Revisiting Stochastic Extragradient

Konstantin Mishchenko, Dmitry Kovalev, Egor Shulgin, Peter Richtárik, Yura Malitsky

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

EPFL

MIPT
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

Dmitry Kovalev

Egor Shulgin

Yura Malitsky

Peter Richtárik

Variational inequality

$$\min_{x \in \mathbb{R}^d} f(x)$$

Variational inequality

$$\min_{x \in \mathbb{R}^d} f(x)$$

Find x^*: $F(x) = 0$, where $F(x) = \nabla f(x)$

Variational inequality

$$\min_{x \in \mathbb{R}^d} f(x)$$

Find x^*: $F(x) = 0$, where $F(x) = \nabla f(x)$

$$\langle F(x) - F(y), x - y \rangle \geq 0, \quad \text{for all } x, y \in \mathbb{R}^d$$

Variational inequality

$$\min_{x \in \mathbb{R}^d} f(x) + g(x)$$

Variational inequality

$$\min_{x \in \mathbb{R}^d} f(x) + g(x)$$

$$g(x) - g(x^*) + \langle \nabla f(x), x - x^* \rangle \geq 0, \text{ for all } x \in \mathbb{R}^d$$

Variational inequality

$$\min_{x \in \mathbb{R}^d} f(x) + g(x)$$

$$g(x) - g(x^*) + \langle \nabla f(x), x - x^* \rangle \geq 0, \text{ for all } x \in \mathbb{R}^d$$

Find x^*

$$g(x) - g\left(x^*\right) + \left\langle F\left(x^*\right), x - x^* \right\rangle \geq 0, \text{ for all } x \in \mathbb{R}^d$$

Minmax problem

$$\min_{\alpha \in \mathcal{A}} \max_{\beta \in \mathcal{B}} \Phi(\alpha, \beta)$$

Minmax problem

$$\min_{\alpha \in \mathcal{A}} \max_{\beta \in \mathcal{B}} \Phi(\alpha, \beta)$$

$$\begin{cases} \alpha^{k+1} = \alpha^k - \gamma_\alpha \nabla_\alpha \Phi(\alpha^k, \beta^k), \\ \beta^{k+1} = \beta^k + \gamma_\beta \nabla_\beta \Phi(\alpha^k, \beta^k) \end{cases}$$

Minmax problem

$$\min_{\alpha \in \mathcal{A}} \max_{\beta \in \mathcal{B}} \Phi(\alpha, \beta)$$

$$\begin{cases} \alpha^{k+1} = \alpha^k - \gamma_\alpha \nabla_\alpha \Phi(\alpha^k, \beta^k), \\ \beta^{k+1} = \beta^k + \gamma_\beta \nabla_\beta \Phi(\alpha^k, \beta^k) \end{cases}$$

$$\min_{\alpha \in \mathbb{R}^m} \max_{\beta \in \mathbb{R}^n} \alpha^\top C \beta, \quad C \in \mathbb{R}^{m \times n}$$

Minmax problem

$$\min_{\alpha \in \mathcal{A}} \max_{\beta \in \mathcal{B}} \Phi(\alpha, \beta)$$

$$\begin{cases} \alpha^{k+1} = \alpha^k - \gamma_\alpha \nabla_\alpha \Phi(\alpha^k, \beta^k) \\ \beta^{k+1} = \beta^k + \gamma_\beta \nabla_\beta \Phi(\alpha^k, \beta^k) \end{cases}$$

DIVERGES

$$\min_{\alpha \in \mathbb{R}^m} \max_{\beta \in \mathbb{R}^n} \alpha^\top C \beta, \quad C \in \mathbb{R}^{m \times n}$$

Extagradient/mirror-prox

$$\text{solve} \quad F(x) = 0$$

$$\|F(x) - F(y)\| \leq L\|x - y\|$$

Extagradient/mirror-prox

$$\text{solve} \quad F(x) = 0$$

$$\|F(x) - F(y)\| \leq L\|x - y\|$$

$$\begin{aligned} y^k &= x^k - \eta F(x^k) \\ x^{k+1} &= x^k - \eta F(y^k). \end{aligned}$$

Extagradient/mirror-prox

$$\text{solve} \quad F(x) = 0$$

$$\|F(x) - F(y)\| \leq L\|x - y\|$$

$$y^k = x^k - \eta F(x^k)$$

$$x^{k+1} = x^k - \eta F(y^k).$$

$$x^{k+1} \approx x^k - \eta F(x^{k+1})$$

Extagradient/mirror-prox

$$\text{solve} \quad F(x) = 0$$

$$\|F(x) - F(y)\| \leq L\|x - y\|$$

$$\begin{aligned} y^k &= x^k - \eta F(x^k) \\ x^{k+1} &= x^k - \eta F(y^k). \end{aligned}$$

Works pretty well! ✓

Stochastic variational inequality

$$\text{Find } x^* : \ \mathbb{E}_\xi[F(x;\xi)] = 0$$

Stochastic variational inequality

$$\text{Find } x^* : \ \mathbb{E}_\xi[F(x;\xi)] = 0$$

$$\min_\alpha \max_\beta \mathbb{E}_\xi[\Phi(\alpha,\beta;\xi)]$$

Stochastic variational inequality

$$\text{Find } x^* : \ \mathbb{E}_\xi[F(x;\xi)] = 0$$

Existing:

$$y^k = x^k - \eta F(x^k;\xi^k)$$
$$x^{k+1} = x^k - \eta F(y^k;\psi^k).$$

Stochastic variational inequality

$$\text{Find } x^* : \; \mathbb{E}_\xi[F(x;\xi)] = 0$$

Existing:

$$y^k = x^k - \eta F(x^k;\xi^k)$$
$$x^{k+1} = x^k - \eta F(y^k;\psi^k).$$

DIVERGES

Stochastic variational inequality

$$\text{Find } x^* : \; \mathbb{E}_\xi[F(x;\xi)] = 0$$

Existing:

$$y^k = x^k - \eta F(x^k;\xi^k)$$
$$x^{k+1} = x^k - \eta F(y^k;\psi^k).$$

Proposed:

$$y^k = x^k - \eta F(x^k;\xi^k)$$
$$x^{k+1} = x^k - \eta F(y^k;\xi^k).$$

Works pretty well! ✓

Stochastic variational inequality

$$\text{Find } x^* : \ \mathbb{E}_\xi[F(x;\xi)] = 0$$

Proposed:

$$y^k = x^k - \eta F(x^k; \xi^k)$$
$$x^{k+1} = x^k - \eta F(y^k; \xi^k).$$

$$x^{k+1} \approx x^k - \eta F(x^{k+1}; \xi^k)$$

Convergence

Let $F(\cdot;\xi)$ be almost surely monotone and L-Lipschitz, its variance at the optimum x^* be finite,

$$\mathbb{E}\left\|F\left(x^*;\xi\right)-F\left(x^*\right)\right\|^2 \leq \sigma^2$$

and the problem be strongly convex. Then for any $\eta \leq \frac{1}{2L}$

$$\mathbb{E}\left\|x^t-x^*\right\|^2 \leq (1-2\eta\mu/3)^t\left\|x^0-x^*\right\|^2+3\eta\sigma^2/\mu.$$

Experiments