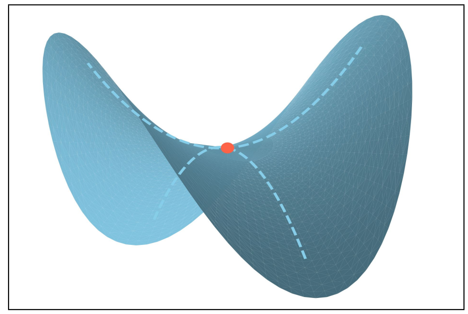
Splitting and contraction methods for convex optimization

Optimization, VI, Algorithmic framework and Generalized-PPA © Bingsheng He

1 Convex optimization and its related variational inequality



We consider the linearly constrained convex optimization

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{1.1}$$

The Lagrangian function of the problem (1.1) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \tag{1.2}$$

which is defined on $\mathcal{U} \times \Re^m$. A pair of $(u^*, \lambda^*) \in \mathcal{U} \times \Re^m$ is called a saddle point of the Lagrangian function (1.2), if

$$L_{u\in\mathcal{U}}(u,\lambda^*) \ge L(u^*,\lambda^*) \ge L_{\lambda\in\Re^m}(u^*,\lambda). \tag{1.3}$$

The two inequalities in (1.3) can be written as $(u^*, \lambda^*) \in \mathcal{U} \times \Re^m$,

$$\begin{cases} L(u,\lambda^*) - L(u^*,\lambda^*) \ge 0, & \forall u \in \mathcal{U}, \\ L(u^*,\lambda^*) - L(u^*,\lambda) \ge 0, & \forall \lambda \in \Re^m. \end{cases} \Longleftrightarrow \begin{cases} \theta(u) - \theta(u^*) + (u-u^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, & \forall u \in \mathcal{U}, \\ (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \ge 0, & \forall \lambda \in \Re^m. \end{cases} \tag{1.4}$$

Combining the above two inequalities, the saddle point is equivalent to the solution of the VI :

$$\text{(VI)} \qquad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{1.5}$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \qquad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{U} \times \Re^m. \tag{1.6}$$

Please notice that $(w - \tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0$. The problem (1.1) is translated to VI (1.5).

For the multi-block separable convex optimization, we take the three-block problem

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) \mid Ax + By + Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\} \tag{1.7}$$

as an example. Its corresponding variational inequality has the form (1.5), where $\Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m$, and

$$u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z), \quad w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix}. \tag{1.8}$$

For the $F(w)$ in (1.8), we still have $(w - \tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0$, for any w and $\tilde{w}$ in the space containing Ω.

2 Algorithmic unified framework for monotone variational inequalities

We focus on how to solve the variational inequality (1.5), following is our algorithmic framework.

Algorithms in a unified framework **(Each iteration of the method consists of a prediction and a correction)**

[Prediction Step]. Start from a given v^k, find a predictor $\tilde{w}^k \in \Omega$, which satisfies

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

where the prediction matrix Q is not necessarily symmetric, but the kernel of $Q^T + Q$ is positive definite.

v is called the essential variable in the iteration which can be equal to w, or a part of the whole vector of w.

[Correction Step]. Find the correction matrix M which satisfied (2.2). The new iteration v^{k+1} is given by

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{2.1b}$$

Convergence Conditions **(It is easy to find the matrix M which satisfies the conditions, see the details in §4)**

[Convergence Conditions]. For the prediction matrix Q in (2.1a) and the correction matrix M in (2.1b), there is a positive definite matrix H, such that

$$HM = Q \quad \text{and} \quad G = Q^T + Q - M^T HM \succ 0. \tag{2.2}$$

3 Convergence proof of the methods in the algorithmic framework

Theorem A. For solving the VI (1.5), let $\{v^k\}$, $\{\tilde{w}^k\}$ be the sequences generated by (2.1). If the conditions (2.2) are satisfied, then we have

$$\tilde{w}^k \in \Omega, \;\; \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge \tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\tfrac{1}{2}\|v^k-\tilde{v}^k\|_G^2, \;\; \forall w \in \Omega. \quad (3.1)$$

and

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \|v^k-\tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \quad (3.2)$$

Proof. Treating the term $Q(v^k-\tilde{v}^k)$ in the RHS of (2.1a) by using $Q = HM$ (see (2.2)) and the correction formula (2.1b), we obtain $Q(v^k-\tilde{v}^k) = HM(v^k-\tilde{v}^k) = H(v^k-v^{k+1})$. Thus we get

$$\tilde{w}^k \in \Omega, \quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T H(v^k-v^{k+1}), \quad \forall w \in \Omega. \quad (3.3)$$

Applying the identity

$$(a-b)^T H(c-d) = \tfrac{1}{2}\big(\|a-d\|_H^2 - \|a-c\|_H^2\big) + \tfrac{1}{2}\big(\|b-c\|_H^2 - \|b-d\|_H^2\big) \quad (3.4)$$

to the RHS of (3.3) with $a = v$, $b = \tilde{v}^k$, $c = v^k$ and $d = v^{k+1}$, we obtain

$$(v-\tilde{v}^k)^T H(v^k-v^{k+1}) = \tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2\big) + \tfrac{1}{2}\big(\|v^k-\tilde{v}^k\|_H^2 - \|v^{k+1}-\tilde{v}^k\|_H^2\big). \quad (3.5)$$

To the second part of the RHS of (3.5), by using $HM = Q$ and $2v^TQv = v^T(Q^T+Q)v$, it follows that

$$\begin{aligned}\|v^k-\tilde{v}^k\|_H^2 - \|v^{k+1}-\tilde{v}^k\|_H^2 &\overset{(2.1b)}{=} \|v^k-\tilde{v}^k\|_H^2 - \|(v^k-\tilde{v}^k) - M(v^k-\tilde{v}^k)\|_H^2 \\ &= (v^k-\tilde{v}^k)^T(Q^T+Q-M^THM)(v^k-\tilde{v}^k) \overset{(2.2)}{=} \|v^k-\tilde{v}^k\|_G^2. \quad (3.6)\end{aligned}$$

Substituting (3.6) in (3.5), and then in (3.3), we get the assertion (3.1) directly. Setting the $w \in \Omega$ in (3.1) by any fixed w^*, then using $(\tilde{w}^k-w^*)^T F(\tilde{w}^k) = (\tilde{w}^k-w^*)^T F(w^*)$ and $\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^T F(w^*) \ge 0$, we obtain (3.2) and the theorem is completely proved. □

This theorem is proved under weak conditions: $Q^T+Q \succeq 0,\; H \succeq 0,\; HM = Q,\; G = Q^T+Q-M^THM \succeq 0.$

Assertion (3.1) is useful for the convergence rate proof of ADMM, see SIAM Numer. Anal. 2012, 50:700-709.

4 The equivalent convergence conditions and the generalized PPA

Under the condition that the prediction matrix Q is nonsingular, it is easy to construct the correction matrix M which satisfies the convergence conditions (2.2). In fact, because $Q^T+Q \succ 0$, we can take

$$D \succ 0, \qquad G \succ 0 \qquad \text{and} \qquad D+G = Q^T+Q. \quad (4.1)$$

Afterwards, we let

$$HM = Q \qquad \text{and} \qquad M^THM = D. \quad (4.2)$$

Because $\left\{\begin{array}{r} HM{=}Q, \\ M^THM{=}D. \end{array}\right. \Leftrightarrow \left\{\begin{array}{r} Q^TM{=}D, \\ HM{=}Q. \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} M{=}Q^{-T}D, \\ H{=}QD^{-1}Q^T, \end{array}\right.$ we get the matrices M, H and G, which satisfy the conditions (2.2). There are infinite combinations of D and G which satisfy conditions (4.1).

Choosing matrix D that satisfies condition (4.1), we get $M = Q^{-T}D$, and $H = QD^{-1}Q^T$ is positive definite.
The correction $v^{k+1} = v^k - M(v^k-\tilde{v}^k)$ can be achieved by solving $Q^T(v^{k+1}-v^k) = D(\tilde{v}^k-v^k)$.

The generalized PPA by choosing a special D. We can take a special pair of D and G in (4.1) by

$$D = G = \tfrac{1}{2}(Q^T+Q). \quad (4.3)$$

In this case, $M = \frac{1}{2}Q^{-T}(Q^T+Q)$. Because $D = G$, the contractive inequality (3.2) becomes

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \|v^k-\tilde{v}^k\|_D^2, \quad \forall v^* \in \mathcal{V}^*. \quad (4.4)$$

Moreover, since $D = M^THM$ (see (4.2)) and $M(v^k-\tilde{v}^k) = v^k-v^{k+1}$ (see (2.1b)), it follows that

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \|v^k-v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \quad (4.5)$$

The inequality (4.5) is just the main convergence result of the classical PPA, a favorable formula ! Because the each iteration consists of a prediction and a correction, we call the related method as a generalized PPA.

In practice, in the generalized PPA, we suggest to take $D = \alpha(Q^T+Q)$ and $\alpha \in [0.5, 1)$.