Convergence rate analysis of the gradient descent-ascent method for convex-concave saddle-point problems

Hadi Abbaszadehpeivasti (joint work with Etienne de Klerk and Moslem Zamani)

February 8, 2023

Tilburg University

Convex-concave saddle-point problems

Saddle point problem (a.k.a. minimax problem)

$$\min_{x\in\mathbb{R}^n} \max_{y\in\mathbb{R}^m} F(x, y),$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$.

Saddle point problem (a.k.a. minimax problem)

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} F(x, y),$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$.

We assume that problem has some solution, that is, there exists $(x^\star, y^\star) \in \mathbb{R}^n \times \mathbb{R}^m$ with

$$F(x^\star, y) \leq F(x^\star, y^\star) \leq F(x, y^\star), \quad \forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^m.$$

Saddle point problem (a.k.a. minimax problem)

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} F(x, y),$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$.

We assume that problem has some solution, that is, there exists $(x^\star, y^\star) \in \mathbb{R}^n \times \mathbb{R}^m$ with

$$F(x^\star, y) \leq F(x^\star, y^\star) \leq F(x, y^\star), \quad \forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^m.$$

and for $\mu_x, \mu_y \geq 0$

$$\begin{array}{ll} i) & F(\cdot, y) - \frac{\mu_x}{2}\|\cdot\|^2 \text{ is convex for any fixed } y \\ ii) & F(x, \cdot) + \frac{\mu_y}{2}\|\cdot\|^2 \text{ is concave for any fixed } x. \end{array}$$

Some assumptions

For some $L_x, L_{xy}, L_y > 0$,

$$
\begin{array}{lll}
i) & \|\nabla_x F(x_2, y) - \nabla_x F(x_1, y)\| \leq L_x \|x_2 - x_1\| & \forall x_1, x_2, y \\
ii) & \|\nabla_y F(x, y_2) - \nabla_y F(x, y_1)\| \leq L_y \|y_2 - y_1\| & \forall x, y_1, y_2 \\
iii) & \|\nabla_x F(x, y_2) - \nabla_x F(x, y_1)\| \leq L_{xy} \|y_2 - y_1\| & \forall x, y_1, y_2 \\
iv) & \|\nabla_y F(x_2, y) - \nabla_y F(x_1, y)\| \leq L_{xy} \|x_2 - x_1\| & \forall x_1, x_2, y \\
v) & (x^\star, y^\star) \text{ denotes the unique saddle point.} &
\end{array}
$$

The gradient descent-ascent method

Algorithm 1 The gradient descent-ascent method (GDA)

Pick $\mathbf{x}^1 \in \mathbb{R}^n$, $\mathbf{y}^1 \in \mathbb{R}^m$ and $N \in \mathbb{N}$.

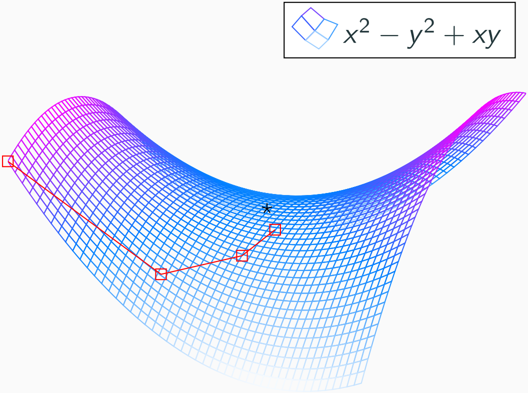
For $k = 1, 2, \ldots, N$ perform the following steps:

1. $\mathbf{x}^{k+1} = \mathbf{x}^k - t\nabla_{\mathbf{x}} F(\mathbf{x}^k, \mathbf{y}^k)$,
2. $\mathbf{y}^{k+1} = \mathbf{y}^k + t\nabla_{\mathbf{y}} F(\mathbf{x}^k, \mathbf{y}^k)$.

Arrow, K. J., Azawa, H., Hurwicz, L., Uzawa, H., Chenery, H. B., Johnson, S. M., & Karlin, S. (1958). Studies in linear and non-linear programming (Vol. 2). Stanford University Press.

Example

Example for gradient descent ascent method



Known results

Theorem

Let $L = \max\{L_x, L_{xy}, L_y\}$ *and* $\mu = \min\{\mu_x, \mu_y\}$*. If* $t \in \left(0, \frac{\mu}{2L^2}\right)$*, then*

$$\|x^2 - x^\star\|^2 + \|y^2 - y^\star\|^2 \leq (1 + 4L^2t^2 - 2\mu t)\left(\|x^1 - x^\star\|^2 + \|y^1 - y^\star\|^2\right).$$

Known results

Theorem

Let $L = \max\{L_x, L_{xy}, L_y\}$ *and* $\mu = \min\{\mu_x, \mu_y\}$. *If* $t \in \left(0, \frac{\mu}{2L^2}\right)$, *then*

$$\|x^2 - x^\star\|^2 + \|y^2 - y^\star\|^2 \leq (1 + 4L^2 t^2 - 2\mu t)\left(\|x^1 - x^\star\|^2 + \|y^1 - y^\star\|^2\right).$$

- By setting $t = \frac{\mu}{4L^2}$, one can infer that the gradient descent-ascent method has a complexity of $\mathcal{O}\left(\frac{L^2}{\mu^2}\ln\left(\frac{1}{\epsilon}\right)\right)$.

See

Beznosikov, A., Polyak, B., Gorbunov, E., Kovalev, D., & Gasnikov, A. (2022). Smooth Monotone Stochastic Variational Inequalities and Saddle Point Problems–Survey. arXiv preprint arXiv:2208.13592.

Performance Estimation (PEP)

Performance Estimation Problem

$$
\begin{aligned}
\max\ & \frac{\|x^2 - x^\star\|^2 + \|y^2 - y^\star\|^2}{\|x^1 - x^\star\|^2 + \|y^1 - y^\star\|^2} \\
\text{s.t. } & (x^2, y^2) \text{ is generated by GDA w.r.t. } F, x^1, y^1, t \\
& (x^\star, y^\star) \text{ is the unique saddle point of minimax problem} \\
& F \in \mathcal{F}(L_x, L_y, L_{xy}, \mu_x, \mu_y) \\
& x^1 \in \mathbb{R}^n, y^1 \in \mathbb{R}^m.
\end{aligned}
$$

Performance Estimation Problem

$$\max \frac{\|x^2 - x^\star\|^2 + \|y^2 - y^\star\|^2}{\|x^1 - x^\star\|^2 + \|y^1 - y^\star\|^2}$$

$$\text{s. t. } (x^2, y^2) \text{ is generated by GDA w.r.t. } F, x^1, y^1, t$$

$$(x^\star, y^\star) \text{ is the unique saddle point of minimax problem}$$

$$F \in \mathcal{F}(L_x, L_y, L_{xy}, \mu_x, \mu_y)$$

$$x^1 \in \mathbb{R}^n, y^1 \in \mathbb{R}^m.$$

- Decision variables: $F, x^1, x^2, x^\star, y^1, y^2, y^\star$.

Performance Estimation Problem

$$
\begin{aligned}
\max\ & \frac{\|x^2 - x^\star\|^2 + \|y^2 - y^\star\|^2}{\|x^1 - x^\star\|^2 + \|y^1 - y^\star\|^2} \\
\text{s. t. } & (x^2, y^2) \text{ is generated by GDA w.r.t. } F, x^1, y^1, t \\
& (x^\star, y^\star) \text{ is the unique saddle point of minimax problem} \\
& F \in \mathcal{F}(L_x, L_y, L_{xy}, \mu_x, \mu_y) \\
& x^1 \in \mathbb{R}^n, y^1 \in \mathbb{R}^m.
\end{aligned}
$$

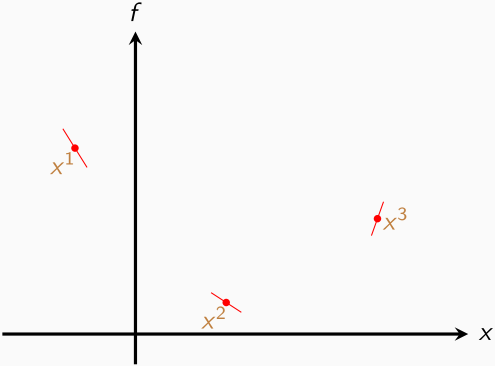
- Decision variables: $F, x^1, x^2, x^\star, y^1, y^2, y^\star$.
- Fixed parameters: $L_x, L_y, L_{xy}, \mu_x, \mu_y, t$

L-smooth and μ-strongly Convex Interpolation Problem

Consider a finite index set I, and given triple $\left\{(\mathbf{x}^k, \mathbf{g}^k, f^k)\right\}_{k \in I}$ where $\mathbf{x}^k \in \mathbb{R}^n$, $\mathbf{g}^k \in \mathbb{R}^n$ and $f^k \in \mathbb{R}$.

L-smooth and μ-strongly Convex Interpolation Problem

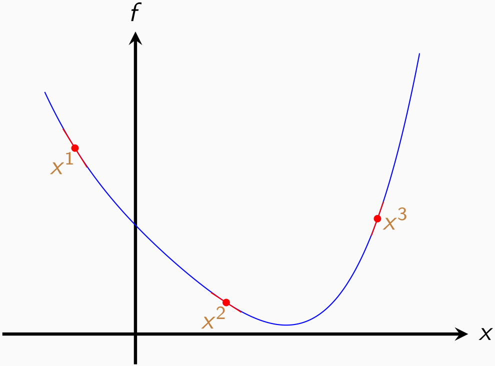
Consider a finite index set I, and given triple $\left\{(\mathbf{x}^k, \mathbf{g}^k, f^k)\right\}_{k\in I}$ where $\mathbf{x}^k \in \mathbb{R}^n$, $\mathbf{g}^k \in \mathbb{R}^n$ and $f^k \in \mathbb{R}$.



$$?\exists\ f \in \mathcal{F}_{\mu,L}(\mathbb{R}^n):\ f(\mathbf{x}^k) = f^k, \quad \text{and} \quad \mathbf{g}^k \in \partial f(\mathbf{x}^k), \qquad \forall k \in I.$$

L-smooth and μ-strongly Convex Interpolation Problem

Consider a finite index set I, and given triple $\left\{(\mathbf{x}^k, \mathbf{g}^k, f^k)\right\}_{k\in I}$ where $\mathbf{x}^k \in \mathbb{R}^n$, $\mathbf{g}^k \in \mathbb{R}^n$ and $f^k \in \mathbb{R}$.



$$?\exists \;\; f \in \mathcal{F}_{\mu,L}(\mathbb{R}^n) \colon \; f(\mathbf{x}^k) = f^k, \quad \text{and} \quad \mathbf{g}^k \in \partial f(\mathbf{x}^k), \qquad \forall k \in I.$$

If yes, we say $\left\{(\mathbf{x}^k, \mathbf{g}^k, f^k)\right\}_{k\in I}$ is $\mathcal{F}_{\mu,L}(\mathbb{R}^n)$-interpolable.

L-smooth and μ-strongly Interpolation

Theorem (Taylor, Hendrickx, and Glineur (2017))

The following statements are equivalent:

1. $\left\{(\mathbf{x}^i, \mathbf{g}^i, f^i)\right\}_{i\in I}$ *is* $\mathcal{F}_{\mu,L}(\mathbb{R}^n)$*-interpolable;*
2. $\forall i, j \in I$*:*

$$\frac{1}{2(1-\frac{\mu}{L})}\left(\frac{1}{L}\left\|g^i - g^j\right\|^2 + \mu\left\|x^i - x^j\right\|^2 - \frac{2\mu}{L}\left\langle g^j - g^i, x^j - x^i\right\rangle\right) \leq f^i - f^j - \left\langle g^j, x^i - x^j\right\rangle.$$

A.B. Taylor, J.M. Hendrickx, and F. Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming* 161.1-2, 307–345 (2017)

L-smooth and μ-strongly Interpolation

Theorem (Taylor, Hendrickx, and Glineur (2017))

The following statements are equivalent:

1. $\left\{(\mathbf{x}^i, \mathbf{g}^i, f^i)\right\}_{i\in I}$ *is* $\mathcal{F}_{\mu,L}(\mathbb{R}^n)$*-interpolable;*
2. $\forall i, j \in I$:

$$\frac{1}{2(1-\frac{\mu}{L})}\left(\frac{1}{L}\left\|g^i - g^j\right\|^2 + \mu\left\|x^i - x^j\right\|^2 - \frac{2\mu}{L}\left\langle g^j - g^i, x^j - x^i\right\rangle\right) \leq$$
$$f^i - f^j - \left\langle g^j, x^i - x^j\right\rangle.$$

A.B. Taylor, J.M. Hendrickx, and F. Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming* 161.1-2, 307–345 (2017)

Performance estimation formulation: Assumptions

Consider variables

$$
\begin{aligned}
F^{i,j} &= F(x^i, y^j) && i,j \in \{1,2,\star\},\\
G_x^{i,j} &= \nabla_x F(x^i, y^j) && i,j \in \{1,2,\star\},\\
G_y^{i,j} &= \nabla_y F(x^i, y^j) && i,j \in \{1,2,\star\},
\end{aligned}
$$

Performance estimation formulation: Assumptions

Consider variables

$$\begin{aligned} F^{i,j} &= F(x^i, y^j) && i,j \in \{1,2,\star\}, \\ G_x^{i,j} &= \nabla_x F(x^i, y^j) && i,j \in \{1,2,\star\}, \\ G_y^{i,j} &= \nabla_y F(x^i, y^j) && i,j \in \{1,2,\star\}, \end{aligned}$$

and the necessary and sufficient conditions for convex-concave saddle point problems

$$G_x^{\star,\star} = 0, \qquad G_y^{\star,\star} = 0,$$

Performance estimation formulation

$$
\begin{aligned}
\max\ & \frac{\|x^2 - x^\star\|^2 + \|y^2 - y^\star\|^2}{\|x^1 - x^\star\|^2 + \|y^1 - y^\star\|^2} \\
\text{s.t. } & \{(x^1; G_x^{1,k}; F^{1,k}), (x^2; G_x^{2,k}; F^{2,k}), (x^\star; G_x^{\star,k}; F^{\star,k})\} \text{ satisfy} \\
& \qquad \text{interpolation constraint for } k \in \{1,2,\star\} \text{ w.r.t. } \mu_x, L_x \\
& \{(y^1; G_y^{k,1}; F^{k,1}), (y^2; G_y^{k,2}; F^{k,2}), (y^\star; G_y^{k,\star}; F^{k,\star})\} \text{ satisfy} \\
& \qquad \text{interpolation constraint for } k \in \{1,2,\star\} \text{ w.r.t. } \mu_y, L_y \\
& \|G_x^{k,i} - G_x^{k,j}\|^2 \leq L_{xy}\|y^i - y^j\|^2, \quad i,j,k \in \{1,2,\star\} \\
& \|G_y^{i,k} - G_y^{j,k}\|^2 \leq L_{xy}\|x^i - x^j\|^2, \quad i,j,k \in \{1,2,\star\} \\
& x^2 = x^1 - t G_x^{1,1} \\
& y^2 = y^1 + t G_y^{1,1}, \\
& G_x^{\star,\star} = 0,\ G_y^{\star,\star} = 0.
\end{aligned}
$$

Performance estimation formulation

$$
\begin{aligned}
\max\ & \frac{\|x^1 - t G_x^{1,1}\|^2 + \|y^1 + t G_y^{1,1}\|^2}{\|x^1\|^2 + \|y^1\|^2} \\
\text{s. t. } & \{(x^1; G_x^{1,k}; F^{1,k}), (x^1 - t G_x^{1,1}; G_x^{2,k}; F^{2,k}), (0; G_x^{\star,k}; F^{\star,k})\} \text{ satisfy} \\
& \qquad \text{interpolation constraint for } k \in \{1, 2, \star\} \text{ w.r.t. } \mu_x, L_x \\
& \{(y^1; G_y^{k,1}; F^{k,1}), (y^1 + t G_y^{1,1}; G_y^{k,2}; F^{k,2}), (0; G_y^{k,\star}; F^{k,\star})\} \text{ satisfy} \\
& \qquad \text{interpolation constraint for } k \in \{1, 2, \star\} \text{ w.r.t. } \mu_y, L_y \\
& \|G_x^{k,i} - G_x^{k,j}\|^2 \le L_{xy} \|y^i - y^j\|^2, \quad i, j, k \in \{1, 2, \star\} \\
& \|G_y^{i,k} - G_y^{j,k}\|^2 \le L_{xy} \|x^i - x^j\|^2, \quad i, j, k \in \{1, 2, \star\} \\
& G_x^{\star,\star} = 0,\ G_y^{\star,\star} = 0.
\end{aligned}
$$

Performance estimation formulation

$$
\begin{aligned}
\max\ & \frac{\|x^1 - tG_x^{1,1}\|^2 + \|y^1 + tG_y^{1,1}\|^2}{\|x^1\|^2 + \|y^1\|^2} \\
\text{s. t. } & \{(x^1; G_x^{1,k}; F^{1,k}), (x^1 - tG_x^{1,1}; G_x^{2,k}; F^{2,k}), (0; G_x^{\star,k}; F^{\star,k})\} \text{ satisfy} \\
& \qquad \text{interpolation constraint for } k \in \{1, 2, \star\} \text{ w.r.t. } \mu_x, L_x \\
& \{(y^1; G_y^{k,1}; F^{k,1}), (y^1 + tG_y^{1,1}; G_y^{k,2}; F^{k,2}), (0; G_y^{k,\star}; F^{k,\star})\} \text{ satisfy} \\
& \qquad \text{interpolation constraint for } k \in \{1, 2, \star\} \text{ w.r.t. } \mu_y, L_y \\
& \|G_x^{k,i} - G_x^{k,j}\|^2 \le L_{xy}\|y^i - y^j\|^2, \quad i, j, k \in \{1, 2, \star\} \\
& \|G_y^{i,k} - G_y^{j,k}\|^2 \le L_{xy}\|x^i - x^j\|^2, \quad i, j, k \in \{1, 2, \star\} \\
& G_x^{\star,\star} = 0,\ G_y^{\star,\star} = 0.
\end{aligned}
$$

- Decision variables: $F^{i,j}, x^i, y^j, G_x^{i,j}, G_y^{i,j}$ for $i, j \in \{1, 2, \star\}$.

Performance estimation formulation

$$
\begin{aligned}
\max\ & \frac{\|x^1 - tG_x^{1,1}\|^2 + \|y^1 + tG_y^{1,1}\|^2}{\|x^1\|^2 + \|y^1\|^2} \\
\text{s. t. } & \{(x^1; G_x^{1,k}; F^{1,k}), (x^1 - tG_x^{1,1}; G_x^{2,k}; F^{2,k}), (0; G_x^{\star,k}; F^{\star,k})\} \text{ satisfy} \\
& \qquad \text{interpolation constraint for } k \in \{1, 2, \star\} \text{ w.r.t. } \mu_x, L_x \\
& \{(y^1; G_y^{k,1}; F^{k,1}), (y^1 + tG_y^{1,1}; G_y^{k,2}; F^{k,2}), (0; G_y^{k,\star}; F^{k,\star})\} \text{ satisfy} \\
& \qquad \text{interpolation constraint for } k \in \{1, 2, \star\} \text{ w.r.t. } \mu_y, L_y \\
& \|G_x^{k,i} - G_x^{k,j}\|^2 \leq L_{xy} \|y^i - y^j\|^2, \quad i, j, k \in \{1, 2, \star\} \\
& \|G_y^{i,k} - G_y^{j,k}\|^2 \leq L_{xy} \|x^i - x^j\|^2, \quad i, j, k \in \{1, 2, \star\} \\
& G_x^{\star,\star} = 0, \ G_y^{\star,\star} = 0.
\end{aligned}
$$

- Decision variables: $F^{i,j}, x^i, y^j, G_x^{i,j}, G_y^{i,j}$ for $i, j \in \{1, 2, \star\}$.
- Fixed parameters: $L_x, L_y, L_{xy}, \mu_x, \mu_y, t$

Convergence rate for strongly convex-strongly concave saddle point problem

Theorem (Zamani, Abbaszadehpeivasti, De Klerk)

Suppose that $L = \max\{L_x, L_y\}$ *and* $\mu = \min\{\mu_x, \mu_y\} > 0$. *If* $t \in \left(0, \frac{2\mu}{\mu L + L_{xy}^2}\right)$, *then*

$$\|x^2 - x^\star\|^2 + \|y^2 - y^\star\|^2 \le \alpha \left(\|x^1 - x^\star\|^2 + \|y^1 - y^\star\|^2\right),$$

where

$$\alpha = 1 + \tfrac{1}{2}\left(L^2 + \mu^2 + 2L_{xy}^2\right)t^2 - (L+\mu)t + \tfrac{1}{2}(L-\mu)t\sqrt{(Lt + \mu t - 2)^2 + 4L_{xy}^2 t^2}.$$

Convergence rate for strongly convex-strongly concave saddle point problem

Theorem (Zamani, Abbaszadehpeivasti, De Klerk)

Suppose that $L = \max\{L_x, L_y\}$ *and* $\mu = \min\{\mu_x, \mu_y\} > 0$. *If* $t \in \left(0, \frac{2\mu}{\mu L + L_{xy}^2}\right)$, *then*

$$\|x^2 - x^\star\|^2 + \|y^2 - y^\star\|^2 \leq \alpha \left(\|x^1 - x^\star\|^2 + \|y^1 - y^\star\|^2\right),$$

where

$$\alpha = 1 + \tfrac{1}{2}\left(L^2 + \mu^2 + 2L_{xy}^2\right)t^2 - (L+\mu)t + \tfrac{1}{2}(L-\mu)t\sqrt{(Lt + \mu t - 2)^2 + 4L_{xy}^2 t^2}.$$

- By setting $t = \frac{2\left((L+\mu)\sqrt{L_{xy}^2 + L\mu} + L_{xy}(\mu - L)\right)}{\left(4L_{xy}^2 + (L+\mu)^2\right)\sqrt{L_{xy}^2 + L\mu}}$, we can infer that the gradient descent-ascent method has a complexity of $O\left(\left(\frac{L}{\mu} + \frac{L_{xy}^2}{\mu^2}\right)\ln\left(\frac{1}{\epsilon}\right)\right)$.

Convergence rate for strongly convex-strongly concave saddle point problem

Theorem (Zamani, Abbaszadehpeivasti, De Klerk)

Suppose that $L = \max\{L_x, L_y\}$ *and* $\mu = \min\{\mu_x, \mu_y\} > 0$*. If* $t \in \left(0, \frac{2\mu}{\mu L + L_{xy}^2}\right)$*, then*

$$\|x^2 - x^\star\|^2 + \|y^2 - y^\star\|^2 \le \alpha \left(\|x^1 - x^\star\|^2 + \|y^1 - y^\star\|^2\right),$$

where

$$\alpha = 1 + \tfrac{1}{2}\left(L^2 + \mu^2 + 2L_{xy}^2\right)t^2 - (L+\mu)t + \tfrac{1}{2}(L-\mu)t\sqrt{(Lt + \mu t - 2)^2 + 4L_{xy}^2 t^2}.$$

- By setting $t = \frac{2\left((L+\mu)\sqrt{L_{xy}^2 + L\mu} + L_{xy}(\mu - L)\right)}{\left(4L_{xy}^2 + (L+\mu)^2\right)\sqrt{L_{xy}^2 + L\mu}}$, we can infer that the gradient descent-ascent method has a complexity of $O\left(\left(\frac{L}{\mu} + \frac{L_{xy}^2}{\mu^2}\right)\ln\left(\frac{1}{\epsilon}\right)\right)$.
- If $L_{xy} = 0$, we get $O\left(\frac{L}{\mu}\ln\left(\frac{1}{\epsilon}\right)\right)$ for the gradient descent method.

Tightness of the bound

Proposition

Let $F \in \mathcal{F}(L_x, L_y, L_{xy}, \mu_x, \mu_y)$*. Suppose that* $L_x = L_y$ *and* $\min\{\mu_x, \mu_y\} > 0$*. If* $t \in \left(0, \frac{2\mu}{\mu L + L_{xy}^2}\right)$*, then the given convergence rate is exact for one iteration.*

Tightness of the bound

Proposition

Let $F \in \mathcal{F}(L_x, L_y, L_{xy}, \mu_x, \mu_y)$. Suppose that $L_x = L_y$ and $\min\{\mu_x, \mu_y\} > 0$. If $t \in \left(0, \frac{2\mu}{\mu L + L_{xy}^2}\right)$, then the given convergence rate is exact for one iteration.

$$\min_{x \in \mathbb{R}^2} \max_{y \in \mathbb{R}^2} \frac{1}{2} x^T \begin{pmatrix} L_x & 0 \\ 0 & \mu_x \end{pmatrix} x + x^T \begin{pmatrix} 0 & L_{xy} \\ L_{xy} & 0 \end{pmatrix} y - \frac{1}{2} y^T \begin{pmatrix} L_y & 0 \\ 0 & \mu_y \end{pmatrix} y,$$

By $L_{xy} = 1$, $L = L_x$, $\mu = \mu_y \leq \mu_x$ and $\beta = \sqrt{(Lt + \mu t - 2)^2 + 4t^2}$, performing the algorithm for the following (x^1, y^1)

$$x_1^1 = 0, \qquad x_2^1 = \sqrt{\frac{2 - t(L+\mu) + \beta}{2\beta}},$$

$$y_1^1 = -t\sqrt{\frac{2}{\beta(2 - t(L+\mu) + \beta)}}, \qquad y_2^1 = 0,$$

generates (x^2, y^2) with the desired equality.

Linear convergence without strong convexity

Definition

Let $\mu_F > 0$. A function F has a quadratic gradient growth if for any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$,

$$\langle \nabla_x F(x,y), x - x^\star \rangle - \langle \nabla_y F(x,y), y - y^\star \rangle \geq \mu_F d^2_{S^\star}((x,y)),$$

where $(x^\star, y^\star) = \Pi_{S^\star}((x,y))$,

Preliminaries

Definition

Let $\mu_F > 0$. A function F has a quadratic gradient growth if for any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$,

$$\langle \nabla_x F(x,y), x - x^\star \rangle - \langle \nabla_y F(x,y), y - y^\star \rangle \geq \mu_F d^2_{S^\star}((x,y)),$$

where $(x^\star, y^\star) = \Pi_{S^\star}((x,y))$,

where

- $S^\star$ denotes nonempty solution set of the saddle point problem.
- for $X \subseteq \mathbb{R}^n$, $d_X(x) := \inf_{\bar{x}\in X} \|x - \bar{x}\|$ denotes the distance function to X
- $\Pi_X(x) := \{y \in X : \|x - y\| = d_X(x)\}$ stands for the projection of x on X.

Necessary and sufficient conditions for linear convergence

Theorem

Let $F \in \mathcal{F}(L_x, L_y, L_{xy}, 0, 0)$ and $L = \max\{L_x, L_y\}$. Assume that F has a quadratic gradient growth with $\mu_F > 0$. If $t \in \left(0, \frac{2\mu_F}{L\mu_F + 2L_{xy}\sqrt{\mu_F(L-\mu_F)} + L_{xy}^2}\right)$, then GDA generates (x^2, y^2) such that

$$d_{S^\star}^2((x^2, y^2)) \leq \alpha d_{S^\star}^2((x^1, y^1)),$$

where

$$\alpha = t\left(2tL_{xy}\sqrt{\mu_F(L-\mu_F)} + \mu_F(Lt-2) + tL_{xy}^2\right) + 1.$$

Necessary and sufficient conditions for linear convergence

Theorem

Let $F \in \mathcal{F}(L_x, L_y, L_{xy}, 0, 0)$ *and* $L = \max\{L_x, L_y\}$. *Assume that* F *has a quadratic gradient growth with* $\mu_F > 0$. *If* $t \in \left(0, \frac{2\mu_F}{L\mu_F + 2L_{xy}\sqrt{\mu_F(L-\mu_F)} + L_{xy}^2}\right)$, *then GDA generates* (x^2, y^2) *such that*

$$d_{S^\star}^2((x^2, y^2)) \leq \alpha d_{S^\star}^2((x^1, y^1)),$$

where

$$\alpha = t\left(2tL_{xy}\sqrt{\mu_F(L-\mu_F)} + \mu_F(Lt - 2) + tL_{xy}^2\right) + 1.$$

If GDA is linearly convergent for any initial point, then F *has a quadratic gradient growth for some* $\mu_F > 0$.

Conclusion

Future work

- Considering the case where the variables x and y in the saddle point problem are constrained to lie in given, compact convex sets.
- Note the interpolation theorem remains open for minimax objective function.

Future work

- Considering the case where the variables x and y in the saddle point problem are constrained to lie in given, compact convex sets.
- Note the interpolation theorem remains open for minimax objective function.

M. Zamani, H. Abbaszadehpeivasti, E. de Klerk. Convergence rate analysis of the gradient descent-ascent method for convex-concave saddle-point problems. *arXiv preprint arXiv:2209.01272*

The End