When ExtraGradient meets VR: bridging two giants to boost VIs part II

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

In previous episodes

Variational inequality:

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

Stochastic setting:

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

Dispersion problem:

$$\mathbb{E}[\|F_\xi(z^t) - F(z^t)\|^2] \nrightarrow 0 \quad with \quad condition \quad z^t \rightarrow z^*$$

ExtraSAGA

$$z^{t+\frac{1}{2}} = z^t - \gamma \frac{1}{n} \sum_{i=1}^{n} y_i^t$$

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

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

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

Convergence with the proper step 1/6L

Descent Lemma

$$\left\|z^{t+1} - z^*\right\|^2 \le \left(1 - \frac{\gamma\mu}{2}\right)\left\|z^t - z^*\right\|^2 - (1-\gamma\mu)\left\|z^{t+\frac{1}{2}} - z^t\right\|^2 +$$
$$+2\gamma\left\langle F(z^{t+\frac{1}{2}}) - G^t, z^{t+\frac{1}{2}} - z^*\right\rangle - \gamma\left\langle F(z^{t+\frac{1}{2}}) - F(z^*), z^{t+\frac{1}{2}} - z^*\right\rangle +$$
$$+\gamma^2\left\|\frac{1}{n}\sum_j y_j^t - G^t\right\|^2$$

Variance Reduction

$$\mathbb{E}\left\|\frac{1}{n}\sum_j y_j^t - G^t\right\|^2 = \mathbb{E}\left\|\frac{1}{n}\sum_j y_j^t - F_{i_t}(z^{t+\frac{1}{2}}) + y_{i_t}^t - \frac{1}{n}\sum_j y_j^t\right\|^2$$
$$= \mathbb{E}\left\|y_{i_t}^t - F_{i_t}(z^*) + F_{i_t}(z^*) - F_{i_t}(z^{t+\frac{1}{2}})\right\|^2$$
$$\le 2\mathbb{E}\left\|y_{i_t}^t - F_{i_t}(z^*)\right\|^2 + 2\mathbb{E}\left\|F_{i_t}(z^{t+\frac{1}{2}}) - F_{i_t}(z^*)\right\|^2$$
$$\le \underbrace{\frac{2}{n}\sum_j \left\|y_j^t - F_j(z^*)\right\|^2}_{\sigma_t^2} + 2L\left\langle F(z^{t+\frac{1}{2}}) - F(z^*), z^{t+\frac{1}{2}} - z^*\right\rangle$$

Final view of linear convergence in a case of mu-strong monotone operator

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

Monotone operator case

$$Gap(w) = \max_{z \in C} \{\langle F(z), w - z \rangle\}$$

Special technic

$$\mathbb{E}\left[Gap(z^K)\right] = O\left(\frac{L}{K}\right)$$

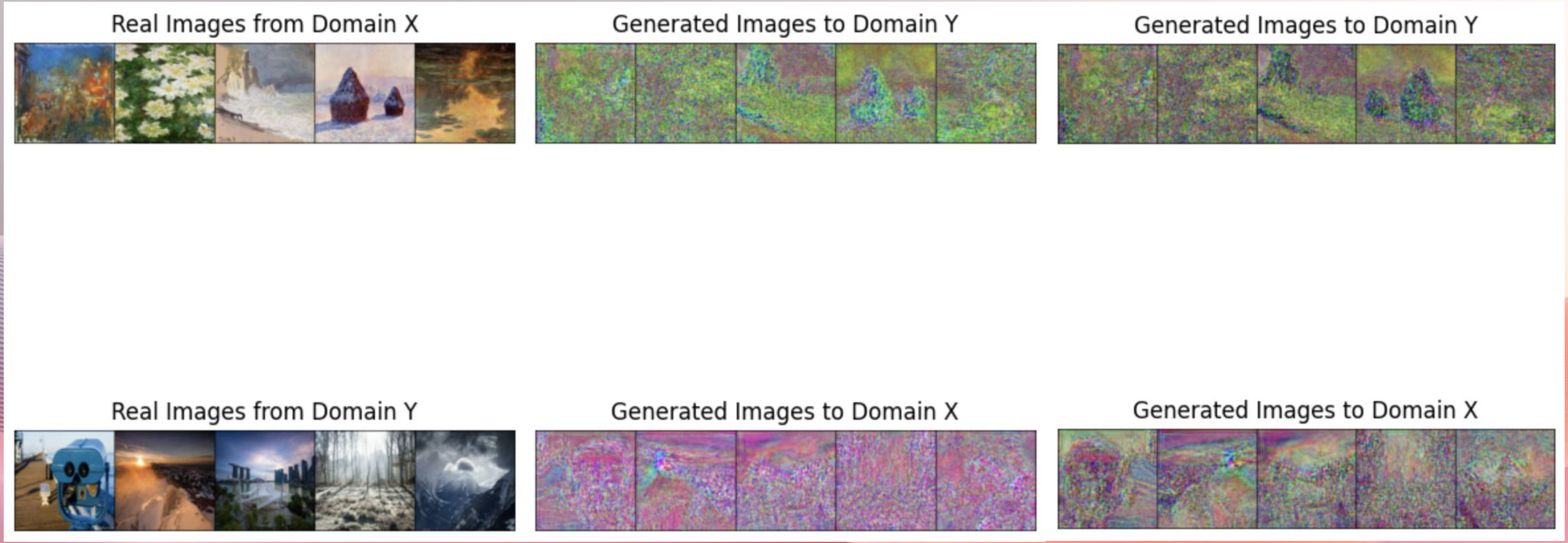
for

$$z^K = \frac{1}{K}\sum_{i=0}^{K-1} z^{i+\frac{1}{2}}$$

$$\gamma K \mathbb{E}\left[Gap(z^K)\right] \leq \max_{z \in C} \|z_0 - z\|^2 + \sigma_K^2$$

! Sublinear Convergence !

Positive Result



Epoch 10

Epoch 160

Idea that coefficient makes ExtraSAGA effective

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

Estimator

$$X^* = X - \alpha(Y - \mathbb{E}Y)$$

Variance through affection

Suggestion for minimizing

$$\alpha^* = \frac{Cov(X, Y)}{Var(Y)}$$

Our intentions:

To make an explorations and experiments with alpha- method;

To achieve positive in Denoising;

GAN and Adversarial training;

Article writing;

Thanks for attention