SIAM J. Optim.
Vol. 34, No. 3, pp. 3033–3063

INTERPOLATION CONDITIONS FOR LINEAR OPERATORS AND APPLICATIONS TO PERFORMANCE ESTIMATION PROBLEMS*

NIZAR BOUSSELMI†, JULIEN M. HENDRICKX†, AND FRANÇOIS GLINEUR‡

Abstract. The performance estimation problem methodology makes it possible to determine the exact worst-case performance of an optimization method. In this work, we generalize this framework to first-order methods involving linear operators. This extension requires an explicit formulation of interpolation conditions for those linear operators. We consider the class of linear operators $\mathcal{M}: x \mapsto Mx$, where matrix M has bounded singular values, and the class of linear operators, where M is symmetric and has bounded eigenvalues. We describe interpolation conditions for these classes, i.e., necessary and sufficient conditions that, given a list of pairs $\{(x_i, y_i)\}$, characterize the existence of a linear operator mapping x_i to y_i for all i. Using these conditions, we first identify the exact worst-case behavior of the gradient method applied to the composed objective $h \circ \mathcal{M}$, and observe that it always corresponds to $\mathcal{M}$ being a scaling operator. We then investigate the Chambolle–Pock method applied to $f + g \circ \mathcal{M}$, and improve the existing analysis to obtain a proof of the exact convergence rate of the primal-dual gap. In addition, we study how this method behaves on Lipschitz convex functions, and obtain a numerical convergence rate for the primal accuracy of the last iterate. We also show numerically that averaging iterates is beneficial in this setting.

Key words. first-order algorithms, performance estimation, exact convergence rates, interpolation conditions, linear operators, primal-dual algorithms, quadratic functions

MSC codes. 90C25, 90C20, 68Q25, 90C22, 49M29

DOI. 10.1137/23M1575391

1. Introduction.

The *performance estimation problem* (PEP) methodology, introduced in [26], computes the exact worst-case performance of a given first-order optimization method on a given class of functions and identifies an instance reaching this worst performance. More precisely, given a method and a performance criterion (lower is better), a PEP is an optimization problem that maximizes this criterion on the application of the method among all possible functions belonging to some class, thus providing the worst possible behavior of the method on the class of functions. It is shown in [63] that PEPs can be reformulated exactly as semidefinite programs for a wide range of function classes and methods. This provided several tight results on the performance of first-order methods (see, e.g., section 1.3 for a short review).

In this work, we extend the PEP framework to first-order methods involving linear operators $\mathcal{M}: x \mapsto Mx$ for several classes of matrices M characterized by their eigenvalues or singular value spectrum. The main tool behind this extension is the development of interpolation conditions for linear operators (see section 3). Therefore, we will be able to study the exact worst-case performance of first-order algorithms involving linear operators.

*Received by the editors May 28, 2023; accepted for publication (in revised form) April 30, 2024; published electronically September 5, 2024.
https://doi.org/10.1137/23M1575391
Funding: Nizar Bousselmi is an FRIA grantee of the Fonds de la Recherche Scientifique - FNRS.
†UCLouvain, ICTEAM, Louvain-la-Neuve, Belgium (nizar.bousselmi@uclouvain.be, julien.hendrickx@uclouvain.be).
‡UCLouvain, ICTEAM, Louvain-la-Neuve, Belgium, and UCLouvain, CORE, Louvain-la-Neuve, Belgium (francois.glineur@uclouvain.be).

1.1. Optimization methods involving linear operators. Many methods aim at solving problems involving linear operators in their objective functions or constraints. Iterations of such methods typically apply the linear operators of the problem themselves. We describe below three motivating examples of optimization problems involving linear operators.

Motivating example 1*:* $\min_x g(Mx)$. To solve this problem, first-order methods evaluate the gradient of the function $F(x) = g(Mx)$ $\forall x \in \mathbb{R}^n$ on some point x_i,

$$\nabla F(x_i) = M^T \nabla g(Mx_i), \tag{1.1}$$

where we can see the applications of the linear operators associated with M and M^T.

Note that the class of functions $g(Mx)$ includes the quadratic functions $F(x) = \frac{1}{2}x^T Qx$ (when $g(y) = \frac{1}{2}||y||^2$ and $M^T M = Q$). In such a case, the gradient is the linear operator $\nabla F(x_i) = Qx_i$. Obtaining the exact worst-case performance of first-order methods on quadratic functions can be done using a known technique (see section 2.1 in [28] or [8] for recent reviews), which consists in analyzing the roots of a polynomial associated with the method of interest. However, this technique can only deal with a single quadratic objective function, whereas our extension of the PEP methodology can analyze more complex formulations, such as $F(x) = f(x) + \frac{1}{2}x^T Qx$ (see [4] for recent work on this class with a nonsmooth function f).

Motivating example 2*:* $\min_x f(x) + g(Mx)$. A possible algorithm to solve the problem $\min_x f(x) + g(Mx)$, when f and g are proximable, is the Chambolle–Pock (CP) algorithm [10]. Given step sizes $\tau, \sigma > 0$ and a pair of primal-dual points (x_i, u_i), the iteration of CP is

$$\begin{cases} x_{i+1} = \operatorname{prox}_{\tau f}\left(x_i - \tau M^T u_i\right), \\ u_{i+1} = \operatorname{prox}_{\sigma g^*}\left(u_i + \sigma M(2x_{i+1} - x_i)\right). \end{cases} \tag{1.2}$$

The method performs proximal steps, which are already handled in the PEP framework [64]. Again this iteration contains products with matrices M and M^T.

Motivating example 3*:* $\min_{x,y} f(x)+g(y)$ *s.t.* $M_1x+M_2y=b$. A famous method to solve the constrained problem $\min_{x,y} f(x)+g(y)$ s.t. $M_1x+M_2y=b$ is the alternating direction method of multipliers (ADMM) [30] with the following iterations:

$$\begin{cases} x_{i+1} \in \arg\min_x f(x) + \frac{\rho}{2}\left\|M_1x + M_2y_i - b + \frac{1}{\rho}z_i\right\|^2, \\ y_{i+1} \in \arg\min_y g(y) + \frac{\rho}{2}\left\|M_1x_{i+1} + M_2y - b + \frac{1}{\rho}z_i\right\|^2, \\ z_{i+1} = z_i + \rho(M_1x_{i+1} + M_2y_{i+1} - b). \end{cases} \tag{1.3}$$

By contrast to the first two motivating examples, linear operators appear here in the constraint of the problem. However, exactly as in the previous motivating examples, the linear operators end up being used in the iterations of the method.

Exploiting additional knowledge about the function structure, e.g., that it can be expressed as $g \circ \mathcal{M}$ or $f + g \circ \mathcal{M}$, improves the analysis of the methods. For an illustrative example, let g be smooth and strongly convex and consider the composed function $g \circ \mathcal{M}$. Despite the fact that $g \circ \mathcal{M}$ only belongs to the class of smooth convex functions (i.e., is not strongly convex), we will see that the gradient method applied to this composed function performs better than on general smooth convex functions (see section 4.1.2 and, e.g., [53]). Likewise, CP and ADMM inherently exploit the structure of the problems and cannot be applied to a generic problem $\min_x f(x)$.

More generally, our extension can analyze any first-order method involving linear operators, for example, primal-dual fixed point [12], Condat–Vũ [17, 66], primal-dual three-operator splitting [67], and proximal alternating predictor-corrector [23] algorithms (see [18] for a recent review on these algorithms). These methods solve a wide range of different classical optimization problems, for example, ℓ_2-regularized robust regression [55], ℓ_1-constrained least squares [29], basis pursuit [13], total variation deblurring [56, 7], and resource allocation [68].

Existing performance guarantees for the above algorithms are often not tight, may use unusual performance criteria or initial conditions for technical reasons, and are thus difficult to compare. Our extension of PEP remedies these issues by providing the exact worst-case performance of these methods for any of the standard performance criteria and initial conditions.

1.2. Outline of the paper. In section 2, we recall the formulation of the PEP as a finite-dimensional optimization problem and point out the missing part that we want to extend, i.e., the interpolation conditions for linear operators with bounded eigenvalues or singular value spectrum. In section 3, as the first main contribution, we derive these needed interpolation conditions in an explicit and tractable way to be able to add them to PEP. As a byproduct, we also present the interpolation conditions for the class of quadratic functions. In section 4, as the second main contribution, we exploit our new extension of PEP, first to analyze the gradient method on the problem $\min_x g(Mx)$ and to provide an expression of its worst-case performance. We then tighten the existing performance guarantees on the CP algorithm and prove the exact convergence rate of the primal-dual gap. We also demonstrate the flexibility of our approach by considering alternate performance criteria and function classes for the same algorithm, for which we obtain numerical performance guarantees. In section 5, we conclude our work and discuss future research directions.

1.3. Prior PEP work. Since its introduction, the PEP framework has evolved in many directions and settings to analyze a lot of different problems and methods including the gradient method (and its accelerated version) on smooth strongly or hypoconvex functions [26, 61, 54], convex functions with a quadratic upper bound [37], functions with lower restricted secant inequality and an upper error bound [40], and nonconvex functions satisfying the Polyak–Łojasiewicz inequality [2]. More variants of the gradient method have been covered by PEP including the gradient method with exact line search [20], the proximal gradient method [64], the inexact gradient method with bounded error on the gradient [21], and the block coordinate descent method [58, 44, 1]. Other methods have also been studied by the framework as decentralized methods [16, 14, 15], difference-of-convex-algorithm [3], the gradient descent-ascent method, the ADMM [69], and Bregman or mirror descent [22]. In addition to the analysis of optimization problems, the PEP framework can also be used to study problems involving operators like the Halpern iteration for fixed point problems [51], the proximal point algorithm for maximal monotone inclusions [38], the relaxed proximal point algorithm for variational inequalities [39], splitting methods [57], and extragradient methods [32, 33, 34]. Some works exploited and extended the PEP framework to develop methodologies that address closely related questions like stochastic optimization [59], continuous-time model version of first-order methods [52], and Lyapunov inequalities for first-order methods [65]. PEP has also been used to design and improve optimization methods for different problems such as an optimal variant of Kelley's cutting-plane method [27], the optimized gradient method for smooth convex functions decreasing the cost function [46, 47, 49] and the gradient norm [50], an optimal gradient method for smooth strongly convex functions [60], optimal methods for nonsmooth and smooth convex functions [25], fast iterative shrinkage/thresholding algorithm [48], and accelerated proximal points method [45]. Recent works also proposed to solve nonconvex PEP to analyze new problems [41, 42]. Simultaneously to their introduction, some of these extensions have been implemented and are available on the *MATLAB* toolbox `PESTO` [62] and *Python* package `PEPit` [36]. The `PEPit` documentation website contains numerous implemented examples of applications of the PEP framework.

Within these works, PEP has been previously used to analyze problems involving linear operators on a few occasions: in [16] for decentralized optimization, in [1] for the random coordinate descent method on nonhomogeneous quadratic functions, and in [69] for ADMM. We will further detail the contribution of these works at the end of section 3.1. In all cases, they present only potential relaxations of the PEP for their specific situation, whereas in this work we propose a provably exact formulation of PEP for general problems with linear operators.

2. PEP formulation. We summarize the PEP methodology as presented in [63] and point out the missing parts to analyze problems involving linear operators. As explained, PEP is a framework that analyzes the worst-case behavior of a given optimization method on a given class of functions. For example, a typical PEP could be formulated as follows (but a lot of variations exist). Given the function class $\mathcal{F}$, the optimization method $\mathcal{A}$ performing N iterations, the initial distance R, the classical performance criterion $f(x_N)-f(x^*)$ (objective function accuracy after N iterations), and $[N]=\{1,\dots,N\}$, the PEP is

$$\text{(PEP)} \qquad \begin{aligned} \max_{x_0,\dots,x_N,x^*,f} \quad & f(x_N)-f(x^*) \\ \text{s.t.} \quad & f\in\mathcal{F}, \\ & x_i \text{ generated by applying } \mathcal{A} \text{ to } f \text{ from } x_{i-1}\ \forall i\in[N], \\ & \|x_0-x^*\|^2\le R^2, \\ & \|\nabla f(x^*)\|^2=0 \text{ (i.e., } x^* \text{ is optimal).} \end{aligned}$$

For simplicity, functions in the class $\mathcal{F}$ considered in this example are convex and smooth, so that the optimality condition for x^* is equivalent to stating $\nabla f(x^*)=0$.

Solving (PEP) yields the worst-case performance that the method $\mathcal{A}$ can exhibit on a function of the class $\mathcal{F}$ for the performance criterion $f(x_N)-f(x^*)$. Moreover, the maximizer will be an example of worst instance reaching that bound. Note that we can use other performance criteria than $f(x_N)-f(x^*)$, e.g., $||x_N-x^*||^2$, $||\nabla f(x_N)||^2$, $\min_{i\in[N]}||\nabla f(x_i)||^2$, $f(\frac{1}{N}\sum_{i\in[N]}x_i)-f(x^*)$, etc.

The conceptual formulation (PEP) is an infinite-dimensional optimization problem as it involves the class of functions $\mathcal{F}$. However, by discretizing and considering function f only on the points actually used by the method $\mathcal{A}$, we can rewrite (PEP) as an equivalent problem with a finite number of variables. Indeed, rather than optimizing over $f\in\mathcal{F}$, we optimize over the points x_i, gradients g_i, and values f_i that are consistent with a function $f\in\mathcal{F}$ and that can be interpolated by this function.

Definition 2.1 ($\mathcal{F}$-interpolability). *Given a function class $\mathcal{F}$, the set of triplets $\{(x_i,g_i,f_i)\}_{i\in[N]}$ is $\mathcal{F}$-interpolable if and only if*

$$\exists f\in\mathcal{F}:\begin{cases} f(x_i)=f_i, \\ \nabla f(x_i)=g_i \end{cases} \quad \forall i\in[N].$$

This definition allows writing the following equivalent discretized formulation of (PEP),

$$\text{(PEP-finite)} \qquad \begin{aligned} \max_{S} \quad & f_N - f^* \\ \text{s.t.} \quad & S \text{ is } \mathcal{F}\text{-interpolable}, \\ & x_i \text{ generated by applying } \mathcal{A} \text{ to } f \text{ from } x_{i-1} \ \forall i \in [N], \\ & \|x_0 - x^*\|^2 \le R^2, \\ & \|g^*\|^2 = 0, \end{aligned}$$

where $S = \{(x_i, g_i, f_i)\}_{i\in\{0\}\cup[N]} \cup \{(x^*, g^*, f^*)\}$. Since we only consider first-order methods, having points x_i, gradients g_i, and function values f_i is enough, as higher-order information on f is not used by the method $\mathcal{A}$.

2.1. Interpolation conditions. It remains to express explicitly the first constraint of (PEP-finite). To do so, we need interpolation conditions on the function class $\mathcal{F}$. These are conditions that must be satisfied by the relevant points x_i, g_i, and f_i to guarantee that there exists a function $f \in \mathcal{F}$ consistent with those points. In other words, constraints on $\{(x_i, g_i, f_i)\}_{i\in[N]}$ are called interpolation conditions of a class $\mathcal{F}$ when they ensure (and are ensured by) $\{(x_i, g_i, f_i)\}_{i\in[N]}$ being $\mathcal{F}$-interpolable. For instance, the following theorem provides interpolation conditions for the class $\mathcal{F}_{\mu,L}$ of L-smooth μ-strongly convex functions.

THEOREM 2.2 ([63, Theorem 4]). *Let $0 \le \mu < L$ and consider the class $\mathcal{F}_{\mu,L}$. The set of triplets $\{(x_i, g_i, f_i)\}_{i\in[N]}$ is $\mathcal{F}_{\mu,L}$-interpolable if and only if $\forall (i,j) \in [N]^2$*

$$2\left(1-\frac{\mu}{L}\right)\left(f_i - f_j - g_j^T(x_i - x_j)\right) \ge \frac{1}{L}\|g_i - g_j\|^2 + \mu\|x_i - x_j\|^2 - 2\frac{\mu}{L}(g_i - g_j)^T(x_i - x_j). \tag{2.1}$$

It turns out that the nature of interpolation conditions allows in many important cases for a tractable formulation of the PEP, this was shown for instance in [63] for the class $\mathcal{F}_{\mu,L}$. This tractable formulation is a semidefinite problem whose variables are the values f_i and the Gram matrix containing the scalar products between all the iterates x_i and gradients g_i. In that case, the explicit formulation of the interpolation conditions can only involve values f_i and scalar products $x_i^T x_j$, $g_i^T g_j$, and $x_i^T g_j$ linearly, e.g., as in (2.1), or in a semidefinite-representable manner. Note that since the variables of the problem are the scalar products between the iterates and the gradients, their dimension no longer appears explicitly in the formulation of the problem. We refer the reader to [63] for more details about the tractable formulation of (PEP) and interpolation conditions.

2.2. Classes of linear operators. In this work, we want to extend PEP to methods involving linear operators. As we saw in the motivating examples of the introduction, such first-order methods typically compute the gradient of the objective function but also products of iterates with M and M^T. More precisely, we can decompose expression (1.1) of the gradient of a composed function $F(x) = g(Mx)$ as

$$v_i = \nabla F(x_i) \quad \Leftrightarrow \quad v_i = M^T \nabla g(Mx_i) \quad \Leftrightarrow \quad \begin{aligned} y_i &= Mx_i, \\ u_i &= \nabla g(y_i), \\ v_i &= M^T u_i. \end{aligned} \tag{2.2}$$

Currently, in the PEP framework, it is known how to incorporate equality $u_i = \nabla g(y_i)$. However, $y_i = Mx_i$ and $v_i = M^T u_i$ need new interpolation conditions. Indeed, analyzing such methods with PEP requires the ability to represent the application of a linear operator to a set of points and possibly also the application of the transpose of the same operator to some other points. We formalize this by defining different classes of linear operators $\mathcal{M}$ of interest together with their respective interpolability. In what follows, we will use matrices M and linear operators $\mathcal{M}$ interchangeably as we only work on finite-dimensional spaces.

We are interested in matrices with bounded eigenvalues or singular value spectrums since the maximal eigenvalue and singular value have an important role in the efficiency of the methods. Such bounds on the spectrum are usually assumed in the literature. We consider general, symmetric, and skew-symmetric matrices, namely,

$$\begin{aligned} \mathcal{L}_L &= \{M : \sigma_{\max}(M) \leq L\} = \{M : M^T M \preceq L^2 I\}, \\ \mathcal{S}_{\mu,L} &= \{Q : Q = Q^T, \mu I \preceq Q \preceq LI\}, \\ \mathcal{T}_L &= \{Q : Q = -Q^T, \sigma_{\max}(Q) \leq L\} = \{Q : Q = -Q^T, Q^T Q \preceq L^2 I\} \end{aligned} \tag{2.3}$$

with $\sigma_{\max}(M)$ the largest singular value of M, and $\preceq$ the Löwner order. Note that $\mathcal{L}_L$ is the class of L-Lipschitz linear operators and $\mathcal{S}_{\mu,L}$ is the class μ-strongly monotone L-Lipschitz symmetric linear operators (see [57] for an analysis on interpolability of general, not necessarily linear operators). We now define operator interpolability in a similar way to function interpolability of Definition 2.1.

DEFINITION 2.3 ($\mathcal{L}_L$-interpolability). *Sets of pairs* $\{(x_i, y_i)\}_{i\in[N_1]}$ *and* $\{(u_j, v_j)\}_{j\in[N_2]}$ *are* $\mathcal{L}_L$*-interpolable if and only if*

$$\exists M \in \mathcal{L}_L \,:\, \begin{cases} y_i = Mx_i & \forall i \in [N_1], \\ v_j = M^T u_j & \forall j \in [N_2]. \end{cases} \tag{2.4}$$

DEFINITION 2.4 ($\mathcal{S}_{\mu,L}$-interpolability). *Set of pairs* $\{(x_i, y_i)\}_{i\in[N]}$ *is* $\mathcal{S}_{\mu,L}$*-interpolable if and only if*

$$\exists Q \in \mathcal{S}_{\mu,L} \,:\, y_i = Qx_i \ \forall i \in [N]. \tag{2.5}$$

DEFINITION 2.5 ($\mathcal{T}_L$-interpolability). *Set of pairs* $\{(x_i, y_i)\}_{i\in[N]}$ *is* $\mathcal{T}_L$*-interpolable if and only if*

$$\exists Q \in \mathcal{T}_L \,:\, y_i = Qx_i \ \forall i \in [N]. \tag{2.6}$$

3. Interpolation conditions for linear operators.

Exploiting PEP to analyze methods involving linear operators requires an explicit formulation of their interpolation conditions. Here we develop tractable necessary and sufficient interpolation conditions for the classes $\mathcal{L}_L$, $\mathcal{T}_L$, and $\mathcal{S}_{\mu,L}$ of linear operators. We also deduce interpolation conditions for the class of quadratic functions.

3.1. Main results.

In the following, to facilitate manipulation of lists of vectors $\{x_i\}_{i\in[N_1]}$, $\{y_i\}_{i\in[N_1]}$, and $\{u_j\}_{j\in[N_2]}, \{v_j\}_{j\in[N_2]}$, we use notations $X = (x_1 \cdots x_{N_1})$, $Y = (y_1 \cdots y_{N_1})$, $U = (u_1 \cdots u_{N_2})$, $V = (v_1 \cdots v_{N_2})$, and write that (X, Y, U, V) is $\mathcal{L}_L$-interpolable when $Y = MX$ and $V = M^T U$ for some $M \in \mathcal{L}_L$. Similarly, we write that (X, Y) is $\mathcal{S}_{\mu,L}$- or $\mathcal{T}_L$-interpolable when $Y = QX$ for some Q in $\mathcal{S}_{\mu,L}$ or $\mathcal{T}_L$. We use $0_{m,n}$ for zero matrices (dimension may be omitted) and $||M|| = \sigma_{\max}(M)$ for the spectral norm of M. We now state our main interpolation theorems.

THEOREM 3.1 ($\mathcal{L}_L$-interpolation conditions). *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 $L \geq 0$.*

Set (X,Y,U,V) is $\mathcal{L}_L$-interpolable if and only if

$$\begin{cases} X^T V = Y^T U, \\ Y^T Y \preceq L^2 X^T X, \\ V^T V \preceq L^2 U^T U. \end{cases} \tag{3.1}$$

Moreover, if $U = X$ and $V = Y$ (resp., $V = -Y$), then the interpolant matrix can be chosen symmetric (resp., skew-symmetric).

COROLLARY 3.2 ($\mathcal{T}_L$-interpolation conditions). *Let $X \in \mathbb{R}^{d\times N}$, $Y \in \mathbb{R}^{d\times N}$, and $L \geq 0$.*

Set (X,Y) is $\mathcal{T}_L$-interpolable if and only if

$$\begin{cases} X^T Y = -Y^T X, \\ Y^T Y \preceq L^2 X^T X. \end{cases} \tag{3.2}$$

THEOREM 3.3 ($\mathcal{S}_{\mu,L}$-interpolation conditions). *Let $X \in \mathbb{R}^{d\times N}$, $Y \in \mathbb{R}^{d\times N}$, and $-\infty < \mu \leq L < \infty$.*

Set (X,Y) is $\mathcal{S}_{\mu,L}$-interpolable if and only if

$$\begin{cases} X^T Y = Y^T X, \\ (Y - \mu X)^T (LX - Y) \succeq 0. \end{cases} \tag{3.3}$$

Crucially, these conditions only involve the scalar products between the columns of $(X\ U)$ and $(Y\ V)$ for $\mathcal{L}_L$ and columns of $(X\ Y)$ for $\mathcal{T}_L$ and $\mathcal{S}_{\mu,L}$, and they are convex semidefinite-representable constraints on the Gram matrices of these scalar products.

Several additional observations can be made about these conditions. First of all, in Theorem 3.1, the condition $X^T V = Y^T U$ is related to the fact that X and Y are linked by the same matrix (but transposed) as U and V. Similarly, in Theorem 3.3 (resp., Corollary 3.2) $X^T Y = Y^T X$ (resp., $X^T Y = -Y^T X$) is related to the symmetry (resp., skew-symmetry) of the operator. Furthermore, a product $X^T X$ is always symmetric positive semidefinite. Finally, in the nonsymmetric and skew-symmetric cases, $Y^T Y \preceq L^2 X^T X$ is related to the fact that the spectral norm is also the induced operator norm in the Euclidean case. In the symmetric case, the situation is a bit more sophisticated due to the presence of a nonzero lower bound on the eigenvalues, therefore, we have $Y = QX$ which cannot be "lower" than μX nor "greater" than LX.

Conditions (3.1) (without the first equation) and (3.3) were presented as necessary conditions for decentralized optimization in Theorem 1 from [16] (in a more specific setting) and in the analysis of ADMM in formulation (13) from [69]. Moreover, in equation (2.7) from [1], the authors further restrict the L-smooth μ-strongly convex interpolation conditions with an additional necessary condition for nonhomogeneous quadratic functions (see section 3.4 for details). Using necessary but not sufficient interpolation conditions in the context of the PEP still allows obtaining bounds on the worst-case performance, but whose tightness is no longer guaranteed. In this work, we show that conditions (3.1) and (3.3) are also sufficient.

3.2. Proofs of the main results. We show that the interpolation conditions are indeed necessary and sufficient for $\mathcal{L}_L$, $\mathcal{T}_L$, and $\mathcal{S}_{\mu,L}$-interpolability. As suggested by the above observations, proving their necessity is much easier than their sufficiency.

To show that conditions (3.1) are sufficient for $\mathcal{L}_L$-interpolability, i.e., for the existence of a linear operator $\mathcal{M}$ for (X,Y,U,V), we will first show that under these conditions there exist factorizations $\left(\begin{smallmatrix} X_R & V_R \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} Y_R & U_R \end{smallmatrix}\right)$ of the Gram matrices $G \triangleq \left(\begin{smallmatrix} X & V \end{smallmatrix}\right)^T \left(\begin{smallmatrix} X & V \end{smallmatrix}\right)$ and $H \triangleq \left(\begin{smallmatrix} Y & U \end{smallmatrix}\right)^T \left(\begin{smallmatrix} Y & U \end{smallmatrix}\right)$, such that (X_R,Y_R,U_R,V_R) is $\mathcal{L}_L$-interpolable (Lemma 3.4). Afterwards, we show how this implies the $\mathcal{L}_L$-interpolability of the actual vectors (X,Y,U,V) (Lemma 3.5). In other words, the proof consists of two steps:

- *Step* 1 (Lemma 3.4)*:* there exist factorizations $\left(\begin{smallmatrix} X_R & V_R \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} Y_R & U_R \end{smallmatrix}\right)$ of the Gram matrices G and H, where (X_R,Y_R,U_R,V_R) is $\mathcal{L}_L$-interpolable;
- *Step* 2 (Lemma 3.5)*:* there exists a rotation from (X_R,Y_R,U_R,V_R) to the initial vectors (X,Y,U,V).

For $\mathcal{S}_{\mu,L}$ and $\mathcal{T}_L$-interpolability, we rely on the fact that the linear operator constructively proposed in the nonsymmetric case is symmetric (resp., skew-symmetric) when we have the symmetric (resp., skew-symmetric) interpolation conditions.

3.2.1. $\mathcal{L}_L$-interpolability of (X_R,Y_R,U_R,V_R) (Step 1). Conditions (3.1) only involve the scalar products between the columns of $\left(\begin{smallmatrix} X & V \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} Y & U \end{smallmatrix}\right)$, in other words, the Gram matrices G and H. Therefore, they apply equally to all sets of vectors leading to the same Gram matrices, i.e., to all factorizations of the Gram matrices. We first show here that at least one of these factorizations is $\mathcal{L}_L$-interpolable.

LEMMA 3.4 (existence of $\mathcal{L}_L$-interpolable factorizations of Gram matrices). *Let two symmetric positive semidefinite matrices $G = \left(\begin{smallmatrix} A_1 & B_1 \\ B_1^T & C_1 \end{smallmatrix}\right)$ and $H = \left(\begin{smallmatrix} A_2 & B_2 \\ B_2^T & C_2 \end{smallmatrix}\right)$, where $A_1, A_2 \in \mathbb{R}^{N_1 \times N_1}$, $C_1, C_2 \in \mathbb{R}^{N_2 \times N_2}$, and $B_1, B_2 \in \mathbb{R}^{N_1 \times N_2}$.*

If G and H satisfy

$$\begin{cases} B_1 = B_2, \\ A_2 \preceq L^2 A_1, \\ C_1 \preceq L^2 C_2, \end{cases} \tag{3.4}$$

then they admit the factorizations

$$G = \begin{pmatrix} X_R & V_R \end{pmatrix}^T \begin{pmatrix} X_R & V_R \end{pmatrix}, \tag{3.5}$$

$$H = \begin{pmatrix} Y_R & U_R \end{pmatrix}^T \begin{pmatrix} Y_R & U_R \end{pmatrix}, \tag{3.6}$$

where

$$X_R = \begin{pmatrix} A_1^{\frac{1}{2}} \\ 0_{N_2,N_1} \end{pmatrix} \in \mathbb{R}^{N_1+N_2 \times N_1}, \quad Y_R = \begin{pmatrix} ({C_2}^{\dagger})^{\frac{1}{2}} B_2^T \\ (A_2 - B_2 {C_2}^{\dagger} B_2^T)^{\frac{1}{2}} \end{pmatrix} \in \mathbb{R}^{N_1+N_2 \times N_1}, \tag{3.7}$$

$$U_R = \begin{pmatrix} C_2^{\frac{1}{2}} \\ 0_{N_1,N_2} \end{pmatrix} \in \mathbb{R}^{N_2+N_1 \times N_2}, \quad V_R = \begin{pmatrix} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 \\ (C_1 - B_1^T {A_1}^{\dagger} B_1)^{\frac{1}{2}} \end{pmatrix} \in \mathbb{R}^{N_2+N_1 \times N_2}, \tag{3.8}$$

such that (X_R,Y_R,U_R,V_R) is $\mathcal{L}_L$-interpolable.

Moreover, if $A_1 = C_2$, $A_2 = C_1$, and $B_1 = B_1^T$ (resp., $B_1 = -B_1^T$), then $U_R = X_R$, $V_R = Y_R$ (resp., $V_R = -Y_R$), and the interpolant matrix is symmetric (resp., skew-symmetric).

Proof. The proof consists in explicitly building a matrix M_R defining the operator such that $Y_R = M_R X_R$, $V_R = M_R^T U_R$, and $||M_R|| \leq L$. The proof is deferred to Appendix B. ∎

This lemma guarantees that, if (X,Y,U,V) satisfies (3.1), then their associated Gram matrices admit a factorization (X_R,Y_R,U_R,V_R) that is $\mathcal{L}_L$-interpolable, but does not yet guarantee the interpolability of the initial vectors (X,Y,U,V) themselves. For example, consider a symmetric case where we just have (X,Y) (hence $N_2=0$) and let

$$X = \begin{pmatrix}1\\1\\1\end{pmatrix},\ Y = \begin{pmatrix}1\\1\\0\end{pmatrix} \qquad \text{(with } N_1 = 1\text{)} \tag{3.9}$$

which satisfy the $\mathcal{S}_{\mu,L}$-interpolation conditions (3.3) for $\mu = 0$ and $L = 1$. If, given these data (X,Y), one computes the Gram matrix $(\,X\ Y\,)^T(\,X\ Y\,) = \left(\begin{smallmatrix}3&2\\2&2\end{smallmatrix}\right)$, it turns out that via Lemma 3.4 one will obtain the following new and different factorization

$$X_R = \begin{pmatrix}\sqrt{3}\\0\end{pmatrix},\ Y_R = \begin{pmatrix}\frac{2\sqrt{3}}{3}\\\frac{\sqrt{6}}{3}\end{pmatrix}. \tag{3.10}$$

Hence Lemma 3.4 does not directly guarantee the $\mathcal{S}_{\mu,L}$-interpolability of (X,Y), but only of (X_R,Y_R), sharing the same Gram matrix but not necessarily equal to or even having the same dimension as (X,Y). The next lemma guarantees that (X,Y) is also $\mathcal{S}_{\mu,L}$-interpolable.

3.2.2. Rotation to (X,Y,U,V) (Step 2).

We now show that if a given factorization of Gram matrices is $\mathcal{L}_L$-interpolable, then all factorizations of these Gram matrices are $\mathcal{L}_L$-interpolable.

LEMMA 3.5 (rotation between (X,Y,U,V) and (X_R,Y_R,U_R,V_R)). *Let $X_R \in \mathbb{R}^{n_R\times N_1}$, $Y_R \in \mathbb{R}^{m_R\times N_1}$, $U_R \in \mathbb{R}^{m_R\times N_2}$, and $V_R \in \mathbb{R}^{n_R\times N_2}$.*

If (X_R,Y_R,U_R,V_R) is $\mathcal{L}_L$-interpolable, then (X,Y,U,V) is $\mathcal{L}_L$-interpolable for all $X \in \mathbb{R}^{n\times N_1}$, $Y \in \mathbb{R}^{m\times N_1}$, $U \in \mathbb{R}^{m\times N_2}$, and $V \in \mathbb{R}^{n\times N_2}$ such that

$$\begin{aligned}\begin{pmatrix}X & V\end{pmatrix}^T\begin{pmatrix}X & V\end{pmatrix} &= \begin{pmatrix}X_R & V_R\end{pmatrix}^T\begin{pmatrix}X_R & V_R\end{pmatrix},\\ \begin{pmatrix}Y & U\end{pmatrix}^T\begin{pmatrix}Y & U\end{pmatrix} &= \begin{pmatrix}Y_R & U_R\end{pmatrix}^T\begin{pmatrix}Y_R & U_R\end{pmatrix}.\end{aligned} \tag{3.11}$$

Moreover, if $U = X$, $V = Y$ (resp., $V = -Y$), $U_R = X_R$, $V_R = Y_R$ (resp., $V_R = -Y_R$), and interpolant matrix M_R is symmetric (resp., skew-symmetric), then interpolant matrix M can be chosen symmetric (resp., skew-symmetric).

Proof. The central idea is that all sets of vectors with the same Gram matrices are equivalent up to a rotation. Some care must however be taken because the dimensions of these vectors are not necessarily the same. The proof is deferred to Appendix C. ∎

We are now able to prove the necessity and sufficiency of interpolation conditions in the nonsymmetric case. The necessity is obtained by a straightforward reasoning and the sufficiency relies only on the successive application of Lemmas 3.4 and 3.5.

3.2.3. Proof of Theorem 3.1.

Proof. (necessity) Let us assume that (X,Y,U,V) is $\mathcal{L}_L$-interpolable. First, $Y = MX$ and $V = M^TU$ yield $X^TV = X^TM^TU = Y^TU$. Moreover, $MM^T \preceq L^2 I$ implies

$U^TMM^TU \preceq L^2U^TU$, i.e., $V^TV \preceq L^2U^TU$ and similarly, $M^TM \preceq L^2I$ implies $X^TM^TMX \preceq L^2X^TX$ i.e. $Y^TY \preceq L^2X^TX$.

(sufficiency) Let us assume that (X,Y,U,V) satisfies conditions (3.1). From Lemma 3.4, there exists (X_R,Y_R,U_R,V_R) which is $\mathcal{L}_L$-interpolable and shares the same Gram matrices as (X,Y,U,V). Thus, by Lemma 3.5, (X,Y,U,V) is $\mathcal{L}_L$-interpolable. Finally, if $U=X$ and $V=Y$ (resp., $V=-Y$), then we can choose $U_R=X_R$, $V_R=Y_R$ (resp., $V_R=-Y_R$), and M_R symmetric (resp., skew-symmetric) in Lemma 3.4 and thus M symmetric (resp., skew-symmetric) in Lemma 3.5. ∎

The results on the symmetric and skew-symmetric cases come from the nonsymmetric result. In the symmetric case, they correspond to matrices with eigenvalues belonging to the centered interval $[-L,L]$. We apply a shift to allow a general interval $[\mu,L]$. In the skew-symmetric case, the sufficiency comes from Theorem 3.1 whereas the necessity follows a similar reasoning to the nonsymmetric case.

3.2.4. Proof of Theorem 3.3.

Proof. First, let us define $\tilde{X}=X$ and $\tilde{Y}=Y-\frac{L+\mu}{2}X$ and show that requiring (X,Y) to satisfy (3.3) is equivalent to

$$\begin{cases}\tilde{X}^T\tilde{Y}=\tilde{Y}^T\tilde{X},\\ \tilde{Y}^T\tilde{Y}\preceq\left(\frac{L-\mu}{2}\right)^2\tilde{X}^T\tilde{X}.\end{cases}\tag{3.12}$$

Indeed, $X^TY=Y^TX \Leftrightarrow X^TY-\frac{L+\mu}{2}X^TX=Y^TX-\frac{L+\mu}{2}X^TX \Leftrightarrow \tilde{X}^T\tilde{Y}=\tilde{Y}^T\tilde{X}$ and

$$\begin{aligned}(Y-\mu X)^T(Y-LX)\preceq 0 &\Leftrightarrow \left(\tilde{Y}+\frac{L-\mu}{2}X\right)^T\left(\tilde{Y}-\frac{L-\mu}{2}X\right)\preceq 0\\ &\Leftrightarrow \tilde{Y}^T\tilde{Y}\preceq\left(\frac{L-\mu}{2}\right)^2X^TX.\end{aligned}\tag{3.13}$$

Second, from Theorem 3.1 with $U=X$ and $V=Y$, conditions (3.12) are equivalent to $(\tilde{X},\tilde{Y})$ being $\mathcal{S}_{-\frac{L-\mu}{2},\frac{L-\mu}{2}}$-interpolable, therefore, $\tilde{Y}=\tilde{Q}\tilde{X}$ for some $\tilde{Q}\in\mathcal{S}_{-\frac{L-\mu}{2},\frac{L-\mu}{2}}$.

Third, $\tilde{Y}=\tilde{Q}\tilde{X}$ with $\tilde{Q}\in\mathcal{S}_{-\frac{L-\mu}{2},\frac{L-\mu}{2}}$ is equivalent to $Y=QX$ with $Q\in\mathcal{S}_{\mu,L}$, indeed, $\tilde{Y}=\tilde{Q}\tilde{X}\Leftrightarrow Y-\frac{L+\mu}{2}X=\tilde{Q}\tilde{X}\Leftrightarrow Y=(\tilde{Q}+\frac{L+\mu}{2}I)X=QX$. ∎

3.3. Limiting cases.

Theorems 3.1 and 3.3 assume finite values of L; we now extend them to "$L\to\infty$," that is unbounded singular values and eigenvalues. Again, we are interested in an explicit convex formulation of the conditions. It is not straightforward to take the limit of conditions (3.1), i.e., $\lim_{L\to\infty}: Y^TY\preceq L^2X^TX$. Indeed, the constraint must still impose in the limit that the nullspace of X^TX is included in the one of Y^TY, which is not directly semidefinite representable. However, it is possible to obtain a tractable formulation of the conditions by considering $\exists L>0: Y^TY\preceq L^2X^TX$ instead of the limit. We define $\mathcal{L}$ the class of matrices with arbitrary real singular values and propose the following $\mathcal{L}$-interpolation conditions.

THEOREM 3.6 ($\mathcal{L}$-interpolation conditions). *Let $X\in\mathbb{R}^{n\times N_1}$, $Y\in\mathbb{R}^{m\times N_1}$, $U\in\mathbb{R}^{m\times N_2}$, and $V\in\mathbb{R}^{n\times N_2}$.*

(X,Y,U,V) is $\mathcal{L}$-interpolable if and only if

$$\exists L>0:\begin{cases}X^TV=Y^TU,\\ \begin{pmatrix}X^TX & Y^TY\\ Y^TY & L^2I\end{pmatrix}\succeq 0,\\ \begin{pmatrix}U^TU & V^TV\\ V^TV & L^2I\end{pmatrix}\succeq 0.\end{cases}\tag{3.14}$$

Proof. By Theorem 3.1, (X,Y,U,V) is $\mathcal{L}$-interpolable if and only if

$$\exists L_1>0:\begin{cases}X^TV=Y^TU,\\ Y^TY\preceq L_1^2X^TX,\\ V^TV\preceq L_1^2U^TU,\end{cases}\quad\overset{\text{Prop. A.5}}{\Leftrightarrow}\exists L_2,L_3>0:\begin{cases}X^TV=Y^TU,\\ (Y^TY)^2\preceq L_2^2X^TX,\\ (V^TV)^2\preceq L_3^2U^TU,\end{cases}$$

$$\overset{\text{Prop. A.1}}{\Leftrightarrow}\exists L_2,L_3>0:\begin{cases}X^TV=Y^TU,\\ \begin{pmatrix}X^TX & Y^TY\\ Y^TY & L_2^2I\end{pmatrix}\succeq 0,\\ \begin{pmatrix}U^TU & V^TV\\ V^TY & L_3^2I\end{pmatrix}\succeq 0,\end{cases}$$

where we can take $L=\max\{L_2,L_3\}$. ∎

Existence of L is not an issue for the PEP framework since we can just add a scalar variable to represent L^2. Conditions (3.14) of Theorem 3.6 are thus convex on the Gram matrices $(\,X\ V\,)^T(\,X\ V\,)$, $(\,Y\ U\,)^T(\,Y\ U\,)$, and L^2.

3.4. Interpolation conditions for quadratic functions. Let $\mathcal{Q}_{\mu,L}$ be the class of homogeneous quadratic functions $f(x)=\frac{1}{2}x^TQx$, where $\mu I\preceq Q\preceq LI$. Our new theorems allow us to write the interpolation conditions of $\mathcal{Q}_{\mu,L}$.

THEOREM 3.7 ($\mathcal{Q}_{\mu,L}$-interpolation conditions). *Let $-\infty<\mu\le L<\infty$. The set of triplets $\{(x_i,g_i,f_i)\}_{i\in[N]}$ is $\mathcal{Q}_{\mu,L}$-interpolable if and only if*

$$\begin{cases}X^TG=G^TX,\\ (G-\mu X)^T(LX-G)\succeq 0,\\ f_i=\frac{1}{2}x_i^Tg_i,\quad\forall i\in[N],\end{cases}\tag{3.15}$$

where $X=(x_1\ \cdots\ x_N)$ and $G=(g_1\ \cdots\ g_N)$.

Proof. We have that $\{(x_i,g_i,f_i)\}_{i\in[N]}$ $\mathcal{Q}_{\mu,L}$-interpolable if and only if

$$\exists Q\in\mathcal{S}_{\mu,L}:\begin{cases}g_i=Qx_i & \forall i\in[N],\\ f_i=\frac{1}{2}x_i^TQx_i & \forall i\in[N],\end{cases}\quad\Leftrightarrow\quad\begin{cases}g_i=Qx_i & \forall i\in[N],\\ f_i=\frac{1}{2}x_i^Tg_i & \forall i\in[N],\end{cases}$$

$$\overset{\text{Th. 3.3}}{\Leftrightarrow}\begin{cases}X^TG=G^TX,\\ (G-\mu X)^T(LX-G)\succeq 0,\\ f_i=\frac{1}{2}x_i^Tg_i\quad\forall i\in[N].\end{cases}$$ ∎

As we have the inclusion $\mathcal{Q}_{\mu,L}\subseteq\mathcal{F}_{\mu,L}$, the quadratic interpolation conditions of Theorem 3.7 must imply the general smooth strongly convex interpolation conditions. Indeed, one can also show algebraically that conditions (3.15) imply conditions (2.1).

As mentioned earlier, the recent work [1] proposed necessary interpolation conditions (originally announced in [24]) for the class of nonhomogeneous quadratic functions $f(x)=\frac{1}{2}x^TQx-b^Tx$ (see their section 2.2). The authors used the L-smooth μ-strongly convex interpolation conditions (2.1) in addition to the following necessary conditions for nonhomogeneous quadratic functions

$$\frac{1}{2}\left(g_i+g_j\right)^T\left(x_i-x_j\right)=f_i-f_j\ \ \forall(i,j)\in[N]^2.\tag{3.16}$$

These conditions are not sufficient. Indeed, there exist sets of points satisfying both conditions (2.1) and (3.16) and that cannot be interpolated by a quadratic function. For example, the following set $\{(x_i, g_i, f_i)\}_{i\in[3]}$

$$\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}0\\0\end{pmatrix},0\right),\ \left(\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}\frac{L+\mu}{2}\\\frac{L-\mu}{2}\end{pmatrix},\frac{L+\mu}{4}\right),\left(\begin{pmatrix}-1\\0\end{pmatrix},\begin{pmatrix}-\frac{(L+\mu)}{2}\\\frac{L-\mu}{2}\end{pmatrix},\frac{L+\mu}{4}\right), \tag{3.17}$$

satisfies conditions (2.1) and (3.16). Yet, there is no function of the form $f(x) = \frac{1}{2}x^TQx - b^Tx$ that interpolates these points. Indeed, the first triplet implies $f(0)=0$ forcing the quadratic to be homogeneous, i.e., $b = 0$, and then the linearity of the gradient imposes that the values of the gradient on the two other points with $x_2 = -x_1$ are opposite (since we have $g_i = Qx_i$ and $x_3 = -x_2$ which implies $g_3 = g_2$). This is impossible unless $\frac{L-\mu}{2} = 0$ (and when $\mu = L$, the class of L-smooth L-strongly convex functions is already the class of nonhomogeneous quadratic functions).

4. Performance estimation of methods involving linear operators.

The new interpolation conditions of Theorems 3.1 and 3.3 allow the analysis of any function class currently available in PEP (smooth strongly convex, proximable, hypoconvex, etc.) combined with a linear operator, leading to the exact worst-case performance of methods applied to these classes. In practice, we used the MATLAB toolbox `PESTO` [62] to which we added our new interpolation conditions. The toolbox and our extension is available in *Python* via the library `PEPit` [36] and the numerical experiments performed in this work can be found at https://github.com/NizarBousselmi/PEP-Linear-Operator-code/tree/main. Semidefinite programs are solved by the *MOSEK* [5] interior-point semidefinite optimization solver.

In this section, we demonstrate the applicability of our extension of the PEP framework. First, we analyze the worst-case performance of the gradient method applied to the first motivating example, $\min_x g(Mx)$ (which covers problem $\min_x \frac{1}{2}x^TQx$ as a special case). Then, we analyze the more recent and practical Chambolle–Pock algorithm [10], which was our second motivating example. For the latter, we show the flexibility of our approach with several types of performance guarantees, both in settings found in the literature and in new, previously unstudied settings.

4.1. Gradient method on $g \circ \mathcal{M}$.

Let us define the class $\mathcal{C}^{0,L_M}_{\mu_g,L_g}$ of functions of the form

$$F = g \circ \mathcal{M}, \tag{4.1}$$

where g is an L_g-smooth μ_g-strongly convex function (where $0 < \mu_g \leq L_g$) and $\mathcal{M} : x \mapsto Mx$ is defined with a general, not necessarily symmetric, matrix M with singular values between 0 and L_M. We also define the class $\mathcal{D}^{\mu_M,L_M}_{\mu_g,L_g}$, where M is symmetric with eigenvalues between μ_M and L_M, and $0 \leq \mu_M \leq L_M$.

By definition of the classes, we have $\mathcal{C}^{0,L_M}_{\mu_g,L_g} = \mathcal{C}^{0,1}_{\mu_g L_M^2, L_g L_M^2}$ and $\mathcal{D}^{\mu_M,L_M}_{\mu_g,L_g} = \mathcal{D}^{\mu_M/L_M,1}_{\mu_g L_M^2, L_g L_M^2}$, therefore, we will only consider the case $L_M = 1$ without loss of generality. For comparison purposes, we will also look at the class $\mathcal{F}_{\mu_f,L_f}$ of functions of the form $F = f$, where f is an L_f-smooth μ_f-strongly convex function.

We analyze the worst-case performance of the gradient method with fixed step

$$x_{k+1} = x_k - \frac{h}{L_F}\nabla F(x_k) \tag{GM}$$

on the problem $\min_x F(x)$, where L_F is the smoothness constant of the class considered, namely, L_g for $\mathcal{C}^{0,1}_{\mu_g,L_g}$ and $\mathcal{D}^{\mu_M,1}_{\mu_g,L_g}$ and L_f for $\mathcal{F}_{\mu_f,L_f}$.

Given a bound R on the initial distance to the solution $||x_0-x^*||$, we are interested in the worst-case performance $w(\mathcal{F}, R, N, \frac{h}{L_F})$ of N iterations of the gradient method with step size $\frac{h}{L_F}$ on the function class $\mathcal{F}$. We define $w(\mathcal{F}, R, N, \frac{h}{L_F})$ as the value of the solution of (PEP) where the method $\mathcal{A}$ is (GM) with step size $\frac{h}{L_F}$, which allows writing the following guarantee

$$F(x_N) - F(x^*) \leq w\left(\mathcal{F}, R, N, \frac{h}{L_F}\right) \quad \forall F \in \mathcal{F} \tag{4.2}$$

with x_N the Nth iterate of (GM) with step size $\frac{h}{L_F}$ on F, and x^* the minimizer of F.

The worst-cases guarantees for classes $\mathcal{F}_{\mu_f, L_f}$ and $\mathcal{C}^{0,1}_{\mu_g, L_g}$ reduce to simpler cases with the following homogeneity relations (see [61, section 4.2.5] for a proof) (semicolons are introduced for readability),

$$\begin{aligned} w\left(\mathcal{F}_{\mu_f, L_f}; R, N, \frac{h}{L_f}\right) &= L_f R^2 w\left(\mathcal{F}_{\frac{\mu_f}{L_f}, 1}; 1, N, h\right), \\ w\left(\mathcal{C}^{0,1}_{\mu_g, L_g}; R, N, \frac{h}{L_g}\right) &= L_g R^2 w\left(\mathcal{C}^{0,1}_{\frac{\mu_g}{L_g}, 1}; 1, N, h\right). \end{aligned} \tag{4.3}$$

Therefore, without loss of generality, we can consider the cases $L_f = L_g = R = 1$, i.e., $w(\mathcal{F}_{\mu_f,1}; 1, N, h)$ and $w(\mathcal{C}^{\mu_M,1}_{\mu_g,1}; 1, N, h)$, from which we will deduce the worst-case for the general cases. We will use the following shortened notations (we have the same results and notations for $\mathcal{D}^{\mu_M,1}_{\mu_g,L_g}$):

$$w\left(\mathcal{F}_{\mu_f,1}; 1, N, h\right) = w\left(\mathcal{F}_{\mu_f}; h\right), \qquad w\left(\mathcal{C}^{0,1}_{\mu_g,1}; 1, N, h\right) = w\left(\mathcal{C}^{0}_{\mu_g}; h\right).$$

Note that $\mathcal{F}_{\mu_g} \subseteq \mathcal{C}^0_{\mu_g} \subseteq \mathcal{F}_0$ and $\mathcal{F}_{\mu_g} \subseteq \mathcal{D}^{\mu_M}_{\mu_g} \subseteq \mathcal{F}_{\mu_g \mu_M^2}$, therefore, $w(\mathcal{F}_{\mu_g}; h) \leq w(\mathcal{C}^0_{\mu_g}; h) \leq w(\mathcal{F}_0; h)$ and $w(\mathcal{F}_{\mu_g}; h) \leq w(\mathcal{D}^{\mu_M}_{\mu_g}; h) \leq w(\mathcal{F}_{\mu_g \mu_M^2}; h)$ will always hold. All inclusions and inequalities become equalities when $\mu_g = 0$.

We are not aware of works addressing the performance of the gradient method on such classes. Reference [53] provided a bound on the performance of a gradient method with unit step size on functions $F = g \circ M$ when M has a nonzero lower bound on its singular values, an interesting case but out of our scope of study.

Solving (PEP) for the new class $\mathcal{C}^0_{\mu_g}$ with $\mu_g = 0.1$ yields numerical results in Figure 1, namely, the worst-case performance of the gradient method after $N = 10$ iterations for varying step size $h \in [0, 2]$ when applied to classes $\mathcal{F}_0$, $\mathcal{C}^0_{0.1}$, and $\mathcal{F}_{0.1}$.

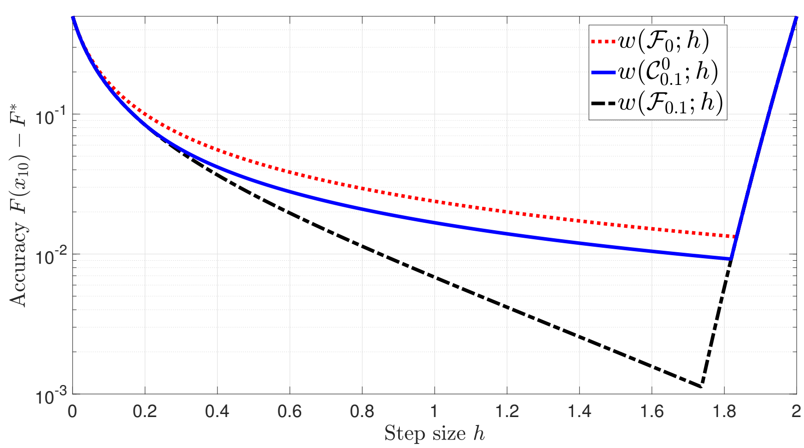

FIG. 1. *Worst-case performance of* 10 *iterations of* (GM) *for varying step size* $h \in [0, 2]$ *on classes* $\mathcal{F}_0$ *of* 1*-smooth convex functions* f *(dotted red line),* $\mathcal{C}^0_{0.1}$ *of* 1*-smooth* 0.1*-strongly convex functions* $g \circ M$ *(solid blue line), and* $\mathcal{F}_{0.1}$ *of* 1*-smooth* 0.1*-strongly convex functions* f *(broken black line). Note: color appears only in the online article.*

The double inequality $w(\mathcal{F}_{0.1};h) \leq w(\mathcal{C}^0_{0.1};h) \leq w(\mathcal{F}_0;h)$ is confirmed, and observed to be strict.

After extensive numerical computations with PESTO and analysis of results such as that depicted in Figure 1, we were able to identify the exact worst-case performances $w(\mathcal{C}^0_{\mu_g};h)$ and $w(\mathcal{D}^{\mu_M}_{\mu_g};h)$ for all values of parameters μ_g, μ_M, and h. Before we show the analytical expressions of these worst-cases in section 4.1.2, we review the results of [63] on the worst-case performance of (GM) on the class $\mathcal{F}_{\mu_f}$.

4.1.1. Performance on $\boldsymbol{\mathcal{F}_\mu}$. Let $N \geq 0$, $h \in [0,2]$, functions ℓ, $q \in \mathcal{F}_\mu$ as

$$\begin{aligned} \ell_{\mu,h}(x) &= \begin{cases} \frac{\mu}{2}x^2 + (1-\mu)\tau_{\mu,h}|x| - \left(\frac{1-\mu}{2}\right)\tau^2_{\mu,h} & \text{if } |x| \geq \tau_{\mu,h}, \\ \frac{1}{2}x^2 & \text{else,} \end{cases} \\ q(x) &= \frac{1}{2}x^2, \end{aligned} \tag{4.4}$$

and $\tau_{\mu,h} = \frac{\mu}{\mu - 1 + (1-\mu h)^{-2N}}$. It is conjectured in [63], with very strong numerical evidence, that the worst-case performance of (GM) on $\mathcal{F}_{\mu_f}$ is given by

$$w\left(\mathcal{F}_{\mu_f};h\right) = \frac{1}{2}\max\left\{\frac{\mu_f}{\mu_f - 1 + (1-\mu_f h)^{-2N}}, (1-h)^{2N}\right\}. \tag{4.5}$$

Moreover, as one can check that this worst-case performance corresponds to that of the one-dimensional functions $\ell_{\mu_f,h}$ and q, this conjecture also implies that the worst-case $w\left(\mathcal{F}_{\mu_f};h\right)$ is attained by one-dimensional functions.

4.1.2. Performance on $\boldsymbol{\mathcal{C}^0_{\mu_g}}$ and $\boldsymbol{\mathcal{D}^{\mu_M}_{\mu_g}}$. Through numerous numerical experiments, we also observed that the worst-case functions are one-dimensional on these two classes. Actually, a more important observation is that, in the worst-case, the operator $\mathcal{M}$ is a pure scaling (a multiple of the identity), i.e., $M = \alpha I$ for some $\alpha \in \mathbb{R}$. Therefore, we denote $\tilde{\mathcal{C}}^0_{\mu_g}$ (resp., $\tilde{\mathcal{D}}^{\mu_M}_{\mu_g}$) the subclass of functions $g \circ M$, where $M = \alpha I$ is a scaling operator and propose the following conjecture, supported by our numerical evidence.

Conjecture 4.1. *Worst-case performances of $\mathcal{C}^0_{\mu_g}$ and $\mathcal{D}^{\mu_M}_{\mu_g}$ are reached by scaling operator $M = \alpha I$, i.e., $w(\mathcal{C}^0_{\mu_g};h) = w(\tilde{\mathcal{C}}^0_{\mu_g};h)$ and $w(\mathcal{D}^{\mu_M}_{\mu_g};h) = w(\tilde{\mathcal{D}}^{\mu_M}_{\mu_g};h)$.*

From now on, given Conjecture 4.1 and the fact that $\tilde{\mathcal{C}}^0_{\mu_g} = \tilde{\mathcal{D}}^0_{\mu_g}$, we will only present the analysis for the symmetric case, as the general case shares the same analysis (but with $\mu_M = 0$). The class of functions $g \circ \alpha I$ can be written as a union of classes of functions f,

$$\tilde{\mathcal{D}}^{\mu_M}_{\mu_g} = \bigcup_{\alpha \in [\mu_M, 1]} \mathcal{F}_{\mu_g \alpha^2, \alpha^2}, \tag{4.6}$$

which allows us to express the worst-case performance $w(\mathcal{D}^{\mu_M}_{\mu_g};h)$ thanks to $w(\mathcal{F}_{\mu_f};h)$ with

$$\begin{aligned} w\left(\mathcal{D}^{\mu_M}_{\mu_g};h\right) &\overset{\text{Conj. 4.1}}{=} w\left(\tilde{\mathcal{D}}^{\mu_M}_{\mu_g};h\right) \\ &\overset{(4.6)}{=} w\left(\bigcup_{\alpha \in [\mu_M, 1]} \mathcal{F}_{\mu_g \alpha^2, \alpha^2};h\right) \\ &= \max_{\alpha \in [\mu_M, 1]} w\left(\mathcal{F}_{\mu_g \alpha^2, \alpha^2};h\right) \\ &\overset{(4.3)}{=} \max_{\alpha \in [\mu_M, 1]} \alpha^2 w\left(\mathcal{F}_{\mu_g};\alpha^2 h\right). \end{aligned} \tag{4.7}$$

This reasoning holds only if we know, or conjecture, that the worst-case operator of a method is a scaling of the form $M = \alpha I$. We conjectured it for (GM) but it is not the case in general. Similarly, we must not infer that the worst-case operator M is symmetric for all methods and settings. Indeed, there exist algorithms where the worst-case performance on symmetric operators is strictly better than its performance on general linear operators.

Given expression (4.5) of $w(\mathcal{F}_{\mu_f}; h)$, it is possible to solve the last maximization problem of (4.7) and end up with the following conjecture featuring an explicit convergence rate.

Conjecture 4.2 (worst-case performance of the gradient method applied to a function in $\mathcal{D}_{\mu_g}^{\mu_M}$). *For all* $0 < \mu_g \leq 1$ *and* $0 \leq \mu_M \leq 1$*, we have*

$$w\left(\mathcal{D}_{\mu_g}^{\mu_M}; h\right) = \frac{1}{2} \max\left\{\frac{\mu_g {\alpha^*}^2}{\mu_g - 1 + \left(1 - \mu_g {\alpha^*}^2 h\right)^{-2N}}, (1-h)^{2N}\right\}, \tag{4.8}$$

where $\alpha^* = \operatorname{proj}_{[\mu_M,1]}\left(\sqrt{\frac{h_0}{h}}\right)$ *with* h_0 *the unique solution in* $[0, \frac{1}{\mu_g}]$ *of equation*

$$(1-\mu_g)(1-\mu_g h_0)^{2N+1} = 1 - (2N+1)\mu_g h_0. \tag{4.9}$$

Moreover, we can exhibit functions reaching this worst-case, therefore, guaranteeing that the worst-case cannot be better (i.e., lower).

Theorem 4.3 (lower bound on the worst-case performance $w(\mathcal{D}_{\mu_g}^{\mu_M}; h)$). *For all* $0 < \mu_g \leq 1$ *and* $0 \leq \mu_M \leq 1$*, we have*

$$w\left(\mathcal{D}_{\mu_g}^{\mu_M}; h\right) \geq \frac{1}{2} \max\left\{\frac{\mu_g {\alpha^*}^2}{\mu_g - 1 + \left(1 - \mu_g {\alpha^*}^2 h\right)^{-2N}}, (1-h)^{2N}\right\} \tag{4.10}$$

with α^* *defined in Conjecture* 4.2.

Proof. Functions ${\alpha^*}^2 \ell_{\mu_g, {\alpha^*}^2 h}(x)$ and $q(x)$, defined in (4.4), belong to $\mathcal{D}_{\mu_g}^{\mu_M}$ and reach the performance (4.10). ∎

The worst-case performance established in Conjecture 4.2 matches exactly the large number of numerical experiments we performed for many different values of parameters μ_g, μ_M, and h. Moreover, development (4.7) shows that Conjecture 4.2 relies only on the weaker Conjecture 4.1 and on the previous conjectures of [63]. We summarize this observation in a corollary.

Corollary 4.4. *If Conjecture* 4.1 *holds and* $w(\mathcal{F}_{\mu_f}; h)$ *is given by* (4.5) *as conjectured in* [63]*, then Conjecture* 4.2 *holds.*

All these observations and conjectures were made possible thanks to the numerical experiments performed on `PESTO` with our extension and the extremely helpful insight and information provided by the solution of the different (PEP) solved.

So far we proposed several partial justifications for Conjecture 4.2 on the worst-case performance of the gradient method. We now give a full proof of the conjecture in a simple case, namely, one iteration of the gradient method with a unit step.

Proof of Conjecture 4.2 *when* $N = 1$, $h = 1$, $\mu_M = 0$. The optimal primal solution of (PEP) yields a function reaching the worst-case performance of interest. In addition, we can extract from the optimal dual solution a combination of inequalities that builds a proof for this performance guarantee (see [35] for more details). Moreover, that type of proof mainly relies on the interpolation conditions. Therefore, solving (PEP) with our new interpolation conditions will provide a proof that exploits these new conditions. We propose a simple example of such a proof for illustration. Analyzing any method with the PEP framework and our interpolation conditions can in principle provide this type of proof.

We consider Conjecture 4.2 in the case $N=1$, $h=1$, $\mu_M=0$, and $\mu_g \le 0.17$. We still consider the case $L_M=1$ but we do not replace it in the proof for clarity and understanding. The proof below, which shows that $F(x_1)-F(x^*) \le \tau ||x_0-x^*||^2$, was identified from the numerical dual solution of the PEP.

$$
\begin{aligned}
&F(x_1) - F(x^*) \\
&= \boxed{\begin{aligned}
&\gamma \Big(F(x_1) - F(x_0) + \nabla g(x_1)^T M(x_0 - x_1) + \tfrac{\mu_g}{2(1-\mu_g)} ||M(x_0 - x_1)||^2 \\
&+ \tfrac{1}{2(1-\mu_g)} ||\nabla g(Mx_0) - \nabla g(Mx_1)||^2 - \tfrac{\mu_g}{1-\mu_g} (\nabla g(Mx_0) - \nabla g(Mx_1))^T M(x_0 - x_1) \Big) \\
&+ \gamma \Big(F(x_0) - F(x^*) + \nabla g(Mx_0)^T M(x^* - x_0) + \tfrac{\mu_g}{2(1-\mu_g)} ||M(x^* - x_0)||^2 \\
&+ \tfrac{1}{2(1-\mu_g)} ||\nabla g(Mx^*) - \nabla g(Mx_0)||^2 - \tfrac{\mu_g}{1-\mu_g} (\nabla g(Mx^*) - \nabla g(Mx_0))^T M(x^* - x_0) \Big) \\
&+ (1-\gamma) \Big(F(x_1) - F(x^*) + \nabla g(Mx_1)^T M(x^* - x_1) + \tfrac{\mu_g}{2(1-\mu_g)} ||M(x^* - x_1)||^2 \\
&+ \tfrac{1}{2(1-\mu_g)} ||\nabla g(Mx^*) - \nabla g(Mx_1)||^2 - \tfrac{\mu_g}{1-\mu_g} (\nabla g(Mx^*) - \nabla g(Mx_1))^T M(x^* - x_1) \Big)
\end{aligned}} \\
&\quad \boxed{\begin{aligned}
&+ \left(\tau - \tfrac{1-\gamma}{4}\right) \left(||M(x^* - x_0)||^2 - L_M^2 ||x^* - x_0||^2 \right) \\
&+ \tfrac{1-\gamma}{4} \left(||M(x_0 + x^* - 2x_1)||^2 - L_M^2 ||x^* + x_0 - 2x_1||^2 \right)
\end{aligned}} \\
&\quad \boxed{-||Pa_{\mu_g}||^2 - ||Pb_{\mu_g}||^2 - ||Pc_{\mu_g}||^2} \\
&\quad + \tau ||x_0 - x^*||^2 \\
&\le \tau ||x_0 - x^*||^2,
\end{aligned} \tag{4.11}
$$

where $P = \left(Mx_0 \; Mx_1 \; Mx^* \; \nabla g(Mx_0) \; \nabla g(Mx_1) \; \nabla g(Mx^*) \right)$, $\gamma = \frac{1-\mu_g}{2-\mu_g}$, $\tau = \frac{\mu_g}{2(\mu_g - 1 + (1-\mu_g)^{-2})}$, $x_1 = x_0 - M^T \nabla g(Mx_0)$, and a_{μ_g}, b_{μ_g}, c_{μ_g} are defined in Appendix D. The expression of the rate τ matches Conjecture 4.2 for this set of parameters. The equality is an algebraic reformulation that is always satisfied (see Appendix D). The terms in the first box are nonpositive thanks to the $\mathcal{F}_{\mu_g,1}$-interpolation conditions applied to the function g, and the terms in the second box are nonpositive thanks to the $\mathcal{L}_{L_M}$-interpolation conditions applied to the linear operator M. Finally, the terms in the third box are a negative sum of squares, thus, always nonpositive.

Comparison between f and $g \circ M$. Table 1 compares the performance of N iterations of (GM) with step size h on the classes $\mathcal{F}_{\mu_f}$ and $\mathcal{D}_{\mu_g}^{\mu_M}$ on the functions ℓ and q, namely,

TABLE 1
Worst-case performances and functions of $\mathcal{F}_{\mu_f}$ and $\mathcal{D}_{\mu_g}^{\mu_M}$.

	$\mathcal{F}_{\mu_f}$	$\mathcal{D}_{\mu_g}^{\mu_M}$
W.c. perf.	$\max\left\{p_1\left(\mu_f, h\right), p_2(h)\right\}$	$\max\left\{{\alpha^*}^2 p_1\left(\mu_g, {\alpha^*}^2 h\right), p_2(h)\right\}$
W.c. fun.	$\begin{cases}\ell_{\mu_f,h}(x)\\ q(x)\end{cases}$	$\begin{cases}{\alpha^*}^2 \ell_{\mu_g, {M^*}^2 h}(x)\\ q(x)\end{cases}$

$$\begin{aligned} p_1(\mu,h) &\triangleq \ell_{\mu,h}(x_N) - \ell^* = \frac{1}{2}\frac{\mu}{\mu - 1 + (1-\mu h)^{-2N}}, \\ p_2(h) &\triangleq q(x_N) - q^* \quad = \frac{1}{2}(1-h)^{2N}. \end{aligned} \tag{4.12}$$

There is an interesting difference of performance between the general and composed cases. Let two convex functions f_1 and $f_2 = g \circ M$ with $\mu_M = 0$. When the worst-case performance w is given by p_1 for f_2 and that $h \geq h_0$, then the performance is $\frac{1}{2}\frac{\mu_g \frac{h_0}{h}}{\mu_g - 1 + (1-\mu_g h_0)^{-2N}} \approx \frac{1}{2}\frac{1}{2Nh+1}e^{-\sqrt{\mu_g}}$ whereas in p_1 for f_1 it is $\frac{1}{2}\frac{1}{2Nh+1}$. Therefore, we see a gain of around a factor $e^{-\sqrt{\mu_g}}$ between the performance of the gradient method on the convex functions f_1 and f_2 in this range of values of the step size h.

Optimal step sizes. Understanding the exact worst-case performance of (GM) on $\mathcal{D}_{\mu_g}^{\mu_M}$ allows us to select the step size that optimizes this worst-case performance. Such an optimal design of (GM) is possible thanks to our extension of PEP. We characterize these optimal step sizes $h \in [0,2]$ minimizing $w(\mathcal{D}_{\mu_g}^{\mu_M}; h)$ from Conjecture 4.2. Optimal steps $h^*(\mu_f)$ for $\mathcal{F}_{\mu_f}$ (see [63]) and $h^*(\mu_g, \mu_M)$ for $\mathcal{D}_{\mu_g}^{\mu_M}$ satisfy

$$h^*(\mu_f) \text{ is the solution of equation } \frac{\mu_f}{\mu_f - 1 + (1-\mu_f h)^{-2N}} = (1-h)^{2N}, \tag{4.13}$$

$$h^*(\mu_g, \mu_M) \text{ is the solution of equation } \frac{\tilde{\mu}(h)}{\mu_g - 1 + (1-h\tilde{\mu}(h))^{-2N}} = (1-h)^{2N}, \tag{4.14}$$

where $\tilde{\mu}(h) = \mu_g \mathrm{proj}_{[\mu_M^2,1]}\left(\frac{h_0}{h}\right)$. Note that both $h^*(\mu_f)$ and $h^*(\mu_g,\mu_M)$ can be easily computed numerically and that they depend on the number of iterations N.

Therefore, when facing a 1-smooth μ_f-strongly convex function F, if we do not know anything else about the function, then we should use $h^*(\mu_f)$. However, if we know that the function F can be written as $F = g \circ \mathcal{M}$, where g is 1-smooth μ_g-strongly convex with $\mu_M \leq ||M|| \leq 1$, then it is preferable to use $h^*(\mu_g, \mu_M)$.

Performance on quadratic functions. The class of L-smooth, μ-strongly convex, not necessarily homogeneous, quadratic functions is easily seen to be equal to $\mathcal{D}_{1,1}^{\sqrt{\mu},\sqrt{L}}$. Moreover, since the gradient method is invariant to translations, this class shares the same worst-case performance as the class of homogeneous quadratic functions $\mathcal{Q}_{\mu,L}$. Therefore, according to Conjecture 4.2, the worst-case performance of the gradient method over the class of quadratic functions is

$$w(\mathcal{Q}_{\mu,L}; h) = w\left(\mathcal{D}_{1,1}^{\sqrt{\mu},\sqrt{L}}; h\right) = \frac{LR^2}{2}\max\left\{q(1-qh)^{2N}, (1-h)^{2N}\right\}, \tag{4.15}$$

where $q = \mathrm{proj}_{[\frac{\mu}{L},1]}(\frac{1}{h(2N+1)})$. Moreover, it is known that the worst-case over the class of quadratic functions must be univariate, a consequence of the polynomial-based analysis mentioned in the introduction (below the first motivating example). Therefore, Conjecture 4.1 holds in the case $L_g = \mu_g$.

4.2. Chambolle–Pock method. We show how to tackle with PEP the analysis of a more sophisticated algorithm, namely, the Chambolle–Pock algorithm [10].

The Chambolle–Pock algorithm solves problems of the form

$$\min_x f(x) + g(Mx), \tag{4.16}$$

where f and g are both convex and proximable and $||M|| \leq L_M$, by computing the following iterations with parameters $\tau > 0$ and $\sigma > 0$,

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

where g^* is the convex conjugate function of g.

4.2.1. Convergence results from the literature. Existing results from the literature sometimes require very specific assumptions, which makes them difficult to exploit and compare. One of our objectives is to show that the PEP framework allows us to obtain guarantees potentially for any set of assumptions, including some that have never been analyzed so far.

For example, the convergence rate stated in the original paper describing the method [10], reproduced below, requires the existence of sets B_1 and B_2 "large enough." We note $\mathcal{L}(x,u) = u^T M x + f(x) - g^*(u)$, the Lagrangian of problem (4.16), and $\bar{x}_N = \frac{1}{N}\sum_{i=1}^N x_i$ and $\bar{u}_N = \frac{1}{N}\sum_{i=1}^N u_i$, the averages of the iterates produced by (CP) starting from x_0 and u_0.

Theorem 4.5 ([10, Theorem 1]). *Let f and g convex, $||M|| \leq L_M$, and B_1 and B_2 large enough to contain all the iterations x_i and u_i, respectively, of* (CP). *If $\tau\sigma L_M^2 < 1$, then after $N \geq 1$ iterations of the Chambolle–Pock algorithm* (CP) *started from x_0 and u_0 we have, for any x and u, that*

$$\mathcal{G}_{B_1 \times B_2}(\bar{x}_N, \bar{u}_N) \leq \frac{D(B_1, B_2)}{N}, \tag{4.17}$$

where performance criteria $\mathcal{G}_{B_1 \times B_2}$ and distance D are defined as follows:

$$\begin{aligned} \mathcal{G}_{B_1 \times B_2}(x,u) &= \max_{u' \in B_2} \mathcal{L}(x,u') - \min_{x' \in B_1} \mathcal{L}(x',u), \\ D(B_1,B_2) &= \sup_{(x,u) \in B_1 \times B_2} \frac{||x - x_0||^2}{2\tau} + \frac{||u - u_0||^2}{2\sigma}. \end{aligned} \tag{4.18}$$

The following result, from the same authors, solves this issue and bounds the performance with a quantity that depends only on the initial iterates x_0 and u_0, but involves some evaluation of the linear operator M of the actual instance of the problem (Remark 2 in [11] adapts the result to remove the dependency in M at the cost of an additional factor).

Theorem 4.6 ([11, Theorem 1]). *Let f and g be convex and $||M|| \leq L_M$. If $\tau\sigma L_M^2 \leq 1$, then after $N \geq 1$ iterations of the Chambolle–Pock algorithm* (CP) *started from x_0 and u_0 we have, for any x and u, that the primal-dual gap satisfies*

$$\mathcal{L}(\bar{x}_N, u) - \mathcal{L}(x, \bar{u}_N) \leq \frac{1}{2N}\left(\frac{\|x - x_0\|^2}{\tau} + \frac{\|u - u_0\|^2}{\sigma} - 2(u - u_0)^T M(x - x_0)\right). \tag{4.19}$$

Note that Theorem 2 of [67] proves a similar result for the primal-dual three-operator splitting method which reduces to the Chambolle-Pock method when the first operator is zero.

Finally, the following result, inspired by a lecture of Professor Beck (slide 29 of part 3 of [6]; see Appendix E for a proof), relies on the previous theorem to bound the primal value accuracy instead of the primal-dual gap. However the proposed performance guarantee involves a point $\tilde{u}_N$ which depends on the average iterate $\bar{x}_N$, and which cannot be easily bounded a priori.

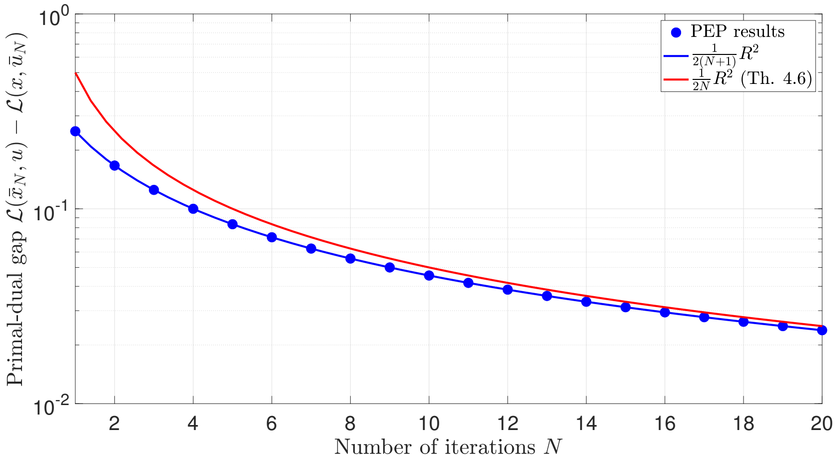

FIG. 2. *Worst-case performance obtained by our extension of PEP for N iterations of the Chambolle–Pock algorithm* (CP) *with any step size parameters σ and τ satisfying $\tau\sigma L_M^2 \leq 1$ on the problem* $\min_x F(x)$, *where $F = f + g \circ M$, f and g are convex proximable, and M is such that $||M|| \leq 1$. The performance criterion is the primal-dual gap $\mathcal{L}(\bar{x}_N, u) - \mathcal{L}(x, \bar{u}_N)$ and the initial distance is $R^2 = \frac{\|x-x_0\|^2}{\tau} + \frac{\|u-u_0\|^2}{\sigma} - 2(u-u_0)^T M(x-x_0)$. PEP results (blue dots) are compared to bound* (4.19) *of Theorem* 4.6 *(red line). Note: color appears only in the online article.*

THEOREM 4.7. *Let f and g be convex and $||M|| \leq L_M$. If $\tau\sigma L_M^2 \leq 1$, then after $N \geq 1$ iterations of the Chambolle–Pock algorithm* (CP) *started from x_0 and u_0 we have, for any x and u, that the primal-dual gap satisfies*

(4.20)
$$F(\bar{x}_N) - F(x^*) \leq \frac{1}{2N}\left(\frac{||x^* - x_0||^2}{\tau} + \frac{||\tilde{u}_N - u_0||^2}{\sigma} - 2(\tilde{u}_N - u_0)^T M(x^* - x_0)\right),$$

where $F(x) = f(x) + g(Mx)$ and $\tilde{u}_N \in \partial g(M\bar{x}_N)$.

4.2.2. Convergence results obtained with the new PEP framework. Theorems 4.6 and 4.7 can be analyzed using our extension of the PEP framework. First, to study the setting of Theorem 4.6, we impose that the initial distance $R^2 = \frac{\|x-x_0\|^2}{\tau} + \frac{\|u-u_0\|^2}{\sigma} - 2(u-u_0)^T M(x-x_0)$ in the right-hand side of (4.19) is bounded by one, and compute the worst-case for the left-hand side (i.e., its maximum), giving us the best possible value of the leading factor. We observe from the tight numerical results provided by PEP that the exact rate is strictly better than the existing bound $\frac{R^2}{2N}$, as shown in Figure 2. Note that these numerical results were identical for all values of τ and σ such that $\tau\sigma L_M^2 \leq 1$. We also observe numerically that the method no longer converges (i.e., the gap is no longer decreasing with N) as soon as $\tau\sigma L_M^2 > 1$.

Moreover, we were able to identify a simple analytical expression matching the tight numerical bound returned by PEP: it corresponds to $\frac{R^2}{2(N+1)}$, i.e., the same bound as (4.19) but with $N+1$ instead of N (shown as a blue line on Figure 2). Based on the hint, provided by PEP, that this better bound actually holds, we were able to adapt the proof of Theorem 4.6 in [11] to obtain the following theorem, that gives the exact convergence rate.

THEOREM 4.8. *Let f and g be convex and $||M|| \leq L_M$. If $\tau\sigma \leq \frac{1}{L_M^2}$, then after $N \geq 1$ iterations of the Chambolle–Pock algorithm* (CP) *started from x_0 and u_0 we have, for any x and u, that the primal-dual gap satisfies*

(4.21)
$$\mathcal{L}(\bar{x}_N, u) - \mathcal{L}(x, \bar{u}_N) \leq \frac{1}{2(N+1)} \left(\frac{\|x - x_0\|^2}{\tau} + \frac{\|u - u_0\|^2}{\sigma} - 2(u - u_0)^T M(x - x_0) \right).$$

Proof. Define $||z||_M^2 = \frac{||z_x||^2}{\tau} + \frac{||z_u||^2}{\sigma} - 2z_u^T M z_x$ with $z = (z_x, z_u)$, so that the initial distance becomes $R^2 = \|z - z_0\|_M^2$ with $z = (x, u)$ and $z_0 = (x_0, u_0)$. Letting now $z_n = (x_n, u_n)$, we sum the following inequality (proved as (15) in [11]),

$$\mathcal{L}(x_{n+1}, u) - \mathcal{L}(x, u_{n+1}) \leq \|z - z_n\|_M^2 - \|z - z_{n+1}\|_M^2 - \|z_{n+1} - z_n\|_M^2$$

from $n = 0$ to $n = N - 1$, without removing the negative terms, yielding

$$\begin{aligned}\sum_{n=0}^{N-1} \mathcal{L}(x_{n+1}, u) - \mathcal{L}(x, u_{n+1}) &\leq \frac{||z_0 - z||_M^2}{2} - \frac{||z - z_N||_M^2}{2} - \sum_{n=0}^{N-1} \frac{||z_{n+1} - z_n||_M^2}{2} \\ &\leq \frac{||z_0 - z||_M^2}{2} - \frac{1}{N+1} \frac{||z_0 - z||_M^2}{2},\end{aligned}$$

where to prove the second inequality we use the following consequence of the convexity of the squared norm, with $s_i = z_i$ for all $0 \leq i \leq N$ and $s_{N+1} = z$:

$$\tfrac{1}{N+1} \|s_0 - s_{N+1}\|^2 = (N+1) \|\tfrac{1}{N+1} \sum_{n=0}^{N} (s_{n+1} - s_n)\|^2 \leq \sum_{n=0}^{N} \|s_{n+1} - s_n\|^2.$$

Finally, by convexity of $\mathcal{L}(\cdot, u) - \mathcal{L}(x, \cdot)$, we have

$$N(\mathcal{L}(\bar{x}_N, u) - \mathcal{L}(x, \bar{u}_N)) \leq \sum_{n=0}^{N-1} \mathcal{L}(x_{n+1}, u) - \mathcal{L}(x, u_{n+1}) \leq \left(1 - \frac{1}{N+1}\right) \frac{||z_0 - z||_M^2}{2}$$

leading finally to $\mathcal{L}(\bar{x}_N, u) - \mathcal{L}(x, \bar{u}_N) \leq \frac{1}{N+1} \frac{||z_0 - z||_M^2}{2}$. ∎

Instead of bounding the quantity $R^2 = \frac{\|x - x_0\|^2}{\tau} + \frac{\|u - u_0\|^2}{\sigma} - 2(u - u_0)^T M(x - x_0)$, it is also possible with PEP to compute rates that depend on the simpler squared distance

$$R_0^2 = \|x - x_0\|^2 + \|u - u_0\|^2. \tag{4.22}$$

In this case, our numerical results indicate that the primal-dual gap depends on the step sizes τ and σ. Figure 3 shows the worst-case performance returned by PEP with initial condition $R_0^2 \leq 1$ for different values of $\tau = \sigma$. When $\tau = \sigma = 1$, the numerical results exactly match the curve $\frac{1}{N+1} R_0^2$ for all tested values of N, which can be compared to the rate $\frac{1}{N} R_0^2$ proved in [11]. We could also identify the rate for all tested values of $\sigma = \tau \leq 1$ when $N = 1$, for which the numerical results match $\frac{\tau+1}{4\tau} R_0^2$. While we were not able to identify the exact rate for general N, τ, and σ, the expression $\frac{\tau+1}{2\tau(N+1)} R_0^2$ appears to be a valid upper bound for all N when $\tau = \sigma$ (and becomes exact when $\tau = \sigma = 1$ or $N = 1$ as mentioned above) according to our numerical observations.

4.2.3. Convergence results on the class of Lipschitz convex functions. The PEP framework allows us to identify performance guarantees in the setting of our choice, i.e., we can study any performance criterion under any initial condition.

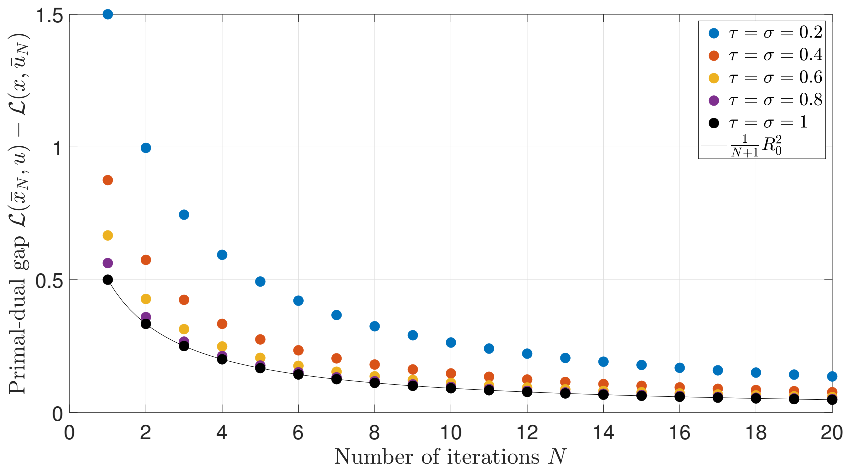

FIG. 3. *Worst-case performance obtained by our extension of PEP for N iterations of the Chambolle–Pock algorithm* (CP) *with different step sizes $\tau = \sigma$ on the problem $\min_x F(x)$, where $F = f + g \circ M$, f and g are convex proximable, and M is such that $||M|| \leq 1$. The performance criterion is the primal-dual gap $\mathcal{L}(\bar{x}_N, u) - \mathcal{L}(x, \bar{u}_N)$ and the initial distance is $R_0^2 = \|x - x_0\|^2 + \|u - u_0\|^2 \leq 1$. Note: color appears only in the online article.*

Theorems 4.5 and 4.6 characterized the primal-dual gap without any boundedness assumptions on the functions f and g. To conclude this section, we propose an alternative analysis, where we consider the primal accuracy, i.e., the value of the (primal) objective function at some iterate. For the corresponding rates to be finite, we must consider a bounded class of functions. We choose the class of Lipschitz convex functions, i.e., convex functions with bounded subgradient. Note that we could have used any other type of bounded classes, e.g., L-smooth functions.

We use $||x_0 - x^*||^2 \leq R_x^2$ and $\|u_0 - u^*\|^2 \leq R_u^2$ as a pair of initial conditions. We have to fix the values of all parameters τ, σ, N, L_M, R_x, and R_u, and we consider not necessarily symmetric matrices M. To show the flexibility of our approach, in addition to computing the rate for the (standard) average iterate, we also perform the analysis for the last iterate, for the best iterate, and for two other averages of iterates, namely, the average of the last $\frac{N}{2}$ iterations, and an average using linearly increasing weights (weight i for iteration i before normalization).

Figure 4 displays the worst-case performance in the above five cases as computed by PEP when minimizing $F = f + g \circ M$ with (CP) for a number of iterations ranging from $N = 1$ to $N = 50$. We bounded the primal and dual initial distances $||x_0 - x^*||^2 \leq R_x^2 = 1$ and $||u_0 - u^*||^2 \leq R_u^2 = 1$ and fixed $\tau = \sigma = 1$. The primal accuracy of the average iterate (blue dots) seems to roughly follow the $\frac{5}{N}$ curve, whereas the last (red squares) and best (green dots) iterates appear closer to the $\frac{1}{\sqrt{N}}$ curve in this example. Moreover, returning the average of the last $\frac{N}{2}$ iterations (magenta dots) or the proposed weighted sum of the iterates (black dots) leads to even better performances of the method.

Numerical guarantees such as those depicted in Figure 4 can also be easily obtained for other performance criteria (e.g., primal-dual gap, dual value accuracy), initial distance conditions, and function classes (e.g., symmetric linear operator with lower bounded eigenvalues). These could be of great help in the analysis and exploration of the algorithm performance and the identification of interesting phenomena. Moreover, any bound obtained in such a manner will be exact, i.e., unimprovable

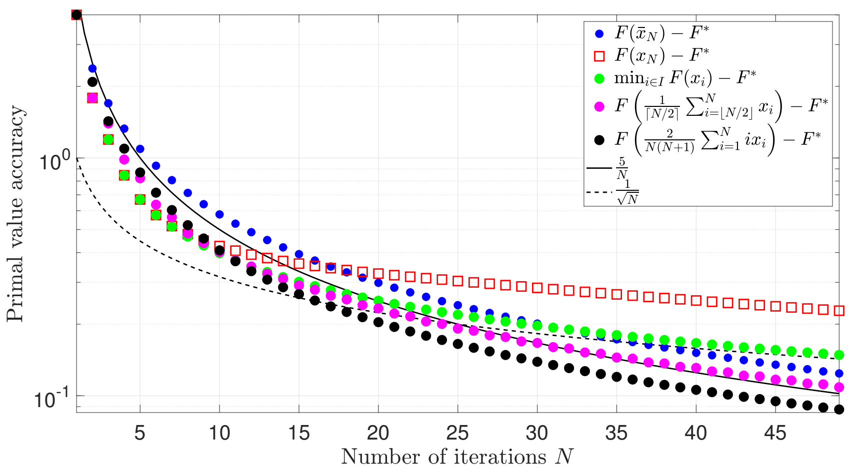

FIG. 4. *Worst-case performance obtained by our extension of PEP for N iterations of the Chambolle–Pock algorithm* (CP) *with step size parameters $\tau = \sigma = 1$ on the problem $\min_x F(x)$, where $F = f + g \circ M$, f and g are* 1*-Lipschitz convex proximable, and M is such that $||M|| \leq 1$. The performance criterion is the objective function accuracy of the average (blue dots), last (red squares), best (green dots), last $\frac{N}{2}$ (magenta dots), and weighted sum (black dots) of iterates. Curves $\frac{5}{N}$ (solid black line) and $\frac{1}{\sqrt{N}}$ (solid black dashed line) are also represented for comparison purposes. Note: color appears only in the online article.*

over the considered function class. By contrast, analytical results typically found in the literature may be subject to nontrivial modification or even have to be redeveloped when changing the framework of evaluation. For example, the convergence of a weighted sum of the iterates as done in Figure 4 would in all likelihood be quite difficult to analyze with standard techniques.

5. Conclusion. Our main contribution is twofold. First, we obtained interpolation conditions for three classes of linear operators corresponding to general, symmetric, and skew-symmetric matrices with bounds on the eigenvalues or singular value spectrum. With these, the powerful and well-developed framework of PEP can be extended to analyze first-order optimization methods applied to problems involving linear operators.

Our extension allowed us to analyze any fixed-step first-order method with any of the standard performance criteria. As a first example, we analyzed the worst-case behavior of the gradient method applied to the problem $\min_x g(Mx)$, where g is smooth and convex. We computed exact performance guarantees and conjectured a closed-form expression for the convergence rate. We also analyzed the more sophisticated Chambolle-Pock algorithm, tightened the existing primal-dual gap convergence rate, and obtained new exact, numerical convergence guarantees when the objective components are convex and Lipschitz.

Throughout this work, we have used the PEP framework to obtain different types of performance guarantees, ranging from purely analytical to purely numerical, depending on the intrinsic complexity of the desired result and our abilities. More precisely, we encountered the following, distinct situations:

1. Design using PEP-identified coefficients of a new, independently verifiable mathematical proof of a performance guarantee (see the proof for one step of the gradient method in section 4.1.2).

2. Identification of the analytical expression of an exact convergence rate, later proved using an improvement of the classic argument (see Theorem 4.8 on the primal-dual gap of (CP)).
3. Identification of new analytical expression for some exact performance guarantees, from which a theoretical explanation is conjectured (see Conjecture 4.2 on the gradient method).
4. Identification of a new analytical expression for a performance guarantee (see Figure 3 for the case $\sigma=\tau=1$ on the primal-dual gap of (CP) with another initial distance).
5. Purely numerical performance guarantee (see Figure 4 for (CP) on Lipschitz convex functions).

For future research, we hope to exploit our new approach to improve our understanding of the large variety of methods tackling problems involving linear operators and compare them in terms of worst-case performance.

Appendix A. Properties of matrices. We review some results of linear algebra used in our proofs. We use I_d and $0_{m,n}$ for identity and zero matrices (dimension may be omitted) and $||M||=\sigma_{\max}(M)$ for the norm of M.

PROPOSITION A.1 ([9]; [31, Theorem 4.3]). *Let $G=\left(\begin{smallmatrix} A & B \\ B^T & C\end{smallmatrix}\right)$ be symmetric. We have the three following equivalences:*

$$G\succeq 0 \Leftrightarrow \begin{cases} C & \succeq 0, \\ A-BC^\dagger B^T & \succeq 0, \\ (I-CC^\dagger)B^T & =0, \end{cases} \Leftrightarrow \begin{cases} A & \succeq 0, \\ C-B^TA^\dagger B & \succeq 0, \\ (I-AA^\dagger)B & =0. \end{cases} \tag{A.1}$$

PROPOSITION A.2. *Let a symmetric matrix $A\succeq 0$. We have $AA^\dagger = A^{\frac{1}{2}}(A^\dagger)^{\frac{1}{2}}$.*

Proof. We have $AA^\dagger = A^{\frac{1}{2}}A^{\frac{1}{2}}(A^{\frac{1}{2}})^\dagger(A^{\frac{1}{2}})^\dagger = A^{\frac{1}{2}}(A^{\frac{1}{2}})^\dagger A^{\frac{1}{2}}(A^{\frac{1}{2}})^\dagger = A^{\frac{1}{2}}(A^{\frac{1}{2}})^\dagger$ by definitions of pseudoinverse and square root matrix. ∎

PROPOSITION A.3. *Let two matrices be C and X. We have $XC=0 \Leftrightarrow XCC^T=0$.*

Proof. If $XCC^T=0$, then $XCC^T(C^\dagger)^T = XCC^\dagger C = XC = 0$ and if $XC=0$, then $XCC^T=0$. ∎

PROPOSITION A.4. *Let two symmetric matrices be $A\succeq 0$ and $C\succeq 0$. We have*

$$\exists \alpha>0 : C\preceq \alpha A \Leftrightarrow AA^\dagger C = C. \tag{A.2}$$

Proof. By application of Proposition A.1, we have $\exists\alpha>0 : \begin{cases}\alpha I\succ 0, \\ C\preceq\alpha A,\end{cases}$

$$\Leftrightarrow \exists\alpha>0 : \begin{pmatrix} A & C^{\frac{1}{2}} \\ C^{\frac{1}{2}} & \alpha I\end{pmatrix}\succeq 0 \Leftrightarrow \exists\alpha>0 : \begin{cases} A\succeq 0, \\ \alpha I\succeq C^{\frac{1}{2}}A^\dagger C^{\frac{1}{2}}, \\ (I-AA^\dagger)C^{\frac{1}{2}}=0, \end{cases} \Leftrightarrow (I-AA^\dagger)C=0,$$

since $A \succeq 0$, $(I - AA^\dagger)C^{\frac{1}{2}} = 0 \Leftrightarrow (I - AA^\dagger)C = 0$ by Proposition A.3, and that we can always find an α (sufficiently large) such that $\alpha I \succeq C^{\frac{1}{2}} A^\dagger C^{\frac{1}{2}}$. ∎

Proposition A.5. *Let two symmetric matrices be $A \succeq 0$ and $C \succeq 0$. We have*

$$\exists L_1 > 0 : C \preceq L_1^2 A \Leftrightarrow \exists L_2 > 0 : C^2 \preceq L_2^2 A. \tag{A.3}$$

Proof. By Propositions A.3 and A.4, we have $\exists L_1 > 0 : C \preceq L_1^2 A \Leftrightarrow AA^\dagger C = C \Leftrightarrow AA^\dagger C^2 = C^2 \Leftrightarrow \exists L_2 > 0 : C^2 \preceq L_2^2 A$. ∎

The following result allows us to extend a block matrix without increasing its maximal singular value. It will be used to prove Lemma 3.4 in Appendix B.

Theorem A.6 ([19, Theorems 1.1 and 1.2]). *Let M_i be conformable matrices. $\exists W$ s.t. $\left\|\begin{pmatrix} M_1 & M_2 \\ M_3 & W \end{pmatrix}\right\| \leq L$ if and only if $\|\begin{pmatrix} M_1 & M_2 \end{pmatrix}\| \leq L$ and $\left\|\begin{pmatrix} M_1 \\ M_3 \end{pmatrix}\right\| \leq L$. Moreover, if M_1 is symmetric (resp., skew-symmetric) and $M_2 = M_3^T$ (resp., $M_2 = -M_3^T$), then there is a W symmetric (resp., skew-symmetric).*

Note that extending a matrix while maintaining its largest singular value under a given bound L is not trivial in general and $W = 0$ does not always work. For example, $\|\begin{pmatrix} 1 & 1 \end{pmatrix}\| = \sqrt{2}$, $\left\|\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\| = \sqrt{2}$, $\left\|\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\right\| = \frac{1+\sqrt{5}}{2} \approx 1.618 > \sqrt{2}$ and $\left\|\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\right\| = \sqrt{2}$, therefore, $W = -1$ is a solution and $W = 0$ is not.

The following result states that vector sets leading to the same Gram matrix are equal up to a rotation when the vectors of the two sets have the same dimension. It will be used to prove Lemma 3.5 in Appendix C.

Theorem A.7 ([43, Theorem 7.3.11]). *A and $B \in \mathbb{R}^{d \times N}$ build the same Gram matrix, i.e., $A^T A = B^T B$, if and only if*

$$\exists V \in \mathbb{R}^{d \times d} \textit{ unitary} : B = VA. \tag{A.4}$$

Appendix B. Proof of Lemma 3.4.

Proof. Let $G = \begin{pmatrix} A_1 & B_1 \\ B_1^T & C_1 \end{pmatrix}$ and $H = \begin{pmatrix} A_2 & B_2 \\ B_2^T & C_2 \end{pmatrix}$ be symmetric and positive semidefinite matrices satisfying

$$\begin{cases} B_1 = B_2, \\ A_2 \preceq L^2 A_1, \\ C_1 \preceq L^2 C_2. \end{cases} \tag{B.1}$$

In the remainder of the proof, we equivalently use B_1 or B_2. We note $S_1 = C_1 - B_1^T {A_1}^\dagger B_1$ and $S_2 = A_2 - B_2 {C_2}^\dagger B_2^T$, where $A_1, C_1, A_2, C_2, S_1, S_2 \succeq 0$ by Proposition A.1 and $G, H \succeq 0$. Moreover, Propositions A.1 and A.2 with $G, H \succeq 0$ provide

$$\begin{aligned} A_1^{\frac{1}{2}} ({A_1}^\dagger)^{\frac{1}{2}} B_1 = A_1 {A_1}^\dagger B_1 = B_1, \\ B_2 ({C_2}^\dagger)^{\frac{1}{2}} C_2^{\frac{1}{2}} = B_2 {C_2}^\dagger C_2 = B_2, \end{aligned} \tag{B.2}$$

and Propositions A.2 and A.4 with (B.1) and (B.2) provide

$$\begin{aligned} S_2^{\frac{1}{2}} A_1^{\frac{1}{2}} ({A_1}^\dagger)^{\frac{1}{2}} = S_2^{\frac{1}{2}}, \\ S_1^{\frac{1}{2}} C_2^{\frac{1}{2}} ({C_2}^\dagger)^{\frac{1}{2}} = S_1^{\frac{1}{2}}. \end{aligned} \tag{B.3}$$

First, we show that

$$(X_R, Y_R, U_R, V_R) = \left(\begin{pmatrix} A_1^{\frac{1}{2}} \\ 0_{N_1,N_1} \end{pmatrix}, \begin{pmatrix} ({C_2}^{\dagger})^{\frac{1}{2}} B_2^T \\ S_2^{\frac{1}{2}} \end{pmatrix}, \begin{pmatrix} C_2^{\frac{1}{2}} \\ 0_{N_1,N_2} \end{pmatrix}, \begin{pmatrix} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 \\ S_1^{\frac{1}{2}} \end{pmatrix} \right) \tag{B.4}$$

is a factorization of G and H. Indeed, we have

$$X_R^T X_R = \begin{pmatrix} A_1^{\frac{1}{2}} & 0_{N_1,N_2} \end{pmatrix} \begin{pmatrix} A_1^{\frac{1}{2}} \\ 0_{N_2,N_1} \end{pmatrix} = A_1, \tag{B.5}$$

$$U_R^T U_R = \begin{pmatrix} C_2^{\frac{1}{2}} & 0_{N_2,N_1} \end{pmatrix} \begin{pmatrix} C_2^{\frac{1}{2}} \\ 0_{N_1,N_2} \end{pmatrix} = C_2, \tag{B.6}$$

$$Y_R^T Y_R = \begin{pmatrix} B_2 ({C_2}^{\dagger})^{\frac{1}{2}} & S_2^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} ({C_2}^{\dagger})^{\frac{1}{2}} B_2^T \\ S_2^{\frac{1}{2}} \end{pmatrix} = B_2 ({C_2}^{\dagger})^{\frac{1}{2}} ({C_2}^{\dagger})^{\frac{1}{2}} B_2^T + S_2 = A_2, \tag{B.7}$$

$$V_R^T V_R = \begin{pmatrix} B_1^T ({A_1}^{\dagger})^{\frac{1}{2}} & S_1^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 \\ S_1^{\frac{1}{2}} \end{pmatrix} = B_1^T ({A_1}^{\dagger})^{\frac{1}{2}} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 + S_1 = C_1, \tag{B.8}$$

$$X_R^T V_R = \begin{pmatrix} A_1^{\frac{1}{2}} & 0_{N_1,N_2} \end{pmatrix} \begin{pmatrix} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 \\ S_1^{\frac{1}{2}} \end{pmatrix} = \overbrace{A_1^{\frac{1}{2}} ({A_1}^{\dagger})^{\frac{1}{2}} B_1}^{=B_1 \text{ by (B.2)}} = B_1, \tag{B.9}$$

$$Y_R^T U_R = \begin{pmatrix} B_2 ({C_2}^{\dagger})^{\frac{1}{2}} & S_2^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} C_2^{\frac{1}{2}} \\ 0_{N_1,N_2} \end{pmatrix} = \overbrace{B_2 ({C_2}^{\dagger})^{\frac{1}{2}} C_2^{\frac{1}{2}}}^{=B_2 \text{ by (B.2)}} = B_2, \tag{B.10}$$

and therefore

$$\begin{aligned} \begin{pmatrix} X_R & V_R \end{pmatrix}^T \begin{pmatrix} X_R & V_R \end{pmatrix} &= \begin{pmatrix} X_R^T X_R & X_R^T V_R \\ V_R^T X_R & V_R^T V_R \end{pmatrix} = \begin{pmatrix} A_1 & B_1 \\ B_1^T & C_1 \end{pmatrix} = G, \\ \begin{pmatrix} Y_R & U_R \end{pmatrix}^T \begin{pmatrix} Y_R & U_R \end{pmatrix} &= \begin{pmatrix} Y_R^T Y_R & Y_R^T U_R \\ U_R^T Y_R & U_R^T U_R \end{pmatrix} = \begin{pmatrix} A_2 & B_2 \\ B_2^T & C_2 \end{pmatrix} = H. \end{aligned} \tag{B.11}$$

Second, we show that (X_R, Y_R, U_R, V_R) is $\mathcal{L}_L$-interpolable by providing a linear operator M_R such that $Y_R = M_R X_R$, $V_R = M_R^T U_R$ and $||M_R|| \leq L$. The expression of M_R is

$$M_R = \begin{pmatrix} ({C_2}^{\dagger})^{\frac{1}{2}} B_1^T ({A_1}^{\dagger})^{\frac{1}{2}} & ({C_2}^{\dagger})^{\frac{1}{2}} S_1^{\frac{1}{2}} \\ S_2^{\frac{1}{2}} ({A_1}^{\dagger})^{\frac{1}{2}} & W \end{pmatrix}, \tag{B.12}$$

where W is a matrix specified later. Indeed, we have

$$\begin{aligned} M_R X_R &= \begin{pmatrix} ({C_2}^{\dagger})^{\frac{1}{2}} B_1^T ({A_1}^{\dagger})^{\frac{1}{2}} & ({C_2}^{\dagger})^{\frac{1}{2}} S_1^{\frac{1}{2}} \\ S_2^{\frac{1}{2}} ({A_1}^{\dagger})^{\frac{1}{2}} & W \end{pmatrix} \begin{pmatrix} A_1^{\frac{1}{2}} \\ 0_{N_2,N_1} \end{pmatrix} \\ &= \begin{pmatrix} ({C_2}^{\dagger})^{\frac{1}{2}} B_1^T ({A_1}^{\dagger})^{\frac{1}{2}} A_1^{\frac{1}{2}} \\ S_2^{\frac{1}{2}} ({A_1}^{\dagger})^{\frac{1}{2}} A_1^{\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} ({C_2}^{\dagger})^{\frac{1}{2}} B_1^T \\ S_2^{\frac{1}{2}} \end{pmatrix} = Y_R, \end{aligned} \tag{B.13}$$

$$
\begin{aligned}
M_R^T U_R &= \begin{pmatrix} ({A_1}^\dagger)^{\frac{1}{2}} B_2^T ({C_2}^\dagger)^{\frac{1}{2}} & ({A_1}^\dagger)^{\frac{1}{2}} S_2^{\frac{1}{2}} \\ S_1^{\frac{1}{2}} ({C_2}^\dagger)^{\frac{1}{2}} & W^T \end{pmatrix} \begin{pmatrix} C_2^{\frac{1}{2}} \\ 0_{N_1,N_2} \end{pmatrix} \\
&= \begin{pmatrix} ({A_1}^\dagger)^{\frac{1}{2}} B_2^T ({C_2}^\dagger)^{\frac{1}{2}} C_2^{\frac{1}{2}} \\ S_1^{\frac{1}{2}} ({C_2}^\dagger)^{\frac{1}{2}} C_2^{\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} ({A_1}^\dagger)^{\frac{1}{2}} B_2^T \\ S_1^{\frac{1}{2}} \end{pmatrix} = V_R.
\end{aligned} \tag{B.14}
$$

It remains to show that the proposed M_R has singular values bounded by L for a suited choice of W. Thanks to Theorem A.6, we must just show that the matrices

$$
M_R^{(\mathrm{up})} = \begin{pmatrix} ({C_2}^\dagger)^{\frac{1}{2}} B_1 ({A_1}^\dagger)^{\frac{1}{2}} & ({C_2}^\dagger)^{\frac{1}{2}} S_1^{\frac{1}{2}} \end{pmatrix} \text{ and } M_R^{(\mathrm{left})} = \begin{pmatrix} ({C_2}^\dagger)^{\frac{1}{2}} B_2^T ({A_1}^\dagger)^{\frac{1}{2}} \\ S_2^{\frac{1}{2}} ({A_1}^\dagger)^{\frac{1}{2}} \end{pmatrix} \tag{B.15}
$$

have singular values lower than L or, equivalently, that the products $M_R^{(\mathrm{up})} {M_R^{(\mathrm{up})}}^T$ and ${M_R^{(\mathrm{left})}}^T M_R^{(\mathrm{left})}$ have eigenvalues lower than L^2, i.e.,

$$
M_R^{(\mathrm{up})} {M_R^{(\mathrm{up})}}^T = ({C_2}^\dagger)^{\frac{1}{2}} B_1 {A_1}^\dagger B_1^T ({C_2}^\dagger)^{\frac{1}{2}} + ({C_2}^\dagger)^{\frac{1}{2}} S_1 ({C_2}^\dagger)^{\frac{1}{2}} = ({C_2}^\dagger)^{\frac{1}{2}} C_1 ({C_2}^\dagger)^{\frac{1}{2}} \preceq L^2 I
$$

and

$$
{M_R^{(\mathrm{left})}}^T M_R^{(\mathrm{left})} = ({A_1}^\dagger)^{\frac{1}{2}} B_2 {C_2}^\dagger B_2^T ({A_1}^\dagger)^{\frac{1}{2}} + ({A_1}^\dagger)^{\frac{1}{2}} S_2 ({A_1}^\dagger)^{\frac{1}{2}} = ({A_1}^\dagger)^{\frac{1}{2}} A_2 ({A_1}^\dagger)^{\frac{1}{2}} \preceq L^2 I
$$

which are both true since $C_1 \preceq L^2 C_2$ implies that

$$
({C_2}^\dagger)^{\frac{1}{2}} C_1 ({C_2}^\dagger)^{\frac{1}{2}} \preceq L^2 ({C_2}^\dagger)^{\frac{1}{2}} C_2^{\frac{1}{2}} C_2^{\frac{1}{2}} ({C_2}^\dagger)^{\frac{1}{2}} = L^2 C_2 {C_2}^\dagger C_2 {C_2}^\dagger = L^2 C_2 {C_2}^\dagger \preceq L^2 I
$$

by definition of the pseudoinverse and Proposition A.2 (the same result for $A_2 \preceq L^2 A_1$).

Finally, we observe on expression (B.12) of M_R that, if $A_1 = C_2$ and $A_2 = C_1$, and $B = B^T$ (resp., $B = -B^T$), then thanks to Theorem A.6, we can choose W symmetric (resp., skew-symmetric) such that M_R is symmetric (resp., skew-symmetric). Moreover, if $A_1 = C_2$, then $U_R = X_R$ and $V_R = Y_R$ (resp., $V_R = -Y_R$). Note that in the skew-symmetric case, we have to add a negative sign on one of the two off-diagonal blocks of M_R in (B.12). ∎

Appendix C. Proof of Lemma 3.5.

Proof. Adding zeros to $(\,X\ V\,)$ or $(\,X_R\ V_R\,)$ such that they have the same number of rows, i.e., $d_n = \max\{n, n_R\}$, will allow us to use Theorem A.7. We use $E_{n,d_n} = \left(\begin{smallmatrix} I_n \\ 0_{(d_n-n),n} \end{smallmatrix}\right)$ and $E_{n_R,d_n} = \left(\begin{smallmatrix} I_{n_R} \\ 0_{(d_n-n_R),n_R} \end{smallmatrix}\right)$ to add the zeros and now work with $E_{n,d_n}(\,X\ V\,)$ and $E_{n_R,d_n}(\,X_R\ V_R\,)$. Moreover, adding the zeros preserves the Gram matrices, indeed, if $(\,X\ V\,)^T(\,X\ V\,) = (\,X_R\ V_R\,)^T(\,X_R\ V_R\,)$, then

$$
\begin{pmatrix} X & V \end{pmatrix}^T E_{n,d_n}^T E_{n,d_n} \begin{pmatrix} X & V \end{pmatrix} = \begin{pmatrix} X_R & V_R \end{pmatrix}^T E_{n_R,d_n}^T E_{n_R,d_n} \begin{pmatrix} X_R & V_R \end{pmatrix}. \tag{C.1}
$$

Therefore, Theorem A.7 applies and yields

$$
E_{n_R,d_n} \begin{pmatrix} X_R & V_R \end{pmatrix} = V_G E_{n,d_n} \begin{pmatrix} X & V \end{pmatrix} \tag{C.2}
$$

and with the same reasoning on $(\,Y\ U\,)$, $(\,Y_R\ U_R\,)$, and $d_m = \max\{m, m_R\}$, it yields

$$
E_{m_R,d_m} \begin{pmatrix} Y_R & U_R \end{pmatrix} = V_H E_{m,d_m} \begin{pmatrix} Y & U \end{pmatrix} \tag{C.3}
$$

for some V_G and V_H unitary. Now, since (X_R, Y_R, U_R, V_R) is $\mathcal{L}_L$-interpolable, there exists an M_R such that $Y_R = M_R X_R$, $V_R = M^T U_R$, and $||M_R|| \leq L$. Therefore,

$$Y_R = M_R X_R \Rightarrow \overbrace{E_{m_R,d_m} Y_R}^{V_H E_{m,d_m} Y} = E_{m_R,d_m} M_R E^T_{n_R,d_n} \overbrace{E_{n_R,d_n} X_R}^{V_G E_{n,d_n} X} \tag{C.4}$$

$$\Rightarrow Y = \overbrace{E^T_{m,d_m} V_H^T E_{m_R,d_m} M_R E^T_{n_R,d_n} V_G E_{n,d_n}}^{M} X, \tag{C.5}$$

$$V_R = M_R^T U_R \Rightarrow \overbrace{E_{n_R,d_n} V_R}^{V_G E_{n,d_n} V} = E_{n_R,d_n} M_R^T E^T_{m_R,d_m} \overbrace{E_{m_R,d_m} U_R}^{V_H E_{m,d_m} X} \tag{C.6}$$

$$\Rightarrow V = \overbrace{E^T_{n,d_n} V_G^T E_{n_R,d_n} M_R^T E^T_{m_R,d_m} V_H E_{m,d_m}}^{M^T} U. \tag{C.7}$$

We have $\|M\| \leq L$ since unitary transformations V_H^T and V_G preserve the maximal singular value and that $E_{i,j}$ can only add zeros singular values.

Finally, when $U = X$, $V = Y$ (resp., $V = -Y$), $U_R = X_R$, and $V_R = Y_R$ (resp., $V_R = -Y_R$), we have $V_G = V_H$. Therefore, M is obtained as a unitary transformation of M_R, in other words, if M_R was symmetric (resp., skew-symmetric), then, M remains symmetric (resp., skew-symmetric). ∎

Appendix D. Proof of the equality in (4.11). The equality in (4.11) holds when $\gamma = \frac{1-\mu_g}{2-\mu_g}$, $\tau = \frac{\mu_g}{2(\mu_g - 1 + (1-\mu_g)^{-2})}$, $x_1 = x_0 - M^T \nabla g(M x_0)$, and

$$\mu_g \leq \frac{1}{6}\left(7 - \frac{7}{\sqrt[3]{44 - 3\sqrt{177}}} - \sqrt[3]{44 - 3\sqrt{177}}\right) \approx 0.17,$$

$$P = \begin{pmatrix} Mx_0 & Mx_1 & Mx^* & \nabla g(Mx_0) & \nabla g(Mx_1) & \nabla g(Mx^*) \end{pmatrix},$$

$$a_{\mu_g} = \begin{pmatrix} a_1 & -(1+\mu_g)a_2 & a_3 & -(1+\mu_g)a_2 & a_2 & \mu_g a_2 \end{pmatrix}^T,$$

$$b_{\mu_g} = \begin{pmatrix} 0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix}^T = \begin{pmatrix} 0 \\ -\sqrt{\frac{\mu_g}{2(1-\mu_g)} + \frac{1}{2-\mu_g} + (1+\mu_g)^2 a_2^2} \\ -b_1 \\ \frac{2(1+\mu_g)^2(2-\mu_g)a_2^2 - (1+\mu_g)}{-2(2-\mu_g)b_1} \\ \frac{-2(1-\mu_g^2)a_2^2 + 1}{-2(1-\mu_g)b_1} \\ \frac{-2\mu_g(1-\mu_g^2)(2-\mu_g)a_2^2 + (1-\mu_g)^2 + \mu_g}{2(1-\mu_g)(2-\mu_g)b_1} \end{pmatrix},$$

$$c_{\mu_g} = \left(0\,0\,0\,\sqrt{\tfrac{1}{2-\mu_g} - (1+\mu_g)^2 a_2^2 - b_2^2}\; - \sqrt{\tfrac{1}{2(1-\mu_g)} - a_2^2 - b_3^2}\; \sqrt{\tfrac{1}{2(1-\mu_g)} - (\mu_g a_2)^2 - b_4^2}\right)^T,$$

where $a_1 = -\sqrt{\frac{\mu_g}{2-\mu_g} + \tau}$, $a_2 = \frac{1}{2(2-\mu_g)a_1}$, $a_3 = \sqrt{\tau - \frac{1}{2-\mu_g} + (1+\mu_g)^2 a_2^2}$. Indeed, given these definitions, elementary algebraic calculations allow us to prove the equality.

Appendix E. Proof of Theorem 4.7. The proof comes from a lecture by Professor Beck [6].

Proof. Since the assumptions of Theorem 4.6 hold, it applies $\forall x, y$, namely,

$$\mathcal{L}(\bar{x}_N, u) - \mathcal{L}(x, \bar{u}_N) \leq \frac{1}{2N}\left(\frac{\|x - x_0\|^2}{\tau} + \frac{\|u - u_0\|^2}{\sigma} - 2(u - u_0)^T M (x - x_0)\right),$$

where $\mathcal{L}(x,u) = x^T M^T u - g^*(u) + f(x)$. Choosing $u = \tilde{u}_N \in \partial g(M\bar{x}_N)$, $x = x^*$, and using the conjugate subgradient theorem and Fenchel's inequality yield

$$\mathcal{L}(\bar{x}_N, \tilde{u}_N) = \underbrace{\bar{x}_N^T M^T \tilde{u}_N - g^*(\tilde{u}_N)}_{=g(M\bar{x}_N) \text{ by (CST)}} + f(\bar{x}_N) = g(M\bar{x}_N) + f(\bar{x}_N) = F(\bar{x}_N),$$
$$\mathcal{L}(x^*, \bar{u}_N) = \underbrace{{x^*}^T M^T \bar{u}_N - g^*(\bar{u}_N)}_{\leq g(Mx^*) \text{ by (FI)}} + f(x^*) \leq g(Mx^*) + f(x^*) = F(x^*),$$

and therefore

$$\begin{aligned} F(\bar{x}_N) - F(x^*) &\leq \mathcal{L}(\bar{x}_N, \tilde{u}_N) - \mathcal{L}(x^*, \bar{u}_N) \\ &\leq \frac{1}{2N}\left(\frac{\|x^* - x_0\|^2}{\tau} + \frac{\|\bar{u}_N - u_0\|^2}{\sigma} - 2(\bar{u}_N - u_0)^T M(x^* - x_0)\right). \quad \square \end{aligned}$$

Acknowledgments. The authors would like to thank Prof. Paul Van Dooren for pointing out the existence of some linear algebra results, Adrien Taylor for very useful exchanges and suggestions, and Anne Rubbens, Moslem Zamani and Teodor Rotaru for fruitful discussions.

REFERENCES

[1] H. Abbaszadehpeivasti, E. de Klerk, and M. Zamani, *Convergence Rate Analysis of Randomized and Cyclic Coordinate Descent for Convex Optimization through Semidefinite Programming*, preprint, arXiv:2212.12384, 2022.

[2] H. Abbaszadehpeivasti, E. de Klerk, and M. Zamani, *Conditions for linear convergence of the gradient method for non-convex optimization*, Optim. Lett., 17 (2023), pp. 1105–1125.

[3] H. Abbaszadehpeivasti, E. de Klerk, and M. Zamani, *On the rate of convergence of the difference-of-convex algorithm (DCA)*, J. Optim. Theory Appl., 202 (2024), pp. 475–496.

[4] A. Aberdam and A. Beck, *An accelerated coordinate gradient descent algorithm for non-separable composite optimization*, J. Optim. Theory Appl., 193 (2022), pp. 219–246.

[5] M. ApS, *The MOSEK Optimization Toolbox for MATLAB Manual*, Version 10.1, 2024, http://docs.mosek.com/latest/toolbox/index.html.

[6] A. Beck, *Proximal Methods Materials*, https://sites.google.com/site/amirbeck314/proximal-methods-materials (2022).

[7] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[8] N. Bousselmi, J. Hendrickx, and F. Glineur, *Performance estimation of first-order methods on quadratic functions*, in Mathematical Theory of Networks and Systems 2022, Springer, Berlin, 2022.

[9] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[10] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40 (2011), pp. 120–145.

[11] A. Chambolle and T. Pock, *On the ergodic convergence rates of a first-order primal–dual algorithm*, Math. Program., 159 (2016), pp. 253–287.

[12] P. Chen, J. Huang, and X. Zhang, *A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions*, Fixed Point Theory Algorithms Sci. Eng., 2016 (2016), 54.

[13] S. Chen and D. Donoho, *Basis pursuit*, in Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers, Vol. 1, IEEE Computer Society, Los Alamitos, CA, 1994, pp. 41–44.

[14] S. Colla and J. M. Hendrickx, *Automated worst-case performance analysis of decentralized gradient descent*, in 2021 60th IEEE Conference on Decision and Control (CDC), IEEE, Piscataway, NJ, 2021, pp. 2627–2633.

[15] S. Colla and J. M. Hendrickx, *Automated performance estimation for decentralized optimization via network size independent problems*, in 2022 IEEE 61st Conference on Decision and Control (CDC), IEEE, Piscataway, NJ, 2022, pp. 5192–5199.

[16] S. Colla and J. M. Hendrickx, *Automatic performance estimation for decentralized optimization*, IEEE Trans. Automat. Control, 68 (2023), pp. 7136–7150.
[17] L. Condat, *A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., 158 (2013), pp. 460–479.
[18] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, *Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists*, SIAM Rev., 65 (2023), pp. 375–435.
[19] C. Davis, W. M. Kahan, and H. F. Weinberger, *Norm-preserving dilations and their applications to optimal error bounds*, SIAM J. Numer. Anal., 19 (1982), pp. 445–469.
[20] E. De Klerk, F. Glineur, and A. B. Taylor, *On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions*, Optim. Lett., 11 (2017), pp. 1185–1199.
[21] E. De Klerk, F. Glineur, and A. B. Taylor, *Worst-case convergence analysis of inexact gradient and Newton methods through semidefinite programming performance estimation*, SIAM J. Optim., 30 (2020), pp. 2053–2082.
[22] R.-A. Dragomir, A. B. Taylor, A. d'Aspremont, and J. Bolte, *Optimal complexity and certification of Bregman first-order methods*, Math. Program., 194 (2022), pp. 41–83.
[23] Y. Drori, S. Sabach, and M. Teboulle, *A simple algorithm for a class of nonsmooth convex–concave saddle-point problems*, Oper. Res. Lett., 43 (2015), pp. 209–214.
[24] Y. Drori and O. Shamir, *Presentation on "performance estimation of stochastic first-order methods*, ICCOPT 2019, Weierstrass Institute, Berlin, 2019.
[25] Y. Drori and A. B. Taylor, *Efficient first-order methods for convex minimization: A constructive approach*, Math. Program., 184 (2020), pp. 183–220.
[26] Y. Drori and M. Teboulle, *Performance of first-order methods for smooth convex minimization: A novel approach*, Math. Program., 145 (2014), pp. 451–482.
[27] Y. Drori and M. Teboulle, *An optimal variant of Kelley's cutting-plane method*, Math. Program., 160 (2016), pp. 321–351.
[28] A. d'Aspremont, D. Scieur, A. Taylor, *Acceleration methods*, Found. Trends Optim., 5 (2021), pp. 1–245.
[29] L. Eldén, *Perturbation theory for the least squares problem with linear equality constraints*, SIAM J. Numer. Anal., 17 (1980), pp. 338–350.
[30] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Comput. Math. Appl., 2 (1976), pp. 17–40.
[31] J. H. Gallier, *Notes on the Schur Complement*, https://repository.upenn.edu/cgi/viewcontent.cgi?article=1637&context=cis_papers (2010).
[32] E. Gorbunov, N. Loizou, and G. Gidel, *Extragradient method: $\mathcal{O}(1/k)$ last-iterate convergence for monotone variational inequalities and connections with cocoercivity*, Proc. Mach. Learn. Res. (PMLR), 151 (2022), pp. 366–402.
[33] E. Gorbunov, A. Taylor, and G. Gidel, *Last-iterate convergence of optimistic gradient method for monotone variational inequalities*, Adv. Neural Inf. Process. Syst., 35 (2022), pp. 21858–21870.
[34] E. Gorbunov, A. Taylor, S. Horváth, and G. Gidel, *Convergence of proximal point and extragradient-based methods beyond monotonicity: The case of negative comonotonicity*, Proc. Mach. Learn. Res. (PMLR), 202 (2023), pp. 11614–11641.
[35] B. Goujaud, A. Dieuleveut, and A. Taylor, *On fundamental proof structures in first-order optimization*, in 2023 62nd IEEE Conference on Decision and Control (CDC), IEEE, Piscataway, NJ, 2023, pp. 3023–3030.
[36] B. Goujaud, C. Moucer, F. Glineur, J. Hendrickx, A. Taylor, and A. Dieuleveut, *PEPit: Computer-assisted worst-case analyses of first-order optimization methods in Python*, Math. Program. Comput., (2024).
[37] B. Goujaud, A. Taylor, and A. Dieuleveut, *Optimal First-Order Methods for Convex Functions with a Quadratic Upper Bound*, preprint, arXiv:2205.15033, 2022.
[38] G. Gu and J. Yang, *Tight sublinear convergence rate of the proximal point algorithm for maximal monotone inclusion problems*, SIAM J. Optim., 30 (2020), pp. 1905–1921.
[39] G. Gu and J. Yang, *Tight ergodic sublinear convergence rate of the relaxed proximal point algorithm for monotone variational inequalities*, J. Optim. Theory Appl., 202 (2024), pp. 373–387.
[40] C. Guille-Escuret, A. Ibrahim, B. Goujaud, and I. Mitliagkas, *Gradient descent is optimal under lower restricted secant inequality and upper error bound*, Adv. Neural Inf. Process. Syst., 35 (2022), pp. 24893–24904.
[41] S. D. Gupta, R. M. Freund, X. A. Sun, and A. Taylor, *Nonlinear conjugate gradient methods: worst-case convergence rates via computer-assisted analyses*, Math. Program., (2024), https://link.springer.com/article/10.1007/s10107-024-02127-7.

[42] S. D. Gupta, B. P. Van Parys, and E. K. Ryu, *Branch-and-bound performance estimation programming: A unified methodology for constructing optimal optimization methods*, Math. Program., 204 (2023), pp. 567–639.

[43] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 2012.

[44] Y. Kamri, J. M. Hendrickx, and F. Glineur, *On the worst-case analysis of cyclic coordinate-wise algorithms on smooth convex functions*, in 2023 European Control Conference (ECC), IEEE, Piscataway, NJ, 2023, pp. 1–8.

[45] D. Kim, *Accelerated proximal point method for maximally monotone operators*, Math. Program., 190 (2021), pp. 57–87.

[46] D. Kim and J. A. Fessler, *Optimized first-order methods for smooth convex minimization*, Math. Program., 159 (2016), pp. 81–107.

[47] D. Kim and J. A. Fessler, *On the convergence analysis of the optimized gradient method*, J. Optim. Theory Appl., 172 (2017), pp. 187–205.

[48] D. Kim and J. A. Fessler, *Another look at the fast iterative shrinkage/thresholding algorithm (FISTA)*, SIAM J. Optim., 28 (2018), pp. 223–250.

[49] D. Kim and J. A. Fessler, *Generalizing the optimized gradient method for smooth convex minimization*, SIAM J. Optim., 28 (2018), pp. 1920–1950.

[50] D. Kim and J. A. Fessler, *Optimizing the efficiency of first-order methods for decreasing the gradient of smooth convex functions*, J. Optim. Theory Appl., 188 (2021), pp. 192–219.

[51] F. Lieder, *On the convergence rate of the Halpern-iteration*, Optim. Lett., 15 (2021), pp. 405–418.

[52] C. Moucer, A. Taylor, and F. Bach, *A systematic approach to Lyapunov analyses of continuous-time models in convex optimization*, SIAM J. Optim., 33 (2023), pp. 1558–1586.

[53] I. Necoara, Y. Nesterov, and F. Glineur, *Linear convergence of first order methods for non-strongly convex optimization*, Math. Program., 175 (2019), pp. 69–107.

[54] T. Rotaru, F. Glineur, and P. Patrinos, *Exact Worst-Case Convergence Rates of Gradient Descent: A Complete Analysis for All Constant Stepsizes over Nonconvex and Convex Functions,* arXiv:2406.17506, 2024.

[55] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, New York, 2005.

[56] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[57] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson, *Operator splitting performance estimation: Tight contraction factors and optimal parameter selection*, SIAM J. Optim., 30 (2020), pp. 2251–2271.

[58] Z. Shi and R. Liu, *A Better Convergence Analysis of the Block Coordinate Descent Method for Large Scale Machine Learning*, preprint, arXiv:1608.04826, 2016.

[59] A. Taylor and F. Bach, *Stochastic first-order methods: Non-asymptotic and computer-aided analyses via potential functions*, Proc. Mach. Learn. Res. (PMLR), 99 (2019), pp. 2934–2992.

[60] A. Taylor and Y. Drori, *An optimal gradient method for smooth strongly convex minimization*, Math. Program., 199 (2022), pp. 557–594.

[61] A. B. Taylor, *Convex Interpolation and Performance Estimation of First-order Methods for Convex Optimization*, Ph.D. thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2017.

[62] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Performance estimation toolbox (pesto): Automated worst-case analysis of first-order optimization methods*, in 2017 IEEE 56th Annual Conference on Decision and Control (CDC), IEEE, Piscataway, NJ, 2017, pp. 1278–1283.

[63] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Smooth strongly convex interpolation and exact worst-case performance of first-order methods*, Math. Program., 161 (2017), pp. 307–345.

[64] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Exact worst-case convergence rates of the proximal gradient method for composite convex minimization*, J. Optim. Theory Appl., 178 (2018), pp. 455–476.

[65] M. Upadhyaya, S. Banert, A. B. Taylor, and P. Giselsson, *Automated tight Lyapunov analysis for first-order methods*, Math. Program., to appear.

[66] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681.

[67] M. Yan, *A new primal–dual algorithm for minimizing the sum of three functions with a linear operator*, J. Sci. Comput., 76 (2018), pp. 1698–1717.

[68] P. Yi, Y. Hong, and F. Liu, *Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems*, Automatica J. IFAC, 74 (2016), pp. 259–269.

[69] M. Zamani, H. Abbaszadehpeivasti, and E. de Klerk, *The exact worst-case convergence rate of the alternating direction method of multipliers*, Math. Program., to appear.