法研究，同样可以追求简单与统一。简单，他人才会看懂使用；统一，自己才有美的享受。

连续优化中一些代表性数学模型

1. 简单约束问题 $\min\{f(x) \mid x \in \mathcal{X}\}$ 其中 $\mathcal{X}$ 是一个凸集.

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. $\mathbf{min} - \mathbf{max}$ 问题 $\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}}\{\Phi(x,y) = \theta_1(x) - y^T Ax - \theta_2(y)\}$

4. 结构型凸优化 $\min\{\theta_1(x) + \theta_2(y) | Ax + By = b,\, x \in \mathcal{X},\, y \in \mathcal{Y}\}$

5. 多块可分离凸优化 $\min\{\sum_{i=1}^p \theta_i(x_i) | \sum_{i=1}^p A_i x_i = b, x_i \in \mathcal{X}_i\}$

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.

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:

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

where $F(x) = \nabla f(x)$. For general VI, F is an operator from $\Re^n$ into itself.

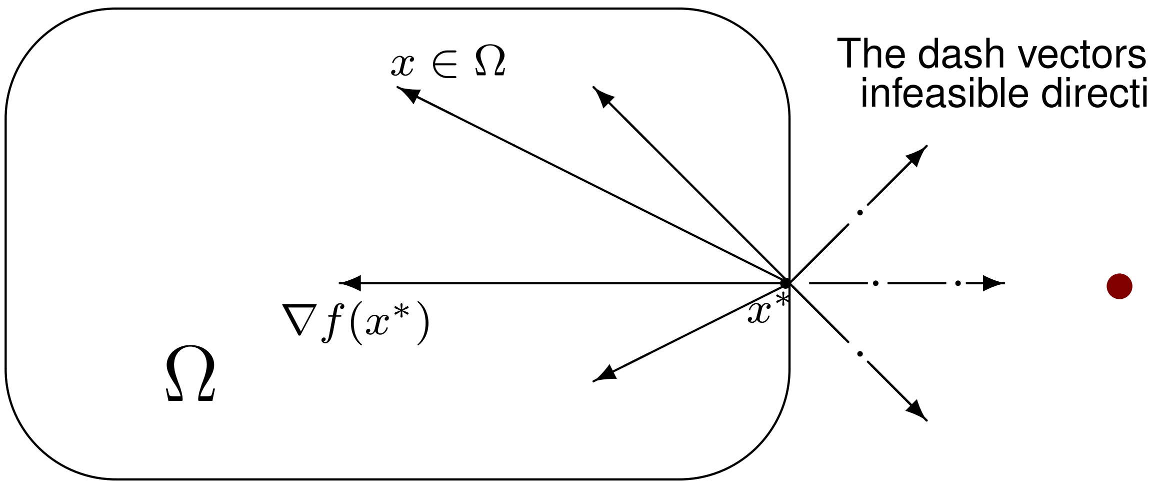


Fig. 1.1 Differential Convex Optimization and VI

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

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

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

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

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

上面的凸优化最优性条件是最基本的, 合在一起就是下面的引理:

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

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

if and only if

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

这样, 我们就把优化问题 (1.3a), 转换成了变分不等式 (1.3b).

引理1的结论是直观的, 证明是初等的, 可以参阅本人主页报告 3 中的引理 1.1 及其证明.

1.2 Linear constrained convex optimization and VI

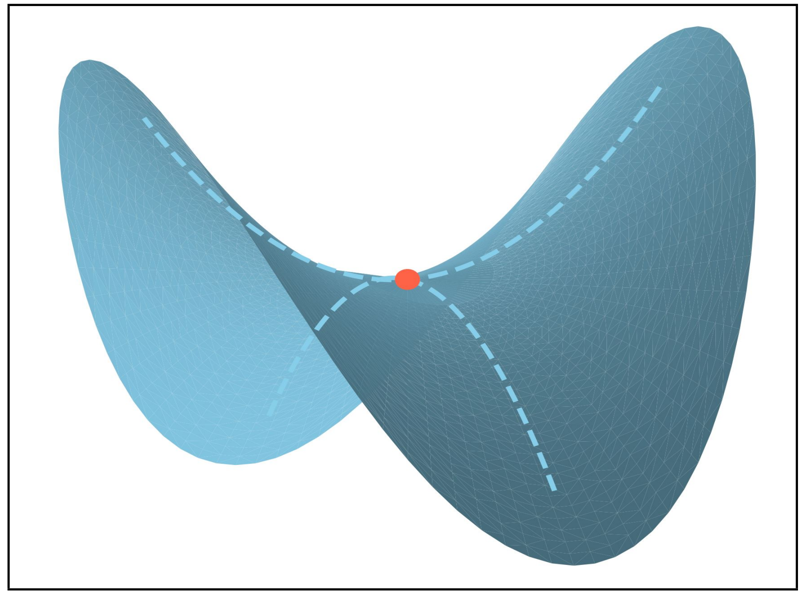
We consider the linearly constrained convex optimization problem

$$\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}. \qquad (1.4)$$

The Lagrangian function of the problem (1.4) is

$$L(x, \lambda) = \theta(x) - \lambda^T(Ax - b), \qquad (1.5)$$

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



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

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

An equivalent expression of the saddle point is the following variational inequality:

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

The optimal condition can be characterized as a monotone variational inequality:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{1.6}$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \text{ and } \Omega = \mathcal{X} \times \Re^m, \tag{1.7}$$

Note that the operator F is monotone. Especially, because

$$(w - \tilde{w})^T(F(w) - F(\tilde{w})) = \begin{pmatrix} x - \tilde{x} \\ \lambda - \tilde{\lambda} \end{pmatrix}^T \begin{pmatrix} -A^T(\lambda - \tilde{\lambda}) \\ A(x - \tilde{x}) \end{pmatrix}$$

$$= \begin{pmatrix} x - \tilde{x} \\ \lambda - \tilde{\lambda} \end{pmatrix}^T \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x - \tilde{x} \\ \lambda - \tilde{\lambda} \end{pmatrix}$$

we have

$$(w - \tilde{w})^T \{F(w) - F(\tilde{w})\} = 0. \tag{1.8}$$

Convex optimization problem with two separable objective functions

Some convex optimization problems have the following separable structure:

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

The Lagrangian function of this problem is

$$L^{(2)}(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b). \tag{1.9}$$

The saddle point $((x^*, y^*), \lambda^*)$ of $L^{(2)}(x, y, \lambda)$ is a solution of the following VI:

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

where

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

and

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

Convex optimization problem with three separable objective functions

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

Its 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 saddle point $((x^*,y^*,z^*),\lambda^*)$ of $L^{(3)}(x,y,z,\lambda)$ is a solution of the VI:

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

where

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

and

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

1.3 Proximal point algorithms for convex optimization

Lemma 2 *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*

$$\|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2. \tag{1.10}$$

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

Convex Optimization

Now, let us consider the *simple* convex optimization

$$\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}, \tag{1.11}$$

where $\theta(x)$ and $f(x)$ are convex but $\theta(x)$ is not necessary smooth, $\mathcal{X}$ is a closed convex set.

For solving (1.11), the k-th iteration of the proximal point algorithm (abbreviated to PPA) [18, 21] begins with a given x^k, offers the new iterate x^{k+1} via the recursion

$$x^{k+1} = \text{Argmin}\{\theta(x) + f(x) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{1.12}$$

Since x^{k+1} is the optimal solution of (1.12), it follows from Lemma 1 that

$$\theta(x)-\theta(x^{k+1})+(x-x^{k+1})^T\{\nabla f(x^{k+1})+r(x^{k+1}-x^k)\}\geq 0,\ \forall\, x\in\mathcal{X}. \tag{1.13}$$

Setting $x=x^*$ in the above inequality, it follows that

$$(x^{k+1}-x^*)^T r(x^k-x^{k+1})\geq \theta(x^{k+1})-\theta(x^*)+(x^{k+1}-x^*)^T\nabla f(x^{k+1}). \tag{1.14}$$

Since

$$(x^{k+1}-x^*)^T\nabla f(x^{k+1})\geq (x^{k+1}-x^*)^T\nabla f(x^*),$$

it follows that the right hand side of (1.14) is nonnegative. And consequently,

$$(x^{k+1}-x^*)^T(x^k-x^{k+1})\geq 0. \tag{1.15}$$

Let $a=x^k-x^*$ and $b=x^{k+1}-x^*$ and using Lemma 2, we obtain

$$\|\boldsymbol{x^{k+1}-x^*}\|^{\mathbf{2}}\leq \|\boldsymbol{x^k-x^*}\|^{\mathbf{2}}-\|\boldsymbol{x^k-x^{k+1}}\|^{\mathbf{2}}, \tag{1.16}$$

which is the nice convergence property of Proximal Point Algorithm.

We write the problem (1.11) and its PPA (1.12) in VI form

The equivalent variational inequality form of the optimization problem (1.11) is

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{1.17a}$$

For solving the problem (1.11), the variational inequality form of the k-th iteration of the PPA (see (1.13)) is:

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad & \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \nabla f(x^{k+1}) \\ & \geq (x - x^{k+1})^T r(x^k - x^{k+1}), \quad \forall\, x \in \mathcal{X}. \end{aligned} \tag{1.17b}$$

PPA reaches the solution of (1.17a) via solving a series of subproblems (1.17b).

采用的是步步为营的策略, 稳扎稳打 !

Using (1.17), we study PPA for VI arising from the constrained optimization

1.4 Preliminaries of PPA for Variational Inequalities

The optimal condition of the problem (1.4) is characterized as a mixed monotone variational inequality:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \qquad (1.18)$$

PPA for VI in Euclidean-norm For given w^k and $r > 0$, find w^{k+1}

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(x) - \theta(x^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &\geq (w - w^{k+1})^T r(w^k - w^{k+1}), \ \forall\, w \in \Omega. \end{aligned} \qquad (1.19)$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (1.18).

✠ **w^k is the solution of (1.18) if and only if $w^k = w^{k+1}$** ✠

Setting $w = w^*$ in (1.19), we obtain

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

Note that (see the structure of $F(w)$ in (1.7))

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

and consequently (by using (1.18)) we obtain

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

and thus

$$(w^{k+1} - w^*)^T (w^k - w^{k+1}) \ge 0. \tag{1.20}$$

Now, by setting $a = w^k - w^*$ and $b = w^{k+1} - w^*$ in the inequality (1.20), it is $b^T(a - b) \ge 0$. Using Lemma2, we obtain

$$\|w^{k+1} - w^*\|^2 \le \|w^k - w^*\|^2 - \|w^k - w^{k+1}\|^2. \tag{1.21}$$

We get the nice convergence property of Proximal Point Algorithm:

The sequence $\{w^k\}$ generated by PPA is Fejér monotone.

PPA for monotone mixed VI in H-norm

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

$$w^{k+1} \in \Omega, \quad \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (w - w^{k+1})^T H(w^k - w^{k+1}), \ \forall \, w \in \Omega, \tag{1.22}$$

where H is a symmetric positive definite matrix.

✠ Again, w^k is the solution of (1.18) if and only if $w^k = w^{k+1}$ ✠

Convergence Property of Proximal Point Algorithm in H-norm

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \tag{1.23}$$

The sequence $\{w^k\}$ is Fejér monotone in H-norm. In customized PPA, via choosing a proper positive definite matrix H, the solution of the subproblem (1.22) has a closed form. An iterative algorithm is called the contraction method, if its generated sequence $\{w^k\}$ satisfies $\|w^{k+1} - w^*\|_H^2 < \|w^k - w^*\|_H^2$.

2 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 [22])

$$\min_x \max_y \{\Phi(x,y) = \theta_1(x) - y^T Ax - \theta_2(y) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{2.1}$$

Let (x^*, y^*) be the solution of (2.1), then we have

根据鞍点的定义

$$\begin{cases} x^* \in \mathcal{X}, \;\; \Phi(x, y^*) - \Phi(x^*, y^*) \geq 0, \;\; \forall\, x \in \mathcal{X}, & \text{(2.2a)} \\ y^* \in \mathcal{Y}, \;\; \Phi(x^*, y^*) - \Phi(x^*, y) \geq 0, \;\; \forall\, y \in \mathcal{Y}. & \text{(2.2b)} \end{cases}$$

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

$$\begin{cases} x^* \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T y^*) \geq 0, \quad \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(Ax^*) \geq 0, \quad \forall\, y \in \mathcal{Y}. \end{cases}$$

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

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T F(u^*) \geq 0, \; \forall\, u \in \Omega, \tag{2.3}$$

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

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

2.1 求解鞍点问题的 原始-对偶混合梯度法 PDHG [23]

For given (x^k, y^k), PDHG [23] produces a pair of (x^{k+1}, y^{k+1}). First,

$$x^{k+1} = \text{argmin}\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, \tag{2.4a}$$

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

$$y^{k+1} = \text{argmax}\{\Phi(x^{k+1}, y) - \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \tag{2.4b}$$

Ignoring the constant term in the objective function, the subproblems (3.5) are reduced to

$$\begin{cases} x^{k+1} = \operatorname{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, & \text{(2.5a)} \\ y^{k+1} = \operatorname{argmin}\{\theta_2(y) + y^T A x^{k+1} + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. & \text{(2.5b)} \end{cases}$$

According to Lemma 1, the optimality condition of (3.6a) is $x^{k+1} \in \mathcal{X}$ and

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T y^k + r(x^{k+1} - x^k)\} \geq 0, \;\; \forall\, x \in \mathcal{X}. \quad \text{(2.6)}$$

Similarly, from (3.6b) we get $y \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{Ax^{k+1} + s(y^{k+1} - y^k)\} \geq 0, \;\; \forall\, y \in \mathcal{Y}. \quad \text{(2.7)}$$

Combining (3.7) and (3.8), we have $(x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$,

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T y^{k+1} \\ Ax^{k+1} \end{pmatrix} + \begin{pmatrix} r(x^{k+1} - x^k) + A^T(y^{k+1} - y^k) \\ s(y^{k+1} - y^k) \end{pmatrix} \right\} \geq 0, \quad \forall (x, y) \in \Omega.$$

The compact form is $u^{k+1} \in \Omega$,

$$u^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T F(u^{k+1}) \\ \geq (u - u^{k+1})^T Q(u^k - u^{k+1}), \quad \forall\, u \in \Omega. \tag{2.8}$$

where

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

It does not be the PPA form (1.22), and we can not expect its convergence.

2.2 Customized Proximal Point Algorithm-Classical Version

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 (3.10) will become the following desirable form:

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

For this purpose, we need only to change (3.8) in PDHG, namely,

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

to

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

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{A[2x^{k+1} - x^k] + s(y^{k+1} - y^k)\} \geq 0. \quad (2.9)$$

Thus, for given (x^k, y^k), producing a proximal point (x^{k+1}, y^{k+1}) via (3.5a) and (2.9) can be summarized as:

$$x^{k+1} = \text{argmin}\Big\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \,\Big|\, x \in \mathcal{X}\Big\}. \quad (2.10a)$$

$$y^{k+1} = \text{argmax}\Big\{\Phi\big([2x^{k+1} - x^k], y\big) - \frac{s}{2}\|y - y^k\|^2\Big\} \quad (2.10b)$$

By ignoring the constant term in the objective function, getting x^{k+1} from (2.10a) 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 (2.10b) 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 (2.10a)-(2.10b).

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

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

Theorem 1 *The sequence* $\{u^k = (x^k, y^k)\}$ *generated by the customized PPA* (2.10) *satisfies*

$$\|u^{k+1} - u^*\|_H^2 \le \|u^k - u^*\|_H^2 - \|u^k - u^{k+1}\|_H^2. \tag{2.11}$$

2.3 Simplicity recognition

Frame of VI and PPA 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

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

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, 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 [13].

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

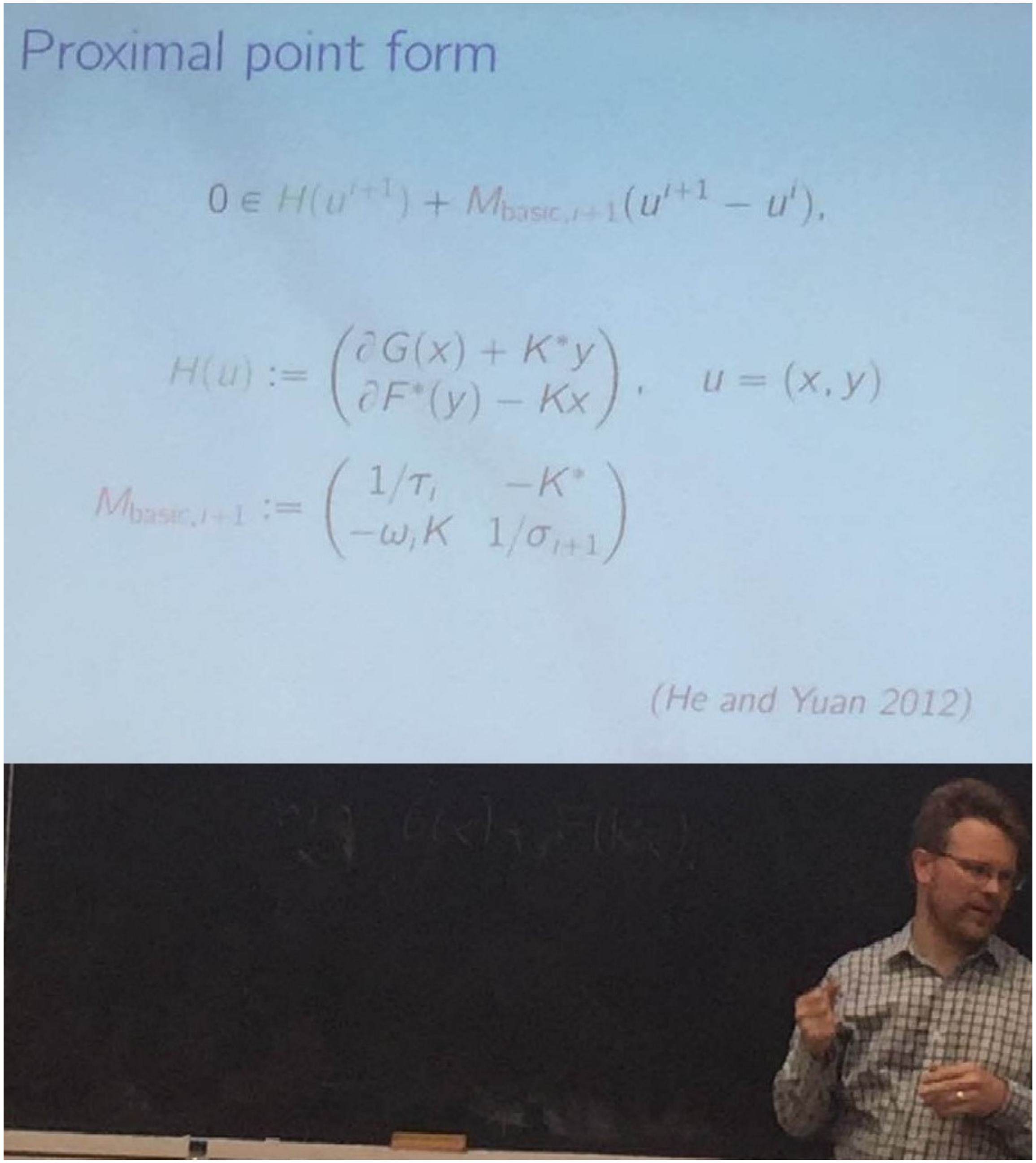
- Input: Operator norm $L := \|K\|$, Lipschitz constant L_f of ∇f, and Bregman distance functions D_x and D_y.
- Initialization: Choose $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$, $\tau, \sigma > 0$
- Iterations: For each $n \geq 0$ let

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

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同事去英国访问. 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.

S. Becker 是 E. Candès 在加州理工的学生, 经过将近八年才挂出这篇短文.

3 The unified framework and its convergence

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

[预测] 第 k-步迭代从给定的核心变量 v^k 开始, 求得预测点 $\tilde{w}^k$, 使得

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

成立. 其中矩阵 $Q^T + Q$ 是正定的.

我们称 Q 为预测矩阵

[校正]. 根据预测得到的 $\tilde{v}^k$, 给出核心变量 v 的新迭代点 v^{k+1} 的公式为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{3.3}$$

我们称 M 为校正矩阵. v 为核心变量, v 可以是 w, 也可以是 w 的部分分量

收敛性条件 对算法框架中的预测矩阵 Q 和校正矩阵 M, 有矩阵

$$H \succ 0 \qquad \text{使得} \qquad HM = Q. \tag{3.4a}$$

此外, 能够保证

$$G = Q^T + Q - M^T H M \succ 0. \tag{3.4b}$$

预测-校正方法的例子

Taking the output of Original PDHG algorithm [23] as predictor

For given (x^k, y^k), PDHG [23] produces a pair of $(\tilde{x}^k, \tilde{y}^k)$. First,

$$\tilde{x}^k = \operatorname{argmin}\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \tag{3.5a}$$

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

$$\tilde{y}^k = \operatorname{argmax}\{\Phi(\tilde{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. \tag{3.5b}$$

Ignoring the constant term in the objective function, the subproblems (3.5) are reduced to

$$\begin{cases} \tilde{x}^k = \operatorname{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(3.6a)} \\ \tilde{y}^k = \operatorname{argmin}\{\theta_2(y) + y^T A\tilde{x}^k + \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. & \text{(3.6b)} \end{cases}$$

According to the basic lemma, the optimality condition of (3.6a) is $\tilde{x}^k \in \mathcal{X}$ and

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T y^k + r(\tilde{x}^k - x^k)\} \geq 0, \ \forall\, x \in \mathcal{X}. \tag{3.7}$$

Similarly, from (3.6b) we get $\tilde{y}^k \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{A\tilde{x}^k + s(\tilde{y}^k - y^k)\} \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{3.8}$$

Combining (3.7) and (3.8), we have

$$\tilde{u}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{y}^k \\ A\tilde{x}^k \end{pmatrix} \right.$$
$$\left. + \begin{pmatrix} r(\tilde{x}^k - x^k) + A^T(\tilde{y}^k - y^k) \\ s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \geq 0, \quad \forall (x, y) \in \Omega. \tag{3.9}$$

Lemma 3 *The compact form is $\tilde{u}^k \in \Omega$,*

$$\theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T F(\tilde{u}^k) \geq (u - \tilde{u}^k)^T Q(u^k - \tilde{u}^k), \ \forall u \in \Omega, \tag{3.10a}$$

where

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}. \tag{3.10b}$$

Proof. Using the notation of (2.3), it follows from (3.9) directly. $\square$

Because

$$Q^T + Q = \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix},$$

the matrix $Q^T + Q$ is positive definite if and only if $rs > \frac{1}{4}\|A^T A\|$.

对于这样的预测, 我们以前考虑比较简单的校正

是想办法去凑的!

$$u^{k+1} = u^k - M(u^k - \tilde{u}^k) \tag{3.11}$$

校正. 其中 M 为单位上三角矩阵或单位下三角矩阵, 同时怎么去满足收敛性条件 (3.4)

- $\exists\, H \succ 0$ 使得 $HM = Q$.
- $G = Q^T + Q - M^T HM \succ 0$.

一. 校正矩阵 M 为单位下三角矩阵

其中的 K 是待定的.

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

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

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

因此,

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

对任意的 $r, s > 0$, 矩阵 H 是正定的.

对条件 (ii),

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

要矩阵 G 正定, 必须有 $rs > \|A^T A\|$.

采用 PDHG 预测, 单位下三角矩阵校正, 需要 $rs > \|A^T A\|$.

二. 校正矩阵 $\boldsymbol{M}$ 为单位上三角矩阵

同样, 其中的 K 是待定的.

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

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),

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

要矩阵 G 正定, 必须有 $rs > \|A^T A\|$.

采用 PDHG 预测, 单位上三角矩阵校正, 需要 $rs > \|A^T A\|$.

把不能保证收敛的 PDHG 方法改造成了收敛的方法.

采用 PDHG, 以后会看到, 有的校正只要求 $rs > \frac{1}{4}\|A^T A\|$.

算法统一框架[4, 5, 14]的收敛性证明非常简单, 对设计算法也有指导意义[6]

Assuming the conditions (3.4) in the unified framework are satisfied, we prove some convergence properties.

Theorem 2 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (3.1) *and* $\tilde{w}^k$ *is obtained in the* k*-th iteration. If* v^k*,* v^{k+1} *and* $\tilde{w}^k$ *satisfy the conditions in the unified framework, then we have*

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ \geq \quad & \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega. \end{aligned} \tag{3.12}$$

Proof. Using $Q = HM$ (see (3.4a)) and the relation (3.3), the right hand side of (3.2) can be written as $(v - \tilde{v}^k)^T H(v^k - v^{k+1})$ and hence

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \forall w \in \Omega. \tag{3.13}$$

Applying the identity

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

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

to the right hand side of (3.13) with

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

we thus obtain

$$2(v - \tilde{v}^k)^T H(v^k - v^{k+1})$$

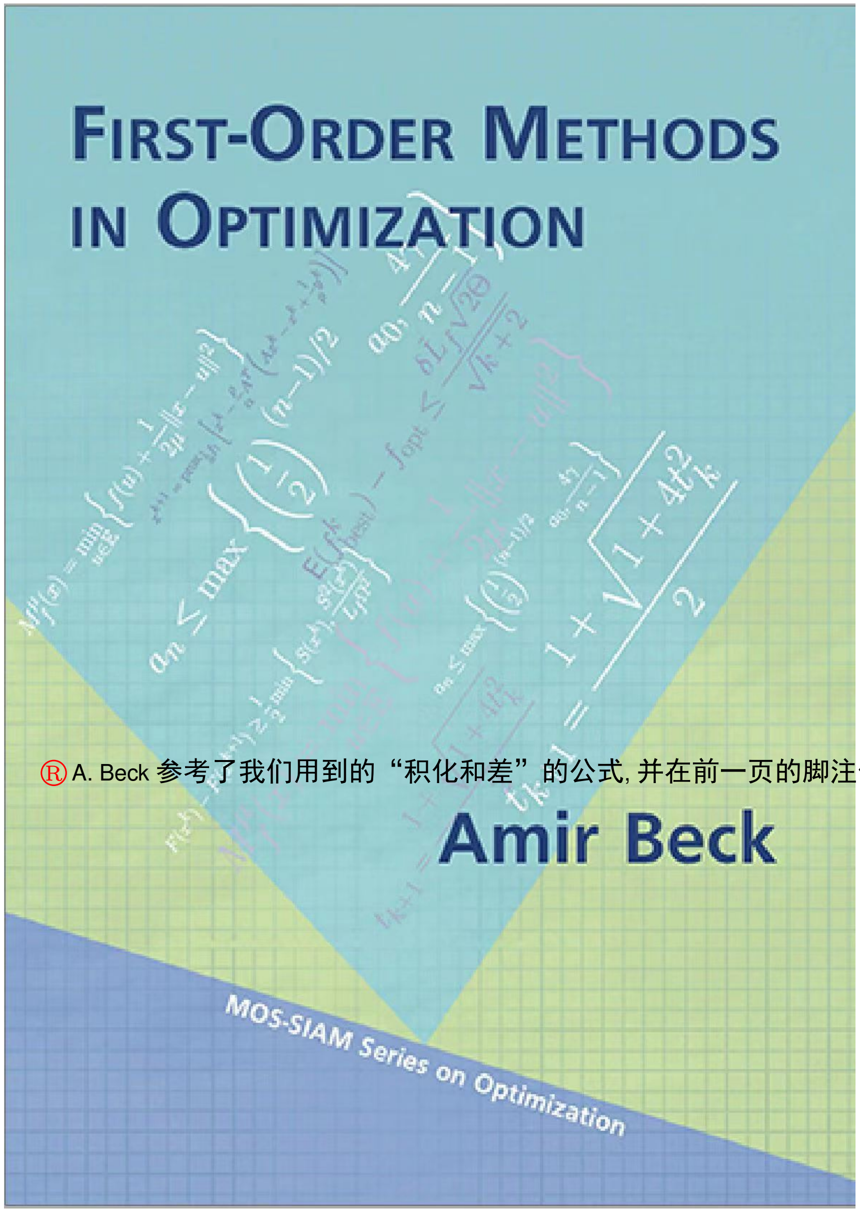
这个积化和差恒等式 A. Beck 专著 [1] 中也采用并予註明

$$= \left(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\right) + \left(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2\right). \quad (3.14)$$

For the last term of (3.14), using $HM = Q$ and $2v^T Qv = v^T(Q^T + Q)v$, we have

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&= \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\
&\overset{(3.3)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
&= 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
&= (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\
&\overset{(3.4b)}{=} \|v^k - \tilde{v}^k\|_G^2. \qquad (3.15)
\end{aligned}$$

Substituting (3.14), (3.15) in (3.13), the assertion of this theorem is proved. □



Ⓡ A. Beck 参考了我们用到的“积化和差”的公式, 并在前一页的脚注做了说明

We will use the following notation:

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

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

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

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 $\mathbf{x} \in \text{dom}(h_1)$, $\mathbf{z} \in \text{dom}(h_2)$, and $\mathbf{y} \in \mathbb{R}^m$,

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

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

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

as well as

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

and

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

Theorem 3 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (3.1) *and $\tilde{w}^k$ is obtained in the k-th iteration. If v^k, v^{k+1} and $\tilde{w}^k$ satisfy the conditions in the unified framework, then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.16}$$

Proof. Setting $w = w^*$ in (3.12), we get

$$\begin{aligned} &\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &\quad \ge \quad \|v^k - \tilde{v}^k\|_G^2 + 2\underline{\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}}. \end{aligned} \tag{3.17}$$

Using (1.8) ($(\tilde{w}^k - w^*)^T (F(\tilde{w}^k) - F(w^*)) = 0$) and the optimality of w^*, we have

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

and thus from (3.17) get

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \|v^k - \tilde{v}^k\|_G^2. \tag{3.18}$$

The assertion (3.16) follows directly. $\square$

校正 $v^{k+1} = v^k - M(v^k - \tilde{v}^k)$, 怎样给出满足收敛性条件的校正矩阵 M ?

$$\begin{cases} \text{预测 (3.2) 提供 } Q: Q^T + Q \succ 0 \\ \text{收敛条件 (3.4)}: \text{选矩阵 } M \text{ 的要求:} \\ \exists\, H \succ 0, \text{ such that } \; HM = Q, \\ G = Q^T + Q - M^T HM \succ 0. \end{cases} \iff \begin{cases} D \succ 0, \qquad G \succ 0, \\ D + G = Q^T + Q, \\ M^T HM = D, \\ HM = Q. \end{cases}$$

$$\iff \begin{cases} D \succ 0, \qquad G \succ 0, \\ D + G = Q^T + Q, \\ Q^T M = D, \\ HM = Q. \end{cases} \iff \begin{cases} D \succ 0, \qquad G \succ 0, \\ D + G = Q^T + Q, \\ M = Q^{-T} D, \\ H = QD^{-1}Q^T. \end{cases}$$

有了预测矩阵 Q, 我们以前的办法是去凑, 凑出 H 和 M, 使之满足条件 (3.4).
现在的做法: 有了预测矩阵 Q, 可以选定 D, 使其满足 $0 \prec D \prec Q^T + Q$.
由 $M = Q^{-T}D$ 得到的校正矩阵 M, 收敛性条件自然满足. D 的选择多种多样!

有了等价的收敛条件, 除了凑, 还可以构造 M, 例如

In (3.10), we get

$$\tilde{u}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T \{F(\tilde{u}^k) + Q(\tilde{u}^k - u^k)\} \geq 0, \ \forall u \in \Omega, \tag{3.19a}$$

where

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}. \tag{3.19b}$$

Because

$$Q^T + Q = \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix},$$

we need only $rs > \frac{1}{4}\|A^TA\|$ to ensure the matrix $Q^T + Q$ to be positive definite. Since

$$Q^{-T} = \begin{pmatrix} \frac{1}{r}I_n & 0 \\ -\frac{1}{rs}A & \frac{1}{s}I_m \end{pmatrix},$$

for any positive definite matrices D and G which satisfy

$$D \succ 0, \quad G \succ 0 \quad \text{and} \quad D + G = Q^T + Q,$$

it is easy to calculate the correction matrix $M = Q^{-T}D$.

对于同一个预测矩阵 Q, 矩阵 D 的选法无穷多, 因此对应的校正方法也无穷多 !

4 Constructing Algorithms for p-block problems

4.1 Motivation from the classical ADMM

Linearly constrained two-blocks separable convex optimization

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

Solving (4.1) by classical ADMM : From (y^k,λ^k) to (y^{k+1},λ^{k+1})

$$\begin{cases} x^{k+1} \in \arg\min\{L^{[2]}(x,y^k,\lambda^k)+\frac{\beta}{2}\|Ax+By^k-b\|^2 \,|\, x\in\mathcal{X}\}, \\ y^{k+1} \in \arg\min\{L^{[2]}(x^{k+1},y,\lambda^k)+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2 \,|\, y\in\mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). \end{cases} \tag{4.2}$$

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

The x-subproblem in ADMM can be written as:

$$x^{k+1}\in\arg\min\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|(Ax^k+By^k-b)+A(x-x^k)\|^2 \,|\, x\in\mathcal{X}\}$$

$$=\arg\min\left\{\begin{array}{c}\theta_1(x)-x^TA^T\underline{[\lambda^k-\beta(Ax^k+By^k-b)]}\\ +\frac{\beta}{2}\|A(x-x^k)\|^2\end{array}\,\middle|\, x\in\mathcal{X}\right\}.$$

Similarly, the y-subproblem in ADMM is equivalent to

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

Thus, ADMM scheme can be written as

$$\left\{ \begin{array}{l} \lambda^{k+\frac{1}{2}} = \underline{\lambda^k - \beta(Ax^k + By^k - b)} \\ x^{k+1} \in \mathrm{argmin}_{x\in\mathcal{X}}\{\theta_1(x) - x^T A^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2}\|A(x - x^k)\|^2\}, \\ y^{k+1} \in \mathrm{argmin}_{y\in\mathcal{Y}}\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2}\|A(x^{k+1} - x^k) + B(y - y^k)\|^2\}, \\ \lambda^{k+1} = \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big). \end{array} \right. \tag{4.3}$$

蓝色部分的形式在后面介绍的方法中要用到

We can also write the form $\lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^k + By^k - b)$ as

$$\lambda^{k+\frac{1}{2}} = P_\Lambda\big[\lambda^k - \beta(Ax^k + By^k - b)\big], \quad \text{where} \quad \Lambda = \Re^m,$$

and P_Λ denotes the projection on Λ. For equality and inequality constraints, we need the different

$$\Lambda = \Re^m \qquad \text{and} \qquad \Lambda = \Re^m_+.$$

For $a \in \Re^m$, $\quad P_{\Re^m}(a) = a \quad$ and $\quad [P_{\Re^m_+}(a)]_i = \max\{a_i, 0\}$.

4.2 Prediction-Correction method for p-block problem

We consider the p-block separable convex optimization

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{p}A_ix_i=b\ (\text{or}\ \ge b),\ \ x_i\in\mathcal{X}_i\Big\}. \tag{4.4}$$

The Lagrangian function is

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

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

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

范围更广：包含等式和不等式约束

便于推广：从一块到多个可分离块

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 (4.4) can be written as the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{4.5a}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \tag{4.5b}$$

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 (4.5).

Prediction

预测公式和交替方向法的公式(4.3)有类似之处

From $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ to $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_p^k, \tilde{\lambda}^k)$:

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

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\
\qquad \vdots \\
\tilde{x}_i^k \in \arg\min_{x_i \in \mathcal{X}_i}\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\
\qquad \vdots \\
\tilde{x}_p^k \in \arg\min_{x_p \in \mathcal{X}_p}\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^p A_j\tilde{x}_j^k - b\big)\big].
\end{array}\right. \tag{4.6}$$

蓝色部分和 ADMM 公式 (4.3) 中蓝色部分相同

预测对可分离的原始子问题逐一按序求解. 犹如 Gauss 消去法往前消元.

Correction

预测步提供的预测点是需要经过进一步校正的.

Correction Step .

Produce $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$ for the next iteration:

Generate the new iterate $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{4.7}$$

校正对新的原始变量按逆序给出. 犹如 Gauss 消去法中往后回代

校正非常简单, 工作量也很小. 把校正公式分开来写就是:

$Ax_i^{k+1}, i = 1, \ldots, p$

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \end{pmatrix} - \nu \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \end{pmatrix}, \tag{4.8}$$

λ^{k+1}

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - [-\nu\beta(A_1x_1^k - A_1\tilde{x}_1^k) + (\lambda^k - \tilde{\lambda}^k)] \\ &= \tilde{\lambda}^k + \nu\beta(A_1x_1^k - A_1\tilde{x}_1^k). \end{aligned} \tag{4.9}$$

这个预测-校正的过程是容易记住的. 但对同一个预测, 校正不是唯一的!

5 Analysis for the prediction step (4.6)

For the analysis, we need only the basic analytical property which is described in Lemma 1

$$x^* \in \arg\min\{\theta(x) + f(x) | x \in \mathcal{X}\} \Leftrightarrow.$$

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

Analysis for the Prediction 我们先看 (4.6) 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

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

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

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \{\underline{-A_i^T \tilde{\lambda}^k} + \beta A_i^T \Big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\Big) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0. \tag{5.1a}$$

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

$$\tilde\lambda^k\in\Lambda,\quad (\lambda-\tilde\lambda^k)^T\big\{\big(\underline{\textstyle\sum_{j=1}^p A_j\tilde x_j^k - b}\big) + \tfrac{1}{\beta}(\tilde\lambda^k-\lambda^k)\big\}\ge 0,\quad \forall\lambda\in\Lambda. \tag{5.1b}$$

(5.1a) 和 (5.1b) 中下划线部分组合在一起就是 $F(\tilde w)$, 所以预测可以写成下面的形式

Summating (5.1a) and (5.1b), for the predictor $\tilde w^k$ generated by (4.6), we have $\tilde w^k\in\Omega$,

$$\theta(x)-\theta(\tilde x^k) + (w-\tilde w^k)^T\underline{F(\tilde w^k)} \ge (w-\tilde w^k)^T Q(w^k-\tilde w^k),\quad \forall\, w\in\Omega, \tag{5.2a}$$

where

$$Q = \begin{pmatrix} \beta A_1^TA_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^TA_1 & \beta A_2^TA_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^TA_1 & \beta A_p^TA_2 & \cdots & \beta A_p^TA_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta}I_m \end{pmatrix}. \tag{5.2b}$$

5.1 变量代换后的预测矩阵

The optimization problem (4.4) has been translated to VI (4.5), namely,

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

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

$$P = \begin{pmatrix} \sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta})I_m \end{pmatrix}, \quad z = Pw = \begin{pmatrix} \sqrt{\beta}A_1x_1 \\ \sqrt{\beta}A_2x_2 \\ \vdots \\ \sqrt{\beta}A_px_p \\ (1/\sqrt{\beta})\lambda \end{pmatrix}. \tag{5.3}$$

Accordingly, we define

$$\mathcal{Z} = \{z \mid z = Pw\},$$

and

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

Using the notation P in (5.3), for the matrix Q in (5.2b), we have

$$Q = P^T \mathcal{Q} P, \qquad \text{where} \qquad \mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{5.4}$$

Thus, for the right hand side of (5.2a), 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 (5.2) that we have the following VI for the P-D prediction:

$$\begin{aligned} \tilde{w}^k \in \Omega, \ \ \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \geq (z - \tilde{z}^k)^T \mathcal{Q} (z^k - \tilde{z}^k), \ \ \forall\, w \in \Omega. & \end{aligned} \tag{5.5}$$

where $\mathcal{Q}$ is given in (5.4).

5.2 变量代换下的算法统一框架

Prediction-Correction Framework for VI (4.5).

1. (Prediction Step) With given w^k and $z^k = Pw^k$, find $\tilde{w}^k \in \Omega$ such that

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ \geq (z - \tilde{z}^k)^T \mathcal{Q}(z^k - \tilde{z}^k), \quad \forall\, w \in \Omega, \end{aligned} \tag{5.6}$$

   with $\mathcal{Q} \in \Re^{(p+1)m \times (p+1)m}$, and the matrix $\mathcal{Q}^T + \mathcal{Q}$ is positive definite.

2. (Correction Step) With the predictor $\tilde{w}^k$ by (5.6) and $\tilde{z}^k = P\tilde{w}^k$, the new iterate z^{k+1} is updated by

$$z^{k+1} = z^k - \mathcal{M}(z^k - \tilde{z}^k), \tag{5.7}$$

   where $\mathcal{M} \in \Re^{(p+1)m \times (p+1)m}$ is a non-singular matrix.

Theorem 4 *For the matrices* $\mathcal{Q}$ (5.6) *and* $\mathcal{M}$ *in* (5.7)*, if there is a positive definite matrix* $\mathcal{H} \in \Re^{(p+1)m\times(p+1)m}$ *such that*

$$\mathcal{H}\mathcal{M} = \mathcal{Q} \tag{5.8a}$$

and

$$\mathcal{G} := \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0, \tag{5.8b}$$

then we have

$$\|z^{k+1} - z^*\|_{\mathcal{H}}^2 \le \|z^k - z^*\|_{\mathcal{H}}^2 - \|z^k - \tilde{z}^k\|_{\mathcal{G}}^2, \quad \forall\, z^* \in \mathcal{Z}^*. \tag{5.9}$$

Proof. Setting w in (5.6) as any fixed $w^* \in \Omega^*$, and using

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

we get

$$(\tilde{z}^k - z^*)^T \mathcal{Q}(z^k - \tilde{z}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*), \;\; \forall\, w^* \in \Omega^*.$$

The right-hand side of the last inequality is non-negative. Thus, we have

$$(\tilde{z}^k - z^*)^T \mathcal{Q}(z^k - \tilde{z}^k) \ge 0 \quad \stackrel{\mathcal{Q}=\mathcal{H}\mathcal{M}}{\Longrightarrow} \quad (\tilde{z}^k - z^*)^T \mathcal{H}\underbrace{\mathcal{M}(z^k - \tilde{z}^k)}_{z^k - z^{k+1}} \ge 0.$$

and thus

$$(z^k - z^{k+1})^T \mathcal{H}(\tilde{z}^k - z^*) \geq 0, \quad \forall\, z^* \in \mathcal{Z}^*. \tag{5.10}$$

Applying the identity

$$2(a-b)^T H(c-d) = \big(\|a-d\|_H^2 - \|b-d\|_H^2\big) - \big(\|a-c\|_H^2 - \|b-c\|_H^2\big)$$

to (5.10) with $a = z^k,\ b = z^{k+1},\ c = \tilde{z}^k$ and $d = z^*$, we get

$$\|z^k - z^*\|_{\mathcal{H}}^2 - \|z^{k+1} - z^*\|_{\mathcal{H}}^2 \geq \underline{\|z^k - \tilde{z}^k\|_{\mathcal{H}}^2 - \|z^{k+1} - \tilde{z}^k\|_{\mathcal{H}}^2}. \tag{5.11}$$

Then, by simple manipulations, we obtain

$$\begin{aligned}
\underline{\|z^k - \tilde{z}^k\|_{\mathcal{H}}^2 - \|z^{k+1} - \tilde{z}^k\|_{\mathcal{H}}^2} &= \|z^k - \tilde{z}^k\|_{\mathcal{H}}^2 - \|[z^k - \mathcal{M}(z^k - \tilde{z}^k)] - \tilde{z}^k\|_{\mathcal{H}}^2 \\
&\overset{(5.7)}{=} \|z^k - \tilde{z}^k\|_{\mathcal{H}}^2 - \|(z^k - \tilde{z}^k) - \mathcal{M}(z^k - \tilde{z}^k)\|_{\mathcal{H}}^2 \\
&= 2(z^k - \tilde{z}^k)^T \mathcal{H}\mathcal{M}(z^k - \tilde{z}^k) - \|\mathcal{M}(z^k - \tilde{z}^k)\|_{\mathcal{H}}^2 \\
&\overset{(5.8)}{=} (z^k - \tilde{z}^k)^T \underbrace{[(\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T \mathcal{H}\mathcal{M}]}_{\mathcal{G}}(z^k - \tilde{z}^k) = \underline{\|z^k - \tilde{z}^k\|_{\mathcal{G}}^2}.
\end{aligned}$$

The assertion of this theorem is proved. □

The inequality (5.9) is the key for the convergence proofs, for details, see [11]

如何给出满足收敛性条件 (5.8) 的校正矩阵 $\mathcal{M}$?

$$\begin{cases} \underline{\text{预测 (5.6) 提供 } \mathcal{Q}} : \mathcal{Q}^T + \mathcal{Q} \succ 0 \\ \underline{\text{收敛条件 (5.8)}} : \text{ 选 } \mathcal{M} \text{ 的要求} \\ \exists\, \mathcal{H} \succ 0, \text{ such that } \mathcal{H}\mathcal{M} = \mathcal{Q}, \\ \mathcal{G} = \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0. \end{cases} \Longleftrightarrow \begin{cases} \mathcal{D} \succ 0, \quad \mathcal{G} \succ 0, \\ \mathcal{D} + \mathcal{G} = \mathcal{Q}^T + \mathcal{Q}, \\ \mathcal{M}^T\mathcal{H}\mathcal{M} = \mathcal{D}, \\ \mathcal{H}\mathcal{M} = \mathcal{Q}. \end{cases}$$

$$\Longleftrightarrow \begin{cases} \mathcal{D} \succ 0, \quad \mathcal{G} \succ 0, \\ \mathcal{D} + \mathcal{G} = \mathcal{Q}^T + \mathcal{Q}, \\ \mathcal{Q}^T\mathcal{M} = \mathcal{D}, \\ \mathcal{H}\mathcal{M} = \mathcal{Q}. \end{cases} \Longleftrightarrow \begin{cases} \mathcal{D} \succ 0, \quad \mathcal{G} \succ 0, \\ \mathcal{D} + \mathcal{G} = \mathcal{Q}^T + \mathcal{Q}, \\ \mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}, \\ \mathcal{H} = \mathcal{Q}\mathcal{D}^{-1}\mathcal{Q}^T. \end{cases}$$

有了预测矩阵 $\mathcal{Q}$, 就可以选多种多样的 $\mathcal{D}$, 使其满足 $0 \prec \mathcal{D} \prec \mathcal{Q}^T + \mathcal{Q}$.
校正矩阵 $\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}$ 就能够得到, 收敛性条件满足, $\mathcal{H}$ 矩阵不用计算.

6 How to choose the correction matrix in (5.7)

构造校正矩阵 $\mathcal{M}$ 的方法并不神秘！如果 $\mathcal{Q}^{-T}$ 结构简单, 计算也是非常容易的.

In order to simplify the notations to be used, we define the following $p \times p$ block matrices:

$$\mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{6.1}$$

We also define the $1 \times p$ block matrix

$$\mathcal{E} = \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}. \tag{6.2}$$

The matrix $\mathcal{Q}$ in (5.4) has the form

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix}.$$

To further analyze the correction steps associated with the correction matrix $\mathcal{M}$, let us take a closer look at the matrix $\mathcal{Q}^{-T}$.

According to the prediction (4.6), the matrix $\mathcal{Q}$ in (5.4), we have

$$\mathcal{Q}^T = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E} & I_m \end{pmatrix}, \qquad \mathcal{Q}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix}. \qquad (6.3)$$

and in detail, 请注意 $\mathcal{E}\mathcal{L}^{-T} = (I_m, 0, \ldots, 0)$

$$\mathcal{Q}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

The calculation $\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}$ is essentially very easy for different $\mathcal{D}$!

Since

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix},$$

it can be decomposed as

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix}.$$

The both matrices in the right hand side are positive definite. If we chose

$$\mathcal{D} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} \qquad \text{and thus} \qquad \mathcal{G} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix}.$$

Indeed,

$$\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix}.$$

$$\mathcal{M} = \underbrace{\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}}_{Q^{-T}} \underbrace{\begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}}_{\mathcal{D}} = \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu I_m & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

Conversely, we can also choose

$$\mathcal{D} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} \quad \text{and} \quad \mathcal{G} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}$$

and thus get the another correction method.

There are many positive definite decompositions of $\mathcal{Q}^T + \mathcal{Q}$. For example,

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} (1-\nu)\mathcal{I} & 0 \\ 0 & (1-\nu)I_m \end{pmatrix} + \begin{pmatrix} \nu\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & (1+\nu)I_m \end{pmatrix},$$

we can set

$$\mathcal{D} = \begin{pmatrix} (1-\nu)\mathcal{I} & 0 \\ 0 & (1-\nu)I_m \end{pmatrix} \text{ and } \mathcal{G} = \begin{pmatrix} \nu\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & (1+\nu)I_m \end{pmatrix}$$

or vice versa.

Another example,

$$\mathcal{Q}^T + \mathcal{Q} = \alpha(\mathcal{Q}^T + \mathcal{Q}) + (1-\alpha)(\mathcal{Q}^T + \mathcal{Q}), \quad \alpha \in (0,1).$$

we can choose $\mathcal{D} = \alpha(\mathcal{Q}^T + \mathcal{Q})$. Thus

$$\begin{aligned} \mathcal{Q}^{-T}\mathcal{D} &= \alpha\left\{\begin{pmatrix} \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix}\begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix}\right\} \\ &= \alpha\begin{pmatrix} 2\mathcal{I} & \mathcal{L}^{-T}\mathcal{E}^T \\ -\mathcal{E} & I_m \end{pmatrix}. \qquad \mathcal{L}^{-T}\mathcal{E}^T = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix} \end{aligned}$$

计算用到了 $\mathcal{E}\mathcal{L}^{-T} = (I_m, 0, \ldots, 0)$, 和 $\mathcal{E}\mathcal{L}^{-T}\mathcal{E}^T = I_m$.

7 Conclusions

- 我的报告中常用的一个题目是“构造凸优化的分裂收缩算法– 用好 VI 和 PPA 两大法宝”, 是指构造变分不等式意义下的PPA 算法, 文章首先发表在SIAM J. Imaging Science [13]. 后来又做了一些人为地将预测矩阵设计成对称正定矩阵的方法 [2, 3], 包括我们 2021 年才提出的均困平衡的增广拉格朗日乘子法 [16]. 有时我们也称这样的方法为按需定制的 PPA - (Customized PPA).
- 对预测矩阵 Q 为非对称的预测-校正方法, 利用统一框架的套路证明收敛性, 最初出现在我和袁晓明 (Xiaoming Yuan) 2012 年 SIAM 数值分析的文章 [12] 中, 后面我们发表的一些论文 [8, 9, 15], 都用这个套路证明收敛性. 以后把它归结为统一框架, 出现在我的主页的讲义和报告的 PPT 中.
- 第一次在正式出版物里提到这个统一框架, 是在 2016 年《高校计算数学学报》的我的中文文章 [4] 中. 2018 年我在《运筹学学报》的综述文章 “我和乘子交替方向法 20 年”[5] 中指出, 我们发表的方法都可以用这个框架非常简单地证明收敛性. 英文出版物中首次出现统一框架的是我和袁晓明 2018 年在 COAP 的文章 [14].
- 从 2018 年开始, 我在自己的报告和论文 [6] 中, 经常讲用统一框架去构造算法主要还是根据简单的原则按收敛条件去“凑”. 如何根据确定的预测矩阵 Q 凑出满足收敛条件的校正矩阵 M. 似乎给人一种难以效仿的神秘感觉.

- 今年初我在南师大、中科大和南航做线上报告, 听众提出一些问题, 教学相长, 促使得到一些新的看法, 觉得有必要将回答整理成下面的材料与听众共享.
- 我们从预测矩阵满足 $Q^T + Q \succ 0$ 出发. 根据条件 $HM = Q$, 我们有

$$H = QM^{-1}.$$

因为 H 是正定矩阵, 必须对称. 从上式又看到, H 有个左因子 Q, 那它必须有个右因子 Q^T, 中间夹一个"待定的"正定矩阵. 我们设这个正定矩阵为 D^{-1}, 则有

$$H = QD^{-1}Q^T.$$

比较上面两式, 我们得到 $M^{-1} = D^{-1}Q^T$, 因此

$$\boldsymbol{M = Q^{-T}D}.$$

这个我们大概在 10 年前就知道. 当时往往考虑选择的 D 应该是个块对角矩阵.

- 至此, 我们还不知道矩阵 D 具体形式是什么. 计算一下收敛性条件中的 M^THM,

$$M^THM = \left(DQ^{-1}\right)\left(QD^{-1}Q^T\right)\left(Q^{-T}D\right) = D.$$

上式已经出现在我 2018 的暑期讲习班的讲义中, 没有向前再迈一步.

- 利用上式和 $G = Q^T + Q - M^THM \succ 0$, 这个待定的正定矩阵 D 只需要满足

$$\boldsymbol{0 \prec D \prec Q^T + Q} \qquad (\text{因此}, 0 \prec G = Q^T + Q - D)$$

就可以了. 明确这一条, 得益于为 2022 年以来在 南师大, 南航和中科大的一些讲座, 迫使我深入思考, 争取把方法讲明白.

- 在选了满足上述条件的矩阵 D 以后, 根据确定的 Q 和 D, 找未知矩阵 H 和 M 使得

$$HM = Q \qquad \text{和} \qquad M^T HM = D,$$

我们的目的就达到了.

- 这样的 M 和 H: 可以通过求解下面的矩阵方程组得到.

$$\left\{ \begin{array}{rcl} HM & = & Q, \\ M^T HM & = & D. \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcl} HM & = & Q, \\ Q^T M & = & D. \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcl} H & = & QD^{-1}Q^T, \\ M & = & Q^{-T}D. \end{array} \right. .$$

- 选择不同的满足条件的矩阵 D (这非常容易), 就有不同的校正方法. 譬如说,

$$D = \alpha[Q^T + Q], \quad \alpha \in (0, 1).$$

- 报告的第 4.2节, 对一般线性约束凸优化问题, 采用 primal-dual 预测, 子问题的求解方式是 ADMM 类型的逐个向前. 我们需要的 $\boldsymbol{Q^{-T}}$ 形式非常简单. 是的, 它需要额外的校正. 可喜的是, 校正花费很少, 又特别容易实现!

- 我们特别推崇“预测-校正”, 尤其是那种代价很小的校正. 生机勃勃的果树, 修剪就是校正. 交替按序预测, 降低了问题难度; 全局整体校正, 把握了收敛方向.

- 预测-校正方法既可以用来求解等式约束的问题, 又可以用来求解不等式约束的问题. 适用从一块到任意多块的可分离问题, 算法结构和收敛性证明完全统一.

- 适用范围广的算法会不会影响效率？对经典 ADMM 擅长的两块可分离的等式约束凸优化问题, 我们用第 4.2 节提到的带校正的交替方向法去求解, 与网上他人提供的 ADMM 代码比较, 发现这种担心是多余的.

- **Question A.** In the prediction step, how to arrange a “good” prediction matrix whose matrix $\mathcal{Q}$ satisfies
$$\mathcal{Q}^T + \mathcal{Q} \succeq \mathcal{I}.$$

- **Question B** For the given prediction matrix $\mathcal{Q}$, what are the criteria for choosing matrix $\mathcal{D}$ which satisfies
$$0 \prec \mathcal{D} \prec \mathcal{Q}^T + \mathcal{Q}.$$

- 对线性约束的多块可分离凸优化问题, 只要了解 Gauss 消去法的基本程式, 就可以非常容易地设计出一簇收缩方法. 过去我们辛辛苦苦 “凑” 出来的每个方法, 都是现在这一簇方法中的一个特例. 我们的研究实践再一次证明, 无需身怀绝技, 只要不 “洗手”, 跟着感觉走, 更好的结果往往就在前面不远处向我们招手.

这个报告的主要参考文献是 [11] 和 [17].

希望各位以质疑的态度审视我的观点, 错误的地方请批评指正.

References

[1] A. Beck, First-Order Methods in Optimization, SIAM, Philadelphia, (2017).

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

[3] 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.

[4] B. S. He, From the projection and contraction methods for variatioonal inequality to the splittingcontraction methods fro convex optimization (in Chinese), Numerical Mathematics, A Journal of Chinese Universities, 38 (2016), 74-96.

[5] B. S. He, My 20 years research on alternating directions method of multipliers (in Chinese), Operations Research Trasactions, 22 (2018), 1-31.

[6] B. S. He, Using a unified framework to design the splitting and contraction methods for convex optimization (in Chinese), Numerical Mathematics, A Journal of Chinese Universities, 44 (2022), 1-35.

[7] B.S. He, F. Ma, S. J. Xu and X. M. Yuan, A rank-two relaxed parallel splitting version of the augmented Lagrangian method with step size in (0, 2) for separable convex programming, arXiv:2205.02723[math.OC].

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

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

[10] 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.

[11] 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]

Handbook of of Numerical Analysis ISSN 1570-8659, https://doi.org/10.1016/bs.hna.2022.08.02

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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]

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

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

[20] 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.

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

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

[23] 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.