构造求解凸优化的分裂收缩算法

—用好变分不等式和邻近点算法两大法宝

Bingsheng HE

Department of Mathematics, SUSTech, and
Department of Mathematics, Nanjing University

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

根据作者 2018 年 8 月在港中文(深圳) 和 2019 年 5 月在华北电力大学的报告整理

连续优化中一些代表性数学模型

- 简单约束问题 $\min\{f(x) \mid x \in \mathcal{X}\}$ 其中 $\mathcal{X}$ 是一个凸集.
- min-max 问题 $\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} \{\mathcal{L}(x,y) = \theta_1(x) - y^T Ax - \theta_2(y)\}$
- 线性约束的凸优化问题 $\min\{\theta(x) | Ax = b$ (or $\geq b$), $x \in \mathcal{X}\}$
- 结构型凸优化 $\min\{\theta_1(x) + \theta_2(y) | Ax + By = b,\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$
- 三个算子的凸优化
  $\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) | Ax+By+Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}$

用“瞎子爬山”判定最优， 靠“步步为营”到达最优.

变分不等式 (VI) 是瞎子爬山判定山顶的数学表达形式

邻近点算法 (PPA) 是步步为营 稳扎稳打的求解方法

Outline

- Preliminaries: Optimization problem and VI
- PPA for monotone variational inequality and its beyond
- P-C Methods with parameters requirements in the prediction
- P-C Methods without parameters requirements in the prediction
- Applications for linearly constrained convex optimization.
- Applications for linearly constrained separable convex optimization.

P-C Method is the abbreviation of Prediction-Correction Method

1 Preliminaries: 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 descent direction.

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 $F(x) = \nabla f(x)$.

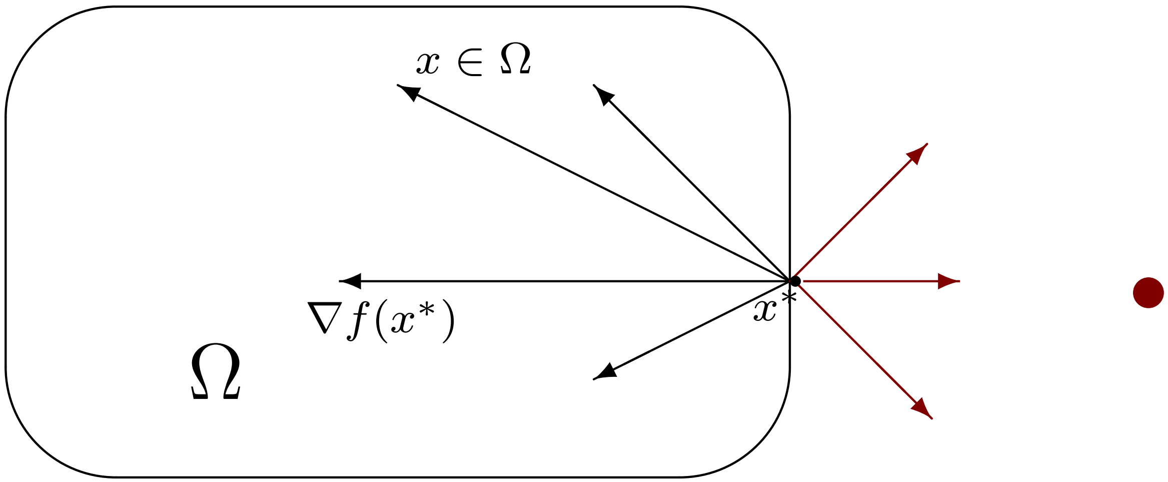


Fig. 1.1 Differential Convex Optimization and VI

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 1.1 *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

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

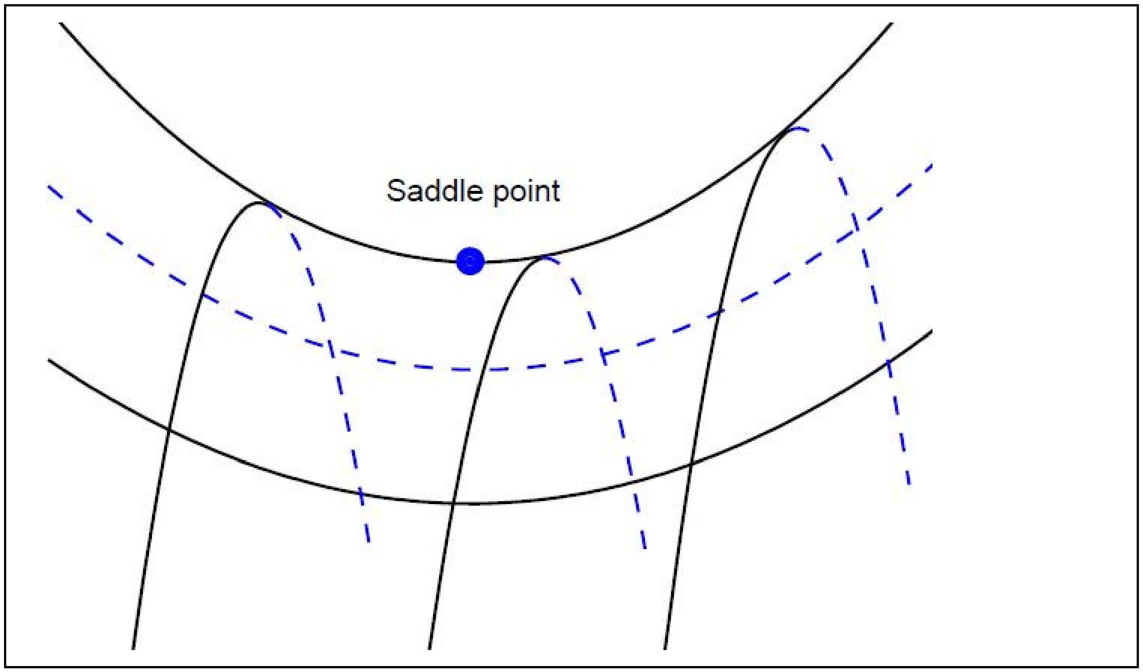
1.2 Linearly constrained Optimization in form of VI

We consider the linearly constrained convex optimization problem

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

The Lagrange function of (1.4) is

$$L(u, \lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \qquad (u, \lambda) \in \mathcal{U} \times \Re^m. \tag{1.5}$$



A pair of (u^*, λ^*) is called a saddle point if

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

The above inequalities can be written as

$$\begin{cases} u^* \in \mathcal{U}, \;\; L(u, \lambda^*) - L(u^*, \lambda^*) \ge 0, \;\; \forall\, u \in \mathcal{U}, & \text{(1.6a)} \\ \lambda^* \in \Lambda, \;\; L(u^*, \lambda^*) - L(u^*, \lambda) \ge 0, \;\; \forall\, \lambda \in \Lambda. & \text{(1.6b)} \end{cases}$$

According to the definition of $L(u, \lambda)$ (see(1.5)),

$$L(u, \lambda^*) - L(u^*, \lambda^*) = [\theta(u) - (\lambda^*)^T(\mathcal{A}u - b)] - [\theta(u^*) - (\lambda^*)^T(\mathcal{A}u^* - b)],$$

it follows from (1.6a) that

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

Similarly, for (1.6b), since

$$L(u^*, \lambda^*) - L(u^*, \lambda) = [\theta(u^*) - (\lambda^*)^T(\mathcal{A}u^* - b)] - [\theta(u^*) - (\lambda)^T(\mathcal{A}u^* - b)],$$

we have

$$\lambda^* \in \Re^m, \quad (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \geq 0, \quad \forall\, \lambda \in \Re^m. \tag{1.8}$$

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

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

Thus, the saddle-point can be characterized as the solution of the following VI:

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

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{U} \times \Re^m. \tag{1.10}$$

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

Since $(w - \tilde{w})^T(F(w) - F(\tilde{w})) \geq 0$ is satisfied, we say F is monotone.

Convex optimization problem with two separable functions

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b, \; x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{1.11}$$

The Lagrangian function 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:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall \, w \in \Omega. \tag{1.12}$$

where

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \tag{1.13a}$$

and

$$\theta(u) = \theta_1(x) + \theta_2(y), \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{1.13b}$$

The variational inequality (1.12)-(1.13) has the same form as (1.9)-(1.10).

Convex optimization problem with three separable functions

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

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:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega. \tag{1.14}$$

where

$$w=\begin{pmatrix} x\\ y\\ z\\ \lambda\end{pmatrix},\; u=\begin{pmatrix} x\\ y\\ z\end{pmatrix},\; F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ -C^T\lambda\\ Ax+By+Cz-b\end{pmatrix}, \tag{1.15a}$$

$$\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z), \qquad \Omega=\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\times\Re^m. \tag{1.15b}$$

The variational inequality (1.14)-(1.15) has the same form as (1.9)-(1.10).

2 Proximal point algorithms and its Beyond

2.1 Proximal point algorithms for convex optimization

Convex Optimization Now, let us consider the *simple* convex optimization

$$\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}, \tag{2.1}$$

where $\theta(x)$ and $f(x)$ are convex but $\theta(x)$ is not necessary smooth, $\mathcal{X}$ is a closed convex set. For solving (2.1), the k-th iteration of the proximal point algorithm (abbreviated to PPA) [13, 16] begins with a given x^k, offers the new iterate x^{k+1} via the recursion

$$x^{k+1} = \text{Argmin}\{\theta(x) + f(x) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{2.2}$$

Since x^{k+1} is the optimal solution of (2.2), it follows from Lemma 1.1 that

$$\theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T\{\nabla f(x^{k+1}) + r(x^{k+1} - x^k)\} \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.3}$$

Setting $x = x^*$ in (2.3), it follows that

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

Since $(x^{k+1} - x^*)^T \nabla f(x^{k+1}) \geq (x^{k+1} - x^*)^T \nabla f(x^*) \geq 0$, it follows that

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

Note that if $b^T(a - b) \geq 0$, then

$$\|a\|^2 = \|b + (a - b)\|^2 \geq \|b\|^2 + \|a - b\|^2.$$

and thus

$$\|b\|^2 \leq \|a\|^2 - \|a - b\|^2. \tag{2.5}$$

Setting $a = x^k - x^*$ and $b = x^{k+1} - x^*$ in (2.4) and using (2.5), we obtain

$$\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^2 \leq \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2 - \|\boldsymbol{x}^k - \boldsymbol{x}^{k+1}\|^2, \tag{2.6}$$

which is the nice convergence property of Proximal Point Algorithm.

In other words, The sequence $\{x^k\}$ generated by PPA is Fejér monotone.

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

Proof. Replacing $k+1$ in (2.3) with k, we get

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

Let $x = x^{k+1}$ in the above inequality, it follows that

$$\theta(x^{k+1}) - \theta(x^k) + (x^{k+1} - x^k)^T \{\nabla f(x^k) + r(x^k - x^{k-1})\} \ge 0. \tag{2.7}$$

Setting $x = x^k$ in (2.3), we become

$$\theta(x^k) - \theta(x^{k+1}) + (x^k - x^{k+1})^T \{\nabla f(x^{k+1}) + r(x^{k+1} - x^k)\} \ge 0. \tag{2.8}$$

Adding (2.7) and (2.8) and using $(x^k - x^{k+1})^T [\nabla f(x^k) - \nabla f(x^{k+1})] \ge 0$,

$$(x^k - x^{k+1})^T \{(x^{k-1} - x^k) - (x^k - x^{k+1})\} \ge 0. \tag{2.9}$$

Setting $a = x^{k-1} - x^k$ and $b = x^k - x^{k+1}$ in (2.9) and using (2.5), we obtain

$$\boldsymbol{\|x^k - x^{k+1}\|^2 \le \|x^{k-1} - x^k\|^2 - \|(x^{k-1} - x^k) - (x^k - x^{k+1})\|^2.} \tag{2.10}$$

We write the problem (2.1) and its PPA (2.2) in VI form

The equivalent variational inequality form of the optimization problem (2.1) is

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.11a}$$

For solving the problem (2.1), the variational inequality form of the k-th iteration of the PPA (see (2.3)) is:

$$\begin{aligned} x^{k+1} \in \Omega, \quad \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \nabla f(x^{k+1}) & \\ \geq (x - x^{k+1})^T r(x^k - x^{k+1}), \quad \forall\, x \in \mathcal{X}. & \end{aligned} \tag{2.11b}$$

PPA 通过求解一系列的 (2.2), 求得 (2.1) 的解, 采用的是步步为营的策略.

Using (2.11), we consider the PPA for the variational inequality (1.9)

2.2 Preliminaries of PPA for Variational Inequalities

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{2.12}$$

PPA for VI (2.12) in Euclidean-norm For given w^k and $r > 0$, find w^{k+1},

$$\begin{array}{r} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (w - w^{k+1})^T r(w^k - w^{k+1}), \ \ \forall\, w \in \Omega. \end{array} \tag{2.13}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.12).

✠ **w^k is the solution of (2.12) if and only if $w^k = w^{k+1}$** ✠

Setting $w = w^*$ in (2.13), we obtain

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

Note that (see the structure of $F(w)$ in (1.10))

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

and consequently (by using (2.12)) we obtain

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

Thus, we have

$$(w^{k+1} - w^*)^T (w^k - w^{k+1}) \geq 0. \tag{2.14}$$

By setting $\boldsymbol{a = w^k - w^*}$ and $\boldsymbol{b = w^{k+1} - w^*}$, the inequality (2.14) means that $\boldsymbol{b^T(a - b) \geq 0}$. Similarly as in §2.1, we obtain

$$\|\boldsymbol{w}^{k+1} - \boldsymbol{w}^*\|^2 \leq \|\boldsymbol{w}^k - \boldsymbol{w}^*\|^2 - \|\boldsymbol{w}^k - \boldsymbol{w}^{k+1}\|^2. \tag{2.15}$$

We get the nice convergence property of Proximal Point Algorithm.

The sequence $\{w^k\}$ generated by PPA is Fejér monotone. **As in** (2.10)**, the residue sequence $\{\|w^k - w^{k+1}\|\}$ is also monotonically no-increasing.**

$$\|\boldsymbol{w}^k - \boldsymbol{w}^{k+1}\|^2 \leq \|\boldsymbol{w}^{k-1} - \boldsymbol{w}^k\|^2 - \|(\boldsymbol{w}^{k-1} - \boldsymbol{w}^k) - (\boldsymbol{w}^k - \boldsymbol{w}^{k+1})\|^2.$$

PPA for monotone mixed VI in H-norm

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

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (w - w^{k+1})^T H(w^k - w^{k+1}), \ \forall\, w \in \Omega, \tag{2.16}$$

where H is a symmetric positive definite matrix.

✠ Again, w^k is the solution of (2.12) if and only if $w^k = w^{k+1}$ ✠

Convergence Property of Proximal Point Algorithm in H-norm

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \tag{2.17}$$

The sequence $\{w^k\}$ is Fejér monotone in H-norm. In primal-dual algorithm [5], via choosing a proper positive definite matrix H, the solution of the subproblem (2.16) has a closed form. **In addition, for the residue sequence, we have**

$$\|w^k - w^{k+1}\|_H^2 \leq \|w^{k-1} - w^k\|_H^2 - \|(w^{k-1} - w^k) - (w^k - w^{k+1})\|_H^2.$$

2.3 Splitting Methods in a Unified Framework [11, 17]

We study the algorithms using the guidance of variational inequality.

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{2.18}$$

Algorithms in a unified framework

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{2.19a}$$

where the matrix Q is not necessary symmetric, but $Q^T + Q$ is positive definite.

[Correction Step.] The new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \tag{2.19b}$$

统一框架算法中的 u 和 v, 可以是 w 本身, 也可以是 w 的部分分量.

- 如果 (2.19a) 中的 Q 对称正定, 将 $\tilde{v}^k$ 和 Q 分别看成 (2.16) 中的 w^{k+1} 和 H, 就是一个标准的 H 模下的 PPA 算法.
- 现在不要求 Q 对称, 但需要 $Q^T + Q$ 正定. 我们可以把 (2.19a) 产生的点当预测点, 通过(2.19b) 校正得到新的迭代点.

Convergence Conditions

For the matrices Q and M in (4.3), there is a positive definite matrix H such that

$$HM = Q. \tag{2.20a}$$

Moreover, the matrix

$$G = Q^T + Q - \alpha M^T HM \tag{2.20b}$$

is positive semi-definite.

The Key Identity in the convergence proofs is

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

Convergence of the algorithms （证明可见 [11, 17]）

Theorem 2.1 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (2.18) *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*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.21}$$

上式是跟 (2.17) 类似的收缩不等式, 所以说这类方法是 PPA Like 方法.

关于统一框架下算法及其收敛性证明可以参考下面的文章：

- 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.
- 何炳生, 我和乘子交替方向法 20 年,《运筹学学报》22 卷第1期, pp. 1-31, 2018.

PPA 类算法步步为营, 稳扎稳打；缺点是思想保守, 影响速度与精度.

3 Prediction-Correction Methods I

Our objective is to solve the variational inequality:

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

For this purpose, we suggest two kinds of prediction-correction methods.

3.1 Algorithms I

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

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

where the matrix H is symmetric and positive definite.

[Correction Step.] The new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0, 2) \qquad (3.2b)$$

H is a symmetric positive definite matrix. 预测往往对参数有要求

3.2 Convergence of the prediction-correction method I

Lemma 3.1 *For given* v^k*, let the predictor* $\tilde{w}^k$ *be generated by* (3.2a)*, then we have*

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_H^2, \tag{3.3}$$

where H *is the positive definite matrix in the right hand side of* (3.2a)*.*

Proof. Set $w = w^*$ in (3.2a), we get

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \tag{3.4}$$

Because

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

and

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

the right hand side of (3.4) is non-negative. Thus, we have

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

Consequently, we get (3.3). The lemma is proved. □

Convergence in a strictly contraction sense

Theorem 3.1 *For given* v^k*, let the predictor* $\tilde{w}^k$ *be generated by* (3.2a)*. If the new iterate* v^{k+1} *is given by*

$$v^{k+1}(\alpha) = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0,2), \tag{3.5}$$

then we have

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - q_k^I(\alpha), \quad \forall v^* \in \mathcal{V}^*, \tag{3.6}$$

where

$$q_k^I(\alpha) = \alpha(2-\alpha)\|v^k - \tilde{v}^k\|_H^2. \tag{3.7}$$

Proof. First, we define the profit function by

$$\vartheta_k^I(\alpha) = \|v^k - v^*\|_H^2 - \|v^{k+1}(\alpha) - v^*\|_H^2. \tag{3.8}$$

Thus, it follows from (3.5) that

$$\begin{aligned} \vartheta_k^I(\alpha) &= \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \alpha(v^k - \tilde{v}^k)\|_H^2 \\ &= 2\alpha(v^k - v^*)^T H(v^k - \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_H^2. \end{aligned}$$

By using (3.3) and (3.7), we get

$$\begin{aligned} \vartheta_k^I(\alpha) &\geq 2\alpha\|v^k - \tilde{v}^k\|_H^2 - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \\ &= \alpha(2-\alpha)\|v^k - \tilde{v}^k\|_H^2 \;=\; q_k^I(\alpha). \qquad \Box \end{aligned}$$

According to (3.6) and (3.7), the sequence $\{v^k\}$ generated by the prediction-correction method (3.2) satisfies

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

The above inequality is the Key for convergence analysis !

上式是和 (2.17) 类似的不等式. 因此, 方法具有 PPA Like 收敛性质.

4 Prediction-Correction Methods II

Recall our objective is to solve the variational inequality:

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

This section presents the second kind of prediction-correction method.

4.1 Algorithms II

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

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

where

$$Q = D + K, \qquad (4.2b)$$

D is a block diagonal positive definite matrix

K is skew-symmetric (反对称) $\quad Q^T + Q = 2D$

[Correction Step.] For the diagonal matrix $D \succ 0$ in (4.2b), the new iterate v^{k+1} is given by

$$v^{k+1} = v^k - \gamma \alpha_k^* M(v^k - \tilde{v}^k), \tag{4.3a}$$

where

$$M = D^{-1}Q, \qquad \gamma \in (0, 2),$$

and the optimal step size is given by

$$\alpha_k^* = \frac{\|v^k - \tilde{v}^k\|_D^2}{\|M(v^k - \tilde{v}^k)\|_D^2}. \tag{4.3b}$$

Since $M^T DM = M^T Q$, we have

$$\|M(v^k - \tilde{v}^k)\|_D^2 = \big[M(v^k - \tilde{v}^k)\big]^T \big[Q(v^k - \tilde{v}^k)\big]$$

and thus

$$\alpha_k^* = \frac{\|v^k - \tilde{v}^k\|_D^2}{[M(v^k - \tilde{v}^k)]^T [Q(v^k - \tilde{v}^k)]}.$$ 数据齐全, 计算并不困难

4.2 Convergence of the prediction-correction method II

Lemma 4.1 *For given* v^k*, let the predictor* $\tilde{w}^k$ *be generated by* (4.2a)*, then we have*

$$(v^k - v^*)^T Q(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_D^2, \tag{4.4}$$

where Q *is given in the right hand side of* (4.2a) *and* D *is given in* (4.2b)*.*

Proof. Set $w = w^*$ in (4.2a), we get

$$(\tilde{v}^k - v^*)^T Q(v^k - \tilde{v}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \tag{4.5}$$

Because

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

and

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

the right hand side of (4.5) is non-negative. Thus, we have

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

and

$$(v^k - v^*)^T Q(v^k - \tilde{v}^k) \geq (v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{4.6}$$

For the right hand side of the above inequality, by using $Q = D + K$ and the skew-symmetry of K, we obtain

$$\begin{aligned} (v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) &= (v^k - \tilde{v}^k)^T (D + K)(v^k - \tilde{v}^k) \\ &= \|v^k - \tilde{v}^k\|_D^2. \end{aligned}$$

The lemma is proved. □

Theorem 4.1 *For given* v^k*, let the predictor* $\tilde{w}^k$ *be generated by* (4.2a)*. If the new iterate* v^{k+1} *is given by*

$$v^{k+1}(\alpha) = v^k - \alpha M(v^k - \tilde{v}^k), \quad \gamma \in (0, 2), \tag{4.7}$$

then we have

$$\|v^{k+1} - v^*\|_D^2 \leq \|v^k - v^*\|_D^2 - q_k^{II}(\alpha), \quad \forall v^* \in \mathcal{V}^*, \tag{4.8}$$

where

$$q_k^{II}(\alpha) = 2\alpha\|w^k - \tilde{w}^k\|_D^2 - \alpha^2\|M(w^k - \tilde{w}^k)\|_D^2. \tag{4.9}$$

Proof. First, we define the profit function by

$$\vartheta_k^{II}(\alpha) = \|v^k - v^*\|_D^2 - \|v^{k+1}(\alpha) - v^*\|_D^2. \tag{4.10}$$

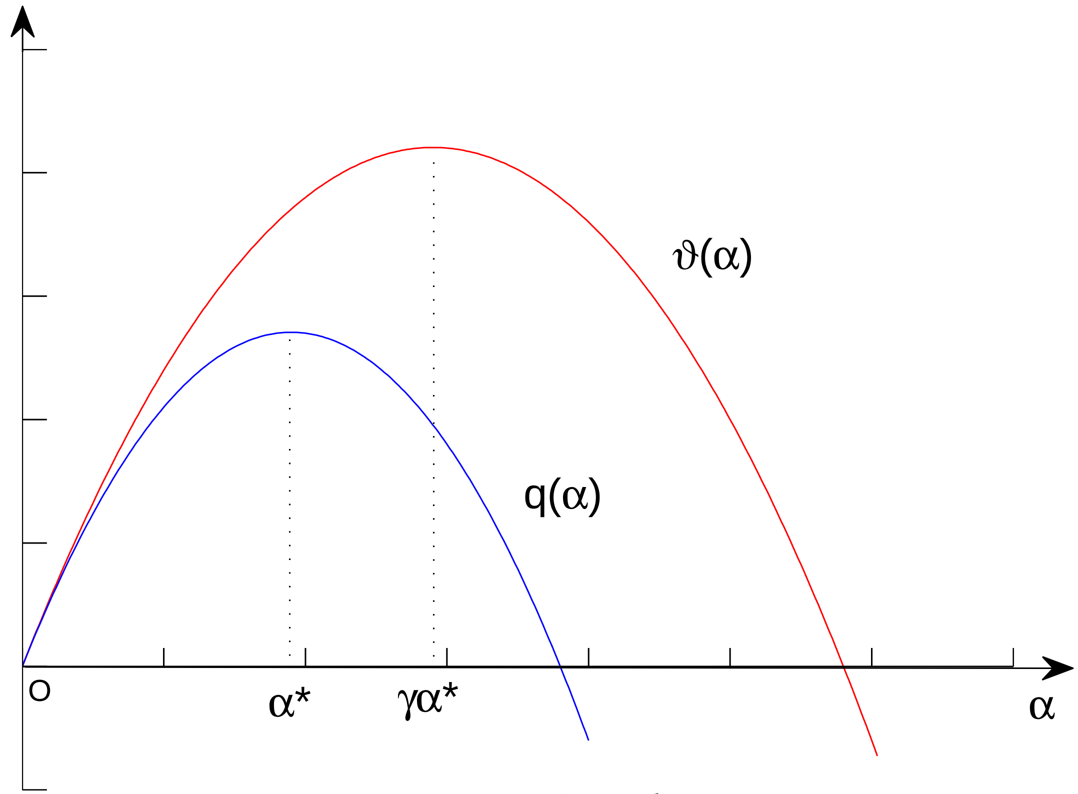
Thus, it follows from (4.7) that

$$\begin{aligned} \vartheta_k^{II}(\alpha) &= \|v^k - v^*\|_D^2 - \|(v^k - v^*) - \alpha M(v^k - \tilde{v}^k)\|_D^2 \\ &= 2\alpha(v^k - v^*)^T DM(v^k - \tilde{v}^k) - \alpha^2\|M(v^k - \tilde{v}^k)\|_D^2. \end{aligned}$$

By using $DM = Q$ and (4.4), we get

$$\vartheta_k^{II}(\alpha) \;\geq\; 2\alpha\|v^k - \tilde{v}^k\|_D^2 - \alpha^2\|M(v^k - \tilde{v}^k)\|_D^2 \;=\; q_k^{II}(\alpha). \qquad \Box$$

$q_k^{II}(\alpha)$ reaches its maximum at α_k^* which is given by $(4.3b)$.



取 $\gamma \in [1,2)$ 的示意图

Since we take $\alpha = \gamma\alpha_k^*$, it follows from (4.9) that

$$q_k^{II}(\alpha) = 2\gamma\alpha_k^*\|v^k - \tilde{v}^k\|_D^2 - \gamma^2(\alpha_k^*)^2\|M(v^k - \tilde{v}^k)\|_D^2. \qquad (4.11)$$

By using (4.3b), we get

$$
\begin{aligned}
(\alpha_k^*)^2\|M(v^k-\tilde{v}^k)\|_D^2 &= \alpha_k^* \frac{\|v^k-\tilde{v}^k\|_D^2}{\|M(v^k-\tilde{v}^k)\|_D^2}\|M(v^k-\tilde{v}^k)\|_D^2 \\
&= \alpha_k^*\|v^k-\tilde{v}^k\|_D^2.
\end{aligned}
$$

Substituting it in (4.11) we get

$$
q_k^{II}(\alpha) \ge \gamma(2-\gamma)\alpha_k^*\|v^k-\tilde{v}^k\|_D^2. \tag{4.12}
$$

According to (4.8) and (4.12), the sequence $\{v^k\}$ generated by the prediction-correction Algorithm II satisfies

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

上式是跟 (2.17) 类似的不等式, 预测-校正方法都具有 PPA Like 收敛性质.

所以, 这个报告中所说的方法, 都是类邻近点 (PPA Like) 算法.

5 Methods for Linearly Constrained Problems

This section presents various applications of the proposed algorithms for the convex optimization (1.4), namely

$$\min\{\theta(u) \mid Au = b,\ u \in \mathcal{U}\}. \tag{5.1}$$

5.1 Augmented Lagrangian Method

Its augmented Lagrangian function is

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

The k-th iteration of the **Augmented Lagrangian Method** [12, 15] begins with a given λ^k, obtain $w^{k+1} = (u^{k+1}, \lambda^{k+1})$ via

$$\text{(ALM)} \quad \begin{cases} \tilde{u}^k = \arg\min\big\{\mathcal{L}_\beta(u,\lambda^k) \;\big|\; u \in \mathcal{U}\big\}, & \text{(5.2a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{u}^k - b). & \text{(5.2b)} \end{cases}$$

In (5.2), $\tilde{u}^k$ is only a computational result of (5.2a) from given λ^k, it is called the intermediate variable. In order to start the k-th iteration of ALM, we need only to have λ^k and thus we call it as the essential variable.

The mathematical form of the subproblem (5.2a) is

$$\min\{\theta(u) + \frac{\beta}{2}\|Au - (b + \frac{1}{\beta}\lambda^k)\|^2 \,|\, u \in \mathcal{U}\} \tag{5.3}$$

Assumption: The solution of problem (5.3) has closed-form solution or can be efficiently computed with a high precision.

The optimal condition of (5.2) (k-th iteration of ALM) can be written as $\tilde{w}^k \in \Omega$,

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

The above relations can be written as

$$\theta(u)-\theta(\tilde{u}^k)+\begin{pmatrix} u-\tilde{u}^k \\ \lambda-\tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{u}^k-b \end{pmatrix} \ge (\lambda-\tilde{\lambda}^k)^T\frac{1}{\beta}(\lambda^k-\tilde{\lambda}^k), \ \forall w\in\Omega. \quad (5.4)$$

Setting $v=\lambda$ in (5.4), it can be written as (3.2a),

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TH(v^k-\tilde{v}^k), \ \forall w\in\Omega,$$

with

$$H=\tfrac{1}{\beta}I.$$

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

增广拉格朗日乘子法(ALM) [13, 15] 可以看作算法 I 的特例

增广拉格朗日乘子法是好算法, 如果求解子问题 (5.2a),

$\min\{\theta(u)+\frac{\beta}{2}\|Au-(b+\frac{1}{\beta}\lambda^k)\|^2 \mid u\in\mathcal{U}\}$ 容易实现的话.

否则, 就要考虑线性化的方法, Primal- Dual 一类子问题简单的算法.

5.2 C-P Algorithm and Customized PPA

Recall the convex optimization problem (1.4), namely,

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

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

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

where

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

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

$$\text{(CPPA)} \quad \left\{ \begin{array}{ll} \tilde{u}^k = \arg\min\big\{L(u, \lambda^k) + \frac{r}{2}\|u - u^k\|^2 \;\big|\; u \in \mathcal{U}\big\}, & \text{(5.5a)} \\ \tilde{\lambda}^k = \arg\max\big\{L\big([2\tilde{u}^k - u^k], \lambda\big) - \frac{s}{2}\|\lambda - \lambda^k\|^2\big\} & \text{(5.5b)} \end{array} \right.$$

The output $\tilde{w}^k \in \Omega$ of the iteration (5.5) satisfies

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

It is a form of (3.2a) where

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

Assumption: To ensure the positiveness of the matrix Q, we have to set $rs > \|A^T A\|$. 此时, 如同(3.2b), 用 $w^{k+1} = w^k - \alpha(w^k - \tilde{w}^k)$ 产生 w^{k+1}.
✠ 求解问题 (5.1), 如果 §5.1 中的子问题不难, 不要用 §5.2 的简易方法.

Chambolle-Pock [3], (CPPA) [5] 可以看作算法 I 的特例

求解 $\min_{u \in \mathcal{U}} \{\theta(u) + \frac{r}{2}\|u - b^k\|^2\}$ 相对容易. 但 $rs > \|A^T A\|$ 会影响速度

5.3 The method does not need $rs > \|A^T A\|$ [9]

Recall the convex optimization problem (1.4), namely,

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

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

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

where

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

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

$$\text{(Prediction)} \quad \begin{cases} \tilde{u}^k = \arg\min\{L(u, \lambda^k) + \frac{r}{2}\|u - u^k\|^2 \mid u \in \mathcal{U}\}, & \text{(5.6a)} \\ \tilde{\lambda}^k = \arg\max\{L(u^k, \lambda) - \frac{s}{2}\|\lambda - \lambda^k\|^2\} & \text{(5.6b)} \end{cases}$$

The output $\tilde{w}^k \in \Omega$ of the iteration (5.6) satisfies

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

It is a form of (4.2a) where

$$Q = \begin{pmatrix} rI & A^T \\ -A & sI \end{pmatrix}. \tag{5.7}$$

Indeed,

$$Q = D + K = \begin{pmatrix} rI & 0 \\ 0 & sI \end{pmatrix} + \begin{pmatrix} 0 & A^T \\ -A & 0 \end{pmatrix}.$$

子问题 $\min_{u\in\mathcal{U}}\{\theta(u) + \frac{r}{2}\|u - b^k\|^2\}$ 类型不变. 预测只需要 $r, s > 0$.

这是算法 II, 用校正 (4.3) 产生新的迭代点. 此法在 [9] 中已经有介绍.

✠ 对问题 (5.1), 如果矩阵 A^TA 的条件数不坏, 就可以用 §5.2 的方法.

6 Applications for separable problems

This section presents various applications of the proposed algorithms for the separable convex optimization problem

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{6.1}$$

Its VI-form is

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{6.2}$$

where

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \tag{6.3a}$$

and

$$\theta(u) = \theta_1(x) + \theta_2(y), \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{6.3b}$$

The augmented Lagrangian Function of the problem (6.1) is

$$\mathcal{L}_\beta(x,y,\lambda) = \theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)+\frac{\beta}{2}\|Ax+By-b\|^2. \quad (6.4)$$

Solving the problem (6.1) by using ADMM, the k-th iteration begins with given (y^k,λ^k), it offers the new iterate (y^{k+1},λ^{k+1}) via

$$\text{(ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x,y^k,\lambda^k) \mid x\in\mathcal{X}\}, & \text{(6.5a)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1},y,\lambda^k) \mid y\in\mathcal{Y}\}, & \text{(6.5b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1}+By^{k+1}-b). & \text{(6.5c)} \end{array} \right.$$

Let

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix}, \quad H = \begin{pmatrix} \beta B^TB & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}$$

and

$$\mathcal{V}^* = \{(y^*,\lambda^*) \,|\, (x^*,y^*,\lambda^*)\in\Omega^*\},$$

The sequence $\{v^k\}$ generated by ADMM has the similar contractive property:

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2. \tag{6.6}$$

This is similar as the contractive property (2.6) of PPA for the “simple” optimization problem (2.1) in §2.1 .

因此, 交替方向法本质上是关于向量 (By, λ) 的邻近点算法.

The residue sequence $\{\|v^k - v^{k+1}\|_H\}$ generated by ADMM is also monotonically no-increasing. In practical, we have

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

✠ For a simple proof, please see [10]: 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.

先考虑根据算法 I 的要求　设计预测公式的方法.

6.1 ADMM in PPA-sense

In order to solve the separable convex optimization problem (6.1), we construct a method whose prediction-step is

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{6.7a}$$

where

$$H = \begin{pmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0\text{, say } \delta = 0.05\text{)}. \tag{6.7b}$$

Since H is positive definite, we can use the update form of Algorithm I to produce the new iterate $v^{k+1} = (y^{k+1}, \lambda^{k+1})$. (In the algorithm [2], we took $\delta = 0$).

The concrete form of (6.7) is

$$\left\{\begin{array}{l} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad\qquad (-A^T\tilde{\lambda}^k) \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad\qquad \{-B^T\tilde{\lambda}^k + (1+\delta)\beta B^T B(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ (A\tilde{x}^k + B\tilde{y}^k - b) \qquad -B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{array}\right.$$

Let $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, the prediction can be implemented by

$$\left\{\begin{array}{ll} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, & \text{(6.8a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & \text{(6.8b)} \\ \tilde{y}^k = \text{Argmin}\left\{ \begin{array}{l} \theta_2(y) - [2\tilde{\lambda}^k - \lambda^k]^T By \\ + \frac{1+\delta}{2}\beta\|B(y - y^k)\|^2 \end{array} \Bigg| \, y \in \mathcal{Y} \right\}. & \text{(6.8c)} \end{array}\right.$$

这个预测与经典的交替方向法 (6.5) 完全相当, 采用(3.2b) 校正, 会加快速度.

6.2 Linearized ADMM-Like Method

当子问题 (6.8c) 求解有困难时, 用 $\frac{s}{2}\|y-y^k\|^2$ 代替 $\frac{1+\delta}{2}\beta\|B(y-y^k)\|^2$.

By using the linearized version of (6.8), the prediction step becomes

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T H(v^k-\tilde{v}^k), \ \forall w \in \Omega, \quad (6.9)$$

where

$$H = \begin{pmatrix} sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix}, \text{ 代替 (6.7) 中的 } \begin{pmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix}. \quad (6.10)$$

The concrete formula of (6.9) is

$$\begin{cases} \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad (-A^T\tilde{\lambda}^k) \ge 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T \\ \qquad \{-B^T\tilde{\lambda}^k + s(\tilde{y}^k-y^k) - B^T(\tilde{\lambda}^k-\lambda^k)\} \ge 0, \\ (A\tilde{x}^k+B\tilde{y}^k-b) - B(\tilde{y}^k-y^k) + (1/\beta)(\tilde{\lambda}^k-\lambda^k) = 0. \end{cases} \quad (6.11)$$

Then, we use the form

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

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

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (6.11),

we need only use the following procedure:

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\theta_2(y) - [2\tilde{\lambda}^k - \lambda^k]^T By + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \end{cases}$$

用 $\frac{s}{2}\|y - y^k\|^2$ 代替 $\frac{1+\delta}{2}\beta\|B(y - y^k)\|^2$, 为保证收敛, 需要 $s > \beta\|B^TB\|$.

对给定的 $\beta > 0$, 要求 $s > \beta\|B^TB\|$, 太大的 s 会影响收敛速度

6.3 Method without $s > \beta\|B^TB\|$

当矩阵 B^TB 的条件不好, 又必须线性化, 就按照算法II 进行预测

For solving the same problem, we give the following prediction:

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{6.12a}$$

where

$$Q = \begin{pmatrix} sI & B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} = D + K. \tag{6.12b}$$

Because

$$D = \begin{pmatrix} sI & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \quad \text{and} \quad K = \begin{pmatrix} 0 & B^T \\ -B & 0 \end{pmatrix},$$

根据这样的预测, 可以用算法 II 的校正公式 (4.3) 产生新的迭代点.

How to implement the prediction? The concrete formula of (6.12) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad\qquad (-A^T\tilde{\lambda}^k) \ge 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad\qquad \{-B^T\tilde{\lambda}^k + s(\tilde{y}^k - y^k) + B^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \\ (A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

This can be implemented by

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\theta_2(y) - (\lambda^k)^T By + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \end{cases}$$

The y-subproblem is easy. 对给定的 $\beta > 0$, 可以取任意的 $s > 0$.

总结：对两类问题, 我们分别在 §5 和 §6 中提出三种预测-校正方法

- 如果子问题中求解过程中, 二次项不带来任何困难的时候, 建议分别采用 §5.1 和 §6.1 中的方法.
- 如果子问题中求解中, 必须对一个子问题中的二次项线性化, 并且矩阵条件好的时候, 建议分别采用 §5.2 和 §6.2 中的方法.
- 如果必须线性化, 矩阵条件又不好的时候, 建议分别采用 §5.3 和 §6.3 中的方法.

希望这些框架能为针对实际问题设计算法提供帮助.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Foundations and Trends in Machine Learning Vol. 3, No. 1 (2010) 1 − 122.

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

[3] 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.

[4] C.H. Chen, B.S. He, Y.Y. Ye and X. M. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessary convergent, Mathematical Programming, 155 (2016) 57-79.

[5] 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, pp. 135-161, 2014.

[6] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach. J. Oper. Res. Soc. China 3 (2015) 391 – 420.

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

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

[9] 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.

[10] 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.

[11] 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.

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

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

[14] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15** (2004), 229–251.

[15] 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.

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

[17] 何炳生, 我和乘子交替方向法 20 年,《运筹学学报》22 卷第1期, pp. 1-31, 2018.

VI 如“瞎子爬山”问是否最优， PPA 以“步步为营”向目标逼近.

Thank you very much for your attention ！