Russian Journal of Nonlinear Dynamics, 2024, vol. 20, no. 5, pp. 875–893.
Full-texts are available at http://nd.ics.org.ru
DOI: 10.20537/nd241215

NONLINEAR SYSTEMS IN ROBOTICS

MSC 2010: 65K15

Convex-Concave Interpolation and Application of PEP to the Bilinear-Coupled Saddle Point Problem

V. O. Krivchenko, A. V. Gasnikov, D. A. Kovalev

In this paper we present interpolation conditions for several important convex-concave function classes: nonsmooth convex-concave functions, conditions for difference of strongly-convex functions in a form that contains oracle information exclusively and smooth convex-concave functions with a bilinear coupling term. Then we demonstrate how the performance estimation problem approach can be adapted to analyze the exact worst-case convergence behavior of first-order methods applied to composite bilinear-coupled min-max problems. Using the performance estimation problem approach, we estimate iteration complexities for several first-order fixed-step methods, Sim-GDA and Alt-GDA, which are applied to smooth convex-concave functions with a bilinear coupling term.

Keywords: saddle point, convex-concave functions, bilinear coupling, performance estimation problem, interpolation conditions

Received October 31, 2024
Accepted December 09, 2024

The research is supported by the Ministry of Science and Higher Education of the Russian Federation (Goszadaniye), project No. FSMG-2024-0011.

Valery O. Krivchenko
krivchenko.vo@phystech.edu
Moscow Institute of Physics ans Technology
Institutskiy per. 9, Dolgoprudny, 141701 Russia

Alexander V. Gasnikov
gasnikov@yandex.ru
Innopolis University
ul. Universitetskaya 1, Innopolis, 420500 Russia
Moscow Institute of Physics ans Technology
Institutskiy per. 9, Dolgoprudny, 141701 Russia
Steklov Mathematical Institute of Russian Academy of Sciences
ul. Gubkina 8, Moscow, 117966 Russia

Dmitry A. Kovalev
dakovalev1@gmail.com
Yandex Research
ul. L'va Tolstogo 16, Moscow, 119021 Russia

1. Introduction

In this paper we consider the following min-max optimization problem:

$$\min_{x\in\mathbb{R}^{d_x}}\max_{y\in\mathbb{R}^{d_y}} f(x, y), \tag{1.1}$$

where $\mathbb{R}^{d_x}$ and $\mathbb{R}^{d_y}$ are Euclidean spaces and $f(x, y)\colon \mathbb{R}^{d_x}\times\mathbb{R}^{d_y}\to\mathbb{R}$ is a real-valued function.

Problem (1.1) is widely studied in optimization and arises not only in mathematics, but also in economics, computer science and machine learning [1, 2]. Because min-max problems arising in practice tend to be high-dimensional, first-order algorithms become a go-to way to solve them in no small part due to low cost of a single iteration. Consequently, there is a need to understand how such methods would perform in the worst possible circumstances of the defined setting and compare their worst-case performance.

Smooth min-max games are currently of great interest for research because the exact worst-case behavior of even the most basic methods remains mostly unknown [3, 4]. Even the asymptotic dependencies of methods' global convergence behavior on smoothness parameters remain a subject of active research [5].

Several SDP-based approaches have been developed for analysis of exact worst-case performance of first-order methods: performance estimation problems (PEP) [6], integral quadratic constraints (IQC) [7] and an approach of automatic generation of Lyapunov functions to verify linear convergence [8] which is roughly a synthesis of the previous two. In this paper we showcase PEP application exclusively since it allows numerical acquisition of exact worst-case linear convergence rates in the most quick and practical manner.

The idea of PEP is to express the exact worst-case performance of an optimization algorithm as the solution of a tractable optimization problem. However, the general formulation of PEP is infinite-dimensional because the performance metric is being optimized over an entire class of functions, which leads to it seemingly being intractable. The obstacle can be overcome by replacing optimization over functions with optimization over all possible oracle outputs for a method applied to a class of functions. In this way the problem becomes finite-dimensional and, in many cases, convex in the form of an SDP. In order for this replacement to be equivalent, one needs to put such constraints on the gradients and functional values that would hold if and only if the gradients and functional values can belong to a function from the class [9]. Such constraints are called interpolation conditions and are a key to computer-assisted derivation of exact worst-case guarantees with PEP approach.

After initial success of PEP in analyzing first-order algorithms applied to smooth convex/strongly-convex(SC) functions [10] and computer-assisted derivation of several optimal methods [11], the attempts have been made to transfer the approach to saddle-point problems [12]. The success of those attempts has remained mostly limited due to the lack of interpolation conditions for smooth convex-concave classes.

Although we do not derive interpolation conditions for the smooth strongly-convex-strongly-concave (SCSC) class in this paper, we still try to partially bridge the existing gap by analyzing special cases of general smooth SCSC class. We also derive conditions for M-Lipshitz convex-concave functions that can be used for numerical analysis of subgradient methods via PEP. In later sections we construct a PEP for first-order fixed-step methods that are applied to a fairly descriptive class of convex-concave functions with bilinear coupling. The PEP setup presented here is reasonably generalized and can be quickly adapted for different first-order fixed-step methods and quadratic performance metrics.

Our contribution is two-fold:

- We obtain interpolation conditions for several saddle function classes: nonsmooth convex-concave functions, conditions for difference of strongly-convex functions in a form containing oracle information exclusively and convex-concave functions with a bilinear coupling term. By that we also hope to set the stage for acquiring interpolation conditions for the general smooth SCSC class which is currently of great interest for min-max optimization research.
- We construct PEP for first-order fixed-step methods that are applied to the bilinear-coupled saddle point problem. Then we solve (dual) PEP to compare iteration complexities of Sim-GDA and Alt-GDA.

2. Interpolation conditions

2.1. Basic definitions

Let us start with definitions. First we define convex-concave functions.

Definition 1. *We say that a function $f(x, y)\colon \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}$ is convex-concave if*

$$f(\cdot, y) \text{ is convex for all } y \in \mathbb{R}^{d_y},$$
$$f(x, \cdot) \text{ is concave for all } x \in \mathbb{R}^{d_x}.$$

We denote this class by S.

Next, we define more frequently studied, smooth, possibly SCSC function class:

Definition 2. *For given constants $(\mu_x, \mu_y, L_x, L_y, L_{xy})$ that satisfy $0 \leqslant \mu_x \leqslant L_x$, $0 \leqslant \mu_y \leqslant L_y$, $L_{xy} \geqslant 0$, we say that a differentiable function $f(x, y)\colon \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}$ is (μ_x, μ_y)-strong-convex-strong-concave (SCSC) and has (L_x, L_y, L_{xy})-Lipshitz gradients if*

$$f(\cdot, y)\colon \mu_x \text{ is strongly convex for all } y \in \mathbb{R}^{d_y},$$
$$f(x, \cdot)\colon \mu_y \text{ is strongly concave for all } x \in \mathbb{R}^{d_x},$$
$$||\nabla_x f(x_1, y) - \nabla_x f(x_0, y)|| \leqslant L_x ||x_1 - x_0|| \quad \text{for all } x_0, x_1 \in \mathbb{R}^{d_x},\ y \in \mathbb{R}^{d_y},$$
$$||\nabla_y f(x, y_1) - \nabla_y f(x, y_0)|| \leqslant L_y ||y_1 - y_0|| \quad \text{for all } y_0, y_1 \in \mathbb{R}^{d_y},\ x \in \mathbb{R}^{d_x},$$
$$||\nabla_x f(x, y_1) - \nabla_x f(x, y_0)|| \leqslant L_{xy} ||y_1 - y_0|| \quad \text{for all } y_0, y_1 \in \mathbb{R}^{d_y},\ x \in \mathbb{R}^{d_x},$$
$$||\nabla_y f(x_1, y) - \nabla_y f(x_0, y)|| \leqslant L_{xy} ||x_1 - x_0|| \quad \text{for all } x_0, x_1 \in \mathbb{R}^{d_x},\ y \in \mathbb{R}^{d_y}.$$

We denote this class by $S_{\mu_x \mu_y L_x L_y L_{xy}}$.

As we explained in the introduction to this paper, the PEP approach relies on reformulation of the originally infinite-dimensional problem in a finite-dimensional fashion which requires interpolation conditions. We now formally define the concept of interpolability.

Definition 3. *Let I be an index set and consider a sequence $A = \{(x_i, y_i, g_i^x, g_i^y, f_i)\}_{i\in I}$ where $x_i, g_i^x \in \mathbb{R}^{d_x}$, $y_i, g_i^y \in \mathbb{R}^{d_y}$, $f_i \in \mathbb{R}$ for all $i \in I$. Consider a set of convex-concave functions F. The sequence A is F-interpolable if and only if there exists a function $f \in F$ such that $g_i^x \in \partial_x f(x_i, y_i)$, $g_i^y \in \partial_y f(x_i, y_i)$, $f_i = f(x_i, y_i)$ for all $i \in I$.*

With those definitions, we proceed to derive interpolation conditions for several convex-concave classes starting with the most general convex-concave class.

2.2. Convex-concave functions

Consider a, generally speaking, nonsmooth convex-concave function $f \in S$ that is also M-Lipshitz:

Definition 4. *A function $f \in S$ is M-Lipshitz if*

$$||f(x_1, y_1) - f(x_2, y_2)|| \leqslant M\,(||x_1 - x_2|| + ||y_1 - y_2||) \tag{2.1}$$

for all $x_1, x_2 \in \mathbb{R}^{d_x}$, $y_1, y_2 \in \mathbb{R}^{d_y}$.

We now prove interpolation conditions for M-Lipshitz convex-concave functions (denoted by S_M). Our approach is constructive, which means we provide a way to construct an interpolating function if certain conditions are satisfied:

Theorem 1. *Let I be an index set and consider a vector sequence $\{(x_i, y_i, g_i^x, g_i^y, f_i)\}_{i\in I}$. This sequence is S_M-interpolable if and only if the following conditions are satisfied: for all $i, j \in I$ the following inequality holds:*

$$f_i \geqslant f_j + g_j^{x\,\mathrm{T}}(x_i - x_j) + g_i^{y\,\mathrm{T}}(y_i - y_j), \tag{2.2}$$

for all $i \in I$ the following inequalities hold:

$$||g_i^x|| \leqslant M, \quad ||g_i^y|| \leqslant M. \tag{2.3}$$

Proof. Necessity. Suppose $f \in S_M$. From the definition of the subdifferentials $\partial_x f(x, y)$ and $\partial_y f(x, y)$ it follows that

$$f_i + g_i^{y\,\mathrm{T}}(y_j - y_i) \geqslant f(x_i, y_j) \geqslant f_j + g_j^{x\,\mathrm{T}}(x^i - x^j),$$

which is inequality (2.2). Inequalities (2.3) are a trivial consequence of f being M-Lipshitz.

Sufficiency. Suppose that inequalities (2.2) and (2.3) are satisfied. Consider a function $f(x, y)$ which is defined as follows:

$$f(x, y) = \sup_{z\in\mathscr{Z}} \min_{i\in[n]} f_i + z^{\mathrm{T}}(x - x_i) + g_i^{y\,\mathrm{T}}(y - y_i), \tag{2.4}$$

where the convex set $\mathscr{Z} \subset \mathbb{R}^{d_x}$ is defined as follows:

$$\mathscr{Z} = \operatorname{conv}\left(\bigcup_{i=1}^{n}\{g_i^x\}\right). \tag{2.5}$$

First, it is easy to observe that the function $f(x, y)$ is convex in x. Indeed, for fixed y, the function $f(x, y)$ is a pointwise supremum of functions that are linear in x:

$$f(x, y) = \sup_{z\in\mathscr{Z}} \left[z^{\mathrm{T}}x + \varphi_y(z)\right],$$

where the function $\varphi_y(z)$ is defined as follows:

$$\varphi_y(z) = \min_{i\in[n]} f_i - z^{\mathrm{T}}x_i + g_i^{y\,\mathrm{T}}(y - y_i). \tag{2.6}$$

Next, one can show that the function $-f(x, y)$ is convex in y. Indeed,

$$-f(x,\, y) = \inf_{z\in\mathscr{Z}} \varphi_x(z,\, y), \tag{2.7}$$

where the function $\varphi_x(z,\, y)$ is defined as follows:

$$\varphi_x(z,\, y) = \max_{i\in[n]} -\left[f_i + z^{\mathrm{T}}(x - x_i) + g_i^{y\,\mathrm{T}}(y - y_i)\right]. \tag{2.8}$$

Note, that the function $\varphi_x(z,\, y)$ is convex in $(z,\, y)$, which implies the convexity of the function $-f(x,\, y)$ in y.

Now, we show that $f(x_k,\, y_k) = f_k$. On the one hand, we get

$$\begin{aligned} f(x_k,\, y_k) &= \sup_{z\in\mathscr{Z}} \min_{i\in[n]} f_i + z^{\mathrm{T}}(x_k - x_i) + g_i^{y\,\mathrm{T}}(y_k - y_i) \leqslant \\ &\leqslant \sup_{z\in\mathscr{Z}} f_k + z^{\mathrm{T}}(x_k - x_k) + g_k^{y\,\mathrm{T}}(y_k - y_k) = f_k. \end{aligned}$$

On the other hand, we get

$$\begin{aligned} f(x_k,\, y_k) &= \sup_{z\in\mathscr{Z}} \min_{i\in[n]} f_i + z^{\mathrm{T}}(x_k - x_i) + g_i^{y\,\mathrm{T}}(y_k - y_i) \geqslant \\ &\geqslant \min_{i\in[n]} f_i + g_k^{x\,\mathrm{T}}(x_k - x_i) + g_i^{y\,\mathrm{T}}(y_k - y_i) \geqslant \min_{i\in[n]} f_k = f_k, \end{aligned}$$

where we have used Eq. (2.2) in the last inequality. Hence, $f(x_k,\, y_k) = f_k$.

Next, we show that $g_k^x \in \partial_x f(x_k,\, y_k)$. Indeed,

$$\begin{aligned} f(x,\, y_k) - f(x_k,\, y_k) - g_k^{x\,\mathrm{T}}(x - x_k) &= \sup_{z\in\mathscr{Z}} \min_{i\in[n]} f_i + z^{\mathrm{T}}(x - x_i) + g_i^{y}(y_k - y_i) - f_k - g_k^{x}(x - x_k) \geqslant \\ &\geqslant \min_{i\in[n]} f_i + g_k^{x\,\mathrm{T}}(x - x_i) + g_i^{y\,\mathrm{T}}(y_k - y_i) - f_k - g_k^{x\,\mathrm{T}}(x - x_k) = \\ &= \min_{i\in[n]} f_i + g_i^{y\,\mathrm{T}}(y_k - y_i) - (f_k + g_k^{x\,\mathrm{T}}(x_i - x_k)) \geqslant 0. \end{aligned}$$

Hence, $g_k^x \in \partial_x f(x_k,\, y_k)$ by the definition of the subdifferential.

Next, we show that $g_k^y \in \partial_y f(x_k,\, y_k)$. Indeed,

$$\begin{aligned} f(x_k,\, y) - f(x_k,\, y_k) - g_k^{y\,\mathrm{T}}(y - y_k) &= \sup_{z\in\mathscr{Z}} \min_{i\in[n]} f_i + z^{\mathrm{T}}(x_k - x_i) + g_i^{y\,\mathrm{T}}(y - y_i) - f_k - g_k^{y\,\mathrm{T}}(y - y_k) \leqslant \\ &\leqslant \sup_{z\in\mathscr{Z}} f_k + z^{\mathrm{T}}(x_k - x_k) + g_k^{y\,\mathrm{T}}(y - y_k) - f_k - g_k^{y\,\mathrm{T}}(y - y_k) = 0. \end{aligned}$$

Hence, $g_k^y \in \partial_y f(x_k,\, y_k)$ by the definition of the subdifferential.

Finally, we show that the function $f(x,\, y)$ satisfies Definition 4. Let $(x_1,\, y_1), (x_2, y_2) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_y}$. Then, we get

$$\begin{aligned} f(x_1,\, y_1) - f(x_2,\, y_2) &= \sup_{z_1\in\mathscr{Z}} \min_{i_1\in[n]} f_{i_1} + {z_1}^{\mathrm{T}}(x_1 - x_{i_1}) + g_{i_1}^{y\ \mathrm{T}}(y_1 - y_{i_1}) - \\ &\quad - \sup_{z_2\in\mathscr{Z}} \min_{i_2\in[n]} f_{i_2} + {z_2}^{\mathrm{T}}(x_2 - x_{i_2}) + g_{i_2}^{y\ \mathrm{T}}(y_2 - y_{i_2}) = \\ &= \sup_{z_1\in\mathscr{Z}} \inf_{z_2\in\mathscr{Z}} \min_{i_1\in[n]} \max_{i_2\in[n]} \left[f^{i_1} + {z_1}^{\mathrm{T}}(x_1 - x_{i_1}) + g_{i_1}^{y\ \mathrm{T}}(y_1 - y_{i_1}) - \right. \\ &\quad \left. - f_{i_2} - {z_2}^{\mathrm{T}}(x_2 - x_{i_2}) - g_{i_2}^{y\ \mathrm{T}}(y_2 - y_{i_2})\right] = \end{aligned}$$

$$
\begin{aligned}
&= \sup_{z_1 \in \mathscr{X}} \inf_{z_2 \in \mathscr{X}} \max_{i_2 \in [n]} \min_{i_1 \in [n]} \big[f_{i_1} + {z_1}^{\mathrm{T}}(x_1 - x_{i_1}) + g^{y\,\mathrm{T}}_{i_1}(y_1 - y_{i_1}) - \\
&\qquad - f_{i_2} - {z_2}^{\mathrm{T}}(x_2 - x_{i_2}) - g^{y\,\mathrm{T}}_{i_2}(y_2 - y_{i_2})\big] \leqslant \\
&\leqslant \sup_{z_1 \in \mathscr{X}} \max_{i_2 \in [n]} \big[f_{i_2} + {z_1}^{\mathrm{T}}(x_1 - x_{i_2}) + g^{y\,\mathrm{T}}_{i_2}(y_1 - y_{i_2}) - \\
&\qquad - f_{i_2} - {z_1}^{\mathrm{T}}(x_2 - x_{i_2}) - g^{y\,\mathrm{T}}_{i_2}(y_2 - y_{i_2})\big] = \\
&= \sup_{z_1 \in \mathscr{X}} \max_{i_2 \in [n]} {z_1}^{\mathrm{T}}(x_1 - x_2) + g^{y\,\mathrm{T}}_{i_2}(y_1 - y_2) \leqslant \\
&\leqslant \sup_{z_1 \in \mathscr{X}} \max_{i_2 \in [n]} ||z_1|| ||x_1 - x_2|| + ||g^y_{i_2}|| ||y_1 - y_2|| \leqslant \\
&\leqslant M\,(||x_1 - x_2|| + ||y_1 - y_2||),
\end{aligned}
$$

where we have used Eq. (2.3) in the last inequality. □

Without the assumption of M-Lipshitsness, the interpolation conditions reduce to Eq. (2.2) which can be seen from the proof structure handling M-Lipshitsness separately.

2.3. Difference of strongly-convex functions

Consider a difference of two smooth strongly-convex functions which can also be refered to as a separable smooth SCSC function.

Definition 5. *For given constants (μ_x, μ_y, L_x, L_y) we say that a differentiable function $f(x, y)\colon \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}$ is a separable smooth SCSC function if there exist differentiable functions $p(x)\colon \mathbb{R}^{d_x} \to \mathbb{R}$ and $q(y)\colon \mathbb{R}^{d_y} \to \mathbb{R}$ such that*

$$
\begin{aligned}
&f(x, y) = p(x) - q(y) \quad \text{for all } x \in \mathbb{R}^{d_x},\ y \in \mathbb{R}^{d_y}, \\
&p(x)\colon L_x \text{ is smooth, } \mu_x \text{ is strongly convex,} \\
&p(y)\colon L_y \text{ is smooth, } \mu_y \text{ is strongly convex.}
\end{aligned}
$$

It's easy to see that this class constitutes a subset of $S_{\mu_x\mu_y L_x L_y L_{xy}}$ when $L_{xy} = 0$, so we denote separable smooth SCSC class as $S_{\mu_x\mu_y L_x L_y 0}$. Although interpolation conditions for $S_{\mu_x\mu_y L_x L_y 0}$ have already been effectively derived in [13], we aim to rewrite the inequalities in a form that only contains min-max oracle information with no additional variables.

First, we will need the conditions for smooth strongly-convex interpolability. We refer to [10] where those conditions were presented along with their application in analysis of exact worst-case performance of first-order optimization methods using the PEP approach.

Theorem 2 ($F_{\mu,L}$-interpolability [10]). *A vector sequence $\{(x_i, g_i, f_i)\}_{i\in I}$ is $F_{\mu,L}$-interpolable if and only if the following inequality holds for all $i, j \in I$:*

$$
f_i \geqslant f_j + {g_j}^{\mathrm{T}}(x_i - x_j) + \frac{L}{2(L-\mu)}\left(\frac{1}{L}||g_i - g_j||^2 + \mu||x_i - x_j||^2 - 2\frac{\mu}{L}(g_i - g_j)^{\mathrm{T}}(x_i - x_j)\right). \quad (2.9)
$$

Lemma 1. *Consider a vector sequence $\{(x_i, y_i, g_i^x, g_i^y, f_i)\}_{i\in I}$. This sequence is $S_{\mu_x\mu_y L_x L_y 0}$-interpolable if and only if there exists a vector $p \in \mathbb{R}^{|I|}$ such that the following inequalities are satisfied for all $i, j \in I$:*

$$
p_i - p_j \geqslant g^{x\,\mathrm{T}}_j(x_i - x_j) +
$$

$$+\frac{L_x}{2(L_x-\mu_x)}\left(\frac{1}{L_x}||g_i^x-g_j^x||^2-\frac{2\mu_x}{L_x}(g_i^x-g_j^x)^{\mathrm{T}}(x_i-x_j)+\mu_x||x_i-x_j||^2\right),$$
$$p_i-p_j\geqslant f_i-f_j-g_j^{y\mathrm{T}}(y_i-y_j)+$$
$$+\frac{L_y}{2(L_y-\mu_y)}\left(\frac{1}{L_y}||g_i^y-g_j^y||^2+\frac{2\mu_y}{L_y}(g_i^y-g_j^y)^{\mathrm{T}}(y_i-y_j)+\mu_y||y_i-y_j||^2\right). \tag{2.10}$$

Proof. Necessity. Suppose there exist $p(x)\in F_{\mu_x,L_x}$ and $q(y)\in F_{\mu_y,L_y}$. Eq. (2.10) immediately follows from Theorem 2 with $p_i=p(x_i)$.

Sufficiency. Suppose conditions (2.10) are satisfied. Applying Theorem 2 twice to vector sequences $\{(x_i,\,g_i^x,\,p_i)\}_{i\in I}$ and $\{(y_i,\,-g_i^y,\,p_i-f_i)\}_{i\in I}$, we prove the existence of functions $p(x)\in F_{\mu_x,L_x}$ and $q(y)\in F_{\mu_y,L_y}$ that interpolate respective vector sequences. From this it follows that a function $f(x,\,y)=p(x)-q(y)$ interpolates $\{(x_i,\,y_i,\,g_i^x,\,g_i^y,\,f_i)\}_{i\in I}$ sequence, which concludes the proof. □

Now we want to rewrite conditions (2.10) without vector $p\in\mathbb{R}^{|I|}$. After combining inequalities (2.10) we get for all $i,\,j\in I$:

$$p_i-p_j\geqslant c^{ij}, \tag{2.11}$$

where

$$c^{ij}=\max\{\alpha^{ij},\,\beta^{ij}\},$$
$$\alpha^{ij}=g_j^{x\mathrm{T}}(x_i-x_j)+\frac{L_x}{2(L_x-\mu_x)}\left(\frac{1}{L_x}||g_i^x-g_j^x||^2-\frac{2\mu_x}{L_x}(g_i^x-g_j^x)^{\mathrm{T}}(x_i-x_j)+\mu_x||x_i-x_j||^2\right),$$
$$\beta^{ij}=f_i-f_j-g_j^{y\mathrm{T}}(y_i-y_j)+\frac{L_y}{2(L_y-\mu_y)}\left(\frac{1}{L_y}||g_i^y-g_j^y||^2+\frac{2\mu_y}{L_y}(g_i^y-g_j^y)^{\mathrm{T}}(y_i-y_j)+\mu_y||y_i-y_j||^2\right).$$

Before determining conditions under which the system of linear inequalities (2.11) has a solution, we introduce several definitions and a lemma.

Definition 6. *Let I be an ordered set of indices. Let $J=(j_0,\,\ldots,\,j_k)$ be an ordered selection of indices from I such that indices are not allowed to repeat except for j_k that is allowed to be equal to j_0. We introduce $J(n)$ as a set of all such selections that have $j_0=0$ and $j_k=n$, where $0,\,n\in I$.*

Definition 7. *Let I be an ordered set of indices. Let $c^{ij}\in\mathbb{R}$ be real numbers indexed by $i,\,j\in I$. Also, let $J=(j_0=0,\,\ldots,\,j_k=n)\in J(n)$. We define a sum C_J of numbers c^{ij} corresponding to J as follows*:

$$C_J=\sum_{l=0}^{k-1}c^{j_{l+1}j_l}.$$

Definition 8. *Let I be an ordered set of indices. Let $c^{ij}\in\mathbb{R}$ be real numbers indexed by $i,\,j\in I$. The maximal sum corresponding to $J(n)$ is*

$$C_{J(n)}=\max_{J\in J(n)}C_J.$$

Definition 9. *Let I be an ordered set of indices. Let $c^{ij} \in \mathbb{R}$ be real numbers indexed by $i, j \in I$. Let $J = (j_0, \ldots, j_k)$ be some ordered subset of I. The cyclic sum of c^{ij} that corresponds to J is*

$$\sum_{l=0}^{k} c^{j_{l+1} j_l},$$

where we define $j_{k+1} = j_0$. If the subset $J = (j_0)$ consists of only one element, we define the cyclic sum as $c^{j_0 j_0}$.

Now we proceed to determining conditions under which the system of linear inequalities (2.11) is feasible. The following Lemma 2 is inspired by the concept of cyclic monotonicity introduced in the classical work of R. Tyrrell Rockafellar [14] and is essentially a variation of Theorem 3.2 from [15]. Though we do not claim Lemma 2 to be a novel result in principle, we find its formulation more convenient for purposes of expressing interpolation conditions.

Lemma 2. *A system of linear inequalities* (2.11) *is feasible* (*in respect to $p \in \mathbb{R}^{|I|}$*) *if and only if all cyclic sums of c^{ij}* (*Definition* 9) *are nonpositive.*

Proof. Let us suppose $p \in \mathbb{R}^{|I|}$ is a solution to the system. Let $J = (j_0, \ldots, j_k)$ be an ordered subset of I. After we take a sum of corresponding inequalities

$$\begin{aligned} p_{j_1} - p_{j_0} &\geqslant c^{j_1 j_0}, \\ p_{j_2} - p_{j_1} &\geqslant c^{j_2 j_1}, \\ &\ldots \\ p_{j_k} - p_{j_{k-1}} &\geqslant c^{j_k j_{k-1}}, \\ p_{j_0} - p_{j_k} &\geqslant c^{j_0 j_k}, \end{aligned}$$

we have zero on the left side and a cyclic sum of c^{ij} that corresponds to the ordered subset J on the right side. Summing over all possible J, we get all cyclic sums of c^{ij} are nonpositive.

Now suppose that all cyclic sums of c^{ij} are nonpositive. We will show that $p_i = C_{J(i)}$ $i \in I$, where $C_{J(i)}$ is a maximal sum from Definition 8, form a feasible point for system (2.11).

After substituting p_i, we need to check that the inequalities hold for all $i, j \in I$:

$$\max_{J \in J(i)} C_J \geqslant \max_{J \in J(j)} C_J + c^{ij}. \tag{2.12}$$

The inequalities obviously hold for $i = j$, so consider $i \neq j$. There are two possible cases.

First case: $\overline{J}_j = \underset{\overline{J} \in J(j)}{\arg\max}\, C_{\overline{J}}$ does not contain i.

Then $(\overline{J}_j, i) \in J(i)$ and the sum of c^{ij} corresponding to $(\overline{J}_j, i)$ is exactly $\max\limits_{J \in J(j)} C_J + c^{ij}$. This sum, however, is not larger than $\max\limits_{J \in J(i)} C_J$ which is a maximal sum for $J(i)$.

Second case: $\overline{J}_j = \underset{\overline{J} \in J(j)}{\arg\max}\, C_{\overline{J}}$ contains i.

From Definition 6 it follows that $\overline{J}_j \in J(j)$ can contain i only once. This implies that

$$\begin{aligned} \overline{J}_j &= (j_0 = 0, \ldots, j_m = i, \ldots, j_k = j), \\ (\overline{J}_j, i) &= (j_0 = 0, \ldots, j_m = i, \ldots, j_k = j, j_{k+1} = i). \end{aligned}$$

The sum of c^{ij} (Definition 7) corresponding to $(\overline{J}_j,\, i)$ can be written as

$$C_{(\overline{J}_j, i)} = C_{\overline{J}_{j1}} + C_{\overline{J}_{j2}},$$

where

$$\overline{J}_{j1} = (j_0 = 0,\, \ldots,\, j_m = i),$$
$$\overline{J}_{j2} = (j_m = i,\, \ldots,\, j_{k+1} = i).$$

It can be seen that $\overline{J}_{j1} \in J(i)$. And since any cyclic sum of c^{ij} is nonpositive, it follows that $C_{\overline{J}_{j2}} \leqslant 0$. Consequently:

$$\max_{J \in J(j)} C_J + c^{ij} = C_{(\overline{J}_j, i)} = C_{\overline{J}_{j1}} + C_{\overline{J}_{j2}} \leqslant C_{\overline{J}_{j1}} \leqslant \max_{J \in J(i)} C_J.$$

This means that the proposed p is indeed a feasible point for the system. □

Finally, we can rewrite interpolation conditions from Lemma 1 in a form that contains oracle information exclusively.

Theorem 3. *Consider a sequence* $\{(x_i,\, y_i,\, g_i^x,\, g_i^y,\, f_i)\}_{i \in I}$. *Consider the numbers* $c^{ij} \in \mathbb{R}$, *indexed by* $i,\, j \in I$ *and calculated from this sequence*:

$$\forall i,\, j \in I\colon \quad c^{ij} = \max\{\alpha^{ij},\, \beta^{ij}\},$$

where

$$\alpha^{ij} = g_j^{x\mathrm{T}}(x_i - x_j) + \frac{L_x}{2(L_x - \mu_x)}\left(\frac{1}{L_x}||g_i^x - g_j^x||^2 - \frac{2\mu_x}{L_x}(g_i^x - g_j^x)^{\mathrm{T}}(x_i - x_j) + \mu_x||x_i - x_j||^2\right),$$
$$\beta^{ij} = f_i - f_j - g_j^{y\mathrm{T}}(y_i - y_j) + \frac{L_y}{2(L_y - \mu_y)}\left(\frac{1}{L_y}||g_i^y - g_j^y||^2 + \frac{2\mu_y}{L_y}(g_i^y - g_j^y)^{\mathrm{T}}(y_i - y_j) + \mu_y||y_i - y_j||^2\right).$$

The sequence is $S_{\mu_x \mu_y L_x L_y 0}$*-interpolable if and only if any cyclic sum* (*Definition* 9) *of* c^{ij} *is nonpositive.*

Proof. The statement of the theorem is a trivial consequence of Lemma 1 and Lemma 2. □

The interpolation conditions for the difference of smooth SC functions are not of great practical interest in themselves because when used in PEP, they would effectively separate into already well-known interpolation conditions of smooth SC functions that are a very well-studied class, both theoretically and numerically via PEP. Still, the interpolation conditions for the difference of SC functions, when written in the form without additional variables, might give a good insight into the potential structure of interpolation conditions of a more general class $S_{\mu_x \mu_y L_x L_y L_{xy}}$.

2.4. Bilinear convex-concave functions

Before moving to saddle functions with bilinear coupling we should deal with functions of the following form:

$$f(x, y) = y^{\mathrm{T}} A x, \tag{2.13}$$

where $A \in \mathbb{R}^{d_y \times d_x}$ is a coupling matrix with the bounded largest singular value $\sigma_{\max}(A) \leqslant L$.

Definition 10. *We consider finite-dimensional linear operators, i.e., matrices, with bounded largest singular value*: $\mathscr{L}_L = \{A\colon \sigma_{\max}(A) \leqslant L\}$.

We will need interpolation conditions for $\mathscr{L}_L$ that were derived in [16].

Definition 11 ($\mathscr{L}_L$-interpolability [16]). *Sets of pairs* $\{(x_i,\, y_i)\}_{i\in[N_1]}$ *and* $\{(u_j,\, v_j)\}_{j\in[N_2]}$ *are* $\mathscr{L}_L$*-interpolable if and only if* $\exists A \in \mathscr{L}_L$ *such that*

$$\begin{cases} y_i = Ax_i, & \forall i \in [N_1],\\ v_j = A^{\mathrm{T}}u_j, & \forall j \in [N_2]. \end{cases}$$

Theorem 4 ($\mathscr{L}_L$-interpolability [16]). *Let* $X \in \mathbb{R}^{n\times N_1}$, $Y \in \mathbb{R}^{m\times N_1}$, $U \in \mathbb{R}^{m\times N_2}$, $V \in \mathbb{R}^{n\times N_2}$, *and* $L \geqslant 0$. $(X,\, Y,\, U,\, V)$ *is* $\mathscr{L}_L$*-interpolable if and only if*

$$\begin{cases} X^{\mathrm{T}}V = Y^{\mathrm{T}}U,\\ Y^{\mathrm{T}}Y \preccurlyeq L^2X^{\mathrm{T}}X,\\ V^{\mathrm{T}}V \preccurlyeq L^2U^{\mathrm{T}}U. \end{cases}$$

To get interpolation conditions for (2.13), we have to slightly modify Theorem 4.

Theorem 5. *Consider a sequence* $\{(x_i,\, y_i,\, g_i^x,\, g_i^y,\, f_i)\}_{i\in I}$. *The sequence is interpolable by functions* (2.13) *if and only if*

$$\begin{cases} f_i = {g_i^x}^{\mathrm{T}}x_i = {g_i^y}^{\mathrm{T}}y_i \quad \forall i \in I,\\ {G_x}^{\mathrm{T}}X = Y^{\mathrm{T}}G_y,\\ {G_x}^{\mathrm{T}}G_x \preccurlyeq L^2Y^{\mathrm{T}}Y,\\ {G_y}^{\mathrm{T}}G_y \preccurlyeq L^2X^{\mathrm{T}}X, \end{cases}$$

where X, Y, G_x, G_y *are matrices with columns formed by* $\{x_i\}_{i\in I}$, $\{y_i\}_{i\in I}$, $\{g_i^x\}_{i\in I}$, $\{g_i^y\}_{i\in I}$, *respectively.*

Proof. Suppose we have a function (2.13). It's easy to see that the conditions are satisfied. Now suppose the conditions are satisfied. With Theorem 4 we can find a linear operator A that interpolates $(X,\, G_y,\, Y,\, G_x)$. We have to check that $y^{\mathrm{T}}Ax$ has correct values in every $(x_i,\, y_i)$:

$$f(x_i,\, y_i) = {y_i}^{\mathrm{T}}Ax_i = {g_i^x}^{\mathrm{T}}x_i = {g_i^y}^{\mathrm{T}}y_i = f_i.$$

This means that $y^{\mathrm{T}}Ax$ interpolates $\{(x_i,\, y_i,\, g_i^x,\, g_i^y,\, f_i)\}_{i\in I}$ sequence correctly. □

We now move to bilinear saddle functions of the form

$$f(x,y) = a^{\mathrm{T}}x + y^{\mathrm{T}}Ax - b^{\mathrm{T}}y, \tag{2.14}$$

where $A \in \mathscr{L}_{L_{xy}}$, $a \in \mathbb{R}^{d_x}$, $b \in \mathbb{R}^{d_y}$.

It is easy to see that functions (2.14) constitute a subset of $S_{\mu_x\mu_yL_xL_yL_{xy}}$ corresponding to $\mu_x = \mu_y = L_x = L_y = 0$ which motivates us to acquire the interpolation conditions for this relatively simple convex-concave class (designated as $S_{0000L_{xy}}$). The next lemma will be useful later for writing conditions in a more compact form.

Lemma 3. *Let $f(x, y) \in S$ (Definition 1). Let $\{(x_i, y_i, g_i^x, g_i^y, f_i)\}_{i\in I}$ be a sequence that is interpolable by $f(x, y)$ that satisfies*

$$f_i = f_j + g_j^{x\mathrm{T}}(x_i - x_j) + g_i^{y\mathrm{T}}(y_i - y_j) \quad \forall i, j \in I, \tag{2.15}$$

then $\forall i, j, k, m \in I$ holds

$$(g_i^x - g_k^x)^\mathrm{T}(x_j - x_m) = (g_j^y - g_m^y)^\mathrm{T}(y_i - y_k).$$

Proof. Write inequalities (2.15) in a cyclic manner and take their sums. For cycles of length 2:

$$f_i = f_j + g_j^{x\mathrm{T}}(x_i - x_j) + g_i^{y\mathrm{T}}(y_i - y_j),$$
$$f_j = f_i + g_i^{x\mathrm{T}}(x_j - x_i) + g_j^{y\mathrm{T}}(y_j - y_i).$$

After addition we get

$$(g_i^x - g_j^x)^\mathrm{T}(x_i - x_j) = (g_i^y - g_j^y)^\mathrm{T}(y_i - y_j). \tag{2.16}$$

After doing the same for a cycle of length 3 and combining with Eq. (2.16):

$$(g_i^x - g_k^x)^\mathrm{T}(x_j - x_k) = (g_j^y - g_k^y)^\mathrm{T}(y_i - y_k). \tag{2.17}$$

Finally, for a cycle of length 4 and combining with both Eq. (2.16) and Eq. (2.17):

$$(g_i^x - g_k^x)^\mathrm{T}(x_j - x_m) = (g_j^y - g_m^y)^\mathrm{T}(y_i - y_k),$$

which is Eq. (3). □

We now prove interpolation conditions for $S_{0000L_{xy}}$.

Theorem 6. *Consider a sequence $\{(x_i, y_i, g_i^x, g_i^y, f_i)\}_{i\in I}$. The sequence is $S_{0000L_{xy}}$-interpolable if and only if*

$$\begin{cases} f_i = f_j + g_j^{x\mathrm{T}}(x_i - x_j) + g_i^{y\mathrm{T}}(y_i - y_j) & \forall i, j \in I, \\ G_x^{p\mathrm{T}} G_x^p \preccurlyeq L_{xy}^2 Y^{p\mathrm{T}} Y^p & \forall p \in I, \\ G_y^{p\mathrm{T}} G_y^p \preccurlyeq L_{xy}^2 X^{p\mathrm{T}} X^p & \forall p \in I, \end{cases}$$

where X^p, Y^p, G_x^p, G_y^p are matrices with columns formed by $\{x_i - x_p\}_{i\in I}$, $\{y_i - y_p\}_{i\in I}$, $\{g_i^x - g_p^x\}_{i\in I}$, $\{g_i^y - g_p^y\}_{i\in I}$, respectively.

Proof. Suppose the sequence is interpolable. It's straightforward to check that the conditions are satisfied. Suppose the conditions are satisfied. Using Lemma 3 we get $\forall p, k, m \in I$:

$$G_y^{m\mathrm{T}} Y^k = X^{m\mathrm{T}} G_x^k,$$
$$G_x^{p\mathrm{T}} G_x^p \preccurlyeq L_{xy}^2 Y^{p\mathrm{T}} Y^p,$$
$$G_y^{p\mathrm{T}} G_y^p \preccurlyeq L_{xy}^2 X^{p\mathrm{T}} X^p.$$

By Theorem 4, for any $p \in I$, there exists $A_p \in \mathscr{L}_{L_{xy}}$ such that $\forall i \in I$:

$$g_i^x - g_p^x = A_p(y_i - y_p),$$
$$g_i^y - g_p^y = A_p^\mathrm{T}(x_i - x_p).$$

We choose matrix A, corresponding to some particular p, $p = 0$, for example, and construct a function:

$$F(x, y) = x^{\mathrm{T}}Ay + (g_0^x - Ay_0)^{\mathrm{T}}x + (g_0^y - A^{\mathrm{T}}x_0)y + (f_0 - g_0^{x\mathrm{T}}x_0 - g_0^{y\mathrm{T}}y_0 + x_0{}^{\mathrm{T}}Ay_0).$$

Now we show that $F(x, y)$ correctly interpolates $(x_i, y_i, g_i^x, g_i^y, f_i)_{i\in I}$:

$$\begin{aligned}F(x_i, y_i) = f_0 + g_0^{x\mathrm{T}}(x_i - x_0) + (x_i{}^{\mathrm{T}}A(y_i - y_0) - x_0{}^{\mathrm{T}}A(y_i - y_0) + g_0^{y\mathrm{T}}(y_i - y_0)) = \\ = f_0 + g_0^{x\mathrm{T}}(x_i - x_0) + g_i^{y\mathrm{T}}(y_i - y_0) = f_i, \\ \nabla_x F(x_i, y_i) = Ay_i - Ay_0 + g_0^x = A(y_i - y_0) + g_0^x = g_i^x, \\ \nabla_y F(x_i, y_i) = A^{\mathrm{T}}x - A^{\mathrm{T}}x_0 + g_0^y = g_i^y,\end{aligned}$$

which concludes the proof. □

Given the interpolation conditions for the main special cases (Theorem 1, Theorem 3, Theorem 6) it should be much easier to derive interpolation conditions for the (most general) smooth SCSC class $S_{\mu_x\mu_y L_x L_y L_{xy}}$. It might also be possible to guess them and then verify numerically using a recently developed approach to verifying interpolation conditions [17]. We leave it for future research.

2.5. Composite convex-concave functions with a bilinear coupling term

In this subsection we consider the following saddle point problem:

$$\min_{x\in\mathbb{R}^{d_x}} \max_{y\in\mathbb{R}^{d_x}} \left[f(x) + y^{\mathrm{T}}Ax - g(y)\right], \tag{2.18}$$

where $f(x) \in F_{\mu_x L_x}$, $g(y) \in F_{\mu_y L_y}$, $A \in \mathscr{L}_{L_{xy}}$, and we denote $K_{\mu_x\mu_y L_x L_y L_{xy}} \subset S_{\mu_x\mu_y L_x L_y L_{xy}}$.

The problem 2.18 is well-studied, has a large number of applications and arises in settings that are seemingly unrelated to min-max optimization: empirical risk minimization [18], reinforcement learning [19], minimization under affine constraints [20] and many others. Multiple methods have even been developed that are optimal in the sense of asymptotic dependence on smoothness parameters [21]. However, the exact worst-case upper bounds remain unknown for the majority of methods including the optimal ones.

From the structure alone, the way to acquire interpolation conditions for this class seems fairly straightforward: it is simply a combination of conditions for smooth strongly-convex functions (Theorem 2) and conditions for the bilinear coupling function (Theorem 5).

Theorem 7. *Sets $\{(x_i, y_i, \phi_i^x, \phi_i^y, f_i)\}_{i\in I}$ are $K_{\mu_x\mu_y L_x L_y L_{xy}}$-interpolable if and only if there exist $\{(h_i^x, h_i^y, g_i^x, g_i^y, f_i^x, f_i^y)\}_{i\in I}$ such that*

$$\begin{aligned}f_i = f_i^x + h_i^{x\mathrm{T}}x_i - f_i^y, \\ \phi_i^x = g_i^x + h_i^x, \\ \phi_i^y = h_i^y - g_i^y,\end{aligned}$$

$$f_i^x \geqslant f_j^x + g_j^{x\mathrm{T}}(x_i - x_j) + \frac{L_x}{2(L_x - \mu_x)}\left(\frac{1}{L_x}||g_i^x - g_j^x||^2 + \mu_x||x_i - x_j||^2 - 2\frac{\mu_x}{L_x}(g_i^x - g_j^x)^{\mathrm{T}}(x_i - x_j)\right),$$

$$f_i^y \geqslant f_j^y + g_j^{y\mathrm{T}}(y_i - y_j) + \frac{L_y}{2(L_y - \mu_y)}\left(\frac{1}{L_y}||g_i^y - g_j^y||^2 + \mu_y||y_i - y_j||^2 - 2\frac{\mu_y}{L_y}(g_i^y - g_j^y)^{\mathrm{T}}(y_i - y_j)\right),$$

$$
\begin{aligned}
{H_x}^{\mathrm{T}} X &= Y^{\mathrm{T}} H_y, \\
{H_x}^{\mathrm{T}} H_x &\preccurlyeq L_{xy}^2 Y^{\mathrm{T}} Y, \\
{H_y}^{\mathrm{T}} H_y &\preccurlyeq L_{xy}^2 X^{\mathrm{T}} X,
\end{aligned}
$$

where X, Y, H_x, H_y are matrices with columns formed by $\{x_i\}_{i\in I}$, $\{y_i\}_{i\in I}$, $\{h_i^x\}_{i\in I}$, $\{h_i^y\}_{i\in I}$, respectively.

Proof. Statement of the theorem follows from combining Theorem 2 and Theorem 5. The logic of the proof is completely analogous to Lemma 1, which presents interpolation conditions for difference of strongly-convex functions. □

3. PEP for min-max problems

3.1. PEP construction

We now move to construct a performance estimation problem for a class of (so-called) N-degree first-order fixed-step methods applied to composite functions with the bilinear coupling term $K_{\mu_x \mu_y L_x L_y L_{xy}}$. The idea of this specific PEP formulation is to find a linear convergence guarantee (for some fixed quadratic metric) that would hold for an arbitrary number of iterations. The approach results in only having to solve a small SDP at the cost of guarantee's tightness [8].

We then numerically solve the (dual) problem for Sim-GDA and Alt-GDA with their best step choices. We refer to [10] for a detailed explanation of the proper and tight PEP approach, containing thorough justifications for every step of PEP's construction. We also assume the problem to have a reasonably high dimension, the need for such an assumption will become clear later.

We consider first-order iterative fixed-step methods of the form

$$
\begin{gathered}
u_k^x = \sum_{j=0}^{N} \theta_j^x x_{k-j}, \quad v_k^y = \sum_{j=0}^{N} \phi_j^y y_{k-j} \qquad \text{(M)} \\
g_k^x = \nabla_x f(u_k^x), \quad h_k^x = A^{\mathrm{T}} v_k^y, \\
x_{k+1} = \sum_{j=0}^{N} \alpha_j^x x_{k-j} - \beta^x g_k^x - \gamma^x h_k^x, \\
u_k^y = \sum_{j=0}^{N} \theta_j^y y_{k-j}, \quad v_k^x = \sum_{j=0}^{N+1} \phi_j^x x_{k+1-j}, \\
g_k^y = \nabla_y g(u_k^y), \quad h_k^y = A v_k^x, \\
y_{k+1} = \sum_{j=0}^{N} \alpha_j^y y_{k-j} - \beta^y g_k^y + \gamma^y h_k^y,
\end{gathered}
$$

where N is a degree (or memory) of method (M). We define methods similar to how it was done in [8] for convenience later on. The algorithm should have a fixed point $z_* = (x_*, y_*)$ so we will demand constant step-sizes to satisfy

$$
\sum_{j=0}^{N} \theta_j^x = \sum_{j=0}^{N} \theta_j^y = \sum_{j=0}^{N+1} \phi_j^x = \sum_{j=0}^{N} \phi_j^y = \sum_{j=0}^{N} \alpha_j^x = \sum_{j=0}^{N} \alpha_j^y = 1.
$$

It also makes sense to demand $\beta^x \neq 0$, $\beta^y \neq 0$, $\gamma^x \neq 0$, $\gamma^y \neq 0$.

We introduced a class of methods with N-step memory that can also take advantage of the problem's bilinear-coupled structure. The class includes the most basic methods like Sim-GDA and Alt-GDA but does not include any of advanced methods which is acceptable since (M) serves primarily the purpose of demonstration. We note that (M) can be freely changed to be any first-order fixed-step method, and we build PEP in a way that its final general structure does not depend on a method. We now introduce a PEP program and move straight to its finite-dimensional formulation:

$$\sup_{\{(x_i, y_i, h_i^x, h_i^y, g_i^x, g_i^y, f_i^x, f_i^y)\}_{i\in I}} ||x_{N+1} - x_*||^2 + ||y_{N+1} - y_*||^2 \qquad \text{(PEP)}$$

such that

$$\begin{gathered}(x,\, y,\, u^x,\, u^y,\, v^x,\, v^y) \text{ are generated by method (M)},\\ \{(u_i^x,\, g_i^x,\, f_i^x)\}_{i\in I} \text{ is } F_{\mu_x L_x} \text{ — interpolable},\\ \{(u_i^y,\, g_i^y,\, f_i^y)\}_{i\in I} \text{ is } F_{\mu_y L_y} \text{ — interpolable},\\ \{(v_i^y,\, h_i^x)\}_{i\in I},\, (v_i^x, h_i^y)_{i\in I} \text{ is } \mathscr{L}_{L_{xy}} \text{ — interpolable},\\ (x_*,\, y_*) = (\mathbf{0}_{d_x},\, \mathbf{0}_{d_y}),\\ g_*^x = -h_*^x = \mathbf{0}_{d_x}, \quad g_*^y = h_*^y = \mathbf{0}_{d_y},\\ ||x_N - x_*||^2 + ||y_N - y_*||^2 \leqslant R^2.\end{gathered}$$

Because the class $K_{\mu_x\mu_y L_x L_y L_{xy}}$ is translation invariant we can let $(x_*,\, y_*) = (\mathbf{0}_{d_x},\, \mathbf{0}_{d_y})$. Due to the problem being unconstrained, and from optimality conditions we also get $g_*^x = -h_*^x = \mathbf{0}_{d_x}$ and $g_*^y = h_*^y = \mathbf{0}_{d_y}$. By shifting the function values we can also have $f_*^x = f_*^y = 0$ though it would not affect the SDP in any manner, which will become apparent later. One can notice that in (PEP) the method performs N steps of (M) which is due to (M) having N-step memory. Also, the performance metric can be changed to be any other quadratic of (current and previous N [8]) distances, gradient and functional values. Our next goal is to rewrite (PEP) as a convex semidefinite program. To achieve that, we will express interpolation conditions in terms of Gram matrices and use them as program's constraints.

We begin with initializing row basis vectors for initial conditions, gradient values and functional values $\overline{x}_k$, $\overline{g}_k^x$, $\overline{h}_k^x \in R^{3N+3}$, $\overline{y}_k$, $\overline{g}_k^y$, $\overline{h}_k^y \in R^{3N+3}$, $\overline{f}_k^x \in R^{N+1}$, $\overline{f}_k^y \in R^{N+1}$:

$$\begin{gathered}\overline{x}_k := \overline{e}_{k+N+1}^{\mathrm{T}},\ k \in -N,\, \ldots,\, 0; \quad \overline{g}_k^x := \overline{e}_{k+N+2}^{\mathrm{T}},\ k \in 0,\, \ldots,\, N; \quad \overline{h}_k^x := \overline{e}_{k+2N+3}^{\mathrm{T}},\ k \in 0,\, \ldots,\, N;\\ \overline{y}_k := \overline{e}_{k+N+1}^{\mathrm{T}},\ k \in -N,\, \ldots,\, 0; \quad \overline{g}_k^y := \overline{e}_{k+N+2}^{\mathrm{T}},\ k \in 0,\, \ldots,\, N; \quad \overline{h}_k^y := \overline{e}_{k+2N+3}^{\mathrm{T}},\ k \in 0,\, \ldots,\, N;\\ \overline{f}_k^x := \overline{e}_{k+1}^{\mathrm{T}},\ k \in 0,\, \ldots,\, N; \quad \overline{f}_k^y := \overline{e}_{k+1}^{\mathrm{T}},\ k \in 0,\, \ldots,\, N;\\ \overline{x}_* := \mathbf{0}_{3N+3}^{\mathrm{T}}; \quad \overline{g}_*^x = -\overline{h}_*^x = \mathbf{0}_{3N+3}^{\mathrm{T}};\\ \overline{y}_* := \mathbf{0}_{3N+3}^{\mathrm{T}}; \quad \overline{g}_*^y = \overline{h}_*^y = \mathbf{0}_{3N+3}^{\mathrm{T}};\\ \overline{f}_*^x = \mathbf{0}_{N+1}^{\mathrm{T}}; \quad \overline{f}_*^y = \mathbf{0}_{N+1}^{\mathrm{T}}.\end{gathered}$$

The rest of row vectors are calculated with (M) iterating for $k = 0,\, \ldots,\, N$ with previously defined vectors:

$$\overline{u}_k^x = \sum_{j=0}^{N} \theta_j^x \overline{x}_{k-j}, \quad \overline{v}_k^y = \sum_{j=0}^{N} \phi_j^y \overline{y}_{k-j},$$

$$\overline{u}_k^y = \sum_{j=0}^{N} \theta_j^y \overline{y}_{k-j}, \quad \overline{v}_k^x = \sum_{j=0}^{N+1} \phi_j^x \overline{x}_{k+1-j},$$
$$\overline{x}_{k+1} = \sum_{j=0}^{N} \alpha_j^x \overline{x}_{k-j} - \beta^x(\overline{g}_k^x + \overline{g}_*^x) - \gamma^x(\overline{h}_k^x + \overline{h}_*^x),$$
$$\overline{y}_{k+1} = \sum_{j=0}^{N} \alpha_j^y \overline{y}_{k-j} - \beta^y(\overline{g}_k^y + \overline{g}_*^y) + \gamma^y(\overline{h}_k^y + \overline{h}_*^y).$$

Now we introduce Gram matrices $W_x = B_x^{\mathrm{T}} B_x \in S^{3N+3}$, $W_y = B_y^{\mathrm{T}} B_y \in S^{3N+3}$ where

$$B_x := [x_{-N} - x_* \quad \cdots \quad x_0 - x_* \quad g_0^x - g_*^x \quad \cdots \quad g_N^x - g_*^x \quad h_0^x - h_*^x \quad \cdots \quad h_N^x - h_*^x],$$
$$B_y := \left[y_{-N} - y_* \quad \cdots \quad y_0 - y_* \quad g_0^y - g_*^y \quad \cdots \quad g_N^y - g_*^y \quad h_0^y - h_*^y \quad \cdots \quad h_N^y - h_*^y\right],$$

and also vectors of function values:

$$\mathbf{f}^x := [f_0^x - f_*^x, \ldots, f_N^x - f_*^x]^{\mathrm{T}},$$
$$\mathbf{f}^y := \left[f_0^y - f_*^y, \ldots, f_N^y - f_*^y\right]^{\mathrm{T}}.$$

We can now express interpolation conditions in terms of Gram matrices. The index set is defined as $I = \{0, \ldots, N, *\}$. With the Gram matrices introduced earlier, the interpolation conditions for strongly-convex functions (Theorem 2) $f(x) \in F_{\mu_x L_x}$, $g(y) \in F_{\mu_y L_y}$ take the form $(\forall i, j \in I)$

$$0 \leqslant {m_{ij}^x}^{\mathrm{T}} \mathbf{f}^x + Tr(W_x M_{ij}^x),$$
$$0 \leqslant {m_{ij}^y}^{\mathrm{T}} \mathbf{f}^y + Tr(W_y M_{ij}^y),$$

where for $m_{ij}^x \in R^{N+1}$, $M_{ij}^x \in S^{3N+3}$:

$$M^x := \frac{1}{2} \begin{bmatrix} -\mu_x L_x & \mu_x L_x & \mu_x & -L_x \\ \mu_x L_x & -\mu_x L_x & -\mu_x & L_x \\ \mu_x & -\mu_x & -1 & 1 \\ -L_x & L_x & 1 & -1 \end{bmatrix},$$
$$m_{ij}^x := (L_x - \mu_x)(\overline{f}_i^x - \overline{f}_j^x)^{\mathrm{T}},$$
$$M_{ij}^x := \begin{bmatrix} \overline{u}_i^x + \overline{x}_* \\ \overline{u}_j^x + \overline{x}_* \\ \overline{g}_i^x + \overline{g}_*^x \\ \overline{g}_j^x + \overline{g}_*^x \end{bmatrix}^{\mathrm{T}} M^x \begin{bmatrix} \overline{u}_i^x + \overline{x}_* \\ \overline{u}_j^x + \overline{x}_* \\ \overline{g}_i^x + \overline{g}_*^x \\ \overline{g}_j^x + \overline{g}_*^x \end{bmatrix},$$

and $m_{ij}^y \in R^{N+1}$, $M_{ij}^y \in S^{3N+3}$ are defined in the same way.

Conditions for the bilinear coupling function (Theorem 5) can be rewritten as

$${P_{Hx}}^{\mathrm{T}} W_x P_{vx} = {P_{vy}}^{\mathrm{T}} W_y P_{Hy},$$
$${P_{Hx}}^{\mathrm{T}} W_x P_{Hx} \preccurlyeq L_{xy}^2 {P_{vy}}^{\mathrm{T}} W_y P_{vy},$$
$${P_{Hy}}^{\mathrm{T}} W_y P_{Hy} \preccurlyeq L_{xy}^2 {P_{vx}}^{\mathrm{T}} W_x P_{vx},$$

where P_{Hx}, $P_{Hy} \in R^{(3N+3)\times(N+1)}$, P_{vx}, $P_{vy} \in R^{(3N+3)\times(N+1)}$:

$$\begin{aligned} P_{Hx} &= \left[(\overline{h}_0^x + \overline{h}_*^x)^{\mathrm{T}}, \quad \ldots, \quad (\overline{h}_N^x + \overline{h}_*^x)^{\mathrm{T}}\right], \\ P_{Hy} &= \left[(\overline{h}_0^y + \overline{h}_*^y)^{\mathrm{T}}, \quad \ldots, \quad (\overline{h}_N^y + \overline{h}_*^y)^{\mathrm{T}}\right], \\ P_{vx} &= \left[(\overline{v}_0^x + \overline{x}_*)^{\mathrm{T}}, \quad \ldots, \quad (\overline{v}_N^x + \overline{x}_*)^{\mathrm{T}}\right], \\ P_{vy} &= \left[(\overline{v}_0^y + \overline{y}_*)^{\mathrm{T}}, \quad ,\ldots, \quad (\overline{v}_N^y + \overline{y}_*)^{\mathrm{T}}\right]. \end{aligned}$$

And finally, the performance metric and its initial bound are

$$\begin{gathered} ||x_{N+1} - x_*||^2 + ||y_{N+1} - y_*||^2 = Tr(W_x A_{N+1}^x) + Tr(W_y A_{N+1}^y), \\ ||x_N - x_*||^2 + ||y_N - y_*||^2 = Tr(W_x A_N^x) + Tr(W_y A_N^y) \leqslant R^2, \end{gathered}$$

where $A_{N+1}^x = {\overline{x}_{N+1}}^{\mathrm{T}} \overline{x}_{N+1}$, $A_{N+1}^y = {\overline{y}_{N+1}}^{\mathrm{T}} \overline{y}_{N+1}$, $A_N^x = {\overline{x}_N}^{\mathrm{T}} \overline{x}_N$, $A_N^y = {\overline{y}_N}^{\mathrm{T}} \overline{y}_N$.

Having rewritten everything in terms of Gram matrices, we can reformulate (PEP) so that optimization is carried out over W_x, W_y, $\mathbf{f}^x$, $\mathbf{f}^y$:

$$\sup_{W_x, W_y, \mathbf{f}^x, \mathbf{f}^y} Tr(W_x A_{N+1}^x) + Tr(W_y A_{N+1}^y) \tag{Gram-PEP}$$

such that $\forall i, j \in I$

$$\begin{gathered} 0 \leqslant {m_{ij}^x}^{\mathrm{T}} \mathbf{f}^x + Tr(W_x M_{ij}^x), \\ 0 \leqslant {m_{ij}^y}^{\mathrm{T}} \mathbf{f}^y + Tr(W_y M_{ij}^y), \\ {P_{Hx}}^{\mathrm{T}} W_x P_{vx} = {P_{vy}}^{\mathrm{T}} W_y P_{Hy}, \\ {P_{Hx}}^{\mathrm{T}} W_x P_{Hx} \preccurlyeq L_{xy}^2 {P_{vy}}^{\mathrm{T}} W_y P_{vy}, \\ {P_{Hy}}^{\mathrm{T}} W_y P_{Hy} \preccurlyeq L_{xy}^2 {P_{vx}}^{\mathrm{T}} W_x P_{vx}, \\ Tr(W_x A_N^x) + Tr(W_y A_N^y) \leqslant R^2, \\ \mathrm{Rank}(W_x) \leqslant d_x, \\ \mathrm{Rank}(W_y) \leqslant d_y. \end{gathered}$$

Problem (Gram-PEP) is still not convex because of the rank constraints. From the rank constraints, it follows that the optimal value of PEP is a nondecreasing function of d_x, dy, that stops growing when $\min(d_x, d_y)$ reaches $3N + 3$.

Using the assumption that problem is high-dimensional, we can ditch the rank constraints for Gram matrices. In this way, the optimal value of (Gram-PEP) remains the same for functions with $\min(d_x, d_y) \geqslant 3N + 3$ and thus, the resulting bound remains tight for those functions. The bound will also hold for functions with $\min(d_x, d_y) < 3N + 3$, but tightness is no longer guaranteed, which we can accept in accordance with high-dimensional assumption. Getting rid of rank constraints turns the program into a convex SDP:

$$\sup_{W_x, W_y, \mathbf{f}^x, \mathbf{f}^y} Tr(W_x A_{N+1}^x) + Tr(W_y A_{N+1}^y) \tag{SDP-PEP}$$

such that $\forall i, j \in I$

$$0 \leqslant {m_{ij}^x}^{\mathrm{T}} \mathbf{f}^x + Tr(W_x M_{ij}^x),$$

$$0 \leqslant {m_{ij}^{y}}^{\mathrm{T}} \mathbf{f}^y + Tr(W_y M_{ij}^y),$$
$$P_{Hx}{}^{\mathrm{T}} W_x P_{vx} = P_{vy}{}^{\mathrm{T}} W_y P_{Hy},$$
$$P_{Hx}{}^{\mathrm{T}} W_x P_{Hx} \preccurlyeq L_{xy}^2 P_{vy}{}^{\mathrm{T}} W_y P_{vy},$$
$$P_{Hy}{}^{\mathrm{T}} W_y P_{Hy} \preccurlyeq L_{xy}^2 P_{vx}{}^{\mathrm{T}} W_x P_{vx},$$
$$Tr(W_x A_N^x) + Tr(W_y A_N^y) \leqslant R^2.$$

Since we aim to get exact upper bounds on (M) worst-case performance, it would be convenient [10, 22] to dualize (SDP-PEP):

$$\inf_{\lambda_{ij},\, \overline{\lambda}_{ij},\, \tau,\, C_1,\, C_2,\, C_3} \tau R^2 \tag{dual-PEP}$$

such that

$$\lambda_{ij},\, \overline{\lambda}_{ij} \geqslant 0 \quad \forall i,\, j \in I,$$
$$\tau \geqslant 0, \quad C_2,\, C_3 \succcurlyeq 0,$$
$$\sum_{i,j \in I} \lambda_{ij} M_{ij}^x - P_{vx} C_1 P_{Hx}{}^{\mathrm{T}} - P_{Hx} C_2 P_{Hx}{}^{\mathrm{T}} + L_{xy}^2 P_{vx} C_3 P_{vx}{}^{\mathrm{T}} - \tau A_N^x + A_{N+1}^x \preccurlyeq 0,$$
$$\sum_{i,j \in I} \overline{\lambda}_{ij} M_{ij}^y + P_{Hy} C_1 P_{vy}{}^{\mathrm{T}} + L_{xy}^2 P_{vy} C_2 P_{vy}{}^{\mathrm{T}} - P_{Hy} C_3 P_{Hy}{}^{\mathrm{T}} - \tau A_N^y + A_{N+1}^y \preccurlyeq 0,$$
$$\sum_{i,j \in I} \lambda_{ij} m_{ij}^x = 0, \quad \sum_{i,j \in I} \overline{\lambda}_{ij} m_{ij}^y = 0.$$

We can set $R = 1$ due to homogeneity. The main advantage of dual-PEP form is that besides the upper bound for the worst-case performance, solution of dual-PEP essentially yields a proof of the said guarantee [22]. We are omitting a proof for strong duality between (SDP-PEP) and (dual-PEP) for the sake of being concise and also note that with PEP, strong duality tends to hold for all methods that satisfy very mild and reasonable assumptions [8, 10].

3.2. PEP experiment

We now consider two specific methods, Sim-GDA:

$$x_{k+1} = x_k - \eta_x \nabla_x f(x_k,\, y_k),$$
$$y_{k+1} = y_k + \eta_y \nabla_x f(x_k,\, y_k)$$

and Alt-GDA:

$$x_{k+1} = x_k - \eta_x \nabla_x f(x_k, ,\, y_k),$$
$$y_{k+1} = y_k + \eta_y \nabla_x f(x_{k+1},\, y_k).$$

Both methods are known to converge linearly (globally) on the class of bilinear composites with iteration complexity $O(k^2)$ for Sim-GDA and $O(k)$ for Alt-GDA where $k = \frac{\max(L_x, L_y, L_{xy})}{\min(\mu_x, \mu_y)}$ is a condition number [4].

We test (dual-PEP) by solving it for Sim-GDA and Alt-GDA with $||z - z_*||$ as a metric of choice. In this way (dual-PEP) basically turns into a standard PEP for 1 iteration. In this

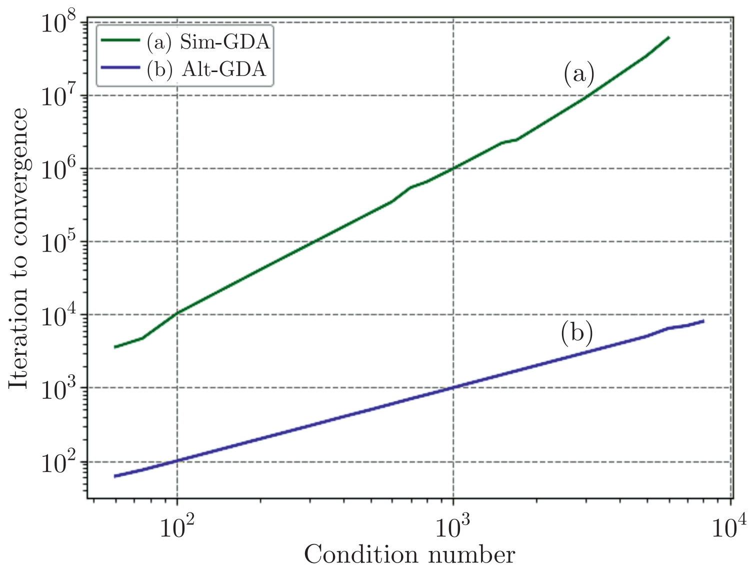

Fig. 1. Verified iteration complexities for Sim-GDA (a) and Alt-GDA (b)

test we let $L_x = L_y = L_{xy} = L$ and $\mu_x = \mu_y = \mu$ and define the condition number $k = \frac{L}{\mu}$. We used grid search to find the best step size for every condition number. Iteration complexity is calculated as $N = -\frac{1}{\log(\rho)} = -\frac{2}{\log(\tau)}$.

The results for iteration complexities are displayed in Fig. 1, they are in agreement with iteration complexities that were acquired with IQC for this same setting [4]. Although the specific PEP example that is presented in the paper does not yield the exact worst-case bounds for a set number of iterations, it was still enough to quickly estimate and compare methods' iteration complexities. We leave thorough analysis of methods' convergence rates via PEP for future research.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., Generative Adversarial Networks, *Commun. ACM,* 2020, vol. 63, no. 11, pp. 139–144.

[2] Li, S., Wu, Y., Cui, X., Dong, H., Fang, F., and Russell, S., Robust Multi-Agent Reinforcement Learning via Minimax Deep Deterministic Policy Gradient, *Proc. AAAI Conf. on Artificial Intelligence,* 2019, vol. 33, no. 01, pp. 4213–4220.

[3] Zhang, G., Bao, X., Lessard, L., and Grosse, R., A Unified Analysis of First-Order Methods for Smooth Games via Integral Quadratic Constraints, *J. Mach. Learn. Res.,* 2021, vol. 22 (103), pp. 1–39.

[4] Zhang, G., Wang, Y., Lessard, L., and Grosse, R., Near-Optimal Local Convergence of Alternating Gradient Descent-Ascent for Minimax Optimization, in *Proc. of the 25th Internat. Conf. on Artificial Intelligence and Statistics (PMLR, Valencia, Spain, Mar 2022),* pp. 7659–7679.

[5] Lee, J., Cho, H., and Yun, C., Fundamental Benefit of Alternating Updates in Minimax Optimization, arXiv:2402.10475 (2024).

[6] Drori, Y. and Teboulle, M., Performance of First-Order Methods for Smooth Convex Minimization: A Novel Approach, *Math. Program.,* 2014, vol. 145, nos. 1–2, pp. 451–482.

[7] Lessard, L., Recht, B., and Packard, A., Analysis and Design of Optimization Algorithms via Integral Quadratic Constraints, *SIAM J. Optim.,* 2016, vol. 26, no. 1, pp. 57–95.

[8] Taylor, A., Van Scoy, B., and Lessard, L., Lyapunov Functions for First-Order Methods: Tight Automated Convergence Guarantees, in *Proc. of the 35th Internat. Conf. on Machine Learning (Stockholm, Sweden, Jul 2018),* pp. 4897–4906.

[9] Rubbens, A., Bousselmi, N., Colla, S., and Hendrickx, J. M., Interpolation Constraints for Computing Worst-Case Bounds in Performance Estimation Problems, in *Proc. of the 62nd IEEE Conf. on Decision and Control (CDC, Marina Bay Sands, Singapore, Dec 2023),* pp. 3015–3022.

[10] Taylor, A. B., Hendrickx, J. M., and Glineur, F., Smooth Strongly Convex Interpolation and Exact Worst-Case Performance of First-Order Methods, *Math. Program.,* 2017, vol. 161, nos. 1–2, pp. 307–345.

[11] Kim, D. and Fessler, J. A., Optimized First-Order Methods for Smooth Convex Minimization, *Math. Program.,* 2016, vol. 159, nos. 1–2, pp. 81–107.

[12] Zamani, M., Abbaszadehpeivasti, H., and de Klerk, E., Convergence Rate Analysis of the Gradient Descent-Ascent Method for Convex-Concave Saddle-Point Problems, *Optim. Methods Softw.,* 2024, vol. 39, no. 5, pp. 967–989.

[13] Dragomir, R. A., Taylor, A. B., d'Aspremont, A., and Bolte, J., Optimal Complexity and Certification of Bregman First-Order Methods, *Math. Program.,* 2022, vol. 194, nos. 1–2, pp. 41–83.

[14] Rockafellar, R. T., Characterization of the Subdifferentials of Convex Functions, *Pacific J. Math.,* 1966, vol. 17, no. 3, pp. 497–510.

[15] Artstein-Avidan, S., Sadovsky, S., and Wyczesany, K., A Rockafellar-Type Theorem for Non-Traditional Costs, *Adv. Math.,* 2022, vol. 395, 108157, 22 pp.

[16] Bousselmi, N., Hendrickx, J. M., and Glineur, F., Interpolation Conditions for Linear Operators and Applications to Performance Estimation Problems, *SIAM J. Optim.,* 2024, vol. 34, no. 3, pp. 3033–3063.

[17] Rubbens, A. and Hendrickx, J. M., A Constraint-Based Approach to Function Interpolation, with Application to Performance Estimation for Weakly Convex Optimisation, arXiv:2405.08405 (2024).

[18] Zhang, Y. and Xiao, L., Stochastic Primal-Dual Coordinate Method for Regularized Empirical Risk Minimization, *J. Mach. Learn. Res.,* 2017, vol. 18, no. 1, pp. 2939–2980.

[19] Du, S. S., Chen, J., Li, L., Xiao, L., and Zhou, D., Stochastic Variance Reduction Methods for Policy Evaluation, in *Proc. of the 34th Internat. Conf. on Machine Learning (ICML, Sydney, NSW, Australia, Aug 2017),* pp. 1049–1058.

[20] Salim, A., Condat, L., Kovalev, D., and Richtárik, P., An Optimal Algorithm for Strongly Convex Minimization under Affine Constraints, in *Proc. of the 25th Internat. Conf. on Artificial Intelligence and Statistics (AISTATS, Valencia, Spain, Mar 2022),* pp. 4482–4498.

[21] Kovalev, D., Gasnikov, A., and Richtárik, P., Accelerated Primal-Dual Gradient Method for Smooth and Convex-Concave Saddle-Point Problems with Bilinear Coupling, in *Proc. of the 36th Internat. Conf. on Neural Information Processing Systems (NIPS, New Orleans, La., USA, 2022),* Art. 1579, 13 pp.

[22] Goujaud, B., Dieuleveut, A., and Taylor, A., On Fundamental Proof Structures in First-Order Optimization, in *Proc. of the 62nd IEEE Conf. on Decision and Control (CDC, Marina Bay Sands, Singapore, Dec 2023),* pp. 3023–3030.