Optimal Algorithms for Decentralized Stochastic Variational Inequalities






Dmitry Kovalev
KAUST

Aleksandr Beznosikov
MIPT, HSE and Yandex

Abdurakhmon Sadiev
MIPT

Michael Persiianov
MIPT

Peter Richtarik
KAUST

Alexander Gasnikov
MIPT, HSE and IITP

KAUST MIPT MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY HIGHER SCHOOL OF ECONOMICS Yandex

NeurIPS 2022

Variational Inequality Problem

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

Variational Inequality Problem

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

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

Variational Inequality Problem

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

- $\min_{z \in \mathbb{R}^d} f(z) + g(z) \longrightarrow F(z) := \nabla f(z)$
- $\min_{x \in \mathbb{R}^{d_x}} \max_{y \in \mathbb{R}^{d_y}} f(x, y) + g_1(x) - g_2(y) \longrightarrow F(z) := [\nabla_x g(x, y), -\nabla_y g(x, y)]$

Distributed Stochastic Setting

$$F(z) := \frac{1}{M} \sum_{m=1}^{M} F_m(z)$$

Distributed Stochastic Setting

$$F(z) := \frac{1}{M} \sum_{m=1}^{M} F_m(z)$$

μ-strongly monotone

Distributed Stochastic Setting

$$F(z) := \frac{1}{M}\sum_{m=1}^{M} F_m(z)$$

μ-strongly monotone

on local devices

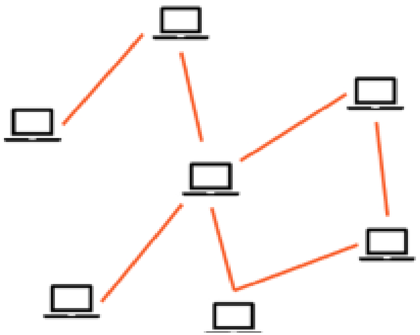
Distributed Stochastic Setting

$$F(z) := \frac{1}{M}\sum_{m=1}^{M} F_m(z)$$

μ-strongly monotone

on local devices

- fixed network



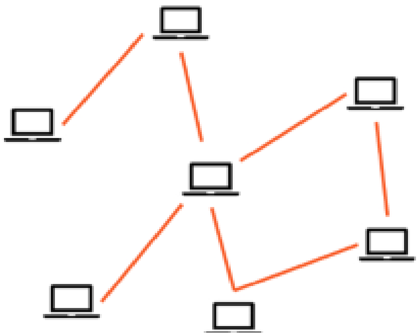
Distributed Stochastic Setting

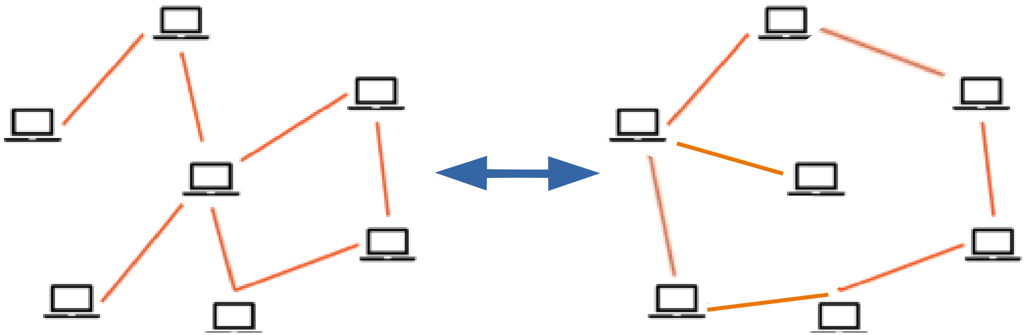
$$F(z) := \frac{1}{M}\sum_{m=1}^{M} F_m(z)$$

μ-strongly monotone

on local devices

- fixed network
- time-varying network





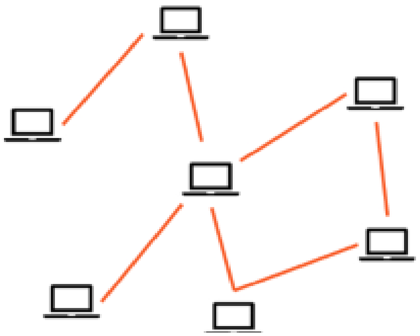
Distributed Stochastic Setting

$$F(z) := \frac{1}{M}\sum_{m=1}^{M} F_m(z) \longrightarrow F_m(z) := \frac{1}{n}\sum_{i=1}^{n} F_{m,i}(z)$$

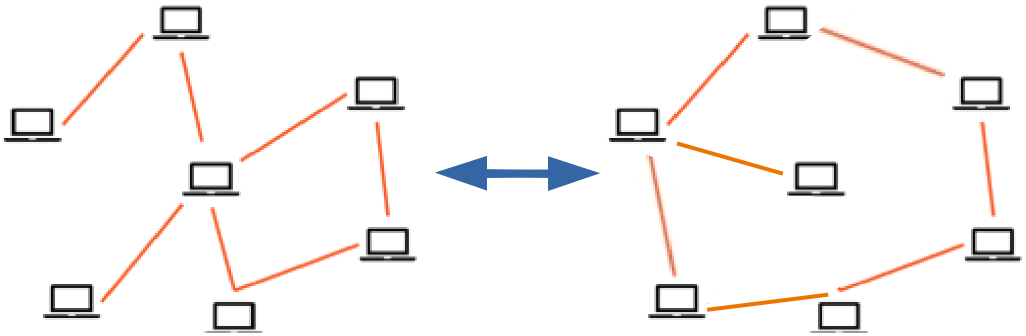
μ-strongly monotone

on local devices

- fixed network



- time-varying network



Distributed Stochastic Setting

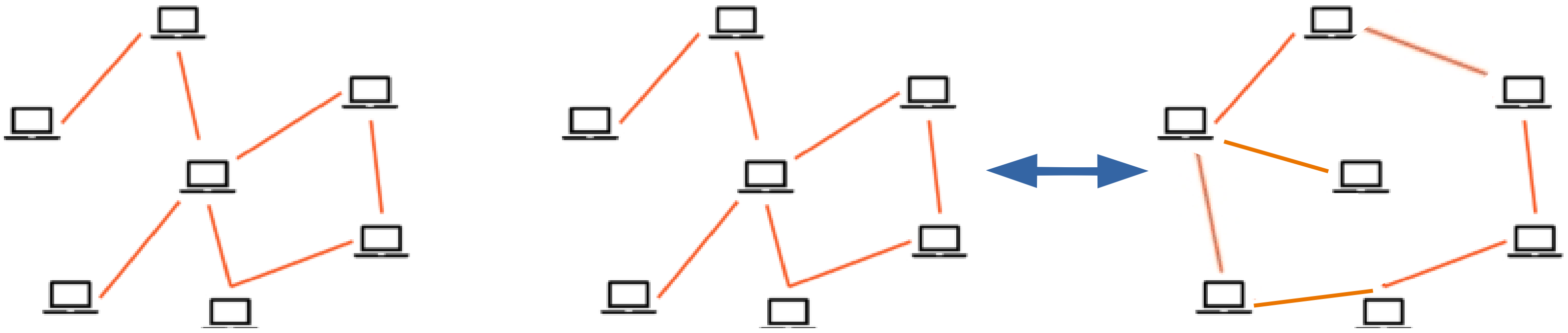
$$F(z) := \frac{1}{M}\sum_{m=1}^{M} F_m(z) \longrightarrow F_m(z) := \frac{1}{n}\sum_{i=1}^{n} F_{m,i}(z)$$

μ-strongly monotone

on local devices

L-Lipschitz continuous

- fixed network
- time-varying network



Lower bounds

Theorem. For any $L \geq \mu > 0$ and $\chi \geq 1$, $n \in \mathbb{N}$, there exist a decentralized variational inequality (satisfying assumptions from previous slides over a fixed network with characteristic number χ, such that the number of communication rounds and local computations required to obtain an ε-solution is lower bounded by

$$\Omega\left(\sqrt{\chi}\left(1+\frac{L}{\mu}\right)\cdot\log\left(\frac{R_0^2}{\varepsilon}\right)\right) \text{ and } \Omega\left(\left(n+\sqrt{n}\cdot\frac{L}{\mu}\right)\cdot\log\left(\frac{R_0^2}{\varepsilon}\right)\right), \text{ respectively.}$$

Theorem. For any $L \geq \mu > 0$ and $\chi \geq 3$, $n \in \mathbb{N}$, there exist a decentralized variational inequality (satisfying assumptions from previous slides) over a time-varying network with characteristic number χ, such that the number of communication rounds and local computations required to obtain an ε-solution is lower bounded by

$$\Omega\left(\chi\left(1+\frac{L}{\mu}\right)\cdot\log\left(\frac{R_0^2}{\varepsilon}\right)\right) \text{ and } \Omega\left(\left(n+\sqrt{n}\cdot\frac{L}{\mu}\right)\cdot\log\left(\frac{R_0^2}{\varepsilon}\right)\right), \text{ respectively.}$$

Upper bounds

- fixed network

Algorithm 1

Parameters: Stepsizes $\eta, \theta > 0$, momentums α, β, γ, batchsize $b \in \{1, \ldots, n\}$, probability $p \in (0,1)$
Initialization: Choose $\mathbf{z}^0 = \mathbf{w}^0 \in (\operatorname{dom} g)^M$, $\mathbf{y}^0 \in \mathcal{L}^\perp$. Put $\mathbf{z}^{-1} = \mathbf{z}^0, \mathbf{w}^{-1} = \mathbf{w}^0, \mathbf{y}^{-1} = \mathbf{y}^0$
for $k = 0, 1, 2 \ldots$ **do**
  Sample $j^k_{m,1}, \ldots, j^k_{m,b}$ independently from $[n]$
  $S^k = \{j^k_{m,1}, \ldots, j^k_{m,b}\}$
  Sample $j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}$ independently from $[n]$
  $S^{k+1/2} = \{j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}\}$
  $\delta^k = \frac{1}{b}\sum_{j \in S^k} \Big(\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{w}^{k-1}) + \alpha[\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{z}^{k-1})]\Big) + \mathbf{F}(\mathbf{w}^{k-1})$
  $\Delta^k = \delta^k - (\mathbf{y}^k + \alpha(\mathbf{y}^k - \mathbf{y}^{k-1}))$
  $\mathbf{z}^{k+1} = \operatorname{prox}_{\eta \mathbf{g}}(\mathbf{z}^k + \gamma(\mathbf{w}^k - \mathbf{z}^k) - \eta\Delta^k)$
  $\Delta^{k+1/2} = \frac{1}{b}\sum_{j \in S^{k+1/2}} (\mathbf{F}_j(\mathbf{z}^{k+1}) - \mathbf{F}_j(\mathbf{w}^k)) + \mathbf{F}(\mathbf{w}^k)$
  $\mathbf{y}^{k+1} = \mathbf{y}^k - \theta(\mathbf{W} \otimes \mathbf{I}_d)(\mathbf{z}^{k+1} - \beta(\Delta^{k+1/2} - \mathbf{y}^k))$
  $\mathbf{w}^{k+1} = \begin{cases} \mathbf{z}^k, & \text{with probability } p \\ \mathbf{w}^k, & \text{with probability } 1-p \end{cases}$
end for

$^*\mathbf{F}_j(\mathbf{z}) = (F_{1,j_{1,l}}(z_1), \ldots, F_{M,j_{M,l}}(z_M))^T$, $l \in \{1, \ldots, b\}$

- time-varying network

Algorithm 2

Parameters: Stepsizes $\eta_z, \eta_y, \eta_x, \theta > 0$, momentums $\alpha, \gamma, \omega, \tau$, parameters ν, β, batchsize $b \in \{1, \ldots, n\}$, probability $p \in (0,1)$
Initialization: Choose $\mathbf{z}^0 = \mathbf{w}^0 \in (\operatorname{dom} g)^M$, $\mathbf{y}^0 \in (\mathbb{R}^d)^M$, $\mathbf{x}^0 \in \mathcal{L}^\perp$. Put $\mathbf{z}^{-1} = \mathbf{z}^0, \mathbf{w}^{-1} = \mathbf{w}^0, \mathbf{y}_f = \mathbf{y}^{-1} = \mathbf{y}^0, \mathbf{x}_f = \mathbf{x}^{-1} = \mathbf{x}^0, m_0 = \mathbf{0}^{dM}$
for $k = 0, 1, 2, \ldots$ **do**
  Sample $j^k_{m,1}, \ldots, j^k_{m,b}$ independently from $[n]$
  $S^k = \{j^k_{m,1}, \ldots, j^k_{m,b}\}$
  Sample $j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}$ independently from $[n]$
  $S^{k+1/2} = \{j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}\}$
  $\delta^k = \frac{1}{b}\sum_{j \in S^k} \Big(\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{w}^{k-1}) + \alpha[\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{z}^{k-1})]\Big) + \mathbf{F}(\mathbf{w}^{k-1})$
  $\Delta^k_z = \delta^k - \nu\mathbf{z}^k - \mathbf{y}^k - \alpha(\mathbf{y}^k - \mathbf{y}^{k-1})$
  $\mathbf{z}^{k+1} = \operatorname{prox}_{\eta_z \mathbf{g}}(\mathbf{z}^k + \omega(\mathbf{w}^k - \mathbf{z}^k) - \eta_z\Delta^k_z)$
  $\mathbf{y}^k_c = \tau\mathbf{y}^k + (1-\tau)\mathbf{y}^k_f$
  $\mathbf{x}^k_c = \tau\mathbf{x}^k + (1-\tau)\mathbf{x}^k_f$
  $\Delta^k_y = \nu^{-1}(\mathbf{y}^k_c + \mathbf{x}^k_c) + \mathbf{z}^{k+1} + \gamma(\mathbf{y}^k + \mathbf{x}^k + \nu\mathbf{z}^k)$
  $\delta^{k+1/2} = \frac{1}{b}\sum_{j \in S^{k+1/2}} (\mathbf{F}_j(\mathbf{z}^{k+1}) - \mathbf{F}_j(\mathbf{w}^k)) + \mathbf{F}(\mathbf{w}^k)$
  $\Delta^k_x = \nu^{-1}(\mathbf{y}^k_c + \mathbf{x}^k_c) + \beta(\mathbf{x}^k + \delta^{k+1/2})$
  $\mathbf{y}^{k+1} = \mathbf{y}^k - \eta_y\Delta^k_y$
  $\mathbf{x}^{k+1} = \mathbf{x}^k - (\mathbf{W}_T(Tk) \otimes \mathbf{I}_d)(\eta_x\Delta^k_x + m^k)$
  $m^{k+1} = \eta_x\Delta^k_x + m^k - (\mathbf{W}_T(Tk) \otimes \mathbf{I}_d)(\eta_x\Delta^k_x + m^k)$
  $\mathbf{y}^{k+1}_f = \mathbf{y}^k_c + \tau(\mathbf{y}^{k+1} - \mathbf{y}^k)$
  $\mathbf{x}^{k+1}_f = \mathbf{x}^k_c - \theta(\mathbf{W}_T(Tk) \otimes \mathbf{I}_d)(\mathbf{y}^k_c + \mathbf{x}^k_c)$
  $\mathbf{w}^{k+1} = \begin{cases} \mathbf{z}^k, & \text{with probability } p \\ \mathbf{w}^k, & \text{with probability } 1-p \end{cases}$
end for

Upper bounds

- fixed network

Algorithm 1

Parameters: Stepsizes $\eta, \theta > 0$, momentums α, β, γ, batchsize $b \in \{1, \ldots, n\}$, probability $p \in (0, 1)$
Initialization: Choose $\mathbf{z}^0 = \mathbf{w}^0 \in (\operatorname{dom} g)^M$, $\mathbf{y}^0 \in \mathcal{L}^\perp$. Put $\mathbf{z}^{-1} = \mathbf{z}^0, \mathbf{w}^{-1} = \mathbf{w}^0, \mathbf{y}^{-1} = \mathbf{y}^0$
for $k = 0, 1, 2 \ldots$ **do**
  Sample $j^k_{m,1}, \ldots, j^k_{m,b}$ independently from $[n]$
  $S^k = \{j^k_{m,1}, \ldots, j^k_{m,b}\}$
  Sample $j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}$ independently from $[n]$
  $S^{k+1/2} = \{j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}\}$
  $\delta^k = \frac{1}{b}\sum_{j \in S^k}\Big(\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{w}^{k-1}) + \alpha[\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{z}^{k-1})]\Big) + \mathbf{F}(\mathbf{w}^{k-1})$
  $\Delta^k = \delta^k - (\mathbf{y}^k + \alpha(\mathbf{y}^k - \mathbf{y}^{k-1}))$
  $\mathbf{z}^{k+1} = \operatorname{prox}_{\eta \mathbf{g}}(\mathbf{z}^k + \gamma(\mathbf{w}^k - \mathbf{z}^k) - \eta\Delta^k)$
  $\Delta^{k+1/2} = \frac{1}{b}\sum_{j \in S^{k+1/2}}(\mathbf{F}_j(\mathbf{z}^{k+1}) - \mathbf{F}_j(\mathbf{w}^k)) + \mathbf{F}(\mathbf{w}^k)$
  $\mathbf{y}^{k+1} = \mathbf{y}^k - \theta(\mathbf{W} \otimes \mathbf{I}_d)(\mathbf{z}^{k+1} - \beta(\Delta^{k+1/2} - \mathbf{y}^k))$
  $\mathbf{w}^{k+1} = \begin{cases} \mathbf{z}^k, & \text{with probability } p \\ \mathbf{w}^k, & \text{with probability } 1 - p \end{cases}$
end for

$^*\mathbf{F}_j(\mathbf{z}) = (F_{1,j_{1,l}}(z_1), \ldots, F_{M,j_{M,l}}(z_M))^T$, $l \in \{1, \ldots, b\}$

- time-varying network

Algorithm 2

Parameters: Stepsizes $\eta_z, \eta_y, \eta_x, \theta > 0$, momentums $\alpha, \gamma, \omega, \tau$, parameters ν, β, batchsize $b \in \{1, \ldots, n\}$, probability $p \in (0, 1)$
Initialization: Choose $\mathbf{z}^0 = \mathbf{w}^0 \in (\operatorname{dom} g)^M$, $\mathbf{y}^0 \in (\mathbb{R}^d)^M$, $\mathbf{x}^0 \in \mathcal{L}^\perp$. Put $\mathbf{z}^{-1} = \mathbf{z}^0, \mathbf{w}^{-1} = \mathbf{w}^0, \mathbf{y}_f = \mathbf{y}^{-1} = \mathbf{y}^0, \mathbf{x}_f = \mathbf{x}^{-1} = \mathbf{x}^0, m_0 = \mathbf{0}^{dM}$
for $k = 0, 1, 2, \ldots$ **do**
  Sample $j^k_{m,1}, \ldots, j^k_{m,b}$ independently from $[n]$
  $S^k = \{j^k_{m,1}, \ldots, j^k_{m,b}\}$
  Sample $j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}$ independently from $[n]$
  $S^{k+1/2} = \{j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}\}$
  $\delta^k = \frac{1}{b}\sum_{j \in S^k}\Big(\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{w}^{k-1}) + \alpha[\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{z}^{k-1})]\Big) + \mathbf{F}(\mathbf{w}^{k-1})$
  $\Delta^k_z = \delta^k - \nu\mathbf{z}^k - \mathbf{y}^k - \alpha(\mathbf{y}^k - \mathbf{y}^{k-1})$
  $\mathbf{z}^{k+1} = \operatorname{prox}_{\eta_z \mathbf{g}}(\mathbf{z}^k + \omega(\mathbf{w}^k - \mathbf{z}^k) - \eta_z\Delta^k_z)$
  $\mathbf{y}^k_c = \tau\mathbf{y}^k + (1 - \tau)\mathbf{y}^k_f$
  $\mathbf{x}^k_c = \tau\mathbf{x}^k + (1 - \tau)\mathbf{x}^k_f$
  $\Delta^k_y = \nu^{-1}(\mathbf{y}^k_c + \mathbf{x}^k_c) + \mathbf{z}^{k+1} + \gamma(\mathbf{y}^k + \mathbf{x}^k + \nu\mathbf{z}^k)$
  $\delta^{k+1/2} = \frac{1}{b}\sum_{j \in S^{k+1/2}}(\mathbf{F}_j(\mathbf{z}^{k+1}) - \mathbf{F}_j(\mathbf{w}^k)) + \mathbf{F}(\mathbf{w}^k)$
  $\Delta^k_x = \nu^{-1}(\mathbf{y}^k_c + \mathbf{x}^k_c) + \beta(\mathbf{x}^k + \delta^{k+1/2})$
  $\mathbf{y}^{k+1} = \mathbf{y}^k - \eta_y\Delta^k_y$
  $\mathbf{x}^{k+1} = \mathbf{x}^k - (\mathbf{W}_T(Tk) \otimes \mathbf{I}_d)(\eta_x\Delta^k_x + m^k)$
  $m^{k+1} = \eta_x\Delta^k_x + m^k - (\mathbf{W}_T(Tk) \otimes \mathbf{I}_d)(\eta_x\Delta^k_x + m^k)$
  $\mathbf{y}^{k+1}_f = \mathbf{y}^k_c + \tau(\mathbf{y}^{k+1} - \mathbf{y}^k)$
  $\mathbf{x}^{k+1}_f = \mathbf{x}^k_c - \theta(\mathbf{W}_T(Tk) \otimes \mathbf{I}_d)(\mathbf{y}^k_c + \mathbf{x}^k_c)$
  $\mathbf{w}^{k+1} = \begin{cases} \mathbf{z}^k, & \text{with probability } p \\ \mathbf{w}^k, & \text{with probability } 1 - p \end{cases}$
end for

Upper bounds matches lower bounds!

Upper bounds

- fixed network

Algorithm 1

Parameters: Stepsizes $\eta, \theta > 0$, momentums α, β, γ, batchsize $b \in \{1, \ldots, n\}$, probability $p \in (0,1)$

Initialization: Choose $\mathbf{z}^0 = \mathbf{w}^0 \in (\mathrm{dom}\, g)^M$, $\mathbf{y}^0 \in \mathcal{L}^\perp$. Put $\mathbf{z}^{-1} = \mathbf{z}^0, \mathbf{w}^{-1} = \mathbf{w}^0, \mathbf{y}^{-1} = \mathbf{y}^0$

for $k = 0, 1, 2 \ldots$ **do**

Sample $j^k_{m,1}, \ldots, j^k_{m,b}$ independently from $[n]$

$S^k = \{j^k_{m,1}, \ldots, j^k_{m,b}\}$

Sample $j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}$ independently from $[n]$

$S^{k+1/2} = \{j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}\}$

$$\delta^k = \tfrac{1}{b} \textstyle\sum_{j \in S^k} \Big(\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{w}^{k-1}) + \alpha[\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{z}^{k-1})]\Big) + \mathbf{F}(\mathbf{w}^{k-1})$$

$\Delta^k = \delta^k - (\mathbf{y}^k + \alpha(\mathbf{y}^k - \mathbf{y}^{k-1}))$

$\mathbf{z}^{k+1} = \mathrm{prox}_{\eta g}(\mathbf{z}^k + \gamma(\mathbf{w}^k - \mathbf{z}^k) - \eta\Delta^k)$

$$\Delta^{k+1/2} = \tfrac{1}{b} \textstyle\sum_{j \in S^{k+1/2}} (\mathbf{F}_j(\mathbf{z}^{k+1}) - \mathbf{F}_j(\mathbf{w}^k)) + \mathbf{F}(\mathbf{w}^k)$$

$\mathbf{y}^{k+1} = \mathbf{y}^k - \theta(\mathbf{W} \otimes \mathbf{I}_d)(\mathbf{z}^{k+1} - \beta(\Delta^{k+1/2} - \mathbf{y}^k))$

$$\mathbf{w}^{k+1} = \begin{cases} \mathbf{z}^k, & \text{with probability } p \\ \mathbf{w}^k, & \text{with probability } 1-p \end{cases}$$

end for

*$\mathbf{F}_j(\mathbf{z}) = (F_{1,j_{1,l}}(z_1), \ldots, F_{M,j_{M,l}}(z_M))^T$, $l \in \{1, \ldots, b\}$

- time-varying network

Algorithm 2

Parameters: Stepsizes $\eta_z, \eta_y, \eta_x, \theta > 0$, momentums $\alpha, \gamma, \omega, \tau$, parameters ν, β, batchsize $b \in \{1, \ldots, n\}$, probability $p \in (0,1)$

Initialization: Choose $\mathbf{z}^0 = \mathbf{w}^0 \in (\mathrm{dom}\, g)^M$, $\mathbf{y}^0 \in (\mathbb{R}^d)^M$, $\mathbf{x}^0 \in \mathcal{L}^\perp$. Put $\mathbf{z}^{-1} = \mathbf{z}^0, \mathbf{w}^{-1} = \mathbf{w}^0, \mathbf{y}_f = \mathbf{y}^{-1} = \mathbf{y}^0, \mathbf{x}_f = \mathbf{x}^{-1} = \mathbf{x}^0, m_0 = \mathbf{0}^{dM}$

for $k = 0, 1, 2, \ldots$ **do**

Sample $j^k_{m,1}, \ldots, j^k_{m,b}$ independently from $[n]$

$S^k = \{j^k_{m,1}, \ldots, j^k_{m,b}\}$

Sample $j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}$ independently from $[n]$

$S^{k+1/2} = \{j^{k+1/2}_{m,1}, \ldots, j^{k+1/2}_{m,b}\}$

$$\delta^k = \tfrac{1}{b} \textstyle\sum_{j \in S^k} \Big(\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{w}^{k-1}) + \alpha[\mathbf{F}_j(\mathbf{z}^k) - \mathbf{F}_j(\mathbf{z}^{k-1})]\Big) + \mathbf{F}(\mathbf{w}^{k-1})$$

$\Delta^k_z = \delta^k - \nu\mathbf{z}^k - \mathbf{y}^k - \alpha(\mathbf{y}^k - \mathbf{y}^{k-1})$

$\mathbf{z}^{k+1} = \mathrm{prox}_{\eta_z g}(\mathbf{z}^k + \omega(\mathbf{w}^k - \mathbf{z}^k) - \eta_z\Delta^k_z)$

$\mathbf{y}^k_c = \tau\mathbf{y}^k + (1-\tau)\mathbf{y}^k_f$

$\mathbf{x}^k_c = \tau\mathbf{x}^k + (1-\tau)\mathbf{x}^k_f$

$\Delta^k_y = \nu^{-1}(\mathbf{y}^k_c + \mathbf{x}^k_c) + \mathbf{z}^{k+1} + \gamma(\mathbf{y}^k + \mathbf{x}^k + \nu\mathbf{z}^k)$

$$\delta^{k+1/2} = \tfrac{1}{b} \textstyle\sum_{j \in S^{k+1/2}} (\mathbf{F}_j(\mathbf{z}^{k+1}) - \mathbf{F}_j(\mathbf{w}^k)) + \mathbf{F}(\mathbf{w}^k)$$

$\Delta^k_x = \nu^{-1}(\mathbf{y}^k_c + \mathbf{x}^k_c) + \beta(\mathbf{x}^k + \delta^{k+1/2})$

$\mathbf{y}^{k+1} = \mathbf{y}^k - \eta_y\Delta^k_y$

$\mathbf{x}^{k+1} = \mathbf{x}^k - (\mathbf{W}_T(Tk) \otimes \mathbf{I}_d)(\eta_x\Delta^k_x + m^k)$

$m^{k+1} = \eta_x\Delta^k_x + m^k - (\mathbf{W}_T(Tk) \otimes \mathbf{I}_d)(\eta_x\Delta^k_x + m^k)$

$\mathbf{y}^{k+1}_f = \mathbf{y}^k_c + \tau(\mathbf{y}^{k+1} - \mathbf{y}^k)$

$\mathbf{x}^{k+1}_f = \mathbf{x}^k_c - \theta(\mathbf{W}_T(Tk) \otimes \mathbf{I}_d)(\mathbf{y}^k_c + \mathbf{x}^k_c)$

$$\mathbf{w}^{k+1} = \begin{cases} \mathbf{z}^k, & \text{with probability } p \\ \mathbf{w}^k, & \text{with probability } 1-p \end{cases}$$

end for

Upper bounds matches lower bounds!

Algorithms in the non-distributed stochastic setting

Thank you!