Tensor methods for strongly convex strongly concave saddle point problems and strongly monotone variational inequalities

Petr Ostroukhov[1,2], Rinat Kamalov[1,3], Pavel Dvurechensky[2,4], Alexander Gasnikov[1,2,4]

[1] Moscow Institute of Physics and Technology, Dolgoprudny, Russia
[2] Institute for Information Transmission Problems RAS, Moscow Russia
[3] V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia
[4] Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

17 July, 2021

Problem statement

1. **Saddle point problem (SPP) and MVI**

$$\min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m} g(x,y),\ \ g:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R},$$

Denote $z=\begin{pmatrix}x\\y\end{pmatrix}$, and operator $F:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^n\times\mathbb{R}^m$:

$$F(z)=F(x,y):=\begin{pmatrix}\nabla_x g(x,y)\\-\nabla_y g(x,y)\end{pmatrix},$$

then if operator $F:\mathcal{Z}\to\mathbb{R}^n$ is monotone over a convex set $\mathcal{Z}\subset\mathbb{R}^n$, find

$$z^*\in\mathcal{Z}:\ \forall z\in\mathcal{Z},\langle F(z),z^*-z\rangle\le 0. \tag{1}$$

2. **Gradient norm minimization**

$$\min_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m}\|\nabla g(x,y)\|_2.$$

Problem statement

1. **Saddle point problem (SPP) and MVI**

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} g(x, y), \;\; g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R},$$

Denote $z = \begin{pmatrix} x \\ y \end{pmatrix}$, and operator $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$:

$$F(z) = F(x, y) := \begin{pmatrix} \nabla_x g(x, y) \\ -\nabla_y g(x, y) \end{pmatrix},$$

then if operator $F : \mathcal{Z} \to \mathbb{R}^n$ is monotone over a convex set $\mathcal{Z} \subset \mathbb{R}^n$, find

$$z^* \in \mathcal{Z} : \;\; \forall z \in \mathcal{Z}, \langle F(z), z^* - z \rangle \leq 0. \tag{1}$$

2. **Gradient norm minimization**

$$\min_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m} \|\nabla g(x, y)\|_2.$$

Additional notations

- **Dual gap**

$$G(x, y) := \max_{y' \in \mathbb{R}^m} g(x, y') - \min_{x' \in \mathbb{R}^n} g(x', y).$$

- **SPP approximate solution**

$$\tilde{x}^* \in \mathbb{R}^n, \tilde{y}^* \in \mathbb{R}^m \Rightarrow G(\tilde{x}^*, \tilde{y}^*) \leq \varepsilon_G.$$

- **Gradient norm minimization approximate solution**

$$\tilde{x}^* \in \mathbb{R}^n, \tilde{y}^* \in \mathbb{R}^m \Rightarrow \|\nabla g(\tilde{x}^*, \tilde{y}^*)\|_2 \leq \varepsilon_\nabla.$$

Assumptions

Assumption 1.

$g(x, y)$ is μ-strongly convex in x and μ-strongly concave in y.

$$\forall x_1, x_2, y \Rightarrow \langle \nabla_x g(x_1, y) - \nabla_x g(x_2, y), x_1 - x_2 \rangle \geq \mu \|x_1 - x_2\|_2^2,$$
$$\forall y_1, y_2, x \Rightarrow \langle -\nabla_y g(x, y_1) + \nabla_y g(x, y_2), y_1 - y_2 \rangle \geq \mu \|y_1 - y_2\|_2^2.$$

Assumptions

Assumption 1.

$g(x, y)$ is μ-strongly convex in x and μ-strongly concave in y.

$$\forall x_1, x_2, y \Rightarrow \langle \nabla_x g(x_1, y) - \nabla_x g(x_2, y), x_1 - x_2 \rangle \geq \mu \|x_1 - x_2\|_2^2,$$
$$\forall y_1, y_2, x \Rightarrow \langle -\nabla_y g(x, y_1) + \nabla_y g(x, y_2), y_1 - y_2 \rangle \geq \mu \|y_1 - y_2\|_2^2.$$

Assumption 2.

$F(z)$ satisfies first order Lipschitz condition:

$$\|F(z_1) - F(z_2)\|_2 \leq L_1 \|z_1 - z_2\|_2$$
$$\Leftrightarrow \|\nabla g(z_1) - \nabla g(z_2)\|_2 \leq L_1 \|z_1 - z_2\|_2.$$

Assumptions

Assumption 3.

$F(z)$ satisfies second order Lipschitz condition:

$$\|\nabla F(z_1) - \nabla F(z_2)\|_2 \leq L_2\|z_1 - z_2\|_2$$
$$\Leftrightarrow \|\nabla^2 g(z_1) - \nabla^2 g(z_2)\|_2 \leq L_2\|z_1 - z_2\|_2.$$

Assumptions

Assumption 3.

$F(z)$ satisfies second order Lipschitz condition:

$$\|\nabla F(z_1) - \nabla F(z_2)\|_2 \leq L_2\|z_1 - z_2\|_2$$
$$\Leftrightarrow \|\nabla^2 g(z_1) - \nabla^2 g(z_2)\|_2 \leq L_2\|z_1 - z_2\|_2.$$

Assumption 4.

$F(z)$ satisfies p-th order Lipschitz condition:

$$\|\nabla^{p-1} F(z_1) - \nabla^{p-1} F(z_2)\|_2 \leq L_p\|z_1 - z_2\|_2$$
$$\Leftrightarrow \|\nabla^p g(z_1) - \nabla^p g(z_2)\|_2 \leq L_p\|z_1 - z_2\|_2.$$

Name	Order	Assumptions	Complexity
Nemirovski 2004	1	L_1	$O(L_1/\varepsilon_G)$
Gasnikov et al. 2020	1	μ, L_1	$O\left(L_1/\mu \cdot \log(1/\varepsilon_G)\right)$
Bullins and Lai 2020	p	L_p	$O(L_p/\varepsilon_G)^{\frac{2}{p+1}}$

Table 1: Existing results

Restarts + HighOrderMirrorProx from Bullins and Lai 2020

Theorem 1.
Suppose $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$*, that is defined in* (1)*, is p-th order Lipschitz and* μ*-strongly monotone (Assumptions 1 and 4 hold). Denote* R *such that* $R \geq \|z_1 - z^*\|_2$*. Then algorithm complexity is*

$$O\left(\left(\frac{L_p R^{p-1}}{\mu}\right)^{\frac{2}{p+1}} \log \frac{\mu R^2}{\varepsilon_G}\right). \tag{2}$$

Restarted HighOrderMirrorProx + CRN-SPP from Huang, J. Zhang, and S. Zhang 2020

Theorem 2.
Suppose $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$, *that is defined in* (1), *is* μ*-strongly monotone, first, second and p-th order Lipschitz operator (all assumptions 1, 2, 3, 4 hold). Denote* $R : R \geq \|z_1 - z^*\|_2$ *and* $\xi = \max\left\{1, \frac{L_1}{\mu}\right\}$. *Then the complexity of algorithm is*

$$O\left(\left(\frac{L_p R^{p-1}}{\mu}\right)^{\frac{2}{p+1}} \log \frac{L_2 \xi R}{\mu} + \log\log \frac{1}{\varepsilon_G}\right). \qquad (3)$$

Summary

Name	Order	Assumptions	Complexity
Nemirovski 2004	1	L_1	$O(L_1/\varepsilon_G)$
Gasnikov et al. 2020	1	μ, L_1	$O\left(L_1/\mu \cdot \log(1/\varepsilon_G)\right)$
Bullins and Lai 2020	p	L_p	$O(L_p/\varepsilon_G^{\frac{2}{p+1}})$
Theorem 1	p	μ, L_p	$O\left((L_p/\mu)^{\frac{2}{p+1}} \log(1/\varepsilon_G)\right)$
Theorem 2	p	μ, L_1, L_2, L_p	$O\left((L_p/\mu)^{\frac{2}{p+1}} + \log\log(1/\varepsilon_G)\right)$

Table 2: Complexity results

Restarted HighOrderMirrorProx + CRN-SPP + framework from Dvurechensky et al. 2019

Theorem 3.

Assume the function $g(x, y) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *is convex-concave,* p *times differentiable on* $\mathbb{R}^n$ *under assumptions 2 – 4. Let* $\tilde{z}$ *be generated by proposed algorithm . Then*

$$\|\nabla g(\tilde{z})\|_2 \leq \varepsilon_\nabla,$$

and the total complexity is

$$O\left(\left(\frac{L_p R^p}{\varepsilon_\nabla}\right)^{\frac{2}{p+1}} \log \frac{L_2 R^2 \xi}{\varepsilon_\nabla}\right), \tag{4}$$

where $\xi = \max\left\{1, \frac{4RL_1}{\varepsilon_\nabla}\right\}$.

Forthcoming research

- Lower bounds for tensor methods for (strongly) convex (strongly) concave SPP.
- Hölder conditions, uniformly convex case.
- In Bullins and Lai 2020 the authors provided implementation details of the HighOrderMirrorProx only for $p = 2$.

References I

[1] Brian Bullins and Kevin A Lai. "Higher-order methods for convex-concave min-max optimization and monotone variational inequalities". In: *arXiv preprint arXiv:2007.04528* (2020).

[2] Pavel Dvurechensky et al. "Near-optimal tensor methods for minimizing the gradient norm of convex function". In: *arXiv preprint arXiv:1912.03381* (2019).

[3] Alexander Gasnikov et al. "Accelerated meta-algorithm for convex optimization". In: *Computational Mathematics and Mathematical Physics* 61.1 (2020).

[4] Kevin Huang, Junyu Zhang, and Shuzhong Zhang. "Cubic Regularized Newton Method for Saddle Point Models: a Global and Local Convergence Analysis". In: *arXiv preprint arXiv:2008.09919* (2020).

References II

[5] Arkadi Nemirovski. "Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems". In: *SIAM Journal on Optimization* 15.1 (2004), pp. 229–251.