Decentralized Local Stochastic Extra-Gradient for Variational Inequalities






Aleksandr Beznosikov
Innopolis, MIPT, HSE and Yandex

Pavel Dvurechensky
WIAS

Anastasia Koloskova
EPFL

Valentin Samokhin
IITP

Sebastian U. Stich
CISPA

Alexander Gasnikov
MIPT, HSE and IITP

NeurIPS 2022

Variational Inequality Problem

$$\text{Find } \; z^* \in \mathbb{R}^d \; \text{ such that } \; \langle F(z), z - z^* \rangle \geq 0, \;\; \forall z \in \mathbb{R}^d$$

Variational Inequality Problem

Find $z^* \in \mathbb{R}^d$ such that $\langle F(z), z - z^* \rangle \geq 0, \;\; \forall z \in \mathbb{R}^d$

- $\min_{z \in \mathbb{R}^d} f(z) \longrightarrow F(z) := \nabla f(z)$

Variational Inequality Problem

Find $z^* \in \mathbb{R}^d$ such that $\langle F(z), z - z^* \rangle \geq 0, \ \forall z \in \mathbb{R}^d$

- $\min_{z \in \mathbb{R}^d} f(z) \longrightarrow F(z) := \nabla f(z)$
- $\min_{x \in \mathbb{R}^{d_x}} \max_{y \in \mathbb{R}^{d_y}} g(x, y) \longrightarrow F(z) := [\nabla_x g(x, y), -\nabla_y g(x, y)]$

Variational Inequality Problem

Find $z^* \in \mathbb{R}^d$ such that $\langle F(z), z - z^* \rangle \geq 0, \ \forall z \in \mathbb{R}^d$

- $\min_{z \in \mathbb{R}^d} f(z) \longrightarrow F(z) := \nabla f(z)$
- $\min_{x \in \mathbb{R}^{d_x}} \max_{y \in \mathbb{R}^{d_y}} g(x, y) \longrightarrow F(z) := [\nabla_x g(x, y), -\nabla_y g(x, y)]$
- Find $z^* \in \mathbb{R}^d$ such that $T(z^*) = z^*$

$$\Downarrow$$

$$F(z) := z - T(z)$$

Distributed Stochastic Setting

$$F(z) := \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}_{\xi_m \sim \mathcal{D}_m} F_m(z, \xi_m)$$

Distributed Stochastic Setting

$$F(z) := \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}_{\xi_m \sim \mathcal{D}_m} F_m(z, \xi_m)$$

(unknown) distribution — on local devices

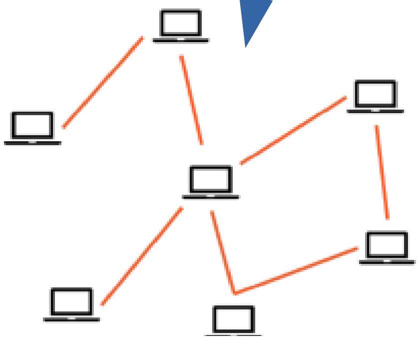
Distributed Stochastic Setting

$$F(z) := \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}_{\xi_m \sim \mathcal{D}_m} F_m(z, \xi_m)$$

(unknown) distribution

on local devices

- Arbitrary network



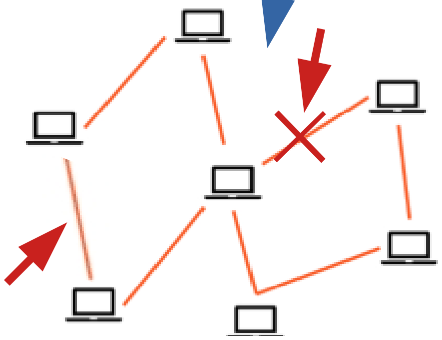
Distributed Stochastic Setting

$$F(z) := \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}_{\xi_m \sim \mathcal{D}_m} F_m(z, \xi_m)$$

(unknown) distribution

on local devices

- Arbitrary network
- Time-varying network



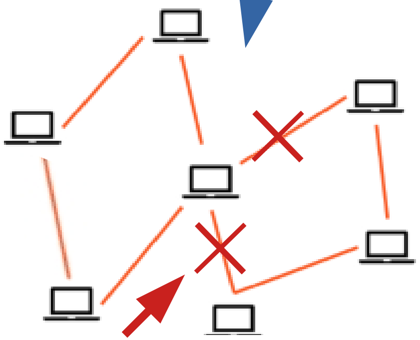
Distributed Stochastic Setting

$$F(z) := \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}_{\xi_m \sim \mathcal{D}_m} F_m(z, \xi_m)$$

(unknown) distribution

on local devices

- Arbitrary network
- Time-varying network
- Disconnected network



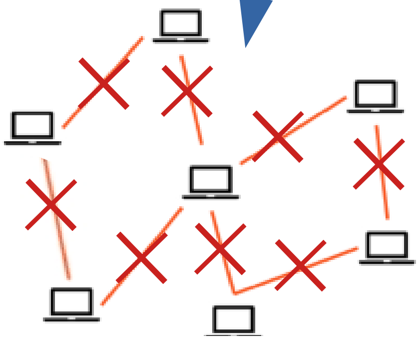
Distributed Stochastic Setting

$$F(z) := \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}_{\xi_m \sim \mathcal{D}_m} F_m(z, \xi_m)$$

(unknown) distribution

on local devices

- Arbitrary network
- Time-varying network
- Disconnected network
- Local updates



Algorithms

Algorithm 1 Extra Step Time-Varying Gossip Method

parameters: stepsize $\gamma > 0$, $\{\mathcal{W}^k\}_{k \geq 0}$ – rules or distributions for mixing matrix in iteration k.
initialize: $z^0 \in \mathcal{Z}, \forall m : z_m^0 = z^0$

for $k = 0, 1, 2, \ldots$ **do**
    Sample matrix W^k from $\mathcal{W}^k$
    for each node m **do**
        Generate independently $\xi_m^k \sim \mathcal{D}_k$, $\xi_m^{k+1/3} \sim \mathcal{D}_k$
        $z_m^{k+1/3} = z_m^k - \gamma F_m(z_m^k, \xi_m^k)$
        $z_m^{k+2/3} = z_m^k - \gamma F_m(z_m^{k+1/3}, \xi_m^{k+1/3})$
        $z_m^{k+1} = \sum_{i \in \mathcal{N}_m^k} w_{m,i}^k z_i^{k+2/3}$
    end for
end for

Algorithms

Algorithm 1 Extra Step Time-Varying Gossip Method

parameters: stepsize $\gamma > 0$, $\{\mathcal{W}^k\}_{k\geq 0}$ – rules or distributions for mixing matrix in iteration k.
initialize: $z^0 \in \mathcal{Z}, \forall m : z_m^0 = z^0$

for $k = 0, 1, 2, \dots$ **do**
    Sample matrix W^k from $\mathcal{W}^k$
    for each node m **do**
        Generate independently $\xi_m^k \sim \mathcal{D}_k$, $\xi_m^{k+1/3} \sim \mathcal{D}_k$
        $z_m^{k+1/3} = z_m^k - \gamma F_m(z_m^k, \xi_m^k)$
        $z_m^{k+2/3} = z_m^k - \gamma F_m(z_m^{k+1/3}, \xi_m^{k+1/3})$
        $z_m^{k+1} = \sum_{i\in \mathcal{N}_m^k} w_{m,i}^k z_i^{k+2/3}$
    end for
end for

Ideas:

- Extragradient

Algorithms

Algorithm 1 Extra Step Time-Varying Gossip Method

parameters: stepsize $\gamma > 0$, $\{\mathcal{W}^k\}_{k \geq 0}$ – rules or distributions for mixing matrix in iteration k.
initialize: $z^0 \in \mathcal{Z}, \forall m : z_m^0 = z^0$

for $k = 0, 1, 2, \ldots$ **do**
  Sample matrix W^k from $\mathcal{W}^k$
  for each node m **do**
    Generate independently $\xi_m^k \sim \mathcal{D}_k$, $\xi_m^{k+1/3} \sim \mathcal{D}_k$
    $z_m^{k+1/3} = z_m^k - \gamma F_m(z_m^k, \xi_m^k)$
    $z_m^{k+2/3} = z_m^k - \gamma F_m(z_m^{k+1/3}, \xi_m^{k+1/3})$
    $z_m^{k+1} = \sum_{i \in \mathcal{N}_m^k} w_{m,i}^k z_i^{k+2/3}$
  end for
end for

Ideas:

- Extragradient
- Gossip step = weighted averaging with network neighbors

Thank you!