Splitting and contraction methods for convex optimization

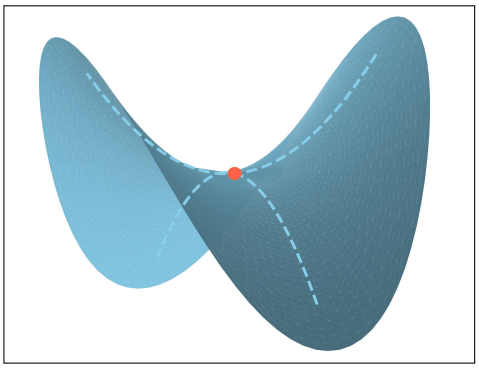
Algorithmic Framework in Variational Inequality and Generalized-PPA

Bingsheng He School of Mathematics, Nanjing University

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

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

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

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

The last inequalities indicate that the saddle point is equivalent to the solution of variational inequality:

$$\text{(VI)} \qquad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{1.4}$$

where

$$w = \left(\begin{array}{c} u \\ \lambda \end{array}\right), \qquad F(w) = \left(\begin{array}{c} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{array}\right) \qquad \text{and} \qquad \Omega = \mathcal{U} \times \Re^m. \tag{1.5}$$

Please notice that $(w - \tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0$. The problem (1.1) is translated to VI (1.4).

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

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) \,|\, Ax + By + Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\} \tag{1.6}$$

as an example. According to the analogous analysis, the saddle point of the Lagrangian function of the problem (1.6) is equivalent to the solution of VI (1.4), where $\Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m$ and

$$u = \left(\begin{array}{c} x \\ y \\ z \end{array}\right), \quad \theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z), \quad w = \left(\begin{array}{c} x \\ y \\ z \\ \lambda \end{array}\right), \quad F(w) = \left(\begin{array}{c} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{array}\right). \tag{1.7}$$

For the $F(w)$ in (1.7), we have $(w - \tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0$ for any w and $\tilde{w}$ in the space containing Ω.

We will focus on constructing algorithmic framework for solving the variational inequality (1.4). For the analysis in this note, we need the following preliminary theorem for convex optimization.

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

$$x^* \in \operatorname{argmin}\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{1.8}$$

if and only if

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{1.9}$$

2 Algorithmic framework for monotone variational inequalities

Algorithms in a unified framework

We take the variational inequality arising from the optimization problem (1.6) as the example. Each iteration of the proposed algorithm consists of a prediction and a correction.

Prediction for (1.6). Start from a given $v^k=(y^k,z^k,\lambda^k)$, find a predictor $\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{z}^k,\tilde{\lambda}^k)$ via

$$\left\{\begin{array}{l}\tilde{x}^k\in\arg\min\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}\beta\|Ax+By^k+Cz^k-b\|^2\mid x\in\mathcal{X}\},\\ \tilde{y}^k\in\arg\min\{\theta_2(y)-y^TB^T\tilde{\lambda}^k+\mu\beta\|B(y-y^k)\|^2|y\in\mathcal{Y}\},\\ \tilde{z}^k\in\arg\min\{\theta_3(z)-z^TC^T\tilde{\lambda}^k+\mu\beta\|C(z-z^k)\|^2|z\in\mathcal{Z}\},\\ \tilde{\lambda}^k=\lambda^k-\beta\big(A\tilde{x}^k+By^k+Cz^k-b\big). \qquad (\mu>2)\end{array}\right. \tag{2.1}$$

The above prediction implemented in order $\tilde{x}^k,\tilde{\lambda}^k,\tilde{y}^k,\tilde{z}^k$. Using $\tilde{\lambda}^k=\lambda^k-\beta\big(A\tilde{x}^k+By^k+Cz^k-b\big)$ and the preliminary theorem, the optimal conditions of the subproblems of (2.1) can be written as

$$\left\{\begin{array}{lll}\underline{\theta_1(x)-\theta_1(\tilde{x}^k)}+(x-\tilde{x}^k)^T\{\underbrace{-A^T\tilde{\lambda}^k}\qquad\qquad\}\geq 0, & & \forall x\in\mathcal{X},\\ \underline{\theta_2(y)-\theta_2(\tilde{y}^k)}+(y-\tilde{y}^k)^T\{\underbrace{-B^T\tilde{\lambda}^k}+\mu\beta B^TB(\tilde{y}^k-y^k)\}\geq 0, & & \forall y\in\mathcal{Y},\\ \underline{\theta_3(z)-\theta_3(\tilde{z}^k)}+(z-\tilde{z}^k)^T\{\underbrace{-C^T\tilde{\lambda}^k}+\mu\beta C^TC(\tilde{z}^k-z^k)\}\geq 0, & & \forall z\in\mathcal{Z},\\ (\lambda-\tilde{\lambda}^k)^T\{\underbrace{(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b)}-B(\tilde{y}^k-y^k)-C(\tilde{z}^k-z^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)\}\geq 0, & & \forall\lambda\in\Re^m.\end{array}\right. \tag{2.2}$$

Because the underline parts are $\theta(u)-\theta(\tilde{u}^k)$ and the wave parts are $F(\tilde{w}^k)$, the compact form of (2.2) is

$$\tilde{w}^k\in\Omega,\quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\quad \forall w\in\Omega, \tag{2.3a}$$

where

$$Q=\left(\begin{array}{ccc}\mu\beta B^TB & 0 & 0\\ 0 & \mu\beta C^TC & 0\\ -B & -C & \frac{1}{\beta}I\end{array}\right). \tag{2.3b}$$

Correction for (1.6). The new essential vector $v^{k+1}=(y^{k+1},z^{k+1},\lambda^{k+1})$ is given by

$$\left(\begin{array}{c}y^{k+1}\\ z^{k+1}\\ \lambda^{k+1}\end{array}\right)=\left(\begin{array}{c}y^{k}\\ z^{k}\\ \lambda^{k}\end{array}\right)-\left(\begin{array}{ccc}I & 0 & 0\\ 0 & I & 0\\ -\beta B & -\beta C & I\end{array}\right)\left(\begin{array}{c}y^k-\tilde{y}^k\\ z^k-\tilde{z}^k\\ \lambda^k-\tilde{\lambda}^k\end{array}\right). \tag{2.4}$$

The unified algorithmic framework for the general variational inequality (1.4)

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

$$\tilde{w}^k\in\Omega,\quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\quad \forall w\in\Omega, \tag{2.5}$$

where the matrix Q is not necessarily symmetric, but the kernel matrix of Q^T+Q is positive definite. [The kernel matrix of Q in (2.3) is obtained by setting B and C to identity matrix I].

[**Correction Step**]. The new essential iterator v^{k+1} is given by

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k). \tag{2.6}$$

Convergence Conditions of the Algorithmic Framework

[**Convergence Conditions**]. For the prediction matrix Q in (2.5) and the correction matrix M in (2.6), there is a positive definite matrix H, such that

$$HM=Q \qquad \text{and} \qquad G=Q^T+Q-M^THM\succ 0. \tag{2.7}$$

For the prediction matrix Q and the correction matrix M in the above motioned example,

$$Q=\left(\begin{array}{ccc}\mu\beta B^TB & 0 & 0\\ 0 & \mu\beta C^TC & 0\\ -B & -C & \frac{1}{\beta}I\end{array}\right),\quad M=\left(\begin{array}{ccc}I & 0 & 0\\ 0 & I & 0\\ -\beta B & -\beta C & I\end{array}\right).\ \text{Set}\ H=\left(\begin{array}{ccc}\mu\beta B^TB & 0 & 0\\ 0 & \mu\beta C^TC & 0\\ 0 & 0 & \frac{1}{\beta}I\end{array}\right),$$

when B and C are full column rank matrices and $\mu > 2$, the convergence conditions (2.7) are satisfied. Indeed, for such H, M and Q, we have

$$HM = Q \quad \text{and} \quad G = Q^T + Q - M^T Q = \begin{pmatrix} (\mu-1)\beta B^T B & -\beta B^T C & 0 \\ -\beta C^T B & (\mu-1)\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix} \succ 0.$$

3 Convergence proof of the methods in the algorithmic framework

Theorem A. For solving the VI (1.4), let $\{v^k\}$, $\{\tilde{w}^k\}$ be the sequences generated by (2.5)-(2.6). If the conditions (2.7) are satisfied, then we have

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

and

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

Proof. Treating the term $Q(v^k - \tilde{v}^k)$ in the RHS of (2.5) by using $Q = HM$ (see (2.7)) and the correction formula (2.6), we obtain $Q(v^k - \tilde{v}^k) = HM(v^k - \tilde{v}^k) = H(v^k - v^{k+1})$. Thus we get

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

Applying the identity

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

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

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

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

$$\begin{aligned} \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 &\overset{(2.6)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ &= (v^k - \tilde{v}^k)^T (Q^T + Q - M^T H M)(v^k - \tilde{v}^k) \overset{(2.7)}{=} \|v^k - \tilde{v}^k\|_G^2. \end{aligned} \tag{3.6}$$

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

This theorem is proved under weak conditions:

$$Q^T + Q \succeq 0, \quad H \succeq 0, \quad HM = Q, \quad G = Q^T + Q - M^T H M \succeq 0.$$

Theorem B. For solving the VI (1.4), let $\{v^k\}$, $\{\tilde{w}^k\}$ be the sequences generated by (2.5)-(2.6). If the conditions (2.7) are satisfied, then we have

$$\|v^{k+1} - v^{k+2}\|_H^2 \leq \|v^k - v^{k+1}\|_H^2 - \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2. \tag{3.7}$$

and consequently

$$\|v^{k+1} - v^{k+2}\|_H^2 \leq \|v^k - v^{k+1}\|_H^2. \tag{3.8}$$

Proof. According to (2.5), we have

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega \tag{3.9}$$

and

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (v - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}), \ \forall w \in \Omega. \tag{3.10}$$

Setting the any $w \in \Omega$ in (3.9) and (3.10) by $\tilde{w}^{k+1}$ and $\tilde{w}^k$, respectively, we get

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

and

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

Adding the above two inequalities and using $(\tilde{w}^k - \tilde{w}^{k+1})^T \big(F(\tilde{w}^k) - F(\tilde{w}^{k+1})\big) = 0$, it follows that

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq 0. \tag{3.11}$$

Adding $\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}$ to the both sides of the inequality (3.11), we get

$$(v^k - v^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \tfrac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}.$$

Using $Q = HM$ and $M(v^k - \tilde{v}^k) = (v^k - v^{k+1})$ (see (2.6)), it follows that

$$(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \geq \tfrac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \tag{3.12}$$

Finally, setting $a = (v^k - v^{k+1})$, $b = (v^{k+1} - v^{k+2})$ in the identity $\|a\|^2_H - \|b\|^2_H = 2a^T H(a-b) - \|a-b\|^2_H$ and using (3.12), we get

$$\begin{aligned}
&\|v^k - v^{k+1}\|^2_H - \|v^{k+1} - v^{k+2}\|^2_H \\
&= 2(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} - \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|^2_H \\
&\geq \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)} - \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|^2_H.
\end{aligned} \tag{3.13}$$

Using $(v^k - v^{k+1}) = M(v^k - \tilde{v}^k)$ and $Q^T + Q - M^T HM = G$, the RHS of (3.13) equals $\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_G$. It follows the assertions (3.7) and thus (3.8). The theorem is proved. □

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 (2.7). In fact, because $Q^T + Q \succ 0$, we can take

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

Afterwards, we let

$$HM = Q \qquad \text{and} \qquad M^T HM = D. \tag{4.2}$$

Because

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

we get the matrices M, H and G, which satisfy the conditions (2.7). Choosing matrix D that satisfies condition (4.1), we get $M = Q^{-T}D$, and $H = QD^{-1}Q^T$ is positive definite.

In this way, the correction $v^{k+1} = v^k - M(v^k - \tilde{v}^k)$ can be achieved by solving

$$Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k). \tag{4.3}$$

Taking a special pair of D and G in (4.1) by $D = G = \frac{1}{2}(Q^T + Q)$, the resulting prediction-correction algorithm is called Generalized Proximal Point Algorithm.

Since $D = G$, the contractive inequality (3.2) becomes

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

Moreover, since $D = M^T HM$ (see (4.2)) and $M(v^k - \tilde{v}^k) = v^k - v^{k+1}$ (see (2.6)), it follows that

$$\|v^k - \tilde{v}^k\|^2_D = \|v^k - \tilde{v}^k\|^2_{M^T HM} = \|M(v^k - \tilde{v}^k)\|^2_H = \|v^k - v^{k+1}\|^2_H.$$

Substituting it in (4.4), we get

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

The inequality (4.5) is just the main convergence result of the classical PPA, a favorable formula! Because the each iteration of the algorithm consists of a prediction and a correction, we call the related method as a generalized PPA. Besides (4.5), the sequence $\{v^k\}$ generated by the generalized PPA possesses another nice property (3.8).

Reference. Bingsheng He and Xiaoming Yuan, On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization, Commum. Optim. Theory 2023 (2023) 28, Available online at http://cot.mathres.org