Solution methods (gradient, extragradient, proximal, Newton) for quasi-variational inequalities

Nevena Mijajlović (University of Montenegro)

University of Montenegro, Department of Mathematics

Vienna, March 2018

Contents

1. Statement
   - Formulation
   - Generalized Nash game
   - Existance of solutions
2. Solution methods
   - General case
   - The moving set case

- E – n-dimensional euclidean vector space,
- $C \subseteq E$,
- $f : E \to \mathbb{R}$ - differentiable function on C

$$f(x) \to inf, x \in C \tag{1}$$

If $x_* \in C$ is a solution of (1) then

$$\langle f'(x_*), x - x_* \rangle \geq 0, \ \forall x \in C. \tag{2}$$

In the case of convexity of f and C, condition (2) is sufficient for optimality of x_* in (1).

Generalization of (2):

Variational inequality (VI)

find $x_* \in C$ such that $\langle F(x_*), x - x_* \rangle \geq 0, \ \forall x \in C$, where $F : E \to E$.

Most methods for solving minimization problem (1) have been adapted for solving VI.

In the case of changeable set C, i.e. set-valued mapping $C : E \to 2^E$ with closed and convex values $C(x) \subseteq E \; \forall x \in E$, problem (1) has a form

$$\text{find } x_* \in C(x_*) \text{ such that } f(x_*) \leq f(x), \forall x \in C(x_*).$$

The corresponding generalized VI is

Quasi variational inequality (QVI)

find $x_* \in C(x_*)$ such that $\langle F(x_*), x - x_* \rangle \geq 0, \; \forall x \in C(x_*)$.

- Let us mention, solving QVI requires that the corresponding VI be solved concurrently with the calculation of a fixed point of the set-valued mapping.

Many important and useful applications of QVI are well known:

- generalized Nash games,
- noncooperative multi-leader-follower game,
- applications to engineering,
- dynamic traffic assignment problem,
- spatial oligopolistic electricity model with Cournot generators and regulated transmission prices,
- applications to some economic and financial models...

It is by now a well-known fact that the NEP where each player solves a convex program can be formulated and solved as a finite-dimensional VI.
The generalized Nash game is a Nash game in which each player's strategy set depends on the other players' strategies.

- N players
- Each player ν controls variables $x^\nu \in \mathbb{R}^{n_\nu}$ (x^ν is a strategy of the player ν)
- $x = (x^\nu)_{\nu=1}^N \in \mathbb{R}^n$ ($n = \sum_{\nu=1}^N n_\nu$) - strategy vector of all the N players in the game
- $n_{-\nu} = n - n_\nu$, $x^{-\nu}$ - the vector formed of all players decision variables except the one of the player ν
- $C^\nu : \mathbb{R}^{n_{-\nu}} \to \mathbb{R}^{n_\nu}$, $\nu = 1, \dots, N$, strategy set of player ν $C^\nu(x^{-\nu}) \subseteq \mathbb{R}^{n_\nu}$, for all $x^{-\nu} \in \mathbb{R}^{n_{-\nu}}$. (represents the ability of players $j \neq \nu$ to affect the feasible strategies of player ν)

Convex assumption:

$g^\nu : \mathbb{R}^n \to \mathbb{R}^{m_\nu}$, $h^\nu : \mathbb{R}^{n_\nu} \to \mathbb{R}^{l_\nu}$ i $\Theta_\nu : \mathbb{R}^n \to \mathbb{R}$ $(m_\nu, l_\nu \in \mathbb{N})$.

- h_j^ν - continuously differentiable and convex on $\mathbb{R}_\nu^n$ for all $j = 1, \ldots, l_\nu$;
- $\forall x^{-\nu} \in \mathbb{R}^{n_{-\nu}}$, $\Theta_\nu(x^{-\nu}, \cdot)$, $g_j^\nu(x^{-\nu}, \cdot)$ are continuously differentiable and convex in the argument x^ν for each $j = 1, \ldots, m_\nu$.

$C^\nu(x^{-\nu}) = \{x^\nu \in \mathbb{R}^{n_\nu} : g^\nu(x) \leq 0, h^\nu(x_\nu) \leq 0\}$ is a closed convex subset of $\mathbb{R}^{n_\nu}$

The generalized Nash game is to find a tuple $x^* = (x^{*,\nu}) \in \mathbb{R}^n$, called a generalized Nash equilibrium (GNE), such that for each $\nu = 1, \ldots, N$, $x^{*,\nu}$ is an optimal solution of the convex optimization problem in the variable x^ν with $x^{-\nu}$ fixed at $x^{*,\nu}$:

$$\Theta_\nu(x^{*,-\nu}, x^\nu) \to min, \quad x^\nu \in C^\nu(x^{*,-\nu}).$$

Defining

$$C(x) = \prod_{\nu=1}^{N} C^{\nu}(x^{-\nu}) \text{ for } x = (x^{\nu})_{\nu=1}^{N} \in \mathbb{R}^n,$$

$$F(x) = (\nabla_{x^{\nu}} \Theta_{\nu}(x))_{\nu=1}^{N} \in \mathbb{R}^n,$$

Bensoussan, 1974, Harker 1989, Pang, Fukushima, 2005

x^* is a GNE if and only if x^* is a solution of QVI

VI and QVI as a fixed point problem

There are several general approaches to obtain existence results for VI and QVI. We will present some results based on preformulation VI (QVI) to fixed point problem.

Theorem

a) $x_* \in C$ *is a solution of VI iff*

$$x_* = \mathcal{P}_C(x_* - \alpha F(x_*)), \quad \alpha > 0.$$

b) $x_* \in C(x_*)$ *is a solution of QVI iff*

$$x_* = \mathcal{P}_{C(x_*)}(x_* - \alpha F(x_*)), \quad \alpha > 0.$$

If previous equalities are satisfied for some $\alpha > 0$ then they are satisfied for all $\alpha > 0$.

Theorem (Facchinei, Pang 2003)

If the map $F : E \to E$ is continuous and strongly monotone and $C \subseteq E$ nonempty closed convex set then VI has a unique solution.

Simple examples for $F(x) = x$.

1. If $C(x) = \{x\}, \forall x \in \mathbb{R}$, QVI has infinitely many solutions.
2. If $C(x) = [x-1, x], \forall x \geq 1$, then QVI has no solution.
3. If $C(x) = [\frac{x}{2}, \frac{x+2}{2}], \forall x \geq 0$, then QVI has a unique solution $x_* = 0$.
4. If $C(x) = \begin{cases} [1/2, 1], & \text{if } \; x \in [0, 1/2) \\ [0, 1]\,, & \text{if } \; x = 1/2 \\ [0, 1/2]\,, & \text{if } \; x \in (1/2, 1] \end{cases}$, then multifunction C has a unique fixed point $\frac{1}{2}$, but it is not a solution of QVI.

The theorem about existence of solutions show a notable difference between VI and QVI.

Theorem (Noor, Oettli, 1994)

If the map F is Lipschitz continuous and strongly monotone on E with constants L and $\mu > 0$, respectively, and C is a set-valued mapping with nonempty closed and convex values such that

$$\|\mathcal{P}_{C(x)}(z) - \mathcal{P}_{C(y)}(z)\| \leq \lambda \|x - y\|, \ \ \forall\, x, y, z \in E. \tag{3}$$

for $\lambda + \sqrt{1 - \mu^2/L^2} < 1$, then QVI problem has a unique solution.

Nesterov and Scrimali (2010) proved that $\lambda < \frac{\mu}{L}$ is sufficient condition for existence and uniqueness of solution of QVI.

Assumption (3) is a kind of strengthening of the contraction property for multifunction $C(x)$. An example of such mapping is given in the following lemma

Lemma (Nesterov, Scrimali 2010)

Let function $c(x) : E \to E$ be Lipschitz continuous with constant l and set C_0 be a closed convex set in E. Then

$$C(x) := c(x) + C_0$$

satisfies

$$\|\mathcal{P}_{C(x)}(z) - \mathcal{P}_{C(y)}(z)\| \leq l\|x - y\|, \quad \forall\, x, y, z \in E.$$

The moving set case is the most studied class of problems in the literature and essentially the only class of problems for which clear convergence conditions are available.

Projection gradient methods

QVI($C(\cdot), F$)

$$x_{k+1} = \mathcal{P}_{C(x_k)}[x_k - \alpha F(x_k)], \quad k \geq 0.$$

Convergence conditions:

- F - Lipschitz continuous and strongly monotone (with constants $L > 0$, $\mu > 0$)
- C is a set-valued mapping with nonempty closed and convex values such that (3) is satisfied with $\lambda < \frac{\mu^2}{L(L+\sqrt{L^2-\mu^2})}$
- $\alpha = \mu/L^2$

Then $x_k \to x_*$ and $\|x_{k+1} - x_*\| \leq \sqrt{(1 - \alpha(2\mu - \alpha L^2))}\|x_k - x_*\|$

Continuous projection gradient methods

QVI($C(\cdot), F$)

$$x'(t) + x(t) = \mathcal{P}_{C(x(t))}[x(t) - \alpha(t)F(x(t)], \ x(0) = x_0.$$

Convergence conditions:

- $\lambda < 1 - \sqrt{L^2 - \mu^2}/L$
- $\alpha(t) \in C[0, +\infty)$ satisfies: $0 < \alpha_0 \leq \alpha(t) \leq \alpha_1, \forall t \geq 0,$
  $\alpha_0 > (\mu - \sqrt{\mu^2 - L^2(2\lambda - \lambda^2)})/L^2,$
  $\alpha_1 < (\mu + \sqrt{\mu^2 - L^2(2\lambda - \lambda^2)})/L^2.$

Then $x(t) \to x_*$ and $\|x(t) - x_*\| \leq$
$\exp\left\{-\left(1 - \left(\lambda + \sqrt{1 - 2\alpha_1\mu + \alpha_0^2 L^2}\right)^2\right) t/2\right\} \|x_0 - x_*\|.$

To speed up the convergence, instead of previous method can be used more general variant of the gradient method:

$$x_{k+1} = x_k + a_k(\mathcal{P}_{C(x_k)}[x_k - \alpha F(x_k)] - x_k)$$
$$= (1 - a_k)x_k + a_k\mathcal{P}_{C(x_k)}[x_k - \alpha F(x_k)], \quad k \geq 0,$$

$\lambda < \frac{1}{2L}(L - \sqrt{L^2 - \mu^2})$ and parameters α and a_k satisfy

$$\left|\alpha - \frac{\mu}{L^2}\right| < \frac{\sqrt{\mu^2 - L^2\lambda(2-\lambda)}}{L^2}, \ \lambda(2-\lambda) < \frac{\mu^2}{L^2}, \ 0 \leq a_k \leq 1, \forall k \geq 0$$

and $\sum_{k=0}^{\infty} a_k = \infty$.
Then method converges to the unique solution of QVI with the following rate:

$$\|x_{k+1} - x_*\| \leq \prod_{i=0}^{k}\left[1 - \left(1 - \left(\lambda + \sqrt{1 - 2\mu\alpha + \alpha^2 L^2}\right)\right) a_i\right]\|x_0 - x_*\|,$$

- The moving set case is the most studied class of problems in the literature and essentially the only class of problems for which clear convergence conditions are available.
- In this class of problems, the feasible mapping $C(\cdot)$ is defined by a closed convex set $C_0 \subseteq H$ and a "trajectory" described by $c : H \to H$ according to: $C(x) = c(x) + C_0$.

Common assumptions:

1. operator F is strongly monotone with constant $\mu > 0$

$$\langle F(x) - F(y), x - y \rangle \geq \mu \|x - y\|^2, \ \ \forall x, y \in E,$$

and Lipschitz continuous with constant $L > 0$

$$\|F(x) - F(y)\| \leq L\|x - y\|, \ \ \forall x, y \in E.$$

2. $C_0 \subseteq E$ is a closed convex set in a space E, $c : E \to E$ is a continuous Lipschitz function with the constant $l < \mu / L$, and $C : E \to 2^E$ is a multifunction of the form $C(x) = c(x) + C_0$.

Second-order iterative method

$$\begin{aligned} x_{k+1} &= \mathcal{P}_{C(x_k)}\left[x_k - \alpha_k F(x_k) - \beta_k(x_{k-1} - x_k)\right] \\ &= c(x_k) + \mathcal{P}_{C_0}\left[x_k - c(x_k) - \alpha_k F(x_k) - \beta_k(x_{k-1} - x_k)\right], \end{aligned}$$

Parameters α_k and β_k satisfy the conditions

$$0 < \underline{\alpha} \leq \alpha_k \leq \bar{\alpha}, \ \beta_{k+1} \geq \beta_k, \ k \geq 0, \ \lim_{k\to\infty} \beta_k = \beta_\infty > 0$$

$$\underline{\alpha} \geq \frac{6l^2}{\mu}, \ \bar{\alpha} < \frac{2}{5}(L+\mu)^{-1}, \ 8\beta_\infty + 2\bar{\alpha}\mu \leq 1 + 6l^2.$$

Then the second-order iterative method converges to the unique solution of QVI with the following rate: $\|x_{k+1} - x_*\|^2 \leq \frac{\left(1-\underline{\alpha}\mu+6l^2\right)^k}{\bar{\alpha}\mu}\left(\left(\frac{1}{2} - \beta_0\right)\|x_1 - x_0\|^2 + \|x_1 - x_*\|^2 - \beta_0\|x_0 - x_*\|^2\right).$

Second-order continuous method

$$\beta(t)x''(t) + x'(t) + x(t) = \mathcal{P}_{C(x(t))}[x(t) - \alpha(t)F(x(t))]$$
$$= c(x(t)) + \mathcal{P}_{C_0}[x(t) - c(x(t)) - \alpha(t)F(x(t))].$$

$$l < \min\left\{\frac{\mu}{L}, \frac{2\mu}{L^2}\left(L + \mu - \sqrt{\mu(2L+\mu)}\right)\right\}$$

and for parameters $\alpha(\cdot) \in C^1[0, +\infty)$ and $\beta(t) = const > 0$ are valid conditions:

$$\alpha(t) > 0, \alpha'(t) \leq 0, t \geq 0, \alpha(0) < 2(L+\mu)^{-1}, \lim_{t\to\infty} \alpha(t) = \alpha_\infty > 0,$$

$$\sqrt{1 - 4A(0)\beta} \geq 4\beta - 1, 1 - 4\mu\alpha_\infty\beta(2-\mu\alpha_\infty) - 4l\beta(1+L\alpha_\infty) > 0,$$

where $b(t) = \frac{1}{2\beta}\left(1 - \sqrt{1 - 4A(t)\beta}\right)$ and
$A(t) = \alpha(t)\mu(2 - \alpha(t)\mu) - l - lL\alpha(t) > 0.$

Then the second-order continuous method converges to the unique solution of QVI with the following rate:

$$\begin{aligned}\|x(t)-x_*\|^2 &\leq \|x(0)-x_*\|^2 \exp\left(-\int_0^t f(s)ds\right)\\ &+ C_0\rho(t)\exp\left(-\int_0^t b(s)ds\right),\end{aligned}$$

where

$$f(t)=\frac{1}{2\beta}\left(1+\sqrt{1-4A(t)\beta}\right)>0,$$

$$C_0=(1-l)\beta h(0)\|x'(0)\|^2+h(0)\left(1-\beta b(0)\right)\|x(0)-x_*\|^2,$$

$$\rho(t)=h(t)H^{-1}(t)\int_0^t H(s)(\beta h(s))^{-1}ds.$$

Extragradient method

QVI$(C(x) = c(x) + C_0, F)$

$$\overline{x}_k = c(x_k) + \pi_{C_0}(x_k - c(x_k) - \alpha F(x_k)),$$

$$x_{k+1} = c(x_k) + \pi_{C_0}(x_k - c(x_k) - \alpha F(\overline{x}_k)), \quad k \geq 0.$$

$$0 < \alpha < \frac{1}{L}, \quad 0 < l < \frac{\mu}{L}\sqrt{\frac{\alpha(1-\alpha^2L^2)}{(\alpha+1)(1-\alpha^2L^2+\alpha\mu)}}$$

Then $x_k \to x_*$ and

$$\|x_{k+1} - x_*\|^2 \leq \left(1 + (\alpha+1)\frac{L^2l^2}{\mu} - \alpha\mu + \frac{(\alpha\mu)^2}{1-\alpha^2L^2+\alpha\mu}\right)^k \|x_0 - x_*\|^2.$$

Proximal operator $pr(x, \alpha)$ satisfies

$$\langle pr(x, \alpha) - x + \alpha F(pr(x, \alpha)), z - pr(x, \alpha) \rangle \geq 0, \ \forall z \in C(x).$$

If we use well known property of projection, we can get the following relationship between proximal operator and projection:

$$pr(x, \alpha) = \mathcal{P}_{C(x)}[x - \alpha F(pr(x, \alpha))], \ \forall x \in H.$$

Because of this relation, it is easy to conclude that x_* is solution of QVI if and only if

$$x_* = pr(x_*, \alpha), \ \forall \alpha > 0.$$

Proximal iterative method

$x_{k+1} \in C(x_k)$ such that $x_{k+1} = pr(x_k, \alpha)$, follow

$$x_{k+1} = \mathcal{P}_{C(x_k)}[x_k - \alpha F(x_{k+1})] = c(x_k) + \mathcal{P}_{C_0}[x_k - c(x_k) - \alpha F(x_{k+1})]$$

$$l < \frac{\sqrt{2}}{2}\frac{\mu}{L}, \quad \left|\alpha - \frac{\mu}{L^2}\right| < \frac{1}{L^2}\sqrt{\mu^2 - 2l^2L^2}.$$

Then, proximal iterative method converges to the unique solution of QVI and

$$\|x_{k+1} - x_*\| \le \left(\sqrt{\frac{1+2l^2}{1+2\alpha\mu - \alpha^2L^2}}\right)^k \|x_0 - x_*\|.$$

Proximal continuous method

$$x' = pr(x, \alpha) - x$$

$$x'(t) + x(t) = c(x(t)) + \mathcal{P}_{C_0}[x(t) - c(x(t)) - \alpha F(x'(t) + x(t))]$$

$$\alpha > \frac{1}{\mu}\left(\sqrt{\frac{\mu^2 - l^2L^2 + \mu^2 l^2}{\mu^2 - l^2L^2}} - 1\right)$$

Then, for each randomly chosen initial approximation $x(0) = x_0 \in C_0$ trajectory $x(t)$, defined by proximal continuous process, converges to a unique solution of QVI and its convergence rate is

$$\|x(t) - x_*\| \leq \exp\left\{-\left(\alpha\mu - \frac{l^2}{2+\alpha\mu} - \frac{\alpha l^2 L^2}{\mu}\right) t \Big/ (1+\alpha\mu)\right\} \|x_0 - x_*\|.$$

Newton's method

Newton method generates a sequence $\{x_k\}$, where x_0 is chosen in E and x_{k+1} is determined to be a solution of QVI obtained by linearizing F at the current iterate x_k, i.e., $x_{k+1} - c(x_{k+1}) \in C_0$ and

$$\langle F(x_k) + F'(x_k)(x_{k+1} - x_k), z - x_{k+1} \rangle \geq 0, \tag{4}$$

for all z such that $z - c(x_{k+1}) \in C_0$.

- The strongly monotonicity of F ensures that the linearized problem (4) always has a unique solution z.
- The linearized problem (4) is usually easier to solve than original problem.

- A lot of assumptions are required to be made in order to guarantee convergence of the method.
- However, Newton's method does have one very attractive feature - under certain assumptions one can prove local quadratic rate of convergence.
- This property essentially means that the number of accuracy digits is doubled at each iteration.
- This is in contrast to the gradient methods in which the convergence theorems are rather independent in the starting point, but only ”relatively” slow linear convergence is assured.

Theorem

Suppose that the following conditions are fulfilled:

1. *Operator* $F : E \to E$ *is strongly monotone, Lipschitz continuous (with constants* $\mu > 0$ *and* L*) and* $\|F'(x)\| \leq L, \ \forall x \in E,$
2. $C_0 \subseteq E$ *is closed, convex set in* E*, function* $c : E \to E$ *is Lipschitz continuous with constant* $l < \mu/L$ *and multifunction* $C : E \to 2^E$ *has a form* $C(x) := c(x) + C_0;$
3. *Initial approximation* $x_0 \in E$ *satisfy*

$$q = \frac{L(1+l)}{2(\mu - lL)}\|x_0 - x_*\| < 1,$$

where x_* *is a solution of QVI.*

Theorem

Then, sequence (x_k) *from (4) exists and converges to the unique solution* x_* *of QVI and the following estimate is valid*

$$\|x_k - x_*\| \leq \frac{2(\mu - lL)}{L(1+l)} q^{2^k}, \quad k = 0, 1, \ldots$$

Previous algorithm is an implicit type Newton's method, which is difficult to implement. It is possible to consider other variant of this method, for example:

- For given $x_0 \in E$, find the approximate solution by solving VI obtained by linearizing F at the current iterate x_k, i.e.,

$x_{k+1} \in C(x_k)$ and

$$\langle F(x_k) + F'(x_k)(x_{k+1} - x_k), z - x_{k+1} \rangle \geq 0, \quad \forall z \in C(x_k) \qquad (5)$$

Gradient method with changeable metric

- Any symmetric positive-definite matrix G defines new, so-called G-scalar product $\langle\cdot,\cdot\rangle_G$ in space E:

$$\langle u,v\rangle_G = \langle Gu,v\rangle, \quad u,v\in E.$$

- We will consider case when matrix $G = G(x)$ is changeable and it depends on current point x.
- Now, relation (3) may be replaced by

$$x_* = \mathcal{P}_{C(x_*)}^{G(x_*)}(x_* - \alpha G^{-1}(x_*)F(x_*)), \quad \forall\alpha > 0.$$

- Gradient method with changeable metric has a form:

$$x_{k+1} = \mathcal{P}_{C(x_k)}^{G(x_k)}(x_k - \alpha G^{-1}(x_k)F(x_k)). \tag{6}$$

- For $\alpha = 1$ and $G = F'$, method (6) turn into Newton's method (5).
- We have proved the convergence of method (6) (Mijajlović, Jaćimović, Comp. Math. and Math. Phys.) when

$$\langle G(x)\xi, \xi\rangle \geq m\|\xi\|^2, \ \ \|G(x)\| \leq L, \ \ \forall x, \xi \in E.$$

- Convergence of Newton method from (5) is consequence of this result.

- Antipin, A.S, Jaćimović, M., Mijajlović, N.: Extragradient method for solving quasivariational inequalities, Optimization, DOI: 10.1080/02331934.2017.1384477 (2018)
- N. Mijajlović, M. Jaćimović, *Continuous methods for solving quasivariational inequalities*, Comp. Math. and Math. Phys, (2018)
- Antipin, A.S., Mijajlović, N., Jaćimović, M.: A Second-Order Iterative Method for Solving Quasi-Variational Inequalities, Computational Mathematics and Mathematical Physics, Vol. 53, No. 3, pp. 258-264, (2013)
- Antipin, A. S., Mijajlović, N., Jaćimović, M.: A Second-Order Continuous Method for Solving Quasi-Variational Inequalities, Computational Mathematics and Mathematical Physics, Vol. 51, No. 11, pp. 1856-1863, (2011)

Thank you for your attention!