

On the Convex Interpolation for Linear Operators

Nizar Bousselmi^{1*†}, Zhicheng Deng^{2*†}, Jie Lu²,
François Glineur^{1,3}, Julien M. Hendrickx¹

¹ICTEAM, UCLouvain, Louvain-la-Neuve, Belgium.

²School of Information Science and Technology, Shanghaitech
University, Shanghai, P. R. China.

³CORE, UCLouvain, Louvain-la-Neuve, Belgium.

*Corresponding author(s). E-mail(s): nizar.bousselmi@uclouvain.be;
dengzhch@shanghaitech.edu.cn;

Contributing authors: lujie@shanghaitech.edu.cn;
francois.glineur@uclouvain.be; julien.hendrickx@uclouvain.be;

[†]These authors contributed equally to this work.

Abstract

The worst-case performance of an optimization method on a problem class can be analyzed using a finite description of the problem class, known as interpolation conditions. In this work, we study interpolation conditions for linear operators given scalar products between discrete inputs and outputs. First, we show that if only convex constraints on the scalar products of inputs and outputs are allowed, it is only possible to characterize classes of linear operators or symmetric linear operator whose all singular values or eigenvalues belong to some subset of $\mathbb{R}$. Then, we propose new interpolation conditions for linear operators with minimal and maximal singular values and linear operators whose eigenvalues or singular values belong to unions of subsets. Finally, we illustrate the new interpolation conditions through the analysis of the Gradient and Chambolle-Pock methods. It allows to obtain new numerical worst-case guarantees on these methods.

Keywords: Linear operator, Convex interpolation, Worst-case analysis, Performance estimation problem

1 Introduction

One recent way to analyze the worst-case performance of a given optimization method on a given function class is the well-developed Performance Estimation Problem (PEP) framework [1, 2]. It formulates the problem of computing the worst-case performance of the method on the class as an optimization problem and then solves it. The framework relies on interpolation conditions of the function classes considered [3, 4] (such interpolation conditions can serve larger purposes). Worst-case guarantees are built up by combining inequalities between specific points of the domain characterizing the function class. In some sense, interpolation conditions of a function class are the best description of this class.

Analyzing the performance of optimization methods on problems of the form $\min_x g(Mx)$ or $\min_x f(x) + g(Mx)$ where $f \in \mathcal{F}$, $g \in \mathcal{G}$, and $M \in \mathcal{M}$ with PEP requires interpolation conditions for the function classes $\mathcal{F}$ and $\mathcal{G}$ and linear operator class $\mathcal{M}$ (see [2, 5] for more details). We will focus on the interpolation of linear operators defined as follows.

Definition 1.1 (Operator interpolability). *Given a class of linear operators $\mathcal{M}$, the sets of pairs $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ are $\mathcal{M}$ -interpolable if*

$$\exists M \in \mathcal{M} : \begin{cases} y_i = Mx_i, & \forall i \in [N_1] \\ v_j = M^T u_j, & \forall j \in [N_2] \end{cases} \quad (1)$$

While given sets of points $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$, this question can often easily be solved by computing M explicitly, the PEP methodology requires an explicit description of the whose sets of points for which the desired interpolable property holds. Interpolation conditions are necessary and sufficient constraints on the set of points $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ guaranteeing their interpolability. A previous work [5], developed interpolation conditions for the classes of linear operators with singular values in $[0, L]$, $\mathcal{L}_{[0, L]}$, and for symmetric linear operators with eigenvalues in $[\mu, L]$, $\mathcal{S}_{[\mu, L]}$, with $\mu \leq L$. Formally, given subset $S \subseteq \mathbb{R}$, we denote

$$\mathcal{L}_S = \{M : \sigma_i(M) \in S, \forall i\} \quad (2)$$

$$\mathcal{S}_S = \{M \text{ is symmetric} : \lambda_i(M) \in S, \forall i\} \quad (3)$$

where $\sigma_i(M)$ and $\lambda_i(M)$ are the singular values and eigenvalues of M .

Given the sets of pairs $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ where $x_i, u_j \in \mathbb{R}^n$, $y_i, v_j \in \mathbb{R}^m$, we denote $X = (x_1 \cdots x_{N_1})$, $Y = (y_1 \cdots y_{N_1})$, $U = (u_1 \cdots u_{N_2})$, $V = (v_1 \cdots v_{N_2})$. Interpolation conditions for $\mathcal{L}_{[0, L]}$ and $\mathcal{S}_{[\mu, L]}$ -interpolability are the following [5] (there also exist interpolation conditions for skew-symmetric linear operators [5, Corollary 3.2]).

Theorem 1.1 ($\mathcal{L}_{[0,L]}$ -interpolation conditions, Theorem 3.1 of [5]). *Given $L \geq 0$. The sets of pairs $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ are $\mathcal{L}_{[0,L]}$ -interpolable if, and only if,*

$$\begin{cases} X^T V = Y^T U, \\ Y^T Y \preceq L^2 X^T X, \\ V^T V \preceq L^2 U^T U. \end{cases} \quad (4)$$

Theorem 1.2 ($\mathcal{S}_{[\mu,L]}$ -interpolation conditions, Theorem 3.3 of [5]). *Given $\mu \leq L$. The set $\{(x_i, y_i)\}_{i \in [N_1]}$ is $\mathcal{S}_{[\mu,L]}$ -interpolable if, and only if,*

$$\begin{cases} X^T Y = Y^T X, \\ (Y - \mu X)^T (LX - Y) \succeq 0. \end{cases} \quad (5)$$

Thanks to these interpolation conditions and PEP, new worst-case guarantees were obtained on the Gradient and Chambolle-Pock methods [5, Section 4], Condat-Vũ methods [6], decentralized optimization [7], min-max optimization [8, 9] and parametric quadratic optimization [10].

Worst-case analysis and Gram representation

Both classical worst-case analysis of optimization methods and PEP require algebraic characterizations of function classes [11, 12], to obtain guaranteed tight and unimprovable bounds. In this work, we focus on the PEP methodology. It appears that in a lot of cases, the problem of computing the worst-case performance of a given method on a given class of functions can be formulated as a convex problem [2]. A necessary condition for the PEP to be efficiently exploitable is that the interpolation conditions must be convex on the scalar products between points. We discuss interpolation conditions for linear operators. This convexity must hold on the elements of the following pair of Gram matrices when considering classes of general linear operators:

$$(G, H) = (\text{Gram}(X \ V), \text{Gram}(Y \ U)) \triangleq ((X \ V)^T (X \ V), (Y \ U)^T (Y \ U)) \quad (6)$$

and the following Gram matrix when considering classes of (square) symmetric linear operators

$$G = \text{Gram}(X \ Y \ U \ V) \triangleq (X \ Y \ U \ V)^T (X \ Y \ U \ V) \quad (7)$$

i.e., all the scalar products whose existence is compatible with the dimensions. For instance, Theorem 1.1 (resp. 1.2) can be written as convex semi-definite constraints on G and H (resp. G).

In this work, we focus on interpolation conditions which are convex on the above Gram matrices G and H . We tackle the following two questions related to operator interpolability:

1. Which classes of linear operators can be represented by convex interpolation conditions on products between the points of the sets $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$?
2. Is there a tractable form of interpolation conditions for $\mathcal{L}_{[\mu,L]}$ when $\mu > 0$?

We here propose answers to these questions.

Contributions and outline of the paper

In Section 2, we show that when restricted to convex constraints on the Gram matrices of the inputs and outputs, we can only characterize classes of linear and symmetric linear operators of the forms (2) and (3) (i.e., classes of linear operators for which the singular values or eigenvalues belong to a given set) (Theorem 2.4). For example, it is impossible to characterize the class of symmetric linear operators for which the sum of singular values is equal to a given number with convex constraints on the scalar products. In Section 3, we present interpolation conditions for the class of linear operators $\mathcal{L}_{[\mu,L]}$ (Theorem 1.1). To obtain them, we rely on the polar decomposition that allows to express any linear operator as a product of a unitary and a symmetric linear operator for which we already knew interpolation conditions. In Section 4, we propose interpolation conditions for linear operators whose singular values or eigenvalues belong to a union of subsets, extending thus Theorem 1.2 and 1.1, and approaching the limitations derived in Section 2. In Section 5, we illustrate the application of our new interpolation conditions to analyze the worst-case behavior of Gradient and Chambolle-Pock methods [13] with the PEP framework. We provide new numerical worst-case guarantees.

Notations and conventions

In this paper, we denote by $\mathcal{L}$ and $\mathcal{S}$ the sets of all linear operators and symmetric linear operators of any finite dimensions. For a nonnegative integer N , let $[N]$ denote the set $\{1, \dots, N\}$. We adopt the following conventions. Let X, Y, U, V be matrices with columns x_i, y_i, u_j, v_j respectively. In the general (rectangular) case, x_i, v_i lie in one space and y_i, u_i lie in a different space, therefore, we have access to the pair of Gram matrices (6). In the (square) symmetric case, all vectors x_i, y_i, u_j, v_j lie in the same space, therefore, we have access to the Gram matrix (7).

2 Limitation of representation of linear operators through Gram matrices

In this section, we present a limitation of representation on classes of linear operators when restricted to convex constraints on the Gram matrices of inputs and outputs of the linear operators. We first introduce some key concepts in the Gram matrices associated with the linear operators in Section 2.1, followed by our main result on the limitation of Gram representation in Section 2.2. The relationship between the Gram representation and the interpolation condition is clarified in Section 2.3.

2.1 Gram matrices associated with classes of linear operators

Definition 2.1 (Gram matrix associated with a linear operator). *Given a matrix G (resp. a pair of Gram matrices (G, H)), we say that G (resp. (G, H)) is associated*

with M if

$$\exists X, Y, U, V : \begin{cases} (G, H) = (\text{Gram}(X \ Y), \text{Gram}(Y \ U)) \quad (\text{resp. } \text{Gram}(X \ Y \ U \ V)) \\ Y = MX \\ V = M^T U. \end{cases} \quad (8)$$

Definition 2.2 (Set of Gram matrices associated with a class of linear operator). *Let classes of linear operators $\mathcal{M} \subseteq \mathcal{L}$ and square linear operators $\mathcal{Q} \subseteq \mathcal{S}$, the set of Gram matrices associated with $\mathcal{M}$ and $\mathcal{Q}$ are*

$$\begin{aligned} \mathcal{P}_{\mathcal{M}}^N &\triangleq \{(G, H) = (\text{Gram}(X \ V), \text{Gram}(Y \ U)) : \exists M \in \mathcal{M} : Y = MX, V = M^T U\}, \\ \mathcal{G}_{\mathcal{Q}}^N &\triangleq \{G = \text{Gram}(X \ Y \ U \ V) : \exists M \in \mathcal{Q} : Y = MX, V = M^T U\}. \end{aligned}$$

All the Gram matrices of a set have the same dimension, i.e., $(G, H) \in \mathbb{R}^{2N \times 2N} \times \mathbb{R}^{2N \times 2N}$ (resp. $G \in \mathbb{R}^{4N \times 4N}$). We often omit the superscript N in the notations $\mathcal{P}_{\mathcal{M}}^N$ and $\mathcal{G}_{\mathcal{Q}}^N$.

A central property of Gram matrices is their unitary invariance, namely, rotating the vectors composing the Gram matrices does not modify them. Formally, given any set of vectors $A = (a_1 \ \cdots \ a_n)$, we have

$$\text{Gram}(A) = \text{Gram}(RA) \quad (9)$$

for all unitary transformations R . Therefore, Gram matrices associated with a linear operator (resp. symmetric linear operator) M are also associated with all linear operators of the form RMS^T (resp. $R^T MR$).

Lemma 2.1 (Unitary invariance of Gram matrices). *Let R and S be any unitary transformations.*

1. *If a Gram matrix G is associated with a symmetric linear operator M then it is also associated with any linear operator $RM R^T$.*
2. *If a pair of Gram matrices (G, H) is associated with a linear operator M then (G, H) is associated with any linear operator RMS^T .*

Proof. For statement 1, we have

$$G = \text{Gram}(X \ Y \ U \ V) \stackrel{(9)}{=} \text{Gram}(R(X \ Y \ U \ V)) = \text{Gram}(X' \ Y' \ U' \ V') \quad (10)$$

where $Y' = RMR^T X'$, $V' = R^T M^T R U'$ for any unitary transformations R . And for statement 2, we have

$$\begin{cases} G = \text{Gram}(X \ V) \stackrel{(9)}{=} \text{Gram}(S(X \ V)) = \text{Gram}(X' \ V') \\ H = \text{Gram}(Y \ U) \stackrel{(9)}{=} \text{Gram}(R(Y \ U)) = \text{Gram}(Y' \ U') \end{cases} \quad (11)$$

where $Y' = RMS^T X'$, $V' = SM^T R^T U'$ for any unitary transformations R and S . $\square$

Therefore, Gram matrices associated with a general (resp. symmetric) linear operator M are also associated with the diagonalized version $M' = RMS^T$ (resp. $M' = RMR^T$) whose diagonal elements are the singular values of M (resp. eigenvalues of symmetric linear operator M) in any order. This implies that the Gram representation only allows characterizing the spectrum of linear operators M , namely, the values and multiplicity of each of their singular values (or eigenvalues for symmetric linear operators), in a sense that will be made clear in Corollary 2.2.

We define spectrums using multiset which is a set where each element has a multiplicity. We use the usual inclusion, sum, and difference of multisets.

Definition 2.3 (Spectrum of matrices). *Let a matrix $M \in \mathcal{L}$ and a square matrix $Q \subseteq \mathcal{S}$. The singular value spectrum of M is the following multiset of $\mathbb{R}$*

$$\sigma(M) \triangleq \{\{ \text{singular values of } M \}\} \quad (12)$$

and, the eigenvalue spectrum of Q is the following multiset of $\mathbb{C}$

$$\lambda(Q) \triangleq \{\{ \text{eigenvalues of } Q \}\}. \quad (13)$$

We will typically use λ to denote an arbitrary multiset of complex or real values with finitely many elements (meant to represent spectrums), and Λ to denote a set of spectrums λ . We denote the set of (eigenvalue or singular value) spectrums whose elements belong to a given subset S by

$$\Delta(S) \triangleq \{\lambda : \forall a \in \lambda, a \in S\}. \quad (14)$$

λ is a multiset and considers the multiplicity of singular values and eigenvalues. We use a multiset and not a sequence since matrices with the same singular values or eigenvalues but not in the same order are the same up to a unitary transformation (and have thus the same Gram matrix by Lemma 2.1).

We can now define Gram matrices associated with spectrums. More precisely, given a set of spectrums Λ , the set of Gram matrices associated with Λ is

$$\mathcal{P}_\Lambda \triangleq \{(G, H) = (\text{Gram}(X \ V), \text{Gram}(Y \ U)) : \exists M : Y = MX, V = M^T U, \text{ and } \sigma(M) \in \Lambda\} \quad (15)$$

and similarly, for symmetric linear operators, we have

$$\mathcal{G}_\Lambda \triangleq \{G = \text{Gram}(X \ Y \ U \ V) : \exists M \text{ symmetric} : Y = MX, V = M^T U, \text{ and } \lambda(M) \in \Lambda\} \quad (16)$$

Due to Lemma 2.1 and the following corollary, we focus on spaces of Gram matrices associated with a set of singular or eigenvalue spectrum Λ .

Corollary 2.2 (Spectrum characterization). *Let $\mathcal{M}$ be a set of linear operators (resp. symmetric linear operators) of any finite dimension. We have $\mathcal{P}_\mathcal{M} = \mathcal{P}_\Lambda$ (resp. $\mathcal{G}_\mathcal{M} = \mathcal{G}_\Lambda$) where $\Lambda = \{\sigma(M) : M \in \mathcal{M}\}$ (resp. $\Lambda = \{\lambda(M) : M \in \mathcal{M}\}$).*

Proof. Thanks to Lemma 2.1, a pair of Gram matrix (G, H) (resp. G) is also associated with its orthogonally diagonalized version and reciprocally, therefore, $(G, H) \in \mathcal{P}_{\mathcal{M}} \Leftrightarrow (G, H) \in \mathcal{P}_{\Lambda}$ (resp. $G \in \mathcal{G}_{\mathcal{M}} \Leftrightarrow G \in \mathcal{G}_{\Lambda}$). $\square$

2.2 Main result

We can now show the two main results of the section, namely, (i) sets of Gram matrices, $\mathcal{P}_{\Delta(S)}$ and $\mathcal{G}_{\Delta(S)}$ associated with linear operators with all singular values in a subset $S \subseteq \mathbb{R}$ or with symmetric linear operators with all eigenvalues in a subset $S \subseteq \mathbb{R}$ are convex, and reciprocally, (ii) if $\mathcal{P}_{\mathcal{M}}$ or $\mathcal{G}_{\mathcal{M}}$ are convex, then they are associated with classes of linear operators and symmetric linear operators with singular values in some subset $S \subseteq \mathbb{R}$ or eigenvalues in some subset $S \subseteq \mathbb{R}$.

Theorem 2.3. *Let $S \subseteq \mathbb{R}$. $\mathcal{P}_{\Delta(S)}$ (resp. $\mathcal{G}_{\Delta(S)}$) is convex.*

Proof. We show that $\mathcal{P}_{\Delta(S)}$ (resp. $\mathcal{G}_{\Delta(S)}$) are convex cones.

Let $G_i \in \mathcal{G}_{\Delta(S)}$ (resp. $(G_i, H_i) \in \mathcal{P}_{\Delta(S)}$) for $i = 1, 2$. We have $G_i = \text{Gram}(X_i \ Y_i \ U_i \ V_i)$ (resp. $(G_i, H_i) = (\text{Gram}(X_i \ V_i), \text{Gram}(Y_i \ U_i))$) with $Y_i = M_i X_i, V_i = M_i^T U_i$ and $\lambda(M_i) \in \Delta(S)$ (resp. $\sigma(M_i) \in \Delta(S)$). Let $(X \ Y \ U \ V) \triangleq \begin{pmatrix} \sqrt{c_1}(X_1 \ Y_1 \ U_1 \ V_1) \\ \sqrt{c_2}(X_2 \ Y_2 \ U_2 \ V_2) \end{pmatrix}$ and $M \triangleq \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}$. One can directly verify that $c_1 G_1 + c_2 G_2 = \text{Gram}(X \ Y \ U \ V)$ (resp. $c_1(G_1, H_1) + c_2(G_2, H_2) = (\text{Gram}(X \ V), \text{Gram}(Y \ U))$, $Y = MX$, $V = M^T U$, $\lambda(M) \in \Delta(S)$ (resp. $\sigma(M) \in \Delta(S)$), and therefore $c_1 G_1 + c_2 G_2 \in \mathcal{G}_{\Delta(S)}$ (resp. $c_1(G_1, H_1) + c_2(G_2, H_2) \in \mathcal{P}_{\Delta(S)}$). $\square$

Theorem 2.4. *Let $\mathcal{M}$ be a set of linear operators (resp. symmetric linear operators). If $\mathcal{P}_{\mathcal{M}}$ (resp. $\mathcal{G}_{\mathcal{M}}$) is convex, then*

$$\mathcal{P}_{\mathcal{M}} = \mathcal{P}_{\Delta(S)} \text{ (resp. } \mathcal{G}_{\mathcal{M}} = \mathcal{G}_{\Delta(S)}) \quad (17)$$

for some $S \subseteq \mathbb{R}$.

Proof. The proof is deferred to Appendix A. $\square$

Observe that it is not possible to control the multiplicity of the singular values or eigenvalues contained in the spectrums of linear operators represented by $\mathcal{P}_{\Delta(S)}$ and $\mathcal{G}_{\Delta(S)}$.

Remark 1. When M is symmetric, $\mathcal{P}_{\mathcal{M}}$ is a projection of $\mathcal{G}_{\mathcal{M}}$, which can be seen from (6) and (7). If $\mathcal{G}_{\mathcal{M}}$ is a convex set, it follows from Theorem 2.4 that $\mathcal{G}_{\mathcal{M}} = \mathcal{G}_{\Delta(S)}$ for some S . Consequently, $\mathcal{P}_{\mathcal{M}}$, being a projection of $\mathcal{G}_{\mathcal{M}}$, is convex, and therefore, $\mathcal{P}_{\mathcal{M}} = \mathcal{P}_{\Delta(T)}$ for some T . Recall that $\mathcal{G}_{\Delta(S)}$ is only characterized by the locations of eigenvalues whereas $\mathcal{P}_{\Delta(T)}$ is only characterized by the locations of singular values. No inconsistency arises since symmetric matrices have the singular values that are the absolute values of their eigenvalues. Therefore, $T = \{|\lambda| : \lambda \in S\}$. Finally, note that $\mathcal{P}_{\Delta(T)}$ is associated with a larger set of linear operators than $\mathcal{G}_{\Delta(S)}$. This is natural since $\mathcal{P}_{\mathcal{M}}$ is merely a projection of $\mathcal{G}_{\mathcal{M}}$.

Remark 2. In this section, we considered Gram matrices constructed with vectors involved by the linear operators. However, we should have also considered other vectors

independent of the linear operator to be complete. In other words, instead of (7) and (6), we should use

$$\bar{G} = \text{Gram}(X \ Y \ U \ V \ W) \quad (18)$$

and

$$(\bar{G}, \bar{H}) = (\text{Gram}(X \ V \ W), (Y \ U \ Z)) \quad (19)$$

where $Y = MX$ and $V = M^T U$ for some linear operator M . Note that since G in (7) (resp. (G, H) in (6)) is a projection of $\bar{G}$ (resp. $(\bar{G}, \bar{H})$), if

$$\begin{aligned} \bar{\mathcal{P}}_{\mathcal{M}} &\triangleq \{(\bar{G}, \bar{H}) = (\text{Gram}(X \ V \ W), (Y \ U \ Z)) : \exists M \in \mathcal{M}, W, Z : Y = MX, V = M^T U\}, \\ \bar{\mathcal{G}}_{\mathcal{Q}} &\triangleq \{\bar{G} = \text{Gram}(X \ Y \ U \ V \ W) : \exists M \in \mathcal{Q}, W : Y = MX, V = M^T U\}. \end{aligned}$$

is convex, their projections $\mathcal{P}_{\mathcal{M}}$ and $\mathcal{G}_{\mathcal{Q}}$ are convex. All the limitation results also hold for extended Gram matrices.

2.3 Set of Gram matrices and interpolation conditions

In this section, we examine the interpolation conditions for the input-output quadruple (X, Y, U, V) and how they relate to the main limitation result concerning Gram representation.

Although the quadruple contains more information than its Gram matrix G (or the pair (G, H)), if one focus only on the locations of eigenvalues or singular values of the linear operators to be interpolated, ignoring multiplicities, then interpolation conditions for (X, Y, U, V) can be derived exclusively from their Gram representations G (or (G, H)).

Theorem 2.5. *Let $S \subseteq \mathbb{R}$. (X, Y, U, V) is $\mathcal{S}_S$ -interpolable (resp. $\mathcal{L}_S$ -interpolable) if, and only if, $\text{Gram}(X \ Y \ U \ V) \in \mathcal{G}_{\Delta(S)}$ (resp. $(\text{Gram}(X \ V), \text{Gram}(Y \ V)) \in \mathcal{P}_{\Delta(S)}$).*

Proof. (Necessity) Let (X, Y, U, V) being $\mathcal{G}_{\Delta(S)}$ -interpolable, then $\exists M$ such that $\lambda(M) \in \Delta(S)$, $Y = MX$, $V = M^T U$ and $G = \text{Gram}(X \ Y \ U \ V) \in \mathcal{G}_{\Delta(S)}$.

(Sufficiency) Let $G \in \mathcal{G}_{\Delta(S)}$, (resp. $(G, H) \in \mathcal{P}_{\Delta(S)}$) then there exists X', Y', U', V', M' such that

$$\begin{aligned} Y' &= M' X', V' = (M')^T U', G = \text{Gram}(X' \ Y' \ U' \ V') \text{ and } \lambda(M') \in \Delta(S). \\ (\text{resp. } G &= \text{Gram}(X' \ V'), H = \text{Gram}(Y' \ U') \text{ and } \sigma(M') \in \Delta(S).) \end{aligned}$$

Therefore, (X, Y, U, V) (resp. $(X \ V), (Y \ U)$) and (X', Y', U', V') (resp. $(X' \ V'), (Y' \ U')$) build the same Gram matrix, i.e., $G = \text{Gram}(X \ Y \ U \ V) = \text{Gram}(X' \ Y' \ U' \ V')$ (resp. $G = \text{Gram}(X \ Y) = \text{Gram}(X' \ Y'), H = \text{Gram}(U \ V) = \text{Gram}(U' \ V')$).

Without loss of generality, we may assume X', Y', U', V' have no more rows than X, Y, U, V respectively. By Theorem 3.2, there exists an isometric matrix R (resp. R, S) such that $RX' = X$, $RY' = Y$, $RU' = U$, $RV' = V$ (resp. $Y = SY'$, $U = SU'$), then $X' = R^T X$, $U' = R^T U$, $Y = RMR^T X$ and $V = R(M')^T R^T U$ (resp. $Y = SM'R^T X$, $V = S(M')^T U$). Since $RM'R^T \in \mathcal{S}_S$ (resp. $SM'R^T \in \mathcal{L}_S$), we have (X, Y, U, V) is $\mathcal{S}_S$ -interpolable (resp. $\mathcal{L}_S$ -interpolable). $\square$

Combined with our main results, if the sets of Gram matrices $\mathcal{G}_{\mathcal{M}}$ or $\mathcal{P}_{\mathcal{M}}$ over the class of linear operators $\mathcal{M}$ are convex, then the Gram matrices inclusion tests provide interpolation conditions for (X, Y, U, V) over the same class of linear operators.

3 Interpolation conditions for $\mathcal{L}_{[\mu, L]}$

In this section, we propose interpolation conditions for the class $\mathcal{L}_{[\mu, L]}$ of linear operators whose singular values lie in the interval $[\mu, L]$. The idea is to exploit the polar decomposition to write any such operator as the product between a isometric transformation and a symmetric positive semidefinite linear operator whose eigenvalues also confined to $[\mu, L]$, for which we already have interpolation conditions ([14, Theorem 7.3.11] and [5, Theorem 3.3]).

The interpolation conditions for the class of unitary and isometric transformations are not new but we propose a new formulation based on Definition 1.1. A linear operator R is called isometric if $R^T R = I$. If, in addition, R also satisfies $RR^T = I$, then R is unitary. Note that an isometric operator need not have an isometric transpose: only when both $R^T R = I$ and $RR^T = I$ hold does R becomes unitary.

Definition 3.1 (Unitary and isometric interpolability). *Given two collections of data pairs $\{(x_i, y_i)\}_{i \in [N_1]} \subseteq \mathbb{R}^n \times \mathbb{R}^m$, $\{(u_j, v_j)\}_{j \in [N_2]} \subseteq \mathbb{R}^m \times \mathbb{R}^n$, we say that the pairs are unitary-interpolable (resp. isometric-interpolable) if there exists a unitary (resp. an isometric) linear operator R such that*

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

Lemma 3.1 (Unitary interpolation conditions). *The sets of pairs $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ are unitary-interpolable if, and only if,*

$$\begin{cases} Y^T Y = X^T X, \\ Y^T U = X^T V, \\ U^T U = V^T V. \end{cases} \quad (20)$$

Proof. We exploit the fact that the transpose of a unitary matrix is its inverse. $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ are unitary-interpolable if and only if there exist an unitary matrix $R \in \mathbb{R}^{n \times n}$, such that $Y = RX$ and $U = RV$. It is equivalent to $(Y \ U)^T (Y \ U) = (X \ V)^T (X \ V)$ by Theorem 3.2. $\square$

Since the following fact on Gram matrices and isometric linear operator is invoked repeatedly in the subsequent proofs, we recall it here for clarity and readability.

Lemma 3.2 (Theorem 7.3.11 of [14]). *Let N, m, n be positive integers with $m \leq n$. Let $A \in \mathbb{R}^{m \times N}$ and $B \in \mathbb{R}^{n \times N}$. Then $A^T A = B^T B$ if and only if there is an isometric matrix $R \in \mathbb{R}^{n \times m}$ such that $B = RA$.*

Proposition 3.3 (Isometric interpolation conditions). *The sets of pairs $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ are isometric-interpolable if, and only if,*

$$\begin{cases} Y^T Y = X^T X, \\ Y^T U = X^T V, \\ U^T U \succeq V^T V. \end{cases} \quad (21)$$

Proof. Given an isometric matrix R , we can find a unitary matrix $\bar{R}$ such that $R = \bar{R} \begin{pmatrix} I \\ 0 \end{pmatrix}$. Therefore, the sets of pairs $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ are isometric-interpolable if, and only if,

$$\begin{aligned} & \exists R \text{ isometric, } Y = RX \text{ and } V = R^T U \\ & \Leftrightarrow \exists \bar{R} \text{ unitary, } Y = \bar{R} \begin{pmatrix} I \\ 0 \end{pmatrix} X \text{ and } V = (I \ 0) \bar{R}^T U \\ & \Leftrightarrow \exists \bar{R} \text{ unitary, } Z, Y = \bar{R} \begin{pmatrix} X \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} V \\ Z \end{pmatrix} = \bar{R}^T U \\ & \stackrel{\text{Th. 3.1}}{\Leftrightarrow} \exists Z, \quad (Y \ U)^T (Y \ U) = \left(\begin{pmatrix} X \\ 0 \end{pmatrix} \begin{pmatrix} V \\ Z \end{pmatrix} \right)^T \left(\begin{pmatrix} X \\ 0 \end{pmatrix} \begin{pmatrix} V \\ Z \end{pmatrix} \right) \\ & \Leftrightarrow \exists Z, \quad \begin{cases} Y^T Y = X^T X, \\ Y^T U = X^T V, \\ U^T U = V^T V + Z^T Z, \end{cases} \\ & \Leftrightarrow \quad \begin{cases} Y^T Y = X^T X, \\ Y^T U = X^T V, \\ U^T U \succeq V^T V. \end{cases} \end{aligned}$$

□

The two interpolation conditions differ only in the third line of systems (20) and (21). In the isometric case, the equality is replaced by an inequality. As a result, the isometric interpolation condition becomes asymmetric: the quadruple (X, Y, U, V) may be isometrically interpolable, while the reversed (U, V, X, Y) need not be. This asymmetry arises because the transpose of an isometric matrix R is not necessarily isometric.

For the sake of completeness, we recall the theorem on the polar decomposition.

Theorem 3.4 (Polar decomposition, Theorem 7.3.1 of [14]). *Let M be a matrix of size $m \times n$ with $m \geq n$. We have*

$$M = RQ \quad (22)$$

where $Q = (M^T M)^{\frac{1}{2}}$ is a symmetric positive semidefinite matrix whose eigenvalues are the singular values of M and R is an isometric matrix.

We denote the class of tall and wide linear operators as follows

$$\mathcal{L}_S^{\text{tall}} = \mathcal{L}_S \cap \{M \in \mathbb{R}^{m \times n} \mid m \geq n\}, \quad \mathcal{L}_S^{\text{wide}} = \mathcal{L}_S \cap \{M \in \mathbb{R}^{m \times n} \mid m \leq n\}.$$

where $S \subseteq [0, +\infty)$. These two classes of linear operators are often explicitly considered in the optimization context. For example, $\mathcal{L}_S^{\text{tall}}$ appears in compressed sensing, while $\mathcal{L}_S^{\text{wide}}$ arises in linearly constrained optimization. The following proposition shows that the interpolation condition of $\mathcal{L}_S^{\text{tall}}$ and $\mathcal{L}_S^{\text{wide}}$ can be derived from each other.

Proposition 3.5. *The sets of pairs $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ are $\mathcal{L}_S^{\text{wide}}$ -interpolable if and only if the sets of pairs $\{(u_j, v_j)\}_{j \in [N_2]}$ and $\{(x_i, y_i)\}_{i \in [N_1]}$ are $\mathcal{L}_S^{\text{tall}}$ -interpolable.*

Proof. Observe that $M \in \mathcal{L}_S^{\text{tall}}$ if and only if $M^T \in \mathcal{L}_S^{\text{wide}}$, since the transpose does not change the singular spectrum. The result then follows directly by applying Definition 1.1 to M and M^T . $\square$

The new theorem on interpolation conditions for $\mathcal{L}_{[\mu, L]}^{\text{tall}}$ and $\mathcal{L}_{[\mu, L]}^{\text{wide}}$ is the following. Note that given the inputs and outputs (X, Y, U, V) , we know whether the linear operators are tall, wide or square and which class should be interpolated.

Theorem 3.6 ($\mathcal{L}_{[\mu, L]}$ -interpolation conditions). *When $m \geq n$, the sets of pairs $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ are $\mathcal{L}_{[\mu, L]}^{\text{tall}}$ -interpolable if, and only if,*

$$\exists Z \in \mathbb{R}^{n \times N_1}, W \in \mathbb{R}^{n \times N_2}, \begin{cases} Y^T Y = Z^T Z \\ Y^T U = Z^T W \\ U^T U \succeq W^T W \\ \begin{pmatrix} X & W \end{pmatrix}^T \begin{pmatrix} Z & V \end{pmatrix} = \begin{pmatrix} Z & V \end{pmatrix}^T \begin{pmatrix} X & W \end{pmatrix} \\ \left(\begin{pmatrix} Z & V \end{pmatrix} - \mu \begin{pmatrix} X & W \end{pmatrix} \right)^T \left(L \begin{pmatrix} X & W \end{pmatrix} - \begin{pmatrix} Z & V \end{pmatrix} \right) \succeq 0. \end{cases} \quad (23)$$

When $m = n$, the third inequality must be satisfied by equality, i.e., $U^T U = W^T W$. When $m \leq n$, the sets of pairs $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ are $\mathcal{L}_{[\mu, L]}^{\text{wide}}$ -interpolable if, and only if, the same conditions (23) hold after the following swaps: $X \leftrightarrow U$, $Y \leftrightarrow V$, $n \leftrightarrow m$, and $N_1 \leftrightarrow N_2$.

Proof. Consider the polar decomposition (Theorem 3.4) of the matrix $M = RQ$ where $Q \in \mathcal{S}_{[\mu, L]}$, and R is an isometric matrix.

$$\begin{aligned} \exists M \in \mathcal{L}_{[\mu, L]}^{\text{tall}} : \begin{cases} Y = MX \\ V = M^T U \end{cases} &\Leftrightarrow \exists R \text{ isometric}, Q \in \mathcal{S}_{[\mu, L]} : \begin{cases} Y = RQX, \\ V = QR^T U \end{cases} \\ &\Leftrightarrow \exists R \text{ isometric}, Q \in \mathcal{S}_{[\mu, L]}, Z, W : \begin{cases} Y = RZ, \\ Z = QX \\ V = QW \\ W = R^T U \end{cases} \\ &\Leftrightarrow \exists Z, W \begin{cases} (Z, Y, U, W) \text{ is isometric-interpolable,} \\ \left(\begin{pmatrix} X & W \end{pmatrix} \begin{pmatrix} Z & V \end{pmatrix} \right) \text{ is } \mathcal{S}_{[\mu, L]} \text{-interpolable,} \end{cases} \end{aligned}$$

$$\text{Th. 1.2 and 3.3} \Leftrightarrow \exists Z, W \begin{cases} Y^T Y = Z^T Z \\ Y^T U = Z^T W \\ U^T U \succeq W^T W \\ \begin{pmatrix} X & W \end{pmatrix}^T \begin{pmatrix} Z & V \end{pmatrix} = \begin{pmatrix} Z & V \end{pmatrix}^T \begin{pmatrix} X & W \end{pmatrix} \\ \left(\begin{pmatrix} Z & V \end{pmatrix} - \mu \begin{pmatrix} X & W \end{pmatrix} \right)^T \left(L \begin{pmatrix} X & W \end{pmatrix} - \begin{pmatrix} Z & V \end{pmatrix} \right) \succeq 0. \end{cases}$$

When $m = n$, R is unitary and we can use Theorem 3.1, therefore, the third inequality becomes the equality $U^T U = W^T W$.

For the case $m \leq n$, by Proposition 3.5, the sets of pairs $\{(x_i, y_i)\}_{i \in [N_1]}$ and $\{(u_j, v_j)\}_{j \in [N_2]}$ are $\mathcal{L}_{[\mu, L]}^{\text{wide}}$ -interpolable if, and only if, the reversed pair $\{(u_j, v_j)\}_{j \in [N_2]}$ and $\{(x_i, y_i)\}_{i \in [N_1]}$ is $\mathcal{L}_{[\mu, L]}^{\text{tall}}$ -interpolable. Applying (23) to the latter statement yields the desired conditions. $\square$

We can make several observations concerning the new interpolation conditions. The overall approach is based on the polar decomposition. The first three lines of (23) are related to the isometric interpolability of Y , U together with auxiliary intermediate variables Z and W . The last two lines of (23) correspond to the $\mathcal{S}_{[\mu, L]}$ -interpolability of some augmented input-output matrices. This theorem extend the previous work [5] Note that these interpolation conditions are convex constraints on Gram matrices of vectors (X, Y, U, V) and the additional vectors (Z, W) , which is of the form (18).

4 Interpolation conditions for union of eigenvalues and singular values

In this section, we propose an expression of interpolation conditions for symmetric linear operators whose eigenvalues belong to a union of subsets of $\mathbb{R}$. The idea is to sum the Gram matrices associated with each subset.

Theorem 4.1. *Let $S_1, S_2 \subseteq \mathbb{R}$. $G \in \mathcal{G}_{\Delta(S_1 \cup S_2)}$ if, and only if, $G = G_1 + G_2$ for some $G_i \in \mathcal{G}_{\Delta(S_i)}$. $(G, H) \in \mathcal{P}_{\Delta(S_1 \cup S_2)}$ if, and only if, $G = G_1 + G_2$ and $H = H_1 + H_2$ for some $(G_i, H_i) \in \mathcal{P}_{\Delta(S_i)}$, $i = 1, 2$.*

Proof. Let $G \in \mathcal{G}_{\Delta(S_1 \cup S_2)}$. Therefore, G is associated with some matrix Q with eigenvalues in $S_1 \cup S_2$ (Definition 2.1) and using Lemma 2.1-1, we can choose Q with the a block-diagonal form

$$\begin{aligned} G &\in \mathcal{G}_{\Delta(S_1 \cup S_2)} \\ \Leftrightarrow G &= \text{Gram} \left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}, \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \right) \text{ for some } X_i, Y_i, U_i, V_i, i = 1, 2, \\ &\text{and } \exists Q_1, Q_2 : \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}^T \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}, \lambda(Q_i) \in \Delta(S_i) \\ \Leftrightarrow G &= \text{Gram}(X_1 \ Y_1 \ U_1 \ V_1) + \text{Gram}(X_2 \ Y_2 \ U_2 \ V_2) \text{ for some } X_i, Y_i, U_i, V_i, i = 1, 2, \\ &\text{and } \exists Q_1, Q_2 : Y_i = Q_i X_i, V_i = Q_i^T U_i, \text{ and } \lambda(Q_i) \in \Delta(S_i) \\ \Leftrightarrow G &= G_1 + G_2 \text{ for some } G_i \in \mathcal{G}_{\Delta(S_i)}, i = 1, 2 \end{aligned}$$

For the second statement, let $(G, H) \in \mathcal{P}_{\Delta(S_1 \cup S_2)}$. Therefore, (G, H) is associated with some matrix Q with singular values in $S_1 \cup S_2$. Using Lemma 2.1-2, similarly,

$$\begin{aligned}
& (G, H) \in \mathcal{P}_{\Delta(S_1 \cup S_2)} \\
& \Leftrightarrow G = \text{Gram} \left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \right), H = \text{Gram} \left(\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \right), \text{ for some } X_i, Y_i, U_i, V_i, \\
& \quad \text{and } \exists Q_1, Q_2 : \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}^T \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}, \lambda(Q_i) \in \Delta(S_i) \\
& \Leftrightarrow G = \text{Gram}(X_1 \ V_1) + \text{Gram}(X_2 \ V_2), H = \text{Gram}(Y_1 \ U_1) + \text{Gram}(Y_2 \ U_2), \\
& \quad \text{for some } X_i, Y_i, U_i, V_i, i = 1, 2 \text{ and } \exists Q_1, Q_2 : \\
& \quad Y_i = Q_i X_i, V_i = Q_i^T U_i, \lambda(Q_i) \in \Delta(S_i), i = 1, 2 \\
& \Leftrightarrow G = G_1 + G_2 \text{ for some } G_i \in \mathcal{G}_{\Delta(S_i)}, i = 1, 2.
\end{aligned}$$

□

We already know interpolations for symmetric linear operators with eigenvalues on a given interval on $\mathbb{R}$, and general linear operators with singular values on a given interval on $[0, +\infty)$. Using the above theorem, Theorem 2.5 and Theorem 1.1, we can develop interpolation conditions for symmetric linear operators with eigenvalues on any (finite) union of intervals on $\mathbb{R}$ and interpolation conditions for general linear operators with singular values on any (finite) union of intervals on $[0, +\infty)$.

For example, this theorem allows to develop interpolation conditions for symmetric linear operators with eigenvalues equal to either $-\mu$ or μ .

Corollary 4.2. *Let $\mu \in \mathbb{R}$. (X, Y) is $\mathcal{S}_{\{-\mu, \mu\}}$ -interpolable if, and only if,*

$$\begin{cases} X^T Y = Y^T X, \\ \mu^2 X^T X = Y^T Y. \end{cases} \quad (24)$$

Proof. The equalities in (24) is equivalent to the statement that the Gram matrix $G = \text{Gram}(X \ Y)$ has the block form

$$G = \begin{pmatrix} A & B \\ B & \mu^2 A \end{pmatrix} \succeq 0, \text{ for some } A \succeq 0, B = B^T. \quad (25)$$

Indeed, the first condition in (24) forces the off-diagonal blocks to be symmetric, while the second condition makes the lower-right block equal to the upper-left block. Setting $A := X^T X$ and $B := X^T Y$ yields (25).

Assume that (X, Y) is $\mathcal{S}_{\{-\mu, \mu\}}$ -interpolable. By Theorems 4.1 and 2.5, we have $G \in \mathcal{G}_{\Delta(\{\mu\} \cup \{-\mu\})}$ if, and only if, $G = G_1 + G_2$ with $G_1 \in \mathcal{G}_{\Delta(\{\mu\})}$ and $G_2 \in \mathcal{G}_{\Delta(\{-\mu\})}$. And, by Theorem 1.2, one can show that for any $\alpha \in \mathbb{R}$,

$$G_i \in \mathcal{G}_{\Delta(\{\alpha\})} \text{ if, and only if } G_i = \begin{pmatrix} A & \alpha A \\ \alpha A & \alpha^2 A \end{pmatrix}, A \succeq 0. \quad (26)$$

Applying (26) with $\alpha = -\mu$ and $\alpha = \mu$ yields

$$G = G_1 + G_2 = \begin{pmatrix} A_1 & \mu A_1 \\ \mu A_1 & \mu^2 A_1 \end{pmatrix} + \begin{pmatrix} A_2 & -\mu A_2 \\ -\mu A_2 & \mu^2 A_2 \end{pmatrix} = \begin{pmatrix} A_1 + A_2 & \mu(A_1 - A_2) \\ \mu(A_1 - A_2) & \mu^2(A_1 + A_2) \end{pmatrix} \quad (27)$$

for some A_1 and A_2 . Hence G can be written as (25) with $A = A_1 + A_2$ and $B = \mu(A_1 - A_2)$.

On the other hand, assume that G can be written as in (25). Define $A_1 = \frac{1}{2} \left(A + \frac{B}{\mu} \right)$ and $A_2 = \frac{1}{2} \left(A - \frac{B}{\mu} \right)$. Because $G \succeq 0$, the matrices A_1, A_2 are positive semidefinite since for any vector x , $\begin{pmatrix} x \\ \pm \frac{1}{\mu} x \end{pmatrix}^T \begin{pmatrix} A & B \\ B & \mu^2 A \end{pmatrix} \begin{pmatrix} x \\ \pm \frac{1}{\mu} x \end{pmatrix} = x^T A_1 x$ or $x^T A_2 x \geq 0$. Using these matrices, we have $G = G_1 + G_2$ with $G_1 = \begin{pmatrix} A_1 & \mu A_1 \\ \mu A_1 & \mu^2 A_1 \end{pmatrix}$ and $G_2 = \begin{pmatrix} A_2 & -\mu A_2 \\ -\mu A_2 & \mu^2 A_2 \end{pmatrix}$. By (26), each summand belongs to $\mathcal{G}_{\Delta(\{\mu\})}$ and $\mathcal{G}_{\Delta(\{-\mu\})}$, respectively. Hence $G \in \mathcal{G}_{\Delta(\{-\mu, \mu\})}$, which means that (X, Y) is $\mathcal{S}_{\{-\mu, \mu\}}$ -interpolable. $\square$

The interpolation condition for symmetric linear operators with eigenvalues in $\{-\mu, \mu\}$ (Corollary 4.2) differs from the corresponding condition for $[-\mu, \mu]$ (Theorem 1.2), which is given by

$$\begin{cases} X^T Y = Y^T X, \\ \mu^2 X^T X \succeq Y^T Y. \end{cases}$$

In the former case, the second condition becomes an equality.

5 Application to Performance Estimation and worst-case analysis

The interpolation conditions derived in Theorems 3.6, 4.1 and 2.5 allow to analyze the worst-case performance of optimization methods with PEP on problems of the forms $\min_x g(Mx)$ or $\min_x f(x) + g(Mx)$ where $M \in \mathcal{L}_I$ or $M \in \mathcal{S}_I$ where I is any subset of $\mathbb{R}$. We propose to analyze the Gradient method on the first problem and the Chambolle-Pock method on the second problem.

5.1 Worst-case performance of Gradient Method on smooth convex functions

The exact convergence rate of Gradient Method (GM) on smooth convex functions is an active research area [1, 2, 11, 15]. The work [5] considered the problem of the form

$$\min_x F(x) \triangleq g(Mx) \quad (28)$$

where g is μ_g -strongly convex and L_g -smooth and $M \in \mathcal{L}_{[\mu_M, L_M]}$ (we say that $F \in \mathcal{C}_{\mu_g, L_g}^{\mu_M, L_M}$) or $M \in \mathcal{S}_{[\mu_M, L_M]}$ (we say that $F \in \mathcal{D}_{\mu_g, L_g}^{\mu_M, L_M}$). They consider GM with fixed stepsize h/L

$$x_{k+1} = x_k - \frac{h}{L} \nabla F(x) = x_k - \frac{h}{L} M^T \nabla g(Mx) \quad (\text{GM})$$

where L is the Lipschitz constant of the gradient of F . They characterized the worst-case behavior of GM on the classes $\mathcal{C}_{\mu_g, L_g}^{0, L_M}$ and $\mathcal{D}_{\mu_g, L_g}^{\mu_M, L_M}$ with $w(\mathcal{F}, h/L)$ being the worst-case behavior of GM with step size h/L on the class $\mathcal{F}$ with respect to the criterion $F(x_N) - F(x^*)$ and bounded initial distance $\|x_0 - x^*\| \leq 1$ on [5, Conjecture 4.2].

Prior to this work, computing the tight performance on $\mathcal{C}_{\mu_g, L_g}^{\mu_M, L_M}$ for $\mu_M > 0$ is impossible due to the unknown interpolation conditions for $\mathcal{L}_{[\mu_M, L_M]}$. Now that we have them (Theorem 3.6), we can extend their analysis of [5] to $\mathcal{C}_{\mu_g, L_g}^{\mu_M, L_M}$.

5.1.1 Performance on linear operators with non-zero minimal singular value

We compute the worst-case performance of GM when the linear operator M is with maximal and nonzero minimal bound on the singular values. We observed that both in the tall and wide case, the worst-case performance of GM is exactly the same as the worst-case performance of GM on $\mathcal{D}_{\mu_g, L_g}^{\mu_M, L_M}$, namely, the symmetry of M has no impact on the worst-case performance.

5.1.2 Performance on linear operators with eigenvalues on a union of intervals

We compute the worst-case performance of GM when the linear operator M has eigenvalues in a union of intervals. Figure 1 compares the worst-case linear operator of GM when varying the step size h when $M \in \mathcal{S}_{[0.6, 1]}$ (red dots) and $M \in \mathcal{S}_{[0.6, 0.65] \cup [0.95, 1]}$ (blue dots) which appears to be both scalar. In other words, either we include the whole interval $[0.6, 1]$, or we exclude the sub-interval $[0.65, 0.95]$.

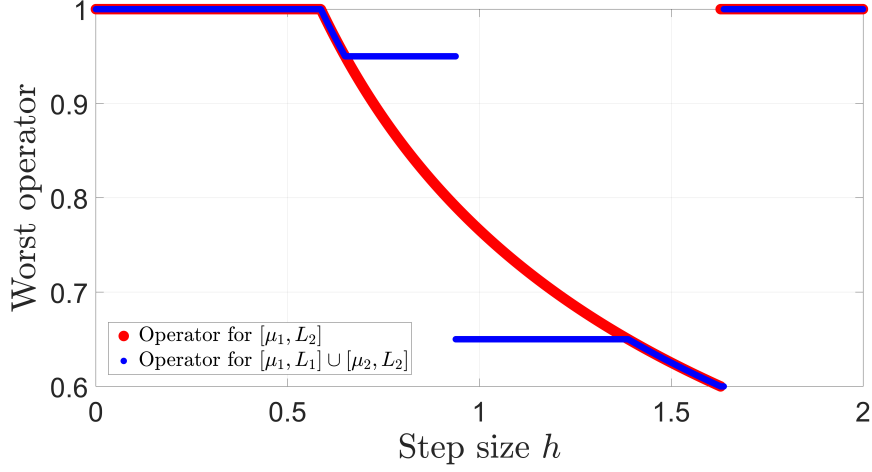


Fig. 1 Comparison of the worst-case linear operators when allowing the whole interval (red curve) or only the union of two sub-intervals (blue curve).

Interestingly, the worst-case linear operator remains a scalar. This could not be observed without our new interpolation conditions.

5.2 Chambolle-Pock Method for composite optimization problem

Another application of our interpolation conditions is the analysis of the Chambolle-Pock method (CP) [13] on the composite optimization problem

$$\min_x f(x) + g(Mx)$$

where f, g are proximable convex functions. The iterations are the following, with parameters $\tau, \sigma > 0$

$$\begin{cases} x_{i+1} = \text{prox}_{\tau f}(x_i - \tau M^T u_i), \\ u_{i+1} = \text{prox}_{\sigma g^*}(u_i + \sigma M(2x_{i+1} - x_i)). \end{cases}$$

The work [5] also analyzed the worst-case behavior of CP on different settings with PEP. More precisely, they improved [16, Theorem 1] the convergence rate of the primal-dual gap of the average iterations produced by CP (see [5, Theorem 4.6]) when $M \in \mathcal{L}_{[0, L_M]}$ for some $L_M > 0$. Note that this classical result requires a bound on

$$\frac{1}{\tau} \|x - x_0\|^2 + \frac{1}{\sigma} \|u - u_0\|^2 - 2(u - u_0)^T M(x - x_0) \quad (29)$$

and [5] also considered the bound

$$\frac{1}{\tau} \|x - x_0\|^2 + \frac{1}{\sigma} \|u - u_0\|^2. \quad (30)$$

Thanks to our new interpolation conditions (Theorem 1.1), we can observe the impact of bounding the minimal singular values of the linear operators M . We found that considering the lower bound of singular value on M does not improve the convergence rate in both settings.

6 Conclusion

Interpolation conditions for linear operators and PEP allow to analyze the performance of optimization methods on problems of the form $\min_x g(Mx)$ and $\min_x f(x) + g(Mx)$ and other problems. We used to have interpolation conditions for linear operators and their transpose characterized by singular values in $[0, L]$ and for symmetric linear operators characterized by eigenvalues in $[\mu, L]$.

In this work, we developed interpolation conditions for the class of linear operators with singular values between μ and L and for symmetric linear operators characterized by any finite union of intervals for eigenvalues. The reasoning to obtain interpolation conditions for linear operators with singular values in a finite union of intervals is the same but we did not present it. We illustrated the application of our new interpolation conditions for the worst-case analysis of the gradient method and Chambolle-Pock methods via the Performance Estimation Problem framework.

We also showed that we can only characterize classes of linear operators whose singular values are in a given subset of $\mathbb{R}$ and classes of symmetric linear operators whose eigenvalues are in a given subset of $\mathbb{R}$ through Gram representation.

Therefore, there are the following open questions:

- Can we develop interpolation conditions for linear operators whose singular values or eigenvalues belong to arbitrary sets that cannot be expressed as a finite union of intervals?
- Can we develop interpolation conditions for general square matrices, and which properties could we require?
- Is there another convex representation of classes of linear operators that does not rely on Gram representations?

A Proofs of Theorem 2.4

Lemma A.1. *Let a set of spectrum Λ . $\mathcal{P}_\Lambda$ and $\mathcal{G}_\Lambda$ are cones, namely,*

$$c(G, H) \in \mathcal{P}_\Lambda \quad \forall c \geq 0, \forall (G, H) \in \mathcal{P}_\Lambda \quad (31)$$

$$cG \in \mathcal{G}_\Lambda \quad \forall c \geq 0, \forall G \in \mathcal{G}_\Lambda. \quad (32)$$

Proof. For $c > 0$, we have

$$\begin{aligned}
(G, H) \in \mathcal{P}_\Lambda &\Rightarrow \exists X, U, M : (G, H) = (\text{Gram}(X \ M^T U), \text{Gram}(MX \ U)) \\
&\quad \text{and } \sigma(M) \in \Lambda \\
&\Rightarrow \exists X, U, M : c(G, H) = (\text{Gram}(\sqrt{c}X \ M^T \sqrt{c}U), \text{Gram}(\sqrt{c}MX \ \sqrt{c}U)) \\
&\quad \text{and } \sigma(M) \in \Lambda \\
&\Rightarrow c(G, H) \in \mathcal{G}_\Lambda.
\end{aligned}$$

and

$$\begin{aligned}
G \in \mathcal{G}_\Lambda &\Rightarrow \exists X, U, \text{symmetric } Q : G = \text{Gram}(X \ QX \ U \ Q^T U) \text{ and } \lambda(Q) \in \Lambda \\
&\Rightarrow \exists X, U, \text{symmetric } Q : cG = \text{Gram}(\sqrt{c}X \ Q \sqrt{c}X \ \sqrt{c}U \ Q^T \sqrt{c}U) \text{ and } \lambda(Q) \in \Lambda \\
&\Rightarrow cG \in \mathcal{G}_\Lambda.
\end{aligned}$$

□

The following lemma formalizes and exploits the fact that when a Gram matrix is associated with a linear operator M_1 and vectors X, Y, U, V , we can always augment the linear operator M_1 as $\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}$ without modifying the Gram matrix by augmenting the vectors with zeros, namely, $\begin{pmatrix} X \\ 0 \end{pmatrix}, \begin{pmatrix} Y \\ 0 \end{pmatrix}, \begin{pmatrix} U \\ 0 \end{pmatrix}, \begin{pmatrix} V \\ 0 \end{pmatrix}$. Therefore, when a Gram matrix is associated with some spectrum λ_1 it is also associated with all larger spectrum λ_2 such that $\lambda_1 \subseteq \lambda_2$.

Lemma A.2 (Extension of spectrums). *Let λ_1 and λ_2 be two spectrums. If $\lambda_1 \subseteq \lambda_2$ then $\mathcal{P}_{\{\lambda_1\}} \subseteq \mathcal{P}_{\{\lambda_2\}}$ and $\mathcal{G}_{\{\lambda_1\}} \subseteq \mathcal{G}_{\{\lambda_2\}}$.*

Proof. Let $G \in \mathcal{G}_{\{\lambda_1\}}$ (resp. $(G, H) \in \mathcal{P}_{\{\lambda_1\}}$). We have

$$\begin{aligned}
G &= \text{Gram}(X \ Y \ U \ V) \text{ (resp. } (G, H) = (\text{Gram}(X \ V), \text{Gram}(Y \ U))) \\
\text{with } \begin{cases} Y &= M_1 X \\ V &= M_1^T U \end{cases} \text{ and } \lambda(M_1) = \lambda_1 \text{ (resp. } \sigma(M_1) = \lambda_1) \\
\Rightarrow G &= \text{Gram}\left(\begin{pmatrix} X \\ 0 \end{pmatrix} \begin{pmatrix} Y \\ 0 \end{pmatrix} \begin{pmatrix} U \\ 0 \end{pmatrix} \begin{pmatrix} V \\ 0 \end{pmatrix}\right) \\
&\quad \text{(resp. } (G, H) = (\text{Gram}\left(\begin{pmatrix} X \\ 0 \end{pmatrix} \begin{pmatrix} V \\ 0 \end{pmatrix}\right), \text{Gram}\left(\begin{pmatrix} Y \\ 0 \end{pmatrix} \begin{pmatrix} U \\ 0 \end{pmatrix}\right))) \\
\text{with } \begin{cases} \begin{pmatrix} Y \\ 0 \end{pmatrix} &= \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \begin{pmatrix} X \\ 0 \end{pmatrix} \\ \begin{pmatrix} V \\ 0 \end{pmatrix} &= \begin{pmatrix} M_1^T & 0 \\ 0 & M_2^T \end{pmatrix} \begin{pmatrix} U \\ 0 \end{pmatrix} \end{cases} \text{ and } \begin{cases} \lambda(M_1) = \lambda_1 & \text{(resp. } \sigma(M_1) = \lambda_1) \\ \lambda(M_2) = \lambda_2 - \lambda_1 & \text{(resp. } \sigma(M_2) = \lambda_2 - \lambda_1) \end{cases} \\
\Rightarrow G &\in \mathcal{G}_{\{\lambda_1 + (\lambda_2 - \lambda_1)\}} = \mathcal{G}_{\{\lambda_2\}} \text{ (resp. } (G, H) \in \mathcal{P}_{\{\lambda_1 + (\lambda_2 - \lambda_1)\}} = \mathcal{P}_{\{\lambda_2\}})
\end{aligned}$$

since $\lambda_1 \subseteq \lambda_2$, we have $\lambda_1 + (\lambda_2 - \lambda_1) = \lambda_2$. $\square$

This property also holds for sets of spectrums.

Corollary A.3. *Let Λ_1 and Λ_2 be sets of spectrums. If for every $\lambda_1 \in \Lambda_1$ there exists $\lambda_2 \in \Lambda_2$ such that $\lambda_1 \subseteq \lambda_2$ then $\mathcal{P}_{\Lambda_1} \subseteq \mathcal{P}_{\Lambda_2}$ and $\mathcal{G}_{\Lambda_1} \subseteq \mathcal{G}_{\Lambda_2}$.*

Theorem A.4 (Theorem 2.4). *Let $\mathcal{M}$ be a set of linear operators (resp. symmetric matrices). If $\mathcal{P}_{\mathcal{M}}$ (resp. $\mathcal{G}_{\mathcal{M}}$) is convex, then*

$$\mathcal{P}_{\mathcal{M}} = \mathcal{P}_{\Delta(S)} \text{ (resp. } \mathcal{G}_{\mathcal{M}} = \mathcal{G}_{\Delta(T)}) \quad (33)$$

for some $S \subseteq \mathbb{R}$.

Proof. By Corollary 2.2, we have $\mathcal{P}_{\mathcal{M}} = \mathcal{P}_{\Lambda}$ (resp. $\mathcal{G}_{\mathcal{M}} = \mathcal{G}_{\Lambda}$) where $\Lambda = \{\sigma(M) : M \in \mathcal{M}\}$ (resp. $\Lambda = \{\lambda(M) : M \in \mathcal{M}\}$). We show that S (resp. T) is the set of all values appearing in the spectrums of Λ , namely, $S_{\Lambda} \triangleq \{\mu : \exists \lambda \in \Lambda \text{ s.t. } \mu \in \lambda\}$. Recall that

$$\Delta(S_{\Lambda}) = \{\lambda : \forall a \in \lambda, a \in S_{\Lambda}\} = \{\lambda : \forall a \in \lambda, a \in \lambda'_a, \lambda'_a \in \Lambda\} \quad (34)$$

is the set of spectrums λ for which each element a belongs to some spectrum λ'_a of Λ .

(Inclusions $\mathcal{P}_{\Lambda} \subseteq \mathcal{P}_{\Delta(S_{\Lambda})}$ and $\mathcal{G}_{\Lambda} \subseteq \mathcal{G}_{\Delta(S_{\Lambda})}$) Let a spectrum $\lambda \in \Lambda$, then clearly all the elements of this spectrum belong to some spectrum of Λ and therefore $\lambda \in \Delta(S_{\Lambda})$. Thus, we have $\Lambda \subseteq \Delta(S_{\Lambda})$ and therefore $\mathcal{P}_{\Lambda} \subseteq \mathcal{P}_{\Delta(S_{\Lambda})}$ and $\mathcal{G}_{\Lambda} \subseteq \mathcal{G}_{\Delta(S_{\Lambda})}$ by Corollary A.3.

(Inclusions $\mathcal{P}_{\Delta(S_{\Lambda})} \subseteq \mathcal{P}_{\Lambda}$ and $\mathcal{G}_{\Delta(S_{\Lambda})} \subseteq \mathcal{G}_{\Lambda}$) First, we show that $\mathcal{G}_{\{\lambda\}} \subseteq \mathcal{G}_{\Lambda}$ (resp. $\mathcal{P}_{\{\lambda\}} \subseteq \mathcal{P}_{\Lambda}$) $\forall \lambda \in \Delta(S_{\Lambda})$. Recall that $\mathcal{G}_{\{\lambda\}}$ (resp. $\mathcal{P}_{\{\lambda\}}$) is the set of Gram matrices associated with linear operators whose eigenvalue (resp. singular value) spectrum is exactly λ . Let $\lambda \in \Delta(S_{\Lambda})$ and $G \in \mathcal{G}_{\{\lambda\}}$ (resp. $(G, H) \in \mathcal{P}_{\{\lambda\}}$). We have $G = \text{Gram}(X \ Y \ U \ V)$ (resp. $(G, H) = (\text{Gram}(X \ V), \text{Gram}(Y \ U))$) with $Y = MX$, $V = M^T U$ and $a \in \lambda'_a$, $\lambda'_a \in \Lambda$, $\forall a \in \lambda(M)$ (resp. $\forall a \in \sigma(M)$). W.l.o.g, let M of dimension $m \times d$ with $d \leq m$ in the rectangle case and of dimension $d \times d$ in the symmetric case. By Lemma 2.1 and eigenvalue (resp. singular value) decomposition of M , we can choose M diagonal with its d eigenvalues (resp. singular values) m_i on its diagonal. Therefore, we have

$$(X, Y, U, V) = \left(\begin{pmatrix} X_1 \\ \vdots \\ X_d \\ X_d \end{pmatrix}, \begin{pmatrix} V_1 \\ \vdots \\ V_d \\ V_d \end{pmatrix}, \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \\ Y_m \end{pmatrix}, \begin{pmatrix} U_1 \\ \vdots \\ U_m \\ U_m \end{pmatrix} \right) \quad (35)$$

and

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} m_1 & & \\ & \ddots & \\ & & m_d \\ - & E & - \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_d \\ X_d \end{pmatrix}, \begin{pmatrix} V_1 \\ \vdots \\ V_d \\ V_d \end{pmatrix} = \begin{pmatrix} m_1 & & | \\ & \ddots & \\ & & E^T \\ & m_d & | \end{pmatrix} \begin{pmatrix} U_1 \\ \vdots \\ U_m \\ U_m \end{pmatrix},$$

where $m_i \in \lambda'_i, \lambda'_i \in \Lambda$, and $E = 0$ with the appropriate size. It is equivalent to

$$\Leftrightarrow Y_i = m_i X_i, V_i = m_i U_i \quad \forall i = 1, \dots, d \text{ and } Y_i = 0 \quad \forall i = d+1, \dots, m, m_i \in \lambda'_i, \lambda'_i \in \Lambda$$

$\Leftrightarrow \begin{pmatrix} Y_i \\ 0 \end{pmatrix} = \overbrace{\begin{pmatrix} m_i & 0 \\ 0 & \bar{M}_i \end{pmatrix}}^{M_i} \begin{pmatrix} X_i \\ 0 \end{pmatrix}, \begin{pmatrix} V_i \\ 0 \end{pmatrix} = \overbrace{\begin{pmatrix} m_i & 0 \\ 0 & \bar{M}_i^T \end{pmatrix}}^{M_i^T} \begin{pmatrix} U_i \\ 0 \end{pmatrix}, \lambda(M_i) = \lambda'_i \text{ (resp. } \sigma(M_i) = \lambda'_i), \forall \lambda'_i \in \Lambda \text{ where } \bar{M}_i \text{ completes the spectrum of } M_i \text{ such that the spectrum of } M_i \text{ is exactly } \lambda'_i. \text{ Let } G_i = \text{Gram} \left(\begin{pmatrix} X_i \\ 0 \end{pmatrix}, \begin{pmatrix} Y_i \\ 0 \end{pmatrix}, \begin{pmatrix} U_i \\ 0 \end{pmatrix}, \begin{pmatrix} V_i \\ 0 \end{pmatrix} \right) \in \mathcal{G}_\Lambda \text{ (resp. } (G_i, H_i) = (\text{Gram} \left(\begin{pmatrix} X_i \\ 0 \end{pmatrix}, \begin{pmatrix} V_i \\ 0 \end{pmatrix} \right), \text{Gram} \left(\begin{pmatrix} Y_i \\ 0 \end{pmatrix}, \begin{pmatrix} U_i \\ 0 \end{pmatrix} \right) \in \mathcal{P}_\Lambda). \text{ We have } G = \sum_{i=1}^d G_i = \sum_{i=1}^d \frac{1}{d}(dG_i) \text{ (resp. } (G, H) = \sum_{i=1}^d (G_i, H_i) = \sum_{i=1}^d \frac{1}{d}(d(G_i, H_i)) \text{ since } Y_i = 0 \text{ for } i = d+1, \dots, m.) \text{ and } dG_i \in \mathcal{G}_\Lambda \text{ (resp. } d(G_i, H_i) \in \mathcal{P}_\Lambda) \text{ by Lemma A.1. And by convexity of } \mathcal{G}_\Lambda \text{ (resp. } \mathcal{P}_\Lambda), \text{ we have } G = \sum_{i=1}^d \frac{1}{d}(dG_i) \in \mathcal{G}_\Lambda \text{ (resp. } (G, H) = \sum_{i=1}^d \frac{1}{d}(d(G_i, H_i)) \in \mathcal{P}_\Lambda). \text{ Finally, let } G \in \mathcal{G}_{\Delta(S_\Lambda)} \text{ (resp. } (G, H) \in \mathcal{P}_{\Delta(S_\Lambda)}). \text{ By definition of } \Delta(S_\Lambda), \text{ there is a } \lambda \in \Delta(S_\Lambda) \text{ such that } G \in \mathcal{G}_{\{\lambda\}} \text{ (resp. } (G, H) \in \mathcal{P}_{\{\lambda\}}) \text{ and therefore } G \in \mathcal{G}_\Lambda \text{ (resp. } (G, H) \in \mathcal{P}_\Lambda). \quad \square$

Acknowledgements. Nizar Bousselmi is an FRIA grantee of the Fonds de la Recherche Scientifique - FNRS. Zhicheng Deng is sponsored by China Scholarship Council.

Delcarations

Data Availability Statement. The datasets generated during the current study are available from the corresponding author on reasonable request.

Conflict of interest. This research was conducted in full compliance with ethical standards, with no Conflict of interest, no involvement of human participants or animals, and no requirement for informed consent.

References

- [1] Drori, Y., Teboulle, M.: Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming* **145**(1), 451–482 (2014)
- [2] Taylor, A.B., Hendrickx, J.M., Glineur, F.: Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming* **161**, 307–345 (2017)
- [3] Rubbens, A., Hendrickx, J.M.: A constraint-based approach to function interpolation, with application to performance estimation for weakly convex optimisation. *arXiv preprint arXiv:2405.08405* (2024)

- [4] Goujaud, B., Dieuleveut, A., Taylor, A.: On fundamental proof structures in first-order optimization. In: 2023 62nd IEEE Conference on Decision and Control (CDC), pp. 3023–3030 (2023). IEEE
- [5] Bousselmi, N., Hendrickx, J.M., Glineur, F.: Interpolation conditions for linear operators and applications to performance estimation problems. *SIAM Journal on Optimization* **34**(3), 3033–3063 (2024)
- [6] Bousselmi, N., Pustelnik, N., Hendrickx, J.M., Glineur, F.: Comparison of proximal first-order primal and primal-dual algorithms via performance estimation. In: 2024 32nd European Signal Processing Conference (EUSIPCO), pp. 2647–2651 (2024). IEEE
- [7] Colla, S.: Computer-aided analysis of decentralized optimization methods. PhD thesis, UCLouvain (2024)
- [8] Vasin, A., Krivchenko, V., Kovalev, D., Stonyakin, F., Tupitsa, N., Dvurechensky, P., Alkousa, M., Kornilov, N., Gasnikov, A.: On solving minimization and min-max problems by first-order methods with relative error in gradients. arXiv preprint arXiv:2503.06628 (2025)
- [9] Krivchenko, V.O., Gasnikov, A.V., Kovalev, D.A.: Convex-concave interpolation and application of PEP to the bilinear-coupled saddle point problem. *Russian Journal of Nonlinear Dynamics* **20**(5), 875–893 (2024)
- [10] Ranjan, V., Stellato, B.: Verification of first-order methods for parametric quadratic optimization. arXiv preprint arXiv:2403.03331 (2024)
- [11] Nesterov, Y., et al.: Lectures on convex optimization. Springer **137** (2018)
- [12] Rubbens, A., Bousselmi, N., Colla, S., Hendrickx, J.M.: Interpolation constraints for computing worst-case bounds in performance estimation problems. In: 2023 62nd IEEE Conference on Decision and Control (CDC), pp. 3015–3022 (2023). IEEE
- [13] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision* **40**, 120–145 (2011)
- [14] Horn, R.A., Johnson, C.R.: Matrix analysis. Cambridge university press (2012)
- [15] Rotaru, T., Glineur, F., Patrinos, P.: Exact worst-case convergence rates of gradient descent: a complete analysis for all constant stepsizes over nonconvex and convex functions. arXiv preprint arXiv:2406.17506 (2024)
- [16] Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal-dual algorithm. *Mathematical Programming* **159**(1), 253–287 (2016)