Single-Call Stochastic Extragradient Methods for Structured Non-monotone Variational Inequalities: Improved Analysis under Weaker Conditions

Sayantan Choudhury [1] Eduard Gorbunov [2] Nicolas Loizou [1]

[1] Johns Hopkins University [2] Mohamed bin Zayed University of Artificial Intelligence

1. The Unconstrained Variational Inequality Problem

Find x^* such that:

$$F(x^*) = \frac{1}{n}\sum_{i=1}^{n} F_i(x^*) = 0 \qquad \text{(VIP)}$$

- $F, F_i : \mathbb{R}^d \to \mathbb{R}^d \quad \forall i \in [n]$ are operators.

Special Cases of VIP:

- For minimization problem $\min_x f(x)$, we have $F(x) = \nabla f(x)$.
- For min-max optimization problem,

$$\min_{x_1\in\mathbb{R}^{d_1}} \max_{x_2\in\mathbb{R}^{d_2}} \frac{1}{n}\sum_{i=1}^{n} g_i(x_1, x_2)$$

we have $x = (x_1; x_2)$ and

$$F_i(x) = (\nabla_{x_1} g_i(x_1, x_2); -\nabla_{x_2} g_i(x_1, x_2)).$$

- These min-max problems are important for their applications in Generative Adversarial Networks [1], Reinforcement Learning [2] and Robust Learning [3] among others.
- Classes of non-monotone VIP considered in our work:

Structured Non-monotone VIP:

- μ-Quasi Strongly Monotone Problem ($\mu > 0$) [4]
$$\langle F(x), x - x^* \rangle \geq \mu \|x - x^*\|^2$$
- Weak Minty Variational Inequality Problem ($\rho > 0$) [5]
$$\langle F(x), x - x^* \rangle \geq -\rho \|F(x)\|^2$$

2. Main Contributions:

- Convergence guarantees of Stochastic Past Extragradient Method (SPEG) without bounded variance assumption. We use instead the Expected Residual (ER) condition and explain its benefits. We show that ER holds for a large class of operators, e.g., whenever F_i are Lipschitz.
- Unified analysis for various sampling strategies, including single-element, minibatch, and importance sampling.
- We can recover the best-known results for deterministic settings from our analysis. This highlights the tightness of our analysis.
- Convergence guarantees with constant (linear convergence to a neighbourhood) and switching (exact convergence at a sublinear rate) step-size choices for solving quasi-strongly monotone VIP.
- Sublinear convergence guarantees for solving weak minty VIP with $\rho < \frac{1}{2L}$. This improves the restriction on ρ for stochastic setting.

Figure: Scan to read the camera-ready version.

3. Algorithms for solving VIP

- Stochastic Past Extragradient (SPEG)[6]:

$$\begin{aligned} \hat{x}_k &= x_k - \gamma_k g(\hat{x}_{k-1}) \\ x_{k+1} &= x_k - \omega_k g(\hat{x}_k) \end{aligned} \qquad \text{(SPEG)}$$

Here, $g(x)$ is an unbiased estimator of $F(x)$. SPEG requires only one oracle call per iteration in contrast to two oracle calls of Stochastic Extragradient (SEG)[7]. This work focuses on convergence guarantees of SPEG.

4. Assumption on Estimator

- In this work, we assume,

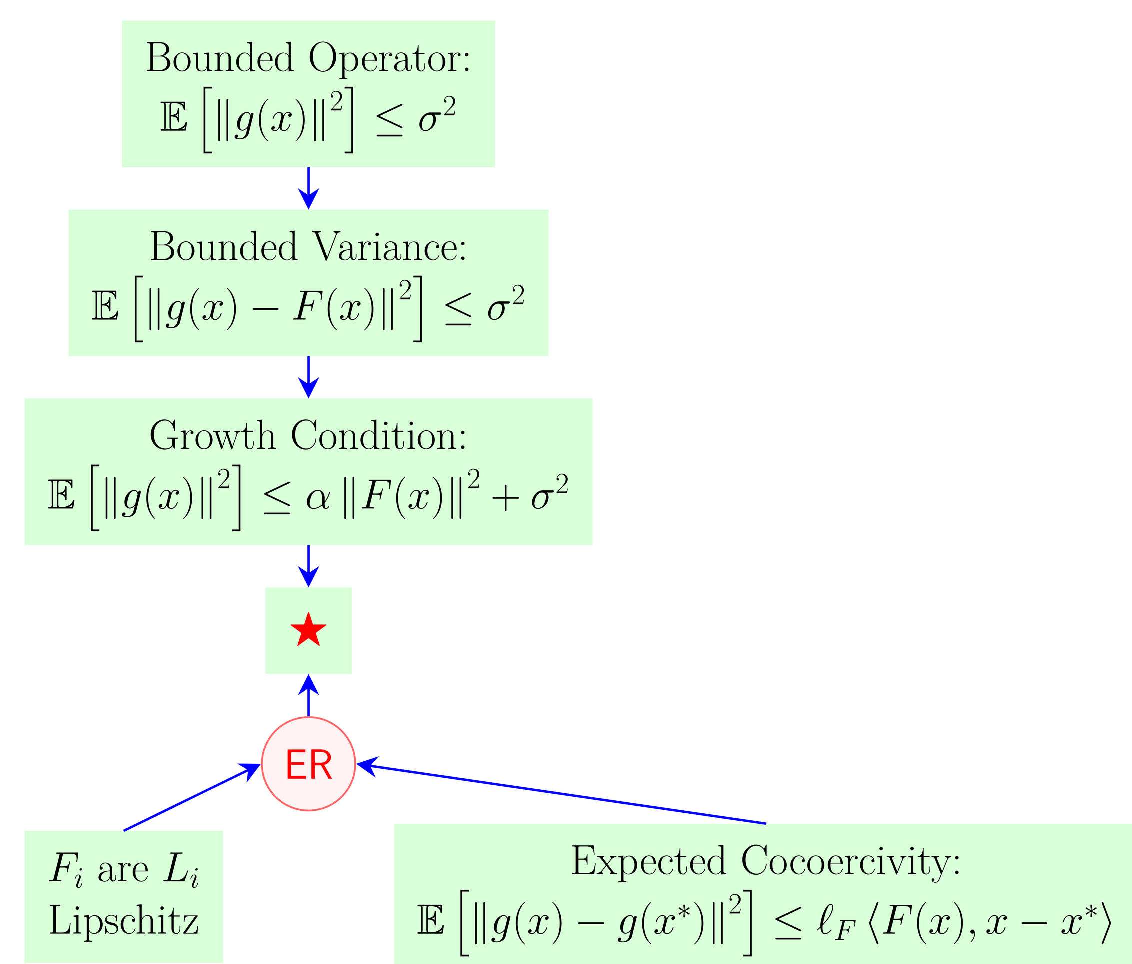
Expected Residual Condition:

$$\mathbb{E}\left[\|(g(x) - g(x^*)) - (F(x) - F(x^*))\|^2\right] \leq \frac{\delta}{2}\|x - x^*\|^2 \qquad \text{(ER)}$$

- For unbiased estimator $g(x)$ satisfying ER, we have

$$\mathbb{E}\left[\|g(x)\|^2\right] \leq \delta\|x - x^*\|^2 + \|F(x)\|^2 + 2\sigma_*^2 \qquad (\star)$$

where $\sigma_*^2 = \mathbb{E}\left[\|g(x^*)\|^2\right]$.



- ER allows us to have the analysis of SPEG under arbitrary sampling paradigm.
- Let F_i are L_i lipschitz operators, then ER condition holds and we can find the closed-form expressions of δ and σ_*^2 for various sampling stategies.

Closed-form expressions:

◇ τ- minibatch sampling:

$$\delta = \frac{2}{n\tau}\frac{n-\tau}{n-1}\sum_{i=1}^{n} L_i^2 \quad \text{and} \quad \sigma_*^2 = \frac{1}{n\tau}\frac{n-\tau}{n-1}\sum_{i=1}^{n}\|F_i(x^*)\|^2$$

◇ Single-element sampling:

$$\delta = \frac{2}{n^2}\sum_{i=1}^{n}\frac{L_i^2}{p_i} \quad \text{and} \quad \sigma_*^2 = \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{p_i}\|F_i(x^*)\|^2$$

where p_i is probability of selecting i th element from $[n]$. For *uniform* and *importance* sampling, we have $p_i = \frac{1}{n}$ and $p_i = \frac{L_i}{\sum_{j=1}^{n} L_j}$, respectively in the above equation.

5. Results for Quasi Strongly Monotone VIP

Constant Stepsize:

Theorem

Let F be L-Lipschitz, μ-quasi strongly monotone, and let ER hold. Choose step-sizes $\gamma_k = \omega_k = \omega$ such that

$$0 < \omega \leq \min\left\{\frac{\mu}{18\delta}, \frac{1}{4L}\right\}$$

for all k. Then the iterates produced by SPEG satisfy

$$R_k^2 \leq \left(1 - \frac{\omega\mu}{2}\right)^k R_0^2 + \frac{24\omega\sigma_*^2}{\mu},$$

where $R_k^2 := \mathbb{E}\left[\|x_k - x^*\|^2 + \|x_k - \hat{x}_{k-1}\|^2\right]$.

- For deterministic setting, $\delta = 0, \sigma_*^2 = 0$ and SPEG converges to the exact solution at a linear rate.

Switching Stepsize:

Theorem

Let F be L-Lipschitz, μ-quasi strongly monotone, and Assumption ER hold. Let

$$\gamma_k = \omega_k := \begin{cases} \bar{\omega}, & \text{if } k \leq k^*, \\ \frac{2k+1}{(k+1)^2}\frac{2}{\mu}, & \text{if } k > k^*, \end{cases}$$

where $\bar{\omega} := \min\{1/(4L), \mu/(18\delta)\}$ and $k^* = \lceil 4/(\mu\bar{\omega}) \rceil$. Then for all $K \geq k^*$ the iterates produced by SPEG with the above step-sizes satisfy

$$R_K^2 \leq \left(\frac{k^*}{K}\right)^2 \frac{R_0^2}{\exp(2)} + \frac{192\sigma_*^2}{\mu^2 K},$$

where $R_K^2 := \mathbb{E}\left[\|x_K - x^*\|^2 + \|x_K - \hat{x}_{K-1}\|^2\right]$.

- For the first k^* iterations, it uses constant step size to reach a neighborhood of the solution, and then the method switches to the decreasing $\mathcal{O}(1/k)$ step-size to converge to the exact solution.

6. Results for Weak Minty VIP

Theorem

Let F be L-Lipschitz and satisfy Weak Minty condition with parameter $\rho < 1/(2L)$. Let Assumption ER hold. Assume that $\gamma_k = \gamma$, $\omega_k = \omega$ such that

$$\max\left\{2\rho, \frac{1}{2L}\right\} < \gamma < \frac{1}{L}, \text{ and } 0 < \omega < \min\left\{\gamma - 2\rho, \frac{1}{4L} - \frac{\gamma}{4}\right\}.$$

Then, for all $K \geq 2$ the iterates produced by mini-batched SPEG with batch-size $\tau \geq \theta(\omega, \gamma, K)$ satisfy

$$\min_{0\leq k\leq K-1} \mathbb{E}\left[\|F(\hat{x}_k)\|^2\right] \leq \frac{C\|x_0 - x^*\|^2}{K-1},$$

where $C = \frac{48}{\omega\gamma(1 - L(\gamma + 4\omega))}$.

- We recover the best-known results for SPEG in deterministic setting [8, 9].
- We improve the restriction on ρ. Previous work by [8] assumes $\rho < \frac{3}{8L}$ with bounded variance.

7. Comparison with Prior Work

setup	method	no bounded variance?	single-call?
Quasi strongly monotone	SEG[10]	✓	✗
	SPEG[11]	✗	✓
	SPEG (This work)	✓	✓
Weak minty	SEG+[5]	✗	✗
	OGDA+ [8]	✗	✓
	SPEG (This work)	✓	✓

8. Numerical Experiments

- We consider a quadratic strongly convex strongly concave problem of the form $\min_x \max_y \frac{1}{n}\sum_{i=1}^{n} f_i(x, y)$ where

$$f_i(x, y) := \frac{1}{2}x^\intercal A_i x + x^\intercal B_i y - \frac{1}{2}y^\intercal C_i y + a_i^\intercal x - c_i^\intercal y. \qquad (■)$$

- Here, A_i, B_i, and C_i are generated such that the quadratic game is strongly monotone and smooth. The vectors a_i and c_i are generated from $\mathcal{N}_d(0, I_d)$.
- On y-axis, we plot relative error i.e. $\frac{\|x_k - x^*\|^2}{\|x_0 - x^*\|^2}$.

Constant vs Switching Stepsize:

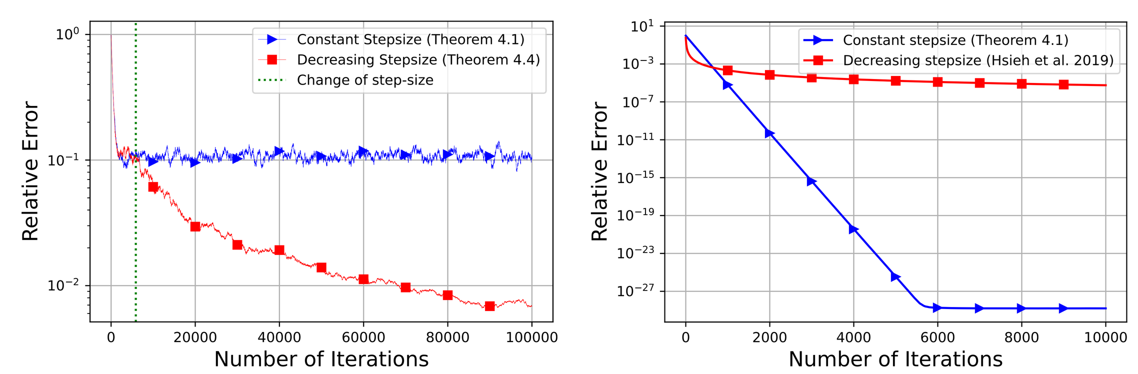
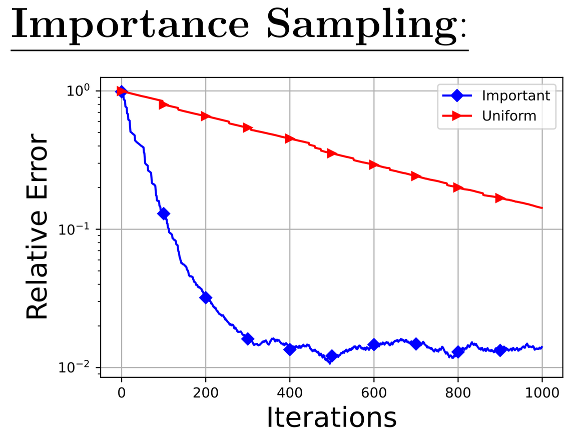


Figure: Comparison of SPEG using our proposed step-size against decreasing step-size of [11] for solving (■). In the left plot, we use the switching step size, while in the right plot, we implement SPEG with constant step size for the interpolated model ($\sigma_*^2 = 0$).

Importance Sampling:



Weak Minty VIP:

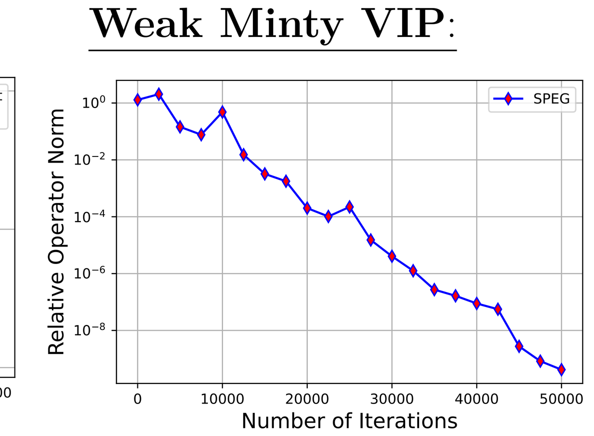


Figure: In the left plot, we demonstrate the advantage of using importance sampling over Uniform sampling for SPEG. In the second plot, we implement SPEG with our proposed step-sizes for solving a Weak Minty VIP of the form $\min_{x\in\mathbb{R}} \max_{y\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n} \xi_i xy + \frac{\zeta_i}{2}(x^2 - y^2)$.

- Code to reproduce our result: https://github.com/isayantan/Single-Call-Stochastic-Extragradient-Methods.

References

[1] G. Gidel, H. Berard, G. Vignoud, P. Vincent, and S. Lacoste-Julien. A variational inequality perspective on generative adversarial networks. ICLR, 2019.

[2]] S. Sokota, R. D'Orazio, J. Z. Kolter, N. Loizou, M. Lanctot, I. Mitliagkas, N. Brown, and C. Kroer. A unified approach to reinforcement learning, quantal response equilibria, and two-player zero-sum games. ICLR, 2023.

[3] Y. Yu, T. Lin, E. V. Mazumdar, and M. Jordan. Fast distributionally robust learning with variance-reduced min-max optimization. In AISTATS, 2022.

[4] N. Loizou, H. Berard, G. Gidel, I. Mitliagkas, and S. Lacoste-Julien. Stochastic gradient descent-ascent and consensus optimization for smooth games: Convergence analysis under expected co-coercivity. NeurIPS, 2021.

[5] J. Diakonikolas, C. Daskalakis, and M. Jordan. Efficient methods for structured nonconvexnonconcave min-max optimization. AISTATS, 2021.

[6] 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, 1980.

[7] G. M. Korpelevich. The extragradient method for finding saddle points and other problems. Matecon, 1976.

[8] A. Böhm. Solving nonconvex-nonconcave min-max problems exhibiting weak minty solutions. TMLR 2022.

[9] E. Gorbunov, A. Taylor, S. Horváth, and G. Gidel. Convergence of proximal point and extragradient-based methods beyond monotonicity: the case of negative comonotonicity. ICML, 2023.

[10] E. Gorbunov, H. Berard, G. Gidel, and N. Loizou. Stochastic extragradient: General analysis and improved rates. AISTATS, 2022.

[11] Y.-G. Hsieh, F. Iutzeler, J. Malick, and P. Mertikopoulos. On the convergence of single-call stochastic extra-gradient methods. NeurIPS, 2019.