

When
ExtraGradient
meets VR:
bridging two giants
to boost VIs

Y. Kabikov, G. Chirkov
D. Medyakov, G. Molodtsov
MIPT, 2025

INTRODUCTION

variational inequality:

$$\forall z \in \mathcal{Z} \hookrightarrow \langle F(z^*), z - z^* \rangle \geq 0$$

finite-sum approach:

$$F(z) = \frac{1}{n} \sum_{i=1}^n F_i(z)$$

$$\min_{x \in \mathbb{R}^d} \left[F(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\xi_i \sim \mathcal{D}_i} [F_{\xi_i}(x)] \right]$$

Stochastic EG and variance problem:

$$z^{t+\frac{1}{2}} = z^t - \gamma F_{i_t}(z^t),$$

$$\mathbf{E}\{\|F_{i_t}(z^t) - F(z^t)\|^2\} \not\rightarrow 0, \quad z^t \rightarrow z^*$$

$$z^{t+1} = z^t - \gamma F_{j_t}(z^{t+\frac{1}{2}})$$

$$z^{t+1/2} = z^t - \gamma G^{t-1}$$

$$G^t = F_{i_t}(z^{t+1/2}) - y_{i_t}^t + \frac{1}{n} \sum_{j=1}^n y_j^t$$

$$y_i^{t+1} = F_{i_t}(z^{t+1/2}) \text{ for } i = i_t$$

$$z^{t+1} = z^t - \gamma G^t$$

ExtraSAGA method

(final)

$$z^{t+1/2} = z^t - \frac{1}{n} \sum_{j=1}^n y_j^t$$

$$G^t = F_{i_t}(z^{t+1/2}) - y_{i_t}^t + \frac{1}{n} \sum_{j=1}^n y_j^t$$

$$y_i^{t+1} = F_{i_t}(z^{t+1/2}) \text{ for } i = i_t$$

$$z^{t+1} = z^t - \gamma G^t$$

ExtraSAGA method

SETUP

-Lipschitzness

$$\mathbb{E} \left[\|F_{\xi}(z_1) - F_{\xi}(z_2)\|^2 \right] \leq L^2 \|z_1 - z_2\|^2$$

-Strong monotonicity

$$\langle F(z_1) - F(z_2), z_1 - z_2 \rangle \geq \mu \|z_1 - z_2\|^2$$

CONVERGENCE

$$\mathbb{E} \left[\sigma_{t+1}^2 | z^{t+\frac{1}{2}} \right] = \mathbb{E} \left[\frac{1}{n} \sum_j \| y_j^{t+1} - F_j(z^*) \|^2 \right]$$

$$\gamma \leq \frac{1}{6L}$$

linear rate

$$\begin{aligned} \mathbb{E} \left[\| z^{t+1} - z^* \| + M \sigma_{t+1}^2 | z^{t+\frac{1}{2}} \right] &\leq \\ &\leq \left(1 - \frac{\gamma \mu}{2} \right) \| z^t - z^* \|^2 + \left(1 - \frac{1}{2n} \right) \sigma_t^2 \end{aligned}$$

CONVERGENCE

Co-coercivity

Stochastic Gradient Descent-Ascent and
Consensus Optimization for Smooth Games:
Convergence Analysis under Expected Co-coercivity.
N. Loizou et al., 2021

Definition 3.1 (Co-coercivity / Co-coercive *around* w^*). We say that an operator ξ is ℓ -co-coercive if there exist $\ell > 0$ such that, $\|\xi(x) - \xi(y)\|^2 \leq \ell \langle \xi(x) - \xi(y), x - y \rangle \quad \forall x, y \in \mathbb{R}^d$.
If there exist $w^* \in \mathbb{R}^d$ and $\ell > 0$ such that $\|\xi(x) - \xi(w^*)\|^2 \leq \ell \langle \xi(x) - \xi(w^*), x - w^* \rangle \quad \forall x \in \mathbb{R}^d$, then we say that the operator ξ is ℓ -co-coercive *around* w^* . Note that in the last definition the point w^* is not necessarily a point where $\xi(w^*) = 0$.

The background is an abstract composition of broad, overlapping brushstrokes in shades of teal, light blue, yellow, and pale green. A dashed white line with small colored dots (pink, orange, purple) meanders across the upper half of the image. The word "EXPERIMENTS" is centered in a bold, dark blue, sans-serif font, with a subtle drop shadow effect.

EXPERIMENTS

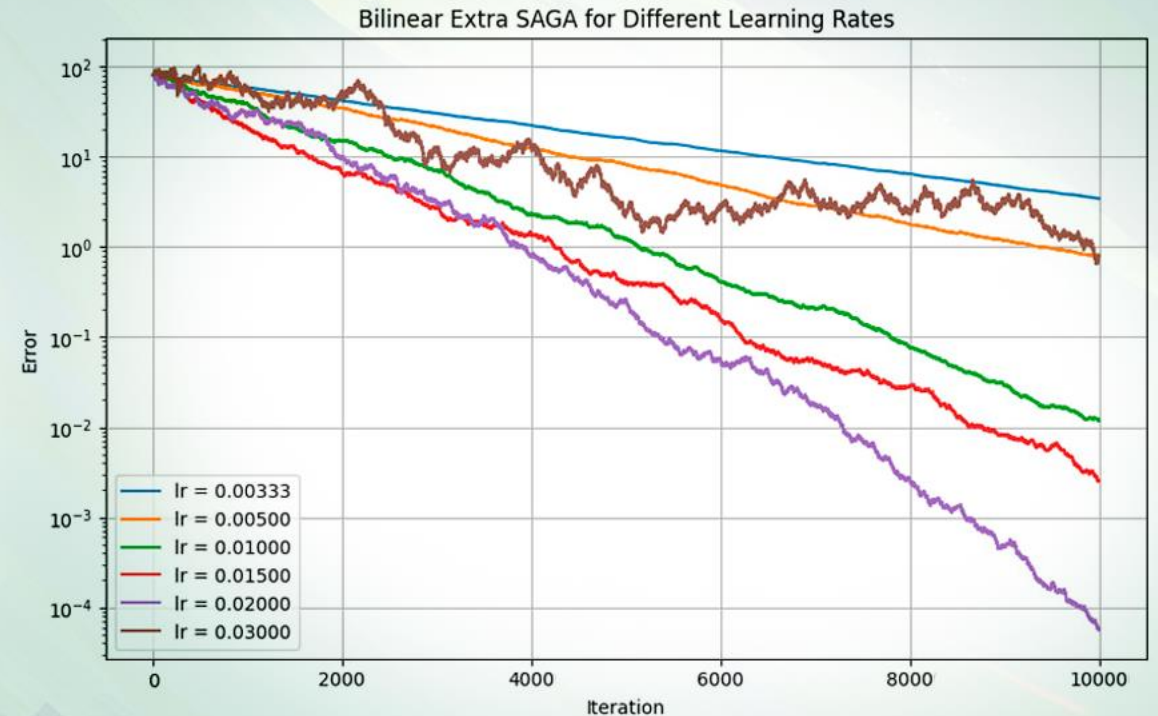
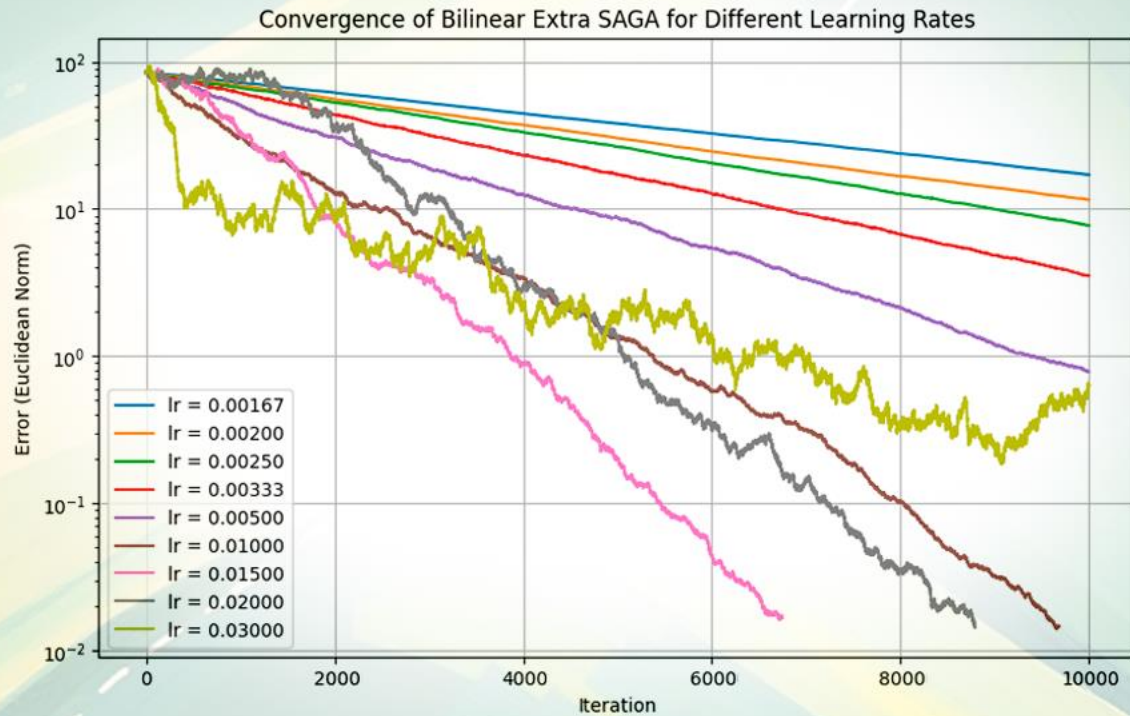
Experiments:

**Bilinear convex-concave min-max
(saddle points)**

$$\min_{x \in \mathbb{R}^{d_x}} \max_{y \in \mathbb{R}^{d_y}} [f(x, y)]$$

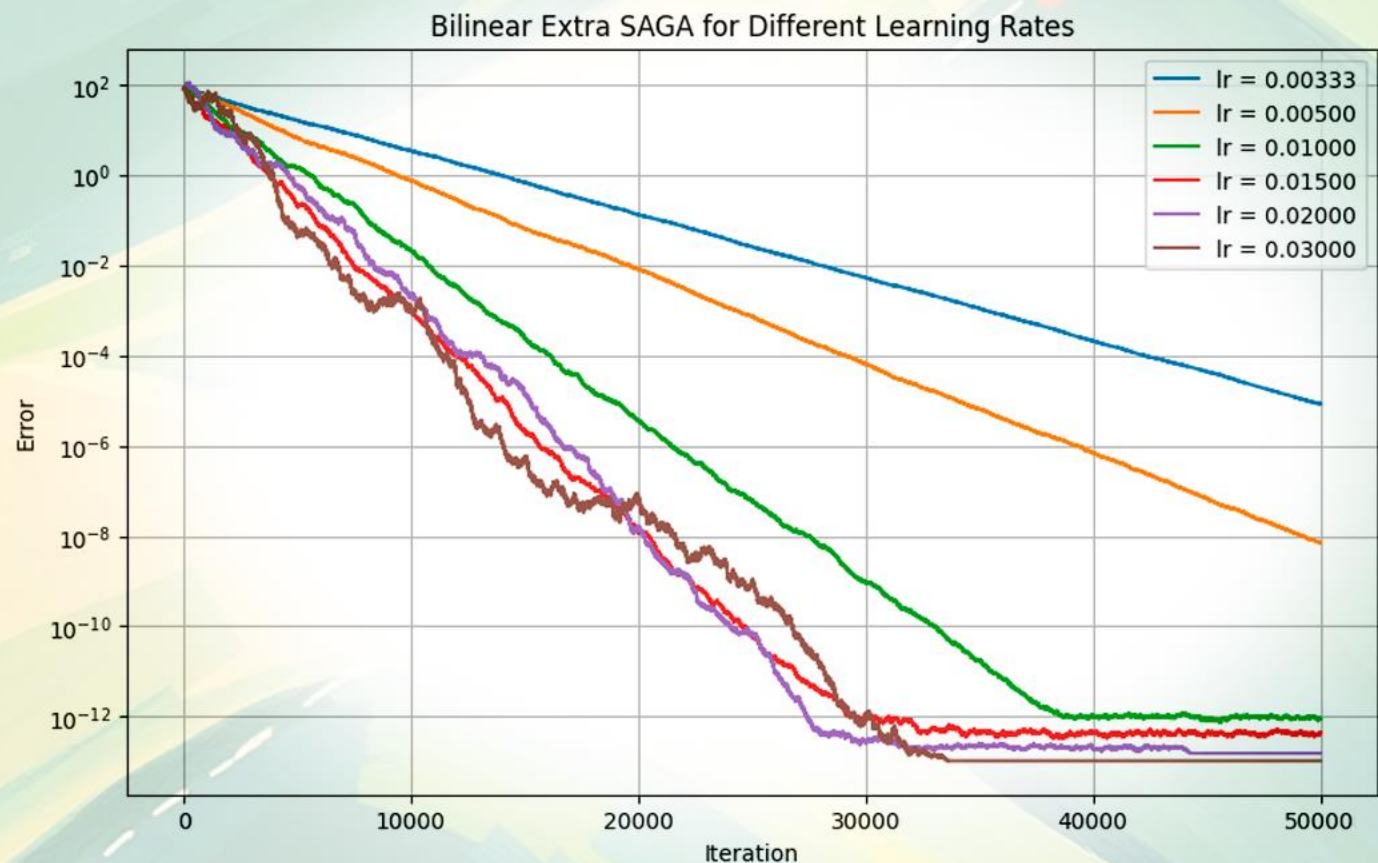
$$f(x, y) = (x - b_x)^T A (y - b_y) + \frac{\lambda}{2} \|x - b_x\|^2 - \frac{\lambda}{2} \|y - b_y\|^2$$

Experiments: Bilinear convex-concave min-max (saddle points)



$L = 100$, $\mu = 0.1$, $n = 100$, $d = 100$

Experiments: Bilinear convex-concave min-max (saddle points)



$L = 100, \mu = 0.1, n = 100, d = 100$

Denoising

$$\{(ih, jh) : 1 \leq i \leq M, 1 \leq j \leq N\}$$

picture = cartesian grid

$$(\nabla u)_{i,j} = \begin{pmatrix} (\nabla u)_{i,j}^1 \\ (\nabla u)_{i,j}^2 \end{pmatrix}, \quad (\nabla u)_{i,j}^1 = \begin{cases} \frac{u_{i+1,j} - u_{i,j}}{h} & \text{if } i < M \\ 0 & \text{if } i = M \end{cases}, \quad (\nabla u)_{i,j}^2 = \begin{cases} \frac{u_{i,j+1} - u_{i,j}}{h} & \text{if } j < N \\ 0 & \text{if } j = N \end{cases}$$

$$\min_x \int_{\Omega} |Du| + \frac{\lambda}{2} \|u - g\|_2^2$$

the ROF model

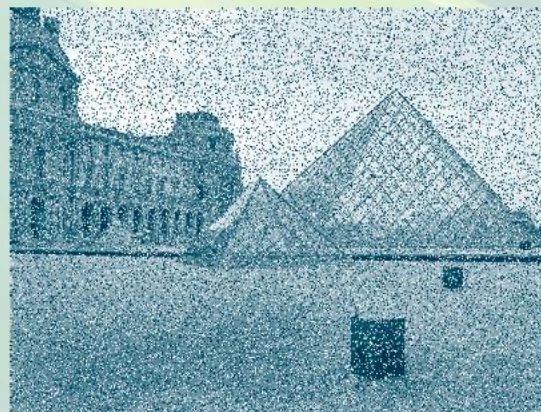
$$\min_{u \in X} \|\nabla u\|_1 + \lambda \|u - g\|_1$$

TV-L1 model

Denoising



Clean image



Noisy image

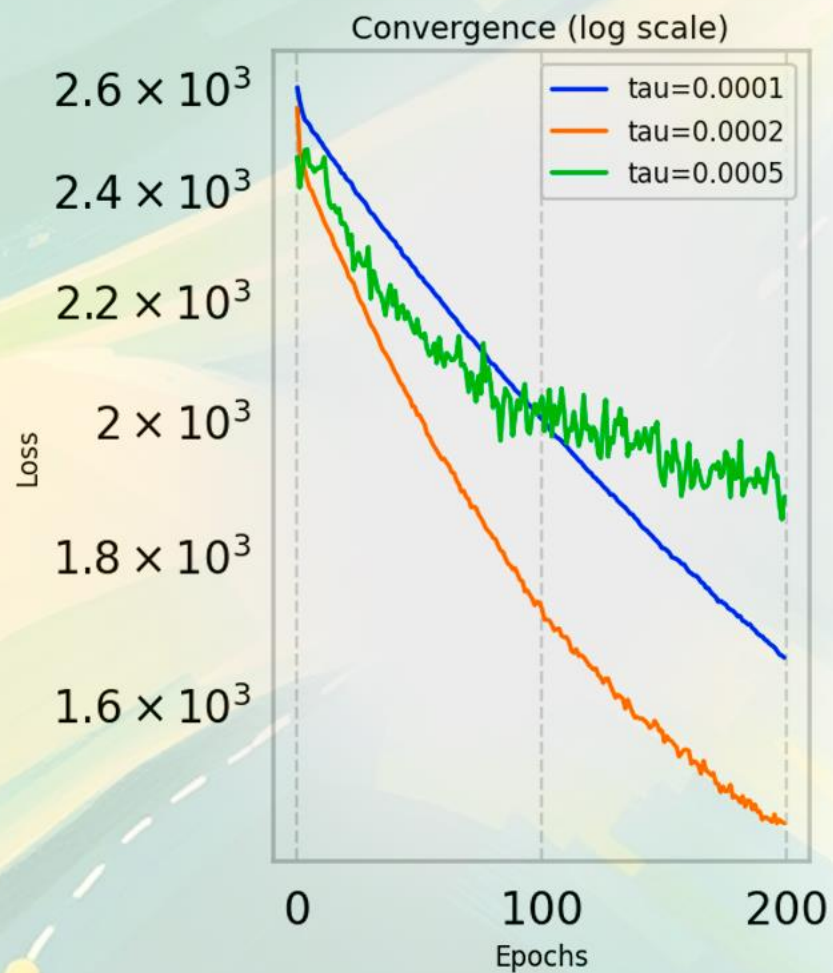


ROF ($\lambda = 8$)



TV-L1 ($\lambda = 1.5$)

DENOISING



Reconstructed (tau=0.0002)



Reconstructed (tau=0.0005)





**THANKS
FOR YOUR
ATTENTION!**