Last-Iterate Convergence of Optimistic Gradient Method for Monotone Variational Inequalities

Eduard Gorbunov[1,2,3], Adrien Taylor[4], Gauthier Gidel[2,5]

[1]MIPT, [2]Mila & UdeM, [3]MBZUAI,
[4]INRIA & ENS & CNRS & PSL,
[5]Canada CIFAR AI Chair

1. Preliminaries

Problem: variational inequality problem (VIP) – find $x^* \in \mathcal{X} \subseteq \mathbb{R}^d$ such that

$$\langle F(x^*), x - x^* \rangle \geq 0, \quad \forall x \in \mathcal{X}$$

Examples:

- Min-max problems $\quad \min_{u \in U} \max_{v \in V} f(u,v)$
- Minimization problems $\quad \min_{x \in \mathcal{X}} f(x)$

Assumptions: for all $x, y \in \mathcal{X}$ we assume

- Lipschitzness $\quad \|F(x) - F(y)\| \leq L\|x - y\|$
- Monotonicity $\quad \langle F(x) - F(y), x - y \rangle \geq 0$

Convergence metrics:

- Restricted gap function: for $R \sim \|x^0 - x^*\|$ it is defined as

$$\mathrm{Gap}_F(x^N) = \max_{y \in \mathcal{X}: \|y - x^*\| \leq R} \langle F(y), x^N - y \rangle$$

- Squared norm of the residual/operator (constrained/unconstrained cases):

$$\|x^N - x^{N-1}\|^2 \qquad \|F(x^N)\|^2$$

By default we report all results in terms of the squared norm of the residual

2. Extragradient and Past Extragradient

Extragradient method (EG) [Korpelevich, 1976]

$$\widetilde{x}^k = \mathrm{proj}\left[x^k - \gamma F(x^k)\right], \quad x^{k+1} = \mathrm{proj}\left[x^k - \gamma F(\widetilde{x}^k)\right]$$

- $\mathrm{proj}[x] = \arg\min_{y \in \mathcal{X}} \|y - x\|$ – projection operator

Past Extragradient/Optimistic Gradient method (PEG) [Popov, 1980]

$$\widetilde{x}^k = \mathrm{proj}\left[x^k - \gamma F(\widetilde{x}^{k-1})\right], \quad x^{k+1} = \mathrm{proj}\left[x^k - \gamma F(\widetilde{x}^k)\right]$$

In contrast to EG, PEG

- Requires only 1 operator call per iteration
- Is implementable as no-regret algorithm

Last-iterate convergence results for

EG

- $\mathcal{O}(1/N)$ bound in the unconstrained case [Gorbunov et al., 2022]
- $\mathcal{O}(1/N)$ bound in the constrained case [Cai et al., 2022]

PEG

- $\mathcal{O}(1/N)$ bound in the unconstrained case *if additionally the Jacobian $\nabla F(x)$ is Λ-Lipschitz* [Golowich et al., 2020]

3. Our Contributions

$\mathcal{O}(1/N)$ last-iterate convergence rate for PEG in terms of the squared norm of the residual for monotone and Lipschitz VIPs in constrained and unconstrained cases

✓ No additional assumptions are used ✓ Potential-based proof obtained via computer

4. Main Results

Unconstrained case

Key lemma: for any $k > 0$ the iterates of **PEG** satisfy

$$\Psi_{k+1} \leq \Psi_k - 3\left(\frac{2}{9} - L^2\gamma^2\right)\left\|F(\widetilde{x}^k) - F(\widetilde{x}^{k-1})\right\|^2$$

for $\Psi_k = \left\|F(x^k)\right\|^2 + 2\left\|F(x^k) - F(\widetilde{x}^{k-1})\right\|^2$

- In contrast, **EG** has much simpler potential: $\Psi_k = \|F(x^k)\|^2$
- As we illustrate in the paper, $\|F(x^k)\|^2$ is not a potential for **PEG**, i.e., $\|F(x^k)\|^2$ can grow for **PEG**

Using this lemma and standard analysis of **PEG**, we derive the following result

Theorem: for any $k > 0$ the iterates of **PEG** with $\gamma \leq 1/3L$ satisfy

$$\Phi_{k+1} \leq \Phi_k, \quad \Phi_k = \left\|x^k - x^*\right\|^2 + \frac{k+32}{3}\gamma^2\Psi_k$$

In particular, this implies

$$\left\|F(x^N)\right\|^2 = \mathcal{O}\left(\frac{R_0^2}{\gamma^2 N}\right) \quad \mathrm{Gap}_F(x^N) = \mathcal{O}\left(\frac{R_0}{\gamma\sqrt{N}}\right)$$

Constrained case

Theorem: for any $k > 1$ the iterates of **PEG** with $\gamma \leq 1/4L$ satisfy

$$\Phi_{k+1} \leq \Phi_k$$

$$\Phi_k = \left\|x^k - x^*\right\|^2 + \frac{1}{16}\left\|\widetilde{x}^{k-1} - \widetilde{x}^{k-2}\right\|^2 + \frac{3k+32}{24}\Psi_k$$

$$\Psi_k = \left\|x^k - x^{k-1}\right\|^2 + \left\|x^k - x^{k-1} - 2\gamma\left(F(x^k) - F(\widetilde{x}^{k-1})\right)\right\|^2$$

In particular, this implies

$$\left\|x^N - x^{N-1}\right\|^2 = \mathcal{O}\left(\frac{R_0^2}{N}\right) \quad \mathrm{Gap}_F(x^N) = \mathcal{O}\left(\frac{R_0}{\gamma\sqrt{N}}\right)$$

- Analysis does not follow straightforwardly from the result in the unconstrained case

5. Path to the Proof

Below we illustrate the non-triviality of the analysis of **PEG** even for unconstrained VIPs. To obtain the results below and the proofs most of the results in the paper we used Performance Estimation Problems technique [Taylor et al., 2017].

- The following problem gives the worst-case last-iterate guarantee for **PEG**

$$G_{\mathrm{PEG}}(\gamma, L, N) = \max_{\substack{F, d, x^* \\ \tilde{x}^0, \ldots, \tilde{x}^N \\ x^0, \ldots, x^N}} \frac{\|F(x^N)\|^2}{\|x^0 - x^*\|^2}$$

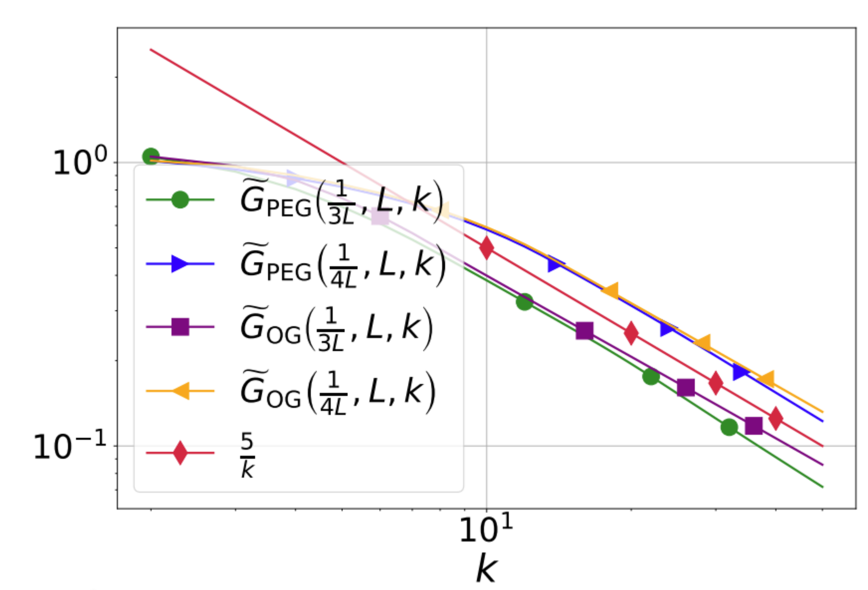
$$\text{s.t.} \quad \begin{aligned} &F \text{ is monotone and } L\text{-Lipschitz}, \\ &\tilde{x}^0 = x^0 \in \mathbb{R}^d, x^1 = x^0 - \gamma F(x^0) \\ &\widetilde{x}^k = x^k - \gamma F(\widetilde{x}^{k-1}), \text{ for } k = 1, \ldots, N, \\ &x^{k+1} = x^k - \gamma F(\widetilde{x}^k), \text{ for } k = 1, \ldots, N-1 \end{aligned}$$

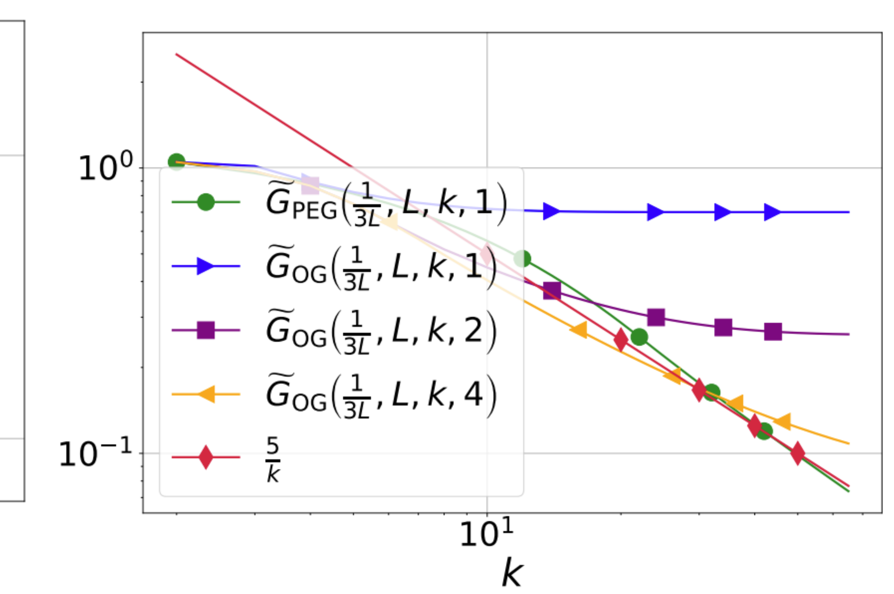
- Bad news: problem is infinitely-dimensional→ hard to solve
- Good news: there exist an SDP relaxation that is easy to solve numerically
- SDP finds pairs $\{x^*, 0\}$, $\{x^k, g^k\}_{k=0}^N$, $\{\tilde{x}^k, \tilde{g}^k\}_{k=0}^N$ such that $g^k \approx F(x^k)$, $\tilde{g}^k \approx F(\tilde{x}^k)$ and all Lipschitzness and monotonicity inequalities between these points hold
- This SDP has optimal value $\tilde{G}_{\mathrm{PEG}}(\gamma, L, N) \geq G_{\mathrm{PEG}}(\gamma, L, N)$. We numerically verified that $\tilde{G}_{\mathrm{PEG}}(\gamma, L, N) = \mathcal{O}(1/N)$

In the unconstrained case, one can rewrite **PEG** in the Optimistic Gradient (**OG**) form

$$\widetilde{x}^{k+1} = \widetilde{x}^k - 2\gamma F(\widetilde{x}^k) + \gamma F(\widetilde{x}^{k-1})$$

- Does not use sequence $\{x^k\}_{k\geq 0}$, looks simpler
- Similarly to **PEG**, one can formulate SDP for **OG** and verify $\tilde{G}_{\mathrm{OG}}(\gamma, L, N) = \mathcal{O}(1/N)$
- However, it is hard to find a simple proof for **OG:** consider SDPs $\tilde{G}_{\mathrm{PEG}}(\gamma, L, N, t)$ and $\tilde{G}_{\mathrm{OG}}(\gamma, L, N, t)$ obtained from $\tilde{G}_{\mathrm{PEG}}(\gamma, L, N)$ and $\tilde{G}_{\mathrm{OG}}(\gamma, L, N)$ via removing the constraints corresponding to the points from steps i, j such that $|i - j| > t$. While $\tilde{G}_{\mathrm{PEG}}(\gamma, L, N, t) = \mathcal{O}(1/N)$, $\tilde{G}_{\mathrm{OG}}(\gamma, L, N, t)$ does not even for $t = 4$.





References

G. M. Korpelevich. The extragradient method for finding saddle points and other problems. Matecon,12:747–756, 1976

L. D. Popov. A modification of the Arrow-Hurwicz method for search of saddle points. Mathematical notes of the Academy of Sciences of the USSR, 28(5):845–848, 1980

E. Gorbunov, N. Loizou, and G. Gidel. Extragradient method: O(1/K) last-iterate convergence for monotone variational inequalities and connections with cocoercivity. In International Conference on Artificial Intelligence and Statistics, pages 366–402. PMLR, 2022

Y. Cai, A. Oikonomou, and W. Zheng. Tight last-iterate convergence of the extragradient method for constrained monotone variational inequalities. arXiv preprint arXiv:2204.09228 version 1, 2022

N. Golowich, S. Pattathil, and C. Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. arXiv preprint arXiv:2010.13724, 2020

A. B. Taylor, J. M. Hendrickx, and F. Glineur. Exact worst-case performance of first-order methods for composite convex optimization. SIAM Journal on Optimization, 27(3):1283–1313, Jan 2017. ISSN 1095-7189.