UCLouvain
Institute for Information
and Communication Technologies,
Electronics and Applied Mathematics

Performance analysis of primal-dual and second-order optimization methods

Nizar Bousselmi

Thesis submitted in partial fulfillment
of the requirements for the degree of
Docteur en sciences de l'ingénieur

Dissertation committee:

Prof. Julien M. Hendrickx (UCLouvain, advisor)
Prof. François Glineur (UCLouvain, advisor)
Prof. Geovani Grapiglia (UCLouvain)
Prof. Etienne de Klerk (Tilburg University, The Netherlands)
Dr. Laurent Condat (KAUST, Saudi Arabia)
Dr. Adrien Taylor (INRIA, France)
Prof. Pierre-Antoine Absil (UCLouvain, chair)

Version of August 11, 2025.

Abstract

Optimization theory develops methods to solve optimization problems. One common way to assess the quality of an optimization method is through its worst-case guarantee over a problem class. It certifies that the method will never perform worse than this guarantee on all instances of the problem class. The Performance Estimation Problem (PEP) methodology makes it possible to determine the exact worst-case performance of an optimization method in a computer-assisted way. It formulates the problem of finding the worst-case instance of the problem class for the method in a tractable way and then solves it analytically or numerically. Importantly, the framework relies on the concept of interpolation conditions, that is, an appropriate characterization of the involved class of functions.

In the first part of this thesis, we generalize the PEP methodology to first-order methods involving linear operators. This extension requires an explicit formulation of interpolation conditions for those linear operators. We also extend the framework to analyze the performance of first-order methods on quadratic functions. We then investigate primal-dual methods such as the Chambolle-Pock and Condat-Vũ methods and improve and extend the existing analysis of their convergence.

In the second part, we develop interpolation conditions for classes of univariate Hessian Lipschitz and univariate generalized self-concordant functions. We exploit these conditions together with the PEP framework to analyze the convergence of second-order optimization methods such as Newton's method and its regularized variants. This allows to obtain new tight analytical and numerical guarantees on these methods for univariate functions.

Remerciements

Beaucoup d'événements ont eu lieu durant ces quatre dernières années. La vie avance et les mathématiques et la recherche auront été une importante et nécessaire distraction.

Je souhaite commencer par remercier mes superviseurs François Glineur et Julien Hendrickx pour leur excellent encadrement et accompagnement, pour le partage de leurs connaissances et expertise, et surtout pour leur bienveillance, compréhension et soutien face aux situations difficiles de la recherche et de la vie. Travailler avec eux a été un honneur et un réel plaisir. Merci de m'avoir guidé et aidé à devenir un meilleur chercheur.

Je remercie mon comité d'encadrement Etienne de Klerk, et Geovani Grapiglia pour leurs précieuses visions et suggestions durant ma confirmation qui ont impacté la suite de ma thèse, ainsi que pour tous les autres conseils en cours de route.

Je remercie aussi les membres de mon jury de thèse. J'ai eu la chance de rencontrer Adrien Taylor, Etienne de Klerk et Laurent Condat à différentes occasions durant ma thèse. C'était à chaque fois très instructif de les écouter. Merci de m'avoir accordé du temps pour discuter et échanger des idées sur nos travaux.

Un tout grand merci à tous mes collègues de l'Euler. Merci à ceux avec qui j'ai pu avoir de nombreuses discussions scientifiques, Anne, Sébastien, Yassine, Sofiane, Teodor, Erwan, merci à mon vieux collègue de bureau Olivier, et mes jeunes collègues de bureau Pierre et Nicolas. Merci à tous ceux avec qui j'ai pu passer de chouettes moments et qui contribuent à rendre la vie agréable à l'Euler, Charles, Antoine, Renato, Rémi, Julien, Martina, Philémon, Clémence, Adrien, Guillaume, Dânâ, Simon, Timothée. Merci aussi à François, Etienne, John, Astrid, Marie-Christine, et Pascale qui m'ont beaucoup facilité l'aspect administratif et technique de la thèse.

Durant ma thèse, j'ai eu l'occasion de faire un échange à l'ENS Lyon dans l'équipe de Nelly Pustelnik. Un tout grand merci à elle et à son département pour leur accueil. J'ai beaucoup apprécié travailler et échanger avec Nelly.

Merci à Zhicheng Deng avec qui j'ai pu collaborer.

J'ai aussi eu l'occasion d'encadrer des cours. C'était très intéressant de travailler et enseigner ces cours avec Geovani Grapiglia, Julien Hendrickx, François Wielant, Estelle Massart, et Frédéric Crevecoeur. Merci pour cet apprentissage.

J'aimerais aussi remercier les professeurs que j'ai eus la chance d'écouter durant mes études avant ma thèse et qui ont influencé mon choix d'en faire une, François Glineur, Julien Hendrickx, Raphaël Jungers, Pierre-Antoine Absil, Vincent Legat, Yurii Nesterov, Jean-Charles Delvenne, Laurent Jacques, et Roland Keunings.

Merci à Philippe Lefèvre pour les aménagements qu'il a mis en place pour m'aider.

Je remercie ma famille, mes parents et mon frère ainsi que ma belle-famille pour tous les bons moments qu'on partage ensemble.

Enfin, je remercie ma femme, Apolline. Ta gentillesse et bienveillance améliorent chaque jour un peu plus notre monde. Ta combativité et ta résilience ont été et sont une énorme source d'inspiration pour moi et ton entourage.

Contents

III Conclusion

List of symbols

$\bar{\mathbb{R}}$ — Projectively extended real line (p. 111)

I_n — Identity matrix of dimension $n \times n$ (dimension may be omitted) (p. 26)

$\mathcal{F}$ — Function class (p. 7)

$\mathcal{F}_{\mu,L}$ — Class of L-smooth μ-strongly convex functions (p. 9)

$\mathcal{H}_{M,+}$ — Strongly convex and Hessian Lipschitz functions (p. 128)

$\mathcal{H}_{M,\alpha,+}$ — Convex generalized self-concordant functions (p. 125)

$\mathcal{H}_M$ — Hessian Lipschitz functions (p. 125)

$\int \mathcal{F}$ — Class of univariate functions whose derivative belongs to $\mathcal{F}$ (p. 113)

$\lambda(M)$ — Eigenvalues of M (p. 32)

$\mathcal{L}$ — Class of linear operators (p. 18)

$\mathcal{L}_{\mu,L}$ — Class of linear operators with singular values in $[\mu, L]$ (p. 18)

$\|M\|$ — Spectral norm of M (p. 24)

$M^\dagger$ — Pseudo-inverse of M (p. 22)

$\preceq$ — Löwner order for symmetric matrices (p. 18)

$\mathcal{Q}_{\mu,L}$ — Class of homogeneous L-smooth μ-strongly convex quadratic functions $f(x) = \frac{1}{2}x^T Qx$ (p. 48)

$\mathcal{T}_{M,+}$ — Quasi-self-concordant functions (p. 125)

$\mathbb{R}$ — Set of real numbers (p. 19)

$\mathbb{R}^{m\times n}$ — Set of real matrices of size $m \times n$ (p. 19)

$\mathcal{S}_{M,+}$ Self-concordant functions (p. 125), Self-concordant functions (p. 129)

$\sigma(M)$ — Singular values of M (p. 37)

$\sigma_{\max}(M)$ — Largest singular value of M (p. 18)

$\sigma_{\min}(M)$ — Smallest singular value of M (p. 18)

$\mathcal{S}_{\mu,L}$ — Class of symmetric linear operators with eigenvalues in $[\mu, L]$ (p. 18)

$\mathcal{T}_{\mu,L}$ Class of skew-symmetric linear operators with singular values in $[\mu, L]$ (p. 18)

$0_{m,n}$ — Zero matrices of dimension $m \times n$ (dimension may be omitted) (p. 19)

$M^{\frac{1}{2}}$ — Square root of positive semidefinite matrix M (p. 22)

x^* or x_* — Local or global optimal solution (p. 7)

List of acronyms

PEP	**P**erformance **E**stimation **P**roblem
GM	**G**radient **M**ethod
FGM	**F**ast **G**radient **M**ethod
CP	**C**hambolle-**P**ock method
CV	**C**ondat-**V**ũ method
ADMM	**A**lternating **D**irections **M**ethod of **M**ultipliers
BB	**B**arzilai **B**orwein method
PESTO	**P**erformance **ES**timation **TO**olbox
PEPit	**P**erformance **E**stimation in **P**ython
FSFOM	**F**ixed-**S**tep **F**irst-**O**rder **M**ethod
FBS	**F**orward-**B**ackward **S**plitting
PRS	**P**eaceman-**R**achford **S**plitting
DRS	**D**ouglas-**R**achford **S**plitting
CNM	**C**ubic regularized **N**ewton **M**ethod
GNM	**G**radient regularized **N**ewton **M**ethod
DNM	**D**amped **N**ewton **M**ethod

1 Introduction

"Nothing whatsoever takes place in the universe in which some relation of maximum and minimum does not appear." - Leonhard Euler

ALL the processes in the world can be viewed through optimization. The most efficient running posture of an animal body, the optimal hexagonal structure of beehives, or the natural shape of rivers carved by the flow of water all consist of minimizing a quantity of interest. We always try to minimize our consumption of time, energy, or other resources.

Optimization theory [NW06, BV04, N$^+$18, Bec17] abstracts real-life problems as *optimization problems* (modelling) and then solves them with *optimization methods* to obtain an approximate solution (solving). An optimization problem has the following form

$$\min_{x \in \mathcal{X}} f(x) \tag{1.1}$$

where x is the decision variable, f the *objective function* to minimize, and $\mathcal{X}$ the feasible sets defined by the constraints of the problem. An optimization method is an algorithmic procedure to be followed in order to produce an approximate solution close to the actual minimizer of (1.1).

Remark 1.1. A fundamental principle in optimization is that a clever modelling of the problem can lead to an efficient solving. Therefore, the modelling and solving steps should not be tackled independently. A clever modelling is a formulation of (1.1) where f and $\mathcal{X}$ have desirable properties or specific structure (e.g., convexity, smoothness, linearity), which can be exploited efficiently by solvers. Two examples of extremely successful exploitation of the structure of problems are the simplex method for linear problems [Dan16] and interior-point methods for convex problems [NN94], which can both solve large-scale problems.

Even if the real-life problems mentioned above can be formulated as optimization problems, these are not the most successful applications of optimization theory because they are processes over which we cannot act. On the contrary, there are domains in science and engineering where optimization theory has achieved astonishing results: supply chain optimization [CM07], signal processing [PE10, CLL17] (including medical imaging), control of dynamical systems [HM12, Ber12] (including robotics and neuroscience), machine learning, and artificial intelligence [SNW11, BCN18, Wri23].

Quality of an optimization method and worst-case guarantees Given a problem class, i.e., a class of problems with a specific structure, we would like to design an optimization method to solve any instance of the problem class. A classical and historical measure of the quality of a method is its *convergence* to the optimal solution, that is, the method eventually finds a solution close to the optimal solution of the problem, possibly after a long time. However, convergence may take an immense amount of time, whereas some applications require a rapid solution. Therefore, a now very popular measure of the quality of a method is its *speed* or *rate of convergence*, namely, how fast the method converges to the optimal solution (if it does).

A standard way to characterize the rate of convergence of a method on a problem class is through a worst-case guarantee of the convergence rate. It is an upper bound on the time or number of iterations that the method will take to solve the hardest instance of the class at a given precision. It guarantees that any instances will be solved at least as fast. To obtain a worst-case performance guarantee, we should build a *proof of performance* by suitably combining the (in-)equalities describing the problem or function class and the method. This work has been done by hand or with computer-assisted tools for many problem classes and methods to provide very insightful knowledge of these classes and methods. It allows us to understand which problems are “easy” to solve and which ones are not, or not yet.

Performance Estimation and tight guarantees The *Performance Estimation Problem* (PEP) methodology, introduced in [DT14] and further improved in [THG17c] and many follow-up works (see, e.g., Section 2.3 for a short review), 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 this class of functions. This provided several tight results on the performance of first-order methods.

If we provide the correct description of the function class, PEP returns, in an automated way, the *exact* or *tight* worst-case guarantee that is attained by a function of the class considered. Therefore, PEP allows to obtain exact convergence rates and can guide us to understand the possible sources of inefficiency and to tune a given method, for example, by choosing its hyperparameters to minimize the rate. In addition, PEP allows to compare different existing optimization methods on the same setting through their exact convergence rates on a given function class, and therefore, to select the best one.

1.1 Outline and contributions

This thesis mainly explores two directions: First, we extend the PEP framework to linear operators and exploit this extension to analyze optimization methods or problems involving linear operators (Part I), which includes primal-dual methods. Then, we exploit the recent non-convex PEP framework to analyze second-order methods on univariate functions (Part II). In both parts, we propose new tight worst-case guarantees.

- **Chapter 2** introduces the Performance Estimation framework and the central concept of interpolation conditions. The framework allows for computing worst-case guarantees on optimization methods. Interpolation conditions of a function class are constraints satisfied by any function of the class. Moreover, if some set of points satisfy these interpolation conditions, then there exists a function of the class that interpolates these points. We also review the different works that extended or exploited the PEP framework to obtain worst-case guarantees on various optimization methods and function classes.

Part I: Performance Estimation Problem and Linear Operators We develop interpolation conditions for linear operators and incorporate them in the PEP framework to analyze optimization methods involving such linear operators. Most of the results of this part are available in [BHG24, BPHG24, BDL+25].

- **Chapter 3** exhibits the interest of developing interpolation conditions for linear operators and proposes examples of methods that could be analyzed with PEP and these interpolation conditions. Many formulations of optimization problems contain linear operators in the objective function or in the constraints, and various optimization methods have been specifically designed to solve such problems.
- **Chapter 4** develops the interpolation conditions for the class of linear operators with singular values in $[\mu, L]$, skew-symmetric linear operators with singular values in $[0, L]$, and symmetric linear operators with eigenvalues in $[\mu, L]$. We also present how we can represent linear operators with singular values or eigenvalues in a union of intervals. These interpolation conditions will then be exploited in Chapters 7 and 8. Most of the results of this chapter are available at [BHG24, BDL+25].
- **Chapter 5** shows a limitation on the representation of the classes of linear operators we can incorporate into the PEP framework. If we only have access to the scalar products between the points, then we can only represent classes of linear operators characterized by their singular value or eigenvalue spectrum belonging to some subset of $\mathbb{R}$. Moreover, the set of Gram matrices associated with such classes of linear operators is always convex. The results of this chapter are available at [BHG24, BDL+25].
- **Chapter 6** first reviews a known technique to analyze fixed-step first-order methods on quadratic functions. This technique is based on the spectral analysis of the polynomial associated with the method. Afterwards, we develop interpolation conditions for quadratic functions thanks to the conditions for linear operators of Chapter 4. These conditions allow to use the PEP framework to analyze the convergence of methods on quadratic functions. Finally, it shows that the performance of a given method on smooth convex functions cannot be deduced *a priori* from its performance on quadratic functions.

- **Chapter 7** exploits the interpolation conditions developed in Chapter 4 to analyze three motivating examples presented in Chapter 3. We consider the worst-case performance of the gradient method on the problem $\min_x g(Mx)$, the Chambolle-Pock method [CP11a] on the problem $\min_x f(x) + g(Mx)$, and the Barzilai-Borwein method [BB88] on the problem $\min_x \frac{1}{2}x^T Qx$. It provides new analytical and numerical worst-case guarantees on these methods. Finally, we show examples where the worst-case linear operator is not just a scaling operator of the form $M = \alpha I$ with $\alpha \in \mathbb{R}$. Most of the results of this chapter are available at [BHG24, BDL+25].

- **Chapter 8** applies the PEP framework to image processing to analyze different methods on the problems $\min_x f(x) + g(x)$ and $\min_x f(x) + g(Mx)$. The former does not require our new interpolation conditions, whereas the second does. We analyze different primal methods for the first problem to obtain their tight contraction factors. Then, we extend this analysis to the second problem and primal-dual methods. The results of this chapter are available at [BPHG24].

- **Appendix A** contains linear algebra results needed in the proofs of Part I. It also contains an alternative proof of a known result on the extension of a matrix that preserves its norm.

Part II: Non-convex Performance Estimation In this part, we exploit the recently introduced non-convex PEP framework to analyze second-order methods. We also develop interpolation conditions for univariate classes of functions studied in the second-order literature, namely, Hessian Lipschitz functions and generalized self-concordant functions [STD19].

- **Chapter 9** presents the non-convex PEP framework and which limitations of the classical convex PEP framework it overcomes. We exhibit some of the new constraints that can be handled together with specific examples of optimization methods that this framework could analyze.

- **Chapter 10** develops a principled technique to obtain exact computer-aided worst-case guarantees on the performance of second-order optimization methods on classes of univariate functions. We first present a generic technique to derive interpolation conditions for a wide range of univariate functions, and rely on this technique to obtain such conditions for generalized self-concordant (including self- and quasi-self-concordant) functions and (strongly convex) functions with Lipschitz Hessian. We then exploit these conditions and the Performance Estimation framework to tightly analyze second-order methods, including (Cubic Regularized) Newton's method and variants on univariate functions. Thereby, we improve on existing convergence rates in the univariate case, exhibit lower bounds valid in the multivariate case, and compare different variants of Newton's method on a fair basis, i.e., with respect to the same setting. The results of this chapter are available at [RBHG25].

- **Appendix B** presents a list of worst cases attained by univariate functions and examples where the worst-case performance is not attained by univariate functions. Moreover, it presents an analysis of the worst-case performance of the fast gradient method on the class of smooth strongly convex functions. Finally, it contains the missing proofs of Chapter 10.

Part III: Conclusion Finally, we conclude with a summary of the research outcomes and a discussion of potential future work in Chapter 11.

2

Performance Estimation Problem

OPTIMIZATION theory relies on performance guarantees of methods. An important type of performance guarantee is the worst-case guarantee. Given a function class and a method, worst-case guarantees describe the worst behavior the method can have on instances of the class. We present how we can obtain such performance guarantees with the Performance Estimation Problem (PEP) framework.

2.1 Conceptual Performance Estimation Problem

We seek a guarantee on the worst behavior of a given optimization method on a given function class. The idea of the Performance Estimation Problem framework is to formulate the problem of finding this worst behavior as an optimization problem and then solve it analytically or numerically. We summarize the PEP methodology as presented in [DT14, THG17c].

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 performance criterion $f(x_N) - f(x^*)$ (objective function accuracy after N iterations), and $[N] = \{1, \ldots, N\}$, the PEP is

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

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.

Remark 2.1. Before solving a (PEP), we have to set numerical values for all the parameters of the function class (e.g., Lipschitz constant), of the method (e.g., step size), and of the analysis (e.g., initial distance R). Solving (PEP) thus only provides the worst-case performance for these fixed parameters, and we have to solve it again to obtain the worst-case performance for other parameters.

The conceptual formulation (PEP) is an infinite-dimensional optimization problem as it involves the class of functions $\mathcal{F}$. However, by discretizing and considering the 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 2.2 ($\mathcal{F}$-interpolability)**.** Let $\mathcal{F}$ be a class of differentiable functions. The set of triplets $S = \{(x_i, g_i, f_i)\}_{i\in[N]}$ is $\mathcal{F}$-interpolable if and only if

$$\exists f \in \mathcal{F} : \begin{cases} f(x_i) = f_i, \\ \nabla f(x_i) = g_i, \end{cases} \quad \forall i \in [N]. \tag{2.1}$$

It allows the following equivalent discretized formulation of (PEP)

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

where $S = \{(x_i, g_i, f_i)\}_{i\in[N]} \cup \{(x_0, g_0, f_0), (x^*, g^*, f^*)\}$. Having points x_i, gradients g_i, and function values f_i is enough if we only consider first-order methods.

2.2 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 2.3 ([THG17c], Theorem 4)**.** *Let $\mu < L$.*
The set of triplets $\{(x_i, g_i, f_i)\}_{i\in[N]}$ is $\mathcal{F}_{\mu,L}$-interpolable if and only if $\forall (i,j) \in [N]^2$

$$\begin{aligned} 2\left(1 - \frac{\mu}{L}\right)\left(f_i - f_j - g_j^T(x_i - x_j)\right) \geq & \frac{1}{L}\left\|g_i - g_j\right\|^2 + \mu\left\|x_i - x_j\right\|^2 \\ & - 2\frac{\mu}{L}(g_i - g_j)^T(x_i - x_j). \end{aligned} \tag{2.2}$$

Remark 2.4. Theorem 2.3 has first been proved for the case $\mu = -L$ in [Wel73]. Later, it was proven for all $\mu \leq L$ and exploited with $0 \leq \mu < L$ in [THG17c]. Finally, it was exploited with $\mu \leq 0$ in [THG17a, AdKZ22, RPG25a]. Moreover, the case $\mu = L$ is covered in [RBCH23].

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 [THG17c] 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 (2.2), 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 [THG17c] for more details about the tractable formulation of (PEP) and interpolation conditions.

2.3 State of the art

Since its introduction, the PEP framework has evolved in many directions and settings to analyze a lot of different problems and methods. In addition to the references listed below, we refer the reader to this excellent recent review [Tay24] and excellent theses [Bar21, Gou24, Abb24, Col24, DG24, Kam25].

Gradient method and related variants Several works studied the gradient method (GM) including GM (and its accelerated version) on smooth strongly or hypoconvex functions [DT14, Tay17, RGP24], convex functions with a quadratic upper bound [GTD22], functions with lower restricted secant inequality and an upper error bound [GEIGM22], and non-convex functions satisfying Łojasiewicz inequality [AdKZ23a]. More variants of GM have been covered by PEP including GM with exact line search [dKGT17], (inexact) proximal GM [THG18, BTB23], inexact GM with bounded error on the gradient [dKGT20, VG24, VG25], subgradient methods [ZG23], and block coordinate descent method [SL16, TB19, KHG23, AdKZ23b].

Different first-order methods Other first-order methods have also been studied by the framework as difference-of-convex algorithm [AdKZ24, RPG25a, RPG25b], the gradient descent-ascent method, the Alternating Direction Method of Multipliers (ADMM) [ZAdK24], the heavy-ball method [GTD23], the Chambolle-Pock and Condat-Vũ methods [BHG24, BPHG24], Bregman or mirror descent [DTdB22], and nonlinear conjugate gradient methods [DGFST24].

Inclusion problem and variational inequalities 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 [Lie21], the proximal point algorithm for maximal monotone inclusions [GY20], the relaxed proximal point algorithm for variational inequalities [GY24], splitting methods [RTBG20], and extragradient methods [GLG22, GTG22, GTHG23].

Design of methods 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 [DT16], the optimized gradient method for smooth convex functions decreasing the cost function [KF16, KF17, KF18b] and the gradient norm [KF21], optimal gradient methods for smooth strongly convex functions [VSFL17, TD23, VSL25], optimal methods for non-smooth and smooth convex functions [DT20], fast iterative shrinkage/thresholding algorithm [KF18a], and accelerated proximal points method [Kim21]. Recent works also proposed to solve non-convex PEP to design optimized methods [DGVPR24a, JDGR23, KHG25, Kam25]. The recent trend to accelerate the convergence of GM by exploiting long steps has also used PEP in some of these works [Alt18, AP25, GSW23, Gri24].

Structure of proofs and interpolation conditions PEP combines interpolation conditions to provide a proof of performance. Therefore, obtaining interpolation conditions for different classes of functions is crucial [RH24, RBCH23, RHT25]. Moreover, the following work analyzed the structure of the proofs in a systematic way [GDT23b].

Other frameworks and applications Some works exploited and extended the PEP framework to develop methodologies that address closely related questions like decentralized optimization [CH23, CH22, Col24], stochastic optimization [TB19, RCH25], continuous-time model version of first-order methods [MTB23], Lyapunov inequalities for first-order methods [LRP16, TVSL18, UBTG25, BUG23], characterizing the set of possible minimum of a sum of functions [ZGH24], identifying if methods can produce cyclic trajectories [GDT23a], analyzing infeasibility detection methods [PR23], and PEP for smooth convex sets [LG24].

Software implementations Simultaneously with their introduction, some of these extensions have been implemented and are available on the *Matlab* toolbox `PESTO` [THG17b] and *Python* package `PEPit` [GMG$^+$24]. The

PEPit documentation website contains numerous implemented examples of applications of the PEP framework.

2.3.1 Linear operators and second-order optimization

We conclude this section by reviewing the different works that addressed the analysis of optimization methods involving linear operators or second-order methods with PEP.

Linear operators In [BHG24], we developed interpolation conditions for linear operators to incorporate them in PEP for the first time (see Chapter 4). However, PEP was used to analyze problems involving linear operators on a few occasions before the introduction of these interpolation conditions: in [CH23] for decentralized optimization, in [AdKZ23b] for the random coordinate descent method on nonhomogeneous quadratic functions, and in [ZAdK24] for ADMM. We will further detail the contribution of these works at the end of Section 4.1. In all cases, they present only potential relaxations of the Performance Estimation Problem for their specific situation, whereas in this work we propose a provably exact formulation of PEP for general problems with linear operators.

After the introduction of our interpolation conditions, different works used or extended them to analyze new problems with PEP [RS25, VKK+25, KGK24].

Finally, the following work proposes a Lyapunov analysis of primal-dual methods on problems involving linear operators without interpolation conditions for linear operators but by reformulating the problem with the singular value decomposition [VSSPL23].

Second-order optimization To the best of our knowledge, the only attempt of second-order computer-aided analysis is the single iteration of second-order methods on the class of self-concordant functions [dKGT20], but these results are based on interpolation conditions of a larger function class, hence not a priori tight, and are inherently restricted to the analysis of a single iteration.

Meanwhile, the approach we proposed in Chapter 10 develops new interpolation conditions for classes of univariate functions and exploits global solvers to solve non-convex formulations of PEP.

PART I
Performance Estimation and Linear Operators

3

Motivation

We propose an extension of 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 values spectrum. The main tool behind this extension is the development of interpolation conditions for linear operators (see Section 4.1). Therefore, we will be able to study the exact worst-case performance of first-order algorithms involving linear operators.

Many methods aim at solving problems involving linear operators in their objective functions or constraints. Iterations of such methods typically apply themselves the linear operators of the problem. We describe below four 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,

$$\nabla F(x_i) = M^T \nabla g(Mx_i) \tag{3.1}$$

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 6.1 or [dST+21, Section 2.1]), 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 [AB22] for recent work on this class with a nonsmooth function f).

Motivating example 2: Adaptive methods on quadratic functions The technique described in the previous motivating example can only deal with fixed-step methods and not with adaptive methods. It is however also possible to analyze some adaptive methods with PEP [BTd20] by the way of non-convex formulations. Therefore, developing interpolation conditions for quadratic functions and incorporating them in PEP will allow to analyze the worst-case performance of such methods on quadratic functions.

Motivating example 3: $\min_x f(x) + g(Mx)$ Let the problem $\min_x f(x) + g(Mx)$ where g is proximable. We can solve this problem with the Chambolle-Pock method (CP) [CP11a] when f is proximable, and with the Condat-Vũ method (CV) [Con13, Vũ13] when f is smooth with the following primal-dual iterations

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

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

These methods perform proximal steps, which are already handled in the PEP framework [THG18]. Again, these iterations contain products with matrices M and M^T.

Motivating example 4: $\min_{x,y} f(x) + g(y)$ s.t. $M_1 x + M_2 y = b$ A famous method to solve the constrained problem $\min_{x,y} f(x) + g(y)$ s.t. $M_1 x + M_2 y = b$ is the Alternating Direction Method of Multipliers (ADMM) [GM76] with the following iterations

$$\begin{cases} x_{i+1} \in \arg\min_x f(x) + \frac{\rho}{2} \left\| M_1 x + M_2 y_i - b + \frac{1}{\rho} z_i \right\|^2, \\ y_{i+1} \in \arg\min_y g(y) + \frac{\rho}{2} \left\| M_1 x_{i+1} + M_2 y - b + \frac{1}{\rho} z_i \right\|^2, \\ z_{i+1} = z_i + \rho(M_1 x_{i+1} + M_2 y_{i+1} - b). \end{cases} \tag{3.2}$$

By contrast to the first three 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, e.g., Section 7.1.2 and [NNG19]). 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 a lot of first-order method involving linear operators, for example, Primal-Dual Fixed Point [CHZ16], Primal-Dual Three-Operator Splitting [Yan18], and Proximal Alternating Predictor-Corrector [DST15] algorithms (see [CKCH23] 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 [RL05], ℓ_1-constrained least squares [Eld80], basis pursuit [CD94], total variation deblurring [ROF92, BT09], and resource allocation [YHL16].

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.

4

Interpolation conditions for linear operators

Some of the results of this chapter were obtained in collaboration with Zhicheng Deng (Theorem 4.6, its proof in Section 4.2.5, and Section 4.3). The existence of Theorem A.7 was pointed out by Paul Van Dooren. Discussions with Adrien Taylor and Anne Rubbens inspired some of the results of this chapter.

We extend the Performance Estimation Problem (PEP) framework to methods involving linear operators. As presented in the motivating examples of Chapter 3, 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 (3.1) of the gradient of a composed function $F(x) = g(Mx)$ as

$$v_i = \nabla F(x_i) \quad \Leftrightarrow \quad v_i = M^T \nabla g(Mx_i) \quad \Leftrightarrow \quad \begin{aligned} y_i &= Mx_i, \\ u_i &= \nabla g(y_i), \\ v_i &= M^T u_i. \end{aligned} \tag{4.1}$$

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 eigenvalue or singular value spectrums since it is a compact way to summarize an important property of those matrices (but not the only one). Such bounds on the spectrum are usually assumed in the literature. We consider general, symmetric, and skew-symmetric matrices, namely,

$$\begin{aligned} \mathcal{L}_{\mu,L} &= \{M : \mu \leq \sigma_{\min}(M) \text{ and } \sigma_{\max}(M) \leq L\}, \\ \mathcal{S}_{\mu,L} &= \{Q : Q = Q^T, \mu I \preceq Q \preceq LI\}, \\ \mathcal{T}_{\mu,L} &= \{Q : Q = -Q^T, \mu \leq \sigma_{\min}(Q) \text{ and } \sigma_{\max}(Q) \leq L\}, \end{aligned} \tag{4.2}$$

with $\sigma_{\min}(M)$ and $\sigma_{\max}(M)$ the smallest and largest singular values of M, and $\preceq$ the Löwner order for symmetric matrices. Note that $\mathcal{L}_{\mu,L}$ is the class of L-Lipschitz μ-strongly linear operators (see [RTBG20] 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 2.2. When the linear operators are not symmetric nor skew-symmetric, we consider two sets of pairs $\{(x_i, y_i)\}_{i\in[N_1]}$ and $\{(u_j, v_j)\}_{j\in[N_2]}$ to represent the applications of the operator and its transpose. In the symmetric and skew-symmetric cases, we only need one set of pairs $\{(x_i, y_i)\}_{i\in[N]}$.

Definition 4.1 ($\mathcal{L}$-interpolability (two sets of pairs))**.** Let $\mathcal{L}$ be a class of linear operators. Sets of pairs $\{(x_i, y_i)\}_{i\in[N_1]}$ and $\{(u_j, v_j)\}_{j\in[N_2]}$ are $\mathcal{L}$-interpolable if and only if

$$\exists M \in \mathcal{L} \ : \ \begin{cases} y_i = Mx_i, & \forall i \in [N_1], \\ v_j = M^T u_j, & \forall j \in [N_2]. \end{cases} \tag{4.3}$$

Definition 4.2 ($\mathcal{L}$-interpolability (one set of pairs))**.** Let $\mathcal{L}$ be a class of symmetric or skew-symmetric linear operators. Set of pairs $\{(x_i, y_i)\}_{i\in[N]}$ is $\mathcal{L}$-interpolable if and only if

$$\exists Q \in \mathcal{L} \ : \ y_i = Qx_i, \ \forall i \in [N]. \tag{4.4}$$

4.1 Main results

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 $\mathcal{L}_{\mu,L}$, $\mathcal{T}_{0,L}$, and $\mathcal{S}_{\mu,L}$ of linear operators.

In the sequel, 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 $\mathcal{L}$-interpolable when $Y = MX$ and $V = M^T U$ for some $M \in \mathcal{L}$. Similarly, we write that (X, Y) is $\mathcal{L}$-interpolable when $Y = QX$ for some $Q \in \mathcal{L}$. We now state our main interpolation theorems.

Theorem 4.3 ($\mathcal{L}_{0,L}$-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 $L \geq 0$.*

(X, Y, U, V) is $\mathcal{L}_{0,L}$-interpolable if and only if

$$\begin{cases} X^T V = Y^T U, \\ Y^T Y \preceq L^2 X^T X, \\ V^T V \preceq L^2 U^T U. \end{cases} \tag{4.5}$$

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

Corollary 4.4 ($\mathcal{T}_{0,L}$-interpolation conditions)**.** *Let $X \in \mathbb{R}^{d\times N}$, $Y \in \mathbb{R}^{d\times N}$, and $L \geq 0$.*

(X, Y) is $\mathcal{T}_{0,L}$-interpolable if and only if

$$\begin{cases} X^T Y = -Y^T X, \\ Y^T Y \preceq L^2 X^T X. \end{cases} \tag{4.6}$$

Theorem 4.5 ($\mathcal{S}_{\mu,L}$-interpolation conditions)**.** *Let $X \in \mathbb{R}^{d\times N}$, $Y \in \mathbb{R}^{d\times N}$, and $-\infty < \mu \leq L < \infty$.*

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

$$\begin{cases} X^T Y = Y^T X, \\ (Y - \mu X)^T (LX - Y) \succeq 0. \end{cases} \tag{4.7}$$

Theorem 4.6 ($\mathcal{L}_{\mu,L}$-interpolation conditions)**.** *Let $X \in \mathbb{R}^{n\times N_1}$, $Y \in \mathbb{R}^{m\times N_1}$, $U \in \mathbb{R}^{m\times N_2}$, $V \in \mathbb{R}^{n\times N_2}$, $m \geq n$, and $-\infty < \mu \leq L < \infty$.*

(X, Y, U, V) is $\mathcal{L}_{\mu,L}$-interpolable if and only if $\exists Z \in \mathbb{R}^{n\times N_1}, W \in \mathbb{R}^{n\times N_2}$ such that

$$\begin{cases} Y^TY = Z^TZ, \\ Y^TU = Z^TW, \\ U^TU \succeq W^TW, \\ \begin{pmatrix} X & W \end{pmatrix}^T \begin{pmatrix} Z & V \end{pmatrix} = \begin{pmatrix} Z & V \end{pmatrix}^T \begin{pmatrix} X & W \end{pmatrix}, \\ \left(\begin{pmatrix} Z & V \end{pmatrix} - \mu \begin{pmatrix} X & W \end{pmatrix}\right)^T \left(L \begin{pmatrix} X & W \end{pmatrix} - \begin{pmatrix} Z & V \end{pmatrix}\right) \succeq 0. \end{cases} \tag{4.8}$$

When $m = n$, the third inequality must be satisfied by equality, i.e., $U^TU = W^TW$.

Remark 4.7. Theorem 4.3 can be seen as a generalization of Douglas' lemma [Dou66, Theorem 1].

Crucially, all these conditions only involve the scalar products between the columns of $(X\ U)$ and $(Y\ V)$ (and additional vectors) for $\mathcal{L}_{0,L}$ and $\mathcal{L}_{\mu,L}$ and columns of $(X\ Y)$ for $\mathcal{T}_{0,L}$ 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 4.3, the condition $X^TV = Y^TU$ is related to the fact that X and Y are linked by the same matrix (but transposed) as U and V. Similarly, in Theorem 4.5 (resp. Corollary 4.4) $X^TY = Y^TX$ (resp. $X^TY = -Y^TX$) is related to the symmetry (resp. skew-symmetry) of the operator. Furthermore, a product X^TX is always symmetric positive semidefinite. Finally, in the nonsymmetric and skew-symmetric cases, $Y^TY \preceq L^2X^TX$ 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 (4.5) (without the first equation) and (4.7) were presented as necessary conditions for decentralized optimization in Theorem 1 from [CH23] (in a more specific setting) and in the analysis of ADMM in formulation (13) from [ZAdK24]. Moreover, in equation (2.7) from [AdKZ23b], the authors further restrict the L-smooth μ-strongly convex interpolation conditions with an additional necessary condition for nonhomogeneous quadratic functions (see Section 6.2 for details). Using necessary but not sufficient interpolation conditions in the context of the Performance Estimation Problem still allows obtaining bounds on the worst-case performance, but whose tightness is no longer guaranteed. In this work, we show that conditions (4.5) and (4.7) are also sufficient.

Finally, we extend the interpolation conditions for $\mathcal{L}_{0,L}$ to $\mathcal{L}_{\mu,L}$ for $\mu \geq 0$ by exploiting the polar decomposition of a matrix and the fact that we can easily interpolate products of matrices.

Remark 4.8. Relying on interpolation conditions guarantees *a priori* to obtain exact and tight worst-case analysis. The previous analysis of ADMM [ZAdK24] did not use interpolation conditions but proved *a posteriori* that their results were tight by exhibiting instances reaching their results.

Remark 4.9. Another similar result on the interpolation of linear operators has further been obtained in [BLDPM24, Proposition 1].

4.2 Proofs of the main results

We show that the interpolation conditions are indeed necessary and sufficient for $\mathcal{L}_{0,L}$, $\mathcal{T}_{0,L}$, and $\mathcal{S}_{\mu,L}$-interpolability. As suggested by the above observations, proving the necessity is much easier than proving the sufficiency.

To show that conditions (4.5) are sufficient for $\mathcal{L}_{0,L}$-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 $(X_R\ V_R)$ and $(Y_R\ U_R)$ of the Gram matrices $G \triangleq (X\ V)^T(X\ V)$ and $H \triangleq (Y\ U)^T(Y\ U)$, such that (X_R, Y_R, U_R, V_R) is $\mathcal{L}_{0,L}$-interpolable (Lemma 4.10). Afterwards, we show how this implies the $\mathcal{L}_{0,L}$-interpolability of the actual vectors (X, Y, U, V) (Lemma 4.11). In other words, the proof consists of two steps:

- **Step 1** (Lemma 4.10): there exist factorizations $(X_R\ V_R)$ and $(Y_R\ U_R)$ of the Gram matrices G and H where (X_R, Y_R, U_R, V_R) is $\mathcal{L}_{0,L}$-interpolable;
- **Step 2** (Lemma 4.11): 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 $\mathcal{T}_{0,L}$-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.

Finally, we prove the case $\mathcal{L}_{\mu,L}$ by exploiting the polar decomposition of a matrix as a product of a semi-unitary and a symmetric matrix.

4.2.1 $\mathcal{L}_{0,L}$-interpolability of (X_R, Y_R, U_R, V_R) (Step 1)

Conditions (4.5) only involve the scalar products between the columns of $\begin{pmatrix} X & V \end{pmatrix}$ and $\begin{pmatrix} Y & U \end{pmatrix}$, 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 $\mathcal{L}_{0,L}$-interpolable.

Lemma 4.10 (Existence of $\mathcal{L}_{0,L}$-interpolable factorizations of Gram matrices). *Let two symmetric positive semidefinite matrices* $G = \begin{pmatrix} A_1 & B_1 \\ B_1^T & C_1 \end{pmatrix}$ *and* $H = \begin{pmatrix} A_2 & B_2 \\ B_2^T & C_2 \end{pmatrix}$ *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*

$$\begin{cases} B_1 = B_2, \\ A_2 \preceq L^2 A_1, \\ C_1 \preceq L^2 C_2, \end{cases} \tag{4.9}$$

then they admit the factorizations

$$G = \begin{pmatrix} X_R & V_R \end{pmatrix}^T \begin{pmatrix} X_R & V_R \end{pmatrix}, \tag{4.10}$$

$$H = \begin{pmatrix} Y_R & U_R \end{pmatrix}^T \begin{pmatrix} Y_R & U_R \end{pmatrix}, \tag{4.11}$$

where

$$X_R = \begin{pmatrix} A_1^{\frac{1}{2}} \\ 0_{N_2,N_1} \end{pmatrix} \in \mathbb{R}^{N_1+N_2 \times N_1}, \quad Y_R = \begin{pmatrix} ({C_2}^{\dagger})^{\frac{1}{2}} B_2^T \\ (A_2 - B_2 {C_2}^{\dagger} B_2^T)^{\frac{1}{2}} \end{pmatrix} \in \mathbb{R}^{N_1+N_2 \times N_1}, \tag{4.12}$$

$$U_R = \begin{pmatrix} C_2^{\frac{1}{2}} \\ 0_{N_1,N_2} \end{pmatrix} \in \mathbb{R}^{N_2+N_1 \times N_2}, \quad V_R = \begin{pmatrix} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 \\ (C_1 - B_1^T {A_1}^{\dagger} B_1)^{\frac{1}{2}} \end{pmatrix} \in \mathbb{R}^{N_2+N_1 \times N_2}, \tag{4.13}$$

such that (X_R, Y_R, U_R, V_R) *is* $\mathcal{L}_{0,L}$*-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. Since $B_1 = B_2$, 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

$$\begin{aligned} A_1^{\frac{1}{2}} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 &= A_1 {A_1}^{\dagger} B_1 = B_1, \\ B_2 ({C_2}^{\dagger})^{\frac{1}{2}} C_2^{\frac{1}{2}} &= B_2 {C_2}^{\dagger} C_2 = B_2, \end{aligned} \tag{4.14}$$

and Propositions A.2 and A.4 with (4.9) and (4.14) provide

$$\begin{aligned} S_2^{\frac{1}{2}} A_1^{\frac{1}{2}} ({A_1}^{\dagger})^{\frac{1}{2}} &= S_2^{\frac{1}{2}}, \\ S_1^{\frac{1}{2}} C_2^{\frac{1}{2}} ({C_2}^{\dagger})^{\frac{1}{2}} &= S_1^{\frac{1}{2}}. \end{aligned} \tag{4.15}$$

Firstly, we show that

$$(X_R, Y_R, U_R, V_R) = \left(\begin{pmatrix} A_1^{\frac{1}{2}} \\ 0_{N_1,N_1} \end{pmatrix}, \begin{pmatrix} ({C_2}^{\dagger})^{\frac{1}{2}} B_2^T \\ S_2^{\frac{1}{2}} \end{pmatrix}, \begin{pmatrix} C_2^{\frac{1}{2}} \\ 0_{N_1,N_2} \end{pmatrix}, \begin{pmatrix} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 \\ S_1^{\frac{1}{2}} \end{pmatrix} \right) \tag{4.16}$$

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

$$X_R^T X_R = \begin{pmatrix} A_1^{\frac{1}{2}} & 0_{N_1,N_2} \end{pmatrix} \begin{pmatrix} A_1^{\frac{1}{2}} \\ 0_{N_2,N_1} \end{pmatrix} = A_1, \tag{4.17}$$

$$U_R^T U_R = \begin{pmatrix} C_2^{\frac{1}{2}} & 0_{N_2,N_1} \end{pmatrix} \begin{pmatrix} C_2^{\frac{1}{2}} \\ 0_{N_1,N_2} \end{pmatrix} = C_2, \tag{4.18}$$

$$Y_R^T Y_R = \begin{pmatrix} B_2 ({C_2}^{\dagger})^{\frac{1}{2}} & S_2^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} ({C_2}^{\dagger})^{\frac{1}{2}} B_2^T \\ S_2^{\frac{1}{2}} \end{pmatrix} = B_2 ({C_2}^{\dagger})^{\frac{1}{2}} ({C_2}^{\dagger})^{\frac{1}{2}} B_2^T + S_2 = A_2, \tag{4.19}$$

$$V_R^T V_R = \begin{pmatrix} B_1^T ({A_1}^{\dagger})^{\frac{1}{2}} & S_1^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 \\ S_1^{\frac{1}{2}} \end{pmatrix} = B_1^T ({A_1}^{\dagger})^{\frac{1}{2}} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 + S_1 = C_1, \tag{4.20}$$

$$X_R^T V_R = \begin{pmatrix} A_1^{\frac{1}{2}} & 0_{N_1,N_2} \end{pmatrix} \begin{pmatrix} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 \\ S_1^{\frac{1}{2}} \end{pmatrix} = A_1^{\frac{1}{2}} ({A_1}^{\dagger})^{\frac{1}{2}} B_1 \overset{(4.14)}{=} B_1, \tag{4.21}$$

$$Y_R^T U_R = \begin{pmatrix} B_2 ({C_2}^{\dagger})^{\frac{1}{2}} & S_2^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} C_2^{\frac{1}{2}} \\ 0_{N_1,N_2} \end{pmatrix} = B_2 ({C_2}^{\dagger})^{\frac{1}{2}} C_2^{\frac{1}{2}} \overset{(4.14)}{=} B_2, \tag{4.22}$$

and therefore

$$\begin{aligned}\begin{pmatrix} X_R & V_R \end{pmatrix}^T \begin{pmatrix} X_R & V_R \end{pmatrix} &= \begin{pmatrix} X_R^T X_R & X_R^T V_R \\ V_R^T X_R & V_R^T V_R \end{pmatrix} = \begin{pmatrix} A_1 & B_1 \\ B_1^T & C_1 \end{pmatrix} = G, \\ \begin{pmatrix} Y_R & U_R \end{pmatrix}^T \begin{pmatrix} Y_R & U_R \end{pmatrix} &= \begin{pmatrix} Y_R^T Y_R & Y_R^T U_R \\ U_R^T Y_R & U_R^T U_R \end{pmatrix} = \begin{pmatrix} A_2 & B_2 \\ B_2^T & C_2 \end{pmatrix} = H. \end{aligned} \tag{4.23}$$

Secondly, we show that (X_R, Y_R, U_R, V_R) is $\mathcal{L}_{0,L}$-interpolable by providing a linear operator M_R such that $Y_R = M_R X_R$, $V_R = M_R^T U_R$ and $\sigma_{\max}(M_R) = ||M_R|| \leq L$. The expression of M_R is

$$M_R = \begin{pmatrix} (C_2{}^\dagger)^{\frac{1}{2}} B_1^T (A_1{}^\dagger)^{\frac{1}{2}} & (C_2{}^\dagger)^{\frac{1}{2}} S_1^{\frac{1}{2}} \\ S_2^{\frac{1}{2}} (A_1{}^\dagger)^{\frac{1}{2}} & W \end{pmatrix} \tag{4.24}$$

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

$$\begin{aligned} M_R X_R &= \begin{pmatrix} (C_2{}^\dagger)^{\frac{1}{2}} B_1^T (A_1{}^\dagger)^{\frac{1}{2}} & (C_2{}^\dagger)^{\frac{1}{2}} S_1^{\frac{1}{2}} \\ S_2^{\frac{1}{2}} (A_1{}^\dagger)^{\frac{1}{2}} & W \end{pmatrix} \begin{pmatrix} A_1^{\frac{1}{2}} \\ 0_{N_2,N_1} \end{pmatrix} \\ &= \begin{pmatrix} (C_2{}^\dagger)^{\frac{1}{2}} B_1^T (A_1{}^\dagger)^{\frac{1}{2}} A_1^{\frac{1}{2}} \\ S_2^{\frac{1}{2}} (A_1{}^\dagger)^{\frac{1}{2}} A_1^{\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} (C_2{}^\dagger)^{\frac{1}{2}} B_1^T \\ S_2^{\frac{1}{2}} \end{pmatrix} = Y_R, \end{aligned} \tag{4.25}$$

$$\begin{aligned} M_R^T U_R &= \begin{pmatrix} (A_1{}^\dagger)^{\frac{1}{2}} B_2^T (C_2{}^\dagger)^{\frac{1}{2}} & (A_1{}^\dagger)^{\frac{1}{2}} S_2^{\frac{1}{2}} \\ S_1^{\frac{1}{2}} (C_2{}^\dagger)^{\frac{1}{2}} & W^T \end{pmatrix} \begin{pmatrix} C_2^{\frac{1}{2}} \\ 0_{N_1,N_2} \end{pmatrix} \\ &= \begin{pmatrix} (A_1{}^\dagger)^{\frac{1}{2}} B_2^T (C_2{}^\dagger)^{\frac{1}{2}} C_2^{\frac{1}{2}} \\ S_1^{\frac{1}{2}} (C_2{}^\dagger)^{\frac{1}{2}} C_2^{\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} (A_1{}^\dagger)^{\frac{1}{2}} B_2^T \\ S_1^{\frac{1}{2}} \end{pmatrix} = V_R. \end{aligned} \tag{4.26}$$

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.7, we must just show that the matrices

$$\begin{aligned} M_R^{(\text{up})} &= \begin{pmatrix} (C_2{}^\dagger)^{\frac{1}{2}} B_1 (A_1{}^\dagger)^{\frac{1}{2}} & (C_2{}^\dagger)^{\frac{1}{2}} S_1^{\frac{1}{2}} \end{pmatrix}, \\ M_R^{(\text{left})} &= \begin{pmatrix} (C_2{}^\dagger)^{\frac{1}{2}} B_2^T (A_1{}^\dagger)^{\frac{1}{2}} \\ S_2^{\frac{1}{2}} (A_1{}^\dagger)^{\frac{1}{2}} \end{pmatrix}, \end{aligned} \tag{4.27}$$

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.,

$$\begin{aligned} M_R^{(\mathrm{up})}{M_R^{(\mathrm{up})}}^T &= ({C_2}^\dagger)^{\frac{1}{2}} B_1 {A_1}^\dagger B_1^T ({C_2}^\dagger)^{\frac{1}{2}} + ({C_2}^\dagger)^{\frac{1}{2}} S_1 ({C_2}^\dagger)^{\frac{1}{2}} \\ &= ({C_2}^\dagger)^{\frac{1}{2}} C_1 ({C_2}^\dagger)^{\frac{1}{2}} \preceq L^2 I, \end{aligned} \tag{4.28}$$

$$\begin{aligned} {M_R^{(\mathrm{left})}}^T M_R^{(\mathrm{left})} &= ({A_1}^\dagger)^{\frac{1}{2}} B_2 {C_2}^\dagger B_2^T ({A_1}^\dagger)^{\frac{1}{2}} + ({A_1}^\dagger)^{\frac{1}{2}} S_2 ({A_1}^\dagger)^{\frac{1}{2}} \\ &= ({A_1}^\dagger)^{\frac{1}{2}} A_2 ({A_1}^\dagger)^{\frac{1}{2}} \preceq L^2 I, \end{aligned} \tag{4.29}$$

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 pseudo inverse and Proposition A.2 (we have the same result for $A_2 \preceq L^2 A_1$).

Finally, we observe on expression (4.24) 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.7, 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 (4.24). □

This lemma guarantees that, if (X, Y, U, V) satisfies (4.5), then their associated Gram matrices admit a factorization (X_R, Y_R, U_R, V_R) that is $\mathcal{L}_{0,L}$-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

$$X = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, Y = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad \text{(with } N_1 = 1) \tag{4.30}$$

which satisfy the $\mathcal{S}_{\mu,L}$-interpolation conditions (4.7) for $\mu = 0$ and $L = 1$. If, given these data (X, Y), one computes the Gram matrix $\begin{pmatrix} X & Y \end{pmatrix}^T \begin{pmatrix} X & Y \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}$, it turns out that via Lemma 4.10 one will obtain the following new and different factorization $X_R = \begin{pmatrix} \sqrt{3} \\ 0 \end{pmatrix}$, $Y_R = \begin{pmatrix} \frac{2\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} \end{pmatrix}$.

Hence Lemma 4.10 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.

4.2.2 Rotation to (X, Y, U, V) (Step 2)

We now show that if a given factorization of Gram matrices is $\mathcal{L}_{0,L}$-interpolable, then all factorizations of these Gram matrices are $\mathcal{L}_{0,L}$-interpolable.

Lemma 4.11 (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 $\mathcal{L}_{0,L}$-interpolable, then, (X, Y, U, V) is $\mathcal{L}_{0,L}$-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

$$\begin{aligned} \begin{pmatrix} X & V \end{pmatrix}^T \begin{pmatrix} X & V \end{pmatrix} &= \begin{pmatrix} X_R & V_R \end{pmatrix}^T \begin{pmatrix} X_R & V_R \end{pmatrix}, \\ \begin{pmatrix} Y & U \end{pmatrix}^T \begin{pmatrix} Y & U \end{pmatrix} &= \begin{pmatrix} Y_R & U_R \end{pmatrix}^T \begin{pmatrix} Y_R & U_R \end{pmatrix}. \end{aligned} \tag{4.31}$$

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. Adding zeros to $\begin{pmatrix} X & V \end{pmatrix}$ or $\begin{pmatrix} X_R & V_R \end{pmatrix}$ such that they have the same number of rows, i.e., $d_n = \max\{n, n_R\}$, will allow us to use Theorem A.8. We use $E_{n,d_n} = \begin{pmatrix} I_n \\ 0_{(d_n-n),n} \end{pmatrix}$ and $E_{n_R,d_n} = \begin{pmatrix} I_{n_R} \\ 0_{(d_n-n_R),n_R} \end{pmatrix}$ to add the zeros and now work with $E_{n,d_n} \begin{pmatrix} X & V \end{pmatrix}$ and $E_{n_R,d_n} \begin{pmatrix} X_R & V_R \end{pmatrix}$. Moreover, adding the zeros preserves the Gram matrices, indeed, if

$$\begin{pmatrix} X & V \end{pmatrix}^T \begin{pmatrix} X & V \end{pmatrix} = \begin{pmatrix} X_R & V_R \end{pmatrix}^T \begin{pmatrix} X_R & V_R \end{pmatrix}, \tag{4.32}$$

then

$$\begin{pmatrix} X & V \end{pmatrix}^T E_{n,d_n}^T E_{n,d_n} \begin{pmatrix} X & V \end{pmatrix} = \begin{pmatrix} X_R & V_R \end{pmatrix}^T E_{n_R,d_n}^T E_{n_R,d_n} \begin{pmatrix} X_R & V_R \end{pmatrix}. \tag{4.33}$$

Therefore, Theorem A.8 applies and yields

$$E_{n_R,d_n} \begin{pmatrix} X_R & V_R \end{pmatrix} = V_G E_{n,d_n} \begin{pmatrix} X & V \end{pmatrix}, \tag{4.34}$$

and with the same reasoning on $\begin{pmatrix} Y & U \end{pmatrix}$, $\begin{pmatrix} Y_R & U_R \end{pmatrix}$, and $d_m = \max\{m, m_R\}$, it yields

$$E_{m_R,d_m}\begin{pmatrix} Y_R & U_R \end{pmatrix} = V_H E_{m,d_m}\begin{pmatrix} Y & U \end{pmatrix}, \tag{4.35}$$

for some V_G and V_H unitary. Now, since (X_R, Y_R, U_R, V_R) is $\mathcal{L}_{0,L}$-interpolable, there exists a M_R such that $Y_R = M_R X_R$, $V_R = M^T U_R$, and $||M_R|| \leq L$. Therefore,

$$\begin{aligned}
Y_R = M_R X_R \Rightarrow \overbrace{E_{m_R,d_m} Y_R}^{=V_H E_{m,d_m} Y} &= E_{m_R,d_m} M_R E^T_{n_R,d_n} \overbrace{E_{n_R,d_n} X_R}^{=V_G E_{n,d_n} X} && (4.36)\\
\Rightarrow Y &= E^T_{m,d_m} V_H^T E_{m_R,d_m} M_R E^T_{n_R,d_n} V_G E_{n,d_n} X \triangleq MX && (4.37)\\
V_R = M_R^T U_R \Rightarrow \overbrace{E_{n_R,d_n} V_R}^{=V_G E_{n,d_n} V} &= E_{n_R,d_n} M_R^T E^T_{m_R,d_m} \overbrace{E_{m_R,d_m} U_R}^{=V_H E_{m,d_m} X} && (4.38)\\
\Rightarrow V &= E^T_{n,d_n} V_G^T E_{n_R,d_n} M_R^T E^T_{m_R,d_m} V_H E_{m,d_m} U \triangleq M^T U. && (4.39)
\end{aligned}$$

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 zero 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). □

We are now able to prove the necessity and sufficiency of interpolation conditions in the nonsymmetric case. The necessity follows from a straightforward reasoning, and the sufficiency relies only on the successive application of Lemmas 4.10 and 4.11.

4.2.3 Proof of Theorem 4.3

Proof of Theorem 4.3. (Necessity) Let us assume that (X, Y, U, V) is $\mathcal{L}_{0,L}$-interpolable. First, $Y = MX$ and $V = M^T U$ yield $X^T V = X^T M^T U = Y^T U$. Moreover, $MM^T \preceq L^2 I$ implies $U^T MM^T U \preceq L^2 U^T U$, i.e., $V^T V \preceq L^2 U^T U$ and similarly, $M^T M \preceq L^2 I$ implies $X^T M^T MX \preceq L^2 X^T X$, i.e., $Y^T Y \preceq L^2 X^T X$.

(Sufficiency) Let us assume that (X, Y, U, V) satisfies conditions (4.5). From Lemma 4.10, there exists (X_R, Y_R, U_R, V_R) which is $\mathcal{L}_{0,L}$-interpolable and shares the same Gram matrices as (X, Y, U, V). Thus, by Lemma 4.11, (X, Y, U, V) is $\mathcal{L}_{0,L}$-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 4.10 and thus M symmetric

(resp. skew-symmetric) in Lemma 4.11. □

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 4.3 whereas the necessity follows a similar reasoning that for the nonsymmetric case.

4.2.4 Proof of Theorem 4.5

Proof of Theorem 4.5. First, let us define $\tilde{X} = X$ and $\tilde{Y} = Y - \frac{L+\mu}{2}X$ and show that requiring (X, Y) to satisfy (4.7) is equivalent to

$$\begin{cases} \tilde{X}^T\tilde{Y} = \tilde{Y}^T\tilde{X}, \\ \tilde{Y}^T\tilde{Y} \preceq \left(\frac{L-\mu}{2}\right)^2 \tilde{X}^T\tilde{X}. \end{cases} \tag{4.40}$$

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

$$\begin{aligned} (Y-\mu X)^T(Y-LX) \preceq 0 &\Leftrightarrow \left(\tilde{Y} + \frac{L-\mu}{2}X\right)^T\left(\tilde{Y} - \frac{L-\mu}{2}X\right) \preceq 0 \\ &\Leftrightarrow \tilde{Y}^T\tilde{Y} \preceq \left(\frac{L-\mu}{2}\right)^2 X^TX. \end{aligned} \tag{4.41}$$

Secondly, from Theorem 4.3 with $U = X$ and $V = Y$, conditions (4.40) 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 $\tilde{Q} \in \mathcal{S}_{-\frac{L-\mu}{2},\frac{L-\mu}{2}}$.
Thirdly, $\tilde{Y} = \tilde{Q}\tilde{X}$ 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,

$$\tilde{Y} = \tilde{Q}\tilde{X} \Leftrightarrow Y - \frac{L+\mu}{2}X = \tilde{Q}X \Leftrightarrow Y = (\tilde{Q} + \frac{L+\mu}{2}I)X = QX. \tag{4.42}$$

□

4.2.5 Proof of Theorem 4.6

We now prove Theorem 4.6 on the interpolation conditions for $\mathcal{L}_{\mu,L}$ of linear operators with singular values between μ and L. The idea is to exploit the polar decomposition (see, e.g., [HJ12, Theorem 7.3.1]) to express a linear operator with singular values between μ and L as the product between a semi-unitary transformation and a symmetric positive semidefinite linear operator with eigenvalues between μ and L. We already have interpolation conditions for symmetric linear operator (Theorem 4.5). We develop interpolation conditions for semi-unitary transformation by exploiting interpolation conditions for unitary transformations, which we also already have [HJ12, Theorem 7.3.11]. The interpolation conditions for the class of unitary transformations are not new but we propose a new formulation based on Definition 4.1.

Theorem 4.12 (Unitary-interpolation conditions, Theorem 7.3.11 of [HJ12])**.** *Let $X \in \mathbb{R}^{n\times N_1}$, $Y \in \mathbb{R}^{m\times N_1}$, $U \in \mathbb{R}^{m\times N_2}$, $V \in \mathbb{R}^{n\times N_2}$.*

(X, Y, U, V) is unitary-interpolable if and only if

$$\begin{pmatrix} Y & U \end{pmatrix}^T \begin{pmatrix} Y & U \end{pmatrix} = \begin{pmatrix} X & V \end{pmatrix}^T \begin{pmatrix} X & V \end{pmatrix}. \tag{4.43}$$

Using these interpolation conditions, we develop interpolation conditions for semi-unitary matrices, i.e., matrices of size $m \times n$ with $m \geq n$ and orthonormal columns.

Theorem 4.13 (Semi-unitary-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 $m \geq n$.*

(X, Y, U, V) is semi-unitary-interpolable if and only if

$$\begin{cases} Y^T Y = X^T X, \\ Y^T U = X^T V, \\ U^T U \succeq V^T V. \end{cases} \tag{4.44}$$

Proof. Given a semi-unitary matrix $\bar{R} \in \mathbb{R}^{m\times n}$ with $m \geq n$, we can find a unitary matrix $R \in \mathbb{R}^{m\times m}$ such that $\bar{R} = R \begin{pmatrix} I_n \\ 0_{m-n,n} \end{pmatrix}$ (see, e.g., [HJ12, Theorem 2.1.18]). Therefore, (X, Y, U, V) is semi-unitary-interpolable if and only if

$$\exists \bar{R} \text{ semi-unitary}: \quad Y = \bar{R}X \text{ and } V = \bar{R}^T U, \tag{4.45}$$

$$\Leftrightarrow \exists R \text{ unitary}: \quad Y = R\begin{pmatrix} I \\ 0 \end{pmatrix} X \text{ and } V = \begin{pmatrix} I & 0 \end{pmatrix} R^T U, \tag{4.46}$$

$$\Leftrightarrow \exists R \text{ unitary}, Z: \quad Y = R\begin{pmatrix} X \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} V \\ Z \end{pmatrix} = R^T U. \tag{4.47}$$

Theorem 4.12 on interpolation conditions for unitary matrices yields

$$\exists Z : \begin{pmatrix} Y & U \end{pmatrix}^T \begin{pmatrix} Y & U \end{pmatrix} = \left(\begin{pmatrix} X \\ 0 \end{pmatrix} \quad \begin{pmatrix} V \\ Z \end{pmatrix} \right)^T \left(\begin{pmatrix} X \\ 0 \end{pmatrix} \quad \begin{pmatrix} V \\ Z \end{pmatrix} \right), \tag{4.48}$$

or equivalently

$$\begin{cases} Y^T Y = X^T X, \\ Y^T U = X^T V, \\ U^T U \succeq V^T V. \end{cases} \tag{4.49}$$

□

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

Theorem 4.14 (Polar decomposition, Theorem 7.3.1 of [HJ12])**.** *Let M be a matrix of size $m \times n$ with $m \geq n$. We have*

$$M = \bar{R}Q \tag{4.50}$$

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

We now formally prove Theorem 4.6.

Proof of Theorem 4.6. Consider the polar decomposition (Theorem 4.14) of the matrix $M = \bar{R}Q$ where $Q \in \mathcal{S}_{\mu,L}$, and $\bar{R}$ is a semi-unitary matrix with orthonormal columns. We have

$$\exists M \in \mathcal{L}_{\mu,L} : \begin{cases} Y = MX, \\ V = M^T U, \end{cases}$$

$$\overset{\text{Th. 4.14}}{\Leftrightarrow} \exists \bar{R} \text{ semi-unitary}, Q \in \mathcal{S}_{\mu,L} : \begin{cases} Y = \bar{R}QX, \\ V = Q\bar{R}^T U, \end{cases}$$

$$\Leftrightarrow \exists \bar{R} \text{ semi-unitary}, Q \in \mathcal{S}_{\mu,L}, Z, W : \begin{cases} Y = \bar{R}Z, \\ V = QW, \\ Z = QX, \\ W = \bar{R}^T U, \end{cases}$$

$$\Leftrightarrow \exists Z, W : \begin{cases} (Z, Y, U, W) \text{ is semi-unitary-interpolable,} \\ ((X\ W), (Z\ V)) \text{ is } \mathcal{S}_{\mu,L}\text{-interpolable.} \end{cases}$$

Theorems 4.5 and 4.13 on the interpolation conditions of semi-unitary and symmetric matrices yields the result, namely,

$$\exists Z, W \begin{cases} Y^T Y = Z^T Z, \\ Y^T U = Z^T W, \\ U^T U \succeq W^T W, \\ \begin{pmatrix} X & W \end{pmatrix}^T \begin{pmatrix} Z & V \end{pmatrix} = \begin{pmatrix} Z & V \end{pmatrix}^T \begin{pmatrix} X & W \end{pmatrix}, \\ \left(\begin{pmatrix} Z & V \end{pmatrix} - \mu \begin{pmatrix} X & W \end{pmatrix}\right)^T \left(L \begin{pmatrix} X & W \end{pmatrix} - \begin{pmatrix} Z & V \end{pmatrix}\right) \succeq 0. \end{cases} \tag{4.51}$$

When $m = n$, $\bar{R}$ is unitary and we can use Theorem 4.12, therefore, the third inequality becomes the equality $U^T U = W^T W$. □

Remark 4.15. Theorem 4.6 provides interpolation conditions of the quadruplet (X, Y, U, V) by a "tall" matrix $M \in \mathbb{R}^{m \times n}$ with $m \geq n$. If we want the interpolation by a "flat" matrix $M \in \mathbb{R}^{m \times n}$ with $n \geq m$, then we should apply the interpolation conditions to the quadruplet (U, V, X, Y).

4.3 Union of spectrums

In this section, we propose an expression of interpolation conditions for symmetric linear operators whose eigenvalues belong to a union of subsets of $\mathbb{R}$. We only present the symmetric case for simplicity, but the skew-symmetric and general cases follow the same reasoning. The idea is to sum the Gram matrices associated with each subset.

Definition 4.16 (Gram matrix associated with a symmetric linear operator)**.** Let a symmetric matrix Q. We say that G is associated with Q if and only if

$$\exists X, Y : \begin{cases} G = \text{Gram}\begin{pmatrix} X & Y \end{pmatrix} \triangleq \begin{pmatrix} X & Y \end{pmatrix}^T \begin{pmatrix} X & Y \end{pmatrix}, \\ Y = QX. \end{cases} \tag{4.52}$$

Given a subset $S \subseteq \mathbb{R}$, we note $\mathcal{G}_{\Delta(S)}$ the set of Gram matrices associated to a symmetric matrix Q whose eigenvalues $\lambda_i(Q)$ all belong to S, namely,

$$\mathcal{G}_{\Delta(S)} := \left\{ G = \text{Gram}\begin{pmatrix} X & Y \end{pmatrix} : \exists Q \text{ symmetric} : Y = QX,\ \lambda_i(Q) \in S,\ \forall i \right\}. \quad (4.53)$$

The following theorem describes the set of Gram matrices associated with symmetric matrices with eigenvalues in a union of subsets, $\mathcal{G}_{\Delta(S_1 \cup S_2)}$, via the sets of Gram matrices associated with symmetric matrices with eigenvalues in these subsets $\mathcal{G}_{\Delta(S_1)}$ and $\mathcal{G}_{\Delta(S_2)}$.

Theorem 4.17. *Let $S_1, S_2 \subseteq \mathbb{R}$. $G \in \mathcal{G}_{\Delta(S_1 \cup S_2)}$ if and only if $G = G_1 + G_2$ for some $G_i \in \mathcal{G}_{\Delta(S_i)}$.*

Proof. Let $G \in \mathcal{G}_{\Delta(S_1 \cup S_2)}$. Therefore, G is associated with some matrix Q with eigenvalues in $S_1 \cup S_2$ (Definition 4.16). Using Lemma 5.6.3 (or Lemma 5.6.4 in the nonsymmetric case), we can choose Q with the following form

$$Q = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \quad (4.54)$$

where the eigenvalues of Q_1 and Q_2 respectively belong to S_1 and S_2. Therefore, we have

$$\begin{aligned} G \in \mathcal{G}_{\Delta(S_1 \cup S_2)} &\Leftrightarrow G = \text{Gram}\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \right), \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \\ &\Leftrightarrow G = \text{Gram}\begin{pmatrix} X_1 & Y_1 \end{pmatrix} + \text{Gram}\begin{pmatrix} X_2 & Y_2 \end{pmatrix}, Y_i = Q_i X_i, \\ &\Leftrightarrow G = G_1 + G_2 \text{ for some } G_i \in \mathcal{G}_{\Delta(S_i)}. \end{aligned}$$

In other words, the sum of Gram matrices associated with linear operators Q_1 and Q_2 is associated with linear operator $\begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}$. □

We now develop interpolation conditions for (X, Y) from the characterization of their Gram matrix $G = \text{Gram}\begin{pmatrix} X & Y \end{pmatrix}$. By definition, when a Gram matrix G belongs to $\mathcal{G}_{\Delta(T)}$, there exists a factorization X' and Y' of G such that $G = \text{Gram}\begin{pmatrix} X' & Y' \end{pmatrix}$ and $Y' = Q'X'$ with eigenvalues of Q in T and we say that (X', Y') is $\mathcal{S}_T$-interpolable. In addition, the following theorem shows that all factorizations (X, Y) of G are also $\mathcal{S}_T$-interpolable.

Theorem 4.18. *Let $T \subseteq \mathbb{R}$. (X, Y) is $\mathcal{S}_T$-interpolable if and only if* $\text{Gram}(X\,Y) \in \mathcal{G}_{\Delta(T)}$.

Proof. The reasoning is similar to Lemma 4.11 and is reproduced here for completeness.
(Necessity) Let (X, Y) being $\mathcal{S}_T$-interpolable, then $\exists Q$ with eigenvalues in T such that $Y = QX$ and $G = \text{Gram}\begin{pmatrix} X & Y \end{pmatrix} \in \mathcal{G}_{\Delta(T)}$.
(Sufficiency) Let $G \in \mathcal{G}_{\Delta(T)}$, then

$$\exists X', Y', Q' \text{ such that } Y' = Q'X', G = \text{Gram}\begin{pmatrix} X' & Y' \end{pmatrix}, \tag{4.55}$$

where the eigenvalues of Q' are in T. Therefore, (X, Y) and (X', Y') build the same Gram matrix in the sense that $G = \text{Gram}(X\ Y) = \text{Gram}(X'\ Y')$. In order to apply Theorem A.8, we add zeros to (X, Y) or (X', Y') such that they have the same number of rows, namely,

$$G = \begin{pmatrix} X & Y \\ 0 & 0 \end{pmatrix}^T \begin{pmatrix} X & Y \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} X' & Y' \\ 0 & 0 \end{pmatrix}^T \begin{pmatrix} X' & Y' \\ 0 & 0 \end{pmatrix}. \tag{4.56}$$

Note that adding such rows of zeros does not modify the Gram matrix G. Theorem A.8 yields

$$\begin{pmatrix} X' & Y' \\ 0 & 0 \end{pmatrix} = V \begin{pmatrix} X & Y \\ 0 & 0 \end{pmatrix} \tag{4.57}$$

for some unitary matrix V. We can augment the equality $Y' = M'X'$ to

$$\begin{pmatrix} Y' \\ 0 \end{pmatrix} = \begin{pmatrix} Q' & 0 \\ 0 & \bar{Q}' \end{pmatrix} \begin{pmatrix} X' \\ 0 \end{pmatrix} \tag{4.58}$$

where $\bar{Q}'$ is a matrix with eigenvalues in T. Exploiting (4.55), (4.57), and (4.58) yields

$$\exists X', Y', Q' : Y' = Q'X' \overset{(4.58)}{\Leftrightarrow} \exists X', Y', Q' : \begin{pmatrix} Y' \\ 0 \end{pmatrix} = \begin{pmatrix} Q' & 0 \\ 0 & \bar{Q}' \end{pmatrix} \begin{pmatrix} X' \\ 0 \end{pmatrix} \tag{4.59}$$

$$\overset{(4.57)}{\Leftrightarrow} \exists X', Y', Q' : V \begin{pmatrix} Y \\ 0 \end{pmatrix} = \begin{pmatrix} Q' & 0 \\ 0 & Q \end{pmatrix} V \begin{pmatrix} X \\ 0 \end{pmatrix} \tag{4.60}$$

$$\Leftrightarrow \exists X', Y', Q' : Y = \overbrace{V^T \begin{pmatrix} Q' & 0 \\ 0 & \bar{Q}' \end{pmatrix} V}^{\triangleq Q} X \tag{4.61}$$

where Q has eigenvalues in T since Q' and $\bar{Q}'$ have and that unitary transformations preserve the spectrum of a symmetric matrix. □

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

Corollary 4.19. *Let $\mu \in \mathbb{R}$. $(X, Y) \in \mathcal{S}_{\{-\mu,\mu\}}$ if and only if*

$$\mathrm{Gram}\,(X, Y) = \begin{pmatrix} A & B \\ B & \mu^2 A \end{pmatrix} \tag{4.62}$$

for some $A \succeq 0$ and B.

Proof. By Theorems 4.17 and 4.18, we have $G = \mathrm{Gram}\,(X, Y) \in \mathcal{G}_{\Delta(\{\mu\}\cup\{-\mu\})}$ if and only if $G = G_1 + G_2$ with $G_1 \in \mathcal{G}_{\Delta(\{\mu\})}$ and $G_2 \in \mathcal{G}_{\Delta(\{-\mu\})}$. And, one can show that $G_i \in \mathcal{G}_{\Delta(\{\alpha\})}$ if and only if

$$G_i = \begin{pmatrix} A & \alpha A \\ \alpha A & \alpha^2 A \end{pmatrix}. \tag{4.63}$$

On one hand, suppose that $G \in \mathcal{G}_{\Delta(\{\mu\}\cup\{-\mu\})}$ and therefore that

$$G = G_1 + G_2 = \begin{pmatrix} A_1 & \mu A_1 \\ \mu A_1 & \mu^2 A_1 \end{pmatrix} + \begin{pmatrix} A_2 & -\mu A_2 \\ -\mu A_2 & \mu^2 A_2 \end{pmatrix} \tag{4.64}$$

$$= \begin{pmatrix} A_1 + A_2 & \mu(A_1 - A_2) \\ \mu(A_1 - A_2) & \mu^2(A_1 + A_2) \end{pmatrix}, \tag{4.65}$$

for some A_1 and A_2. Then, $G = \begin{pmatrix} A & B \\ B & \mu^2 A \end{pmatrix}$ with $A = A_1 + A_2$ and $B = \mu(A_1 - A_2)$.

On the other hand, if $G = \begin{pmatrix} A & B \\ B & \mu^2 A \end{pmatrix}$ for some A and B, then defining $A_1 = \frac{1}{2}\left(A + \frac{B}{\mu}\right)$ and $A_2 = \frac{1}{2}\left(A - \frac{B}{\mu}\right)$, we have $G = G_1 + G_2$ with $G_1 = \begin{pmatrix} A_1 & \mu A_1 \\ \mu A_1 & \mu^2 A_1 \end{pmatrix}$ and $G_2 = \begin{pmatrix} A_2 & -\mu A_2 \\ -\mu A_2 & \mu^2 A_2 \end{pmatrix}$. □

We observe that there is a difference between the interpolation conditions for linear operators with eigenvalues in $[-\mu, \mu]$ (Theorem 4.5) and in $\{-\mu, \mu\}$ (Theorem 4.19).

4.4 Limiting cases

Theorems 4.3 and 4.5 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 (4.5), 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 $\mathcal{L}$ as the class of matrices with arbitrary real singular values and propose the following $\mathcal{L}$-interpolation conditions.

Theorem 4.20 ($\mathcal{L}$-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*

$$\exists L > 0 \ : \ \begin{cases} X^TV = Y^TU, \\ \begin{pmatrix} X^TX & Y^TY \\ Y^TY & L^2I \end{pmatrix} \succeq 0, \\ \begin{pmatrix} U^TU & V^TV \\ V^TV & L^2I \end{pmatrix} \succeq 0. \end{cases} \tag{4.66}$$

Proof. By Theorem 4.3, (X, Y, U, V) is $\mathcal{L}$-interpolable if and only if

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

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

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

Remark 4.21. Existence of L is not an issue for the PEP framework since we can add a scalar variable to represent L^2. Conditions (4.66) of Theorem 4.20 are thus convex on the Gram matrices $(X\ V)^T(X\ V)$, $(Y\ U)^T(Y\ U)$ and L^2.

5

Limitations of representations of linear operators

The results of this chapter were obtained in collaboration with Zhicheng Deng.

In this chapter, we present a limitation on classes of linear operators that can be represented in a convex way. Most of the existing PEP methodology works with convex constraints on the scalar products between the points or equivalently on the Gram matrices to be described. In the general case, x_i, v_i lie in one space and y_i, u_i lie in a different space, therefore, we have access to the pair of Gram matrices

$$\begin{aligned} (G, H) &= \left(\text{Gram}\begin{pmatrix} X & V \end{pmatrix}, \text{Gram}\begin{pmatrix} Y & U \end{pmatrix}\right) \\ &\triangleq \left(\begin{pmatrix} X & V \end{pmatrix}^T \begin{pmatrix} X & V \end{pmatrix}, \begin{pmatrix} Y & U \end{pmatrix}^T \begin{pmatrix} Y & U \end{pmatrix}\right). \end{aligned} \tag{5.1}$$

In the (square) symmetric case, all vectors x_i, y_i, u_j, v_j lie in the same space, therefore, we have access to the single Gram matrix

$$G = \text{Gram}\begin{pmatrix} X & Y & U & V \end{pmatrix} \triangleq \begin{pmatrix} X & Y & U & V \end{pmatrix}^T \begin{pmatrix} X & Y & U & V \end{pmatrix}. \tag{5.2}$$

We show that through convex constraints on the Gram matrices, we can only represent classes of linear operators and symmetric linear operators whose all singular values or eigenvalues belong to a subset of $\mathbb{R}$, i.e., classes of the form of

$$\mathcal{L}_S = \{M : \sigma_i(M) \in S,\ \forall i\}, \tag{5.3}$$
$$\mathcal{S}_S = \{M \text{ is symmetric } : \lambda_i(M) \in S,\ \forall i\}. \tag{5.4}$$

Therefore, analyzing optimization methods with convex formulations of PEP through Gram matrices is only possible for these classes of linear operators.

Remark 5.1. For instance, it is not possible to require a specific structure for the linear operators (triangular, diagonal, sparse, etc) nor to select the multiplicity of the elements of the spectrum.

Remark 5.2. In the context of the Performance Estimation Problem, in addition to the constraints characterizing the linear operators classes, there are also other constraints on the Gram matrices (due to the method analyzed, for example). Therefore, the final and actual subset of Gram matrices could be convex, even though the set of Gram matrices associated with the linear operators is non-convex.

More precisely, we show that spaces of Gram matrices associated with such classes of linear operators are always convex (Theorem 5.9). Moreover, if a set of Gram matrices associated with a set of (resp. symmetric) linear operators is convex, then it is possible to describe this set of (resp. symmetric) linear operators as linear operators with singular values (resp. eigenvalues) in a subset of $\mathbb{R}$.

For the symmetric case, the result also holds for the broader class of normal matrices with complex eigenvalues. Therefore, we will prove the result for normal matrices. Recall that a symmetric matrix is a normal matrix with real eigenvalues and conversely.

We denote $\mathcal{L}$ and $\mathcal{S}$ the sets of all linear operators and normal linear operators of any finite dimensions.

5.1 Gram matrices associated with classes of linear operators

We define the concept of Gram matrices associated with a linear operator and with a class of linear operators.

Definition 5.3 (Gram matrix associated with a linear operator)**.** Given a matrix G (resp. a pair of Gram matrices (G, H)), we say that G (resp. (G, H)) is associated with M if

$$\exists X, Y, U, V : \begin{cases} G = \mathrm{Gram}\begin{pmatrix} X & Y & U & V \end{pmatrix}, \\ (\text{resp. } (G, H) = \left(\mathrm{Gram}\begin{pmatrix} X & V \end{pmatrix}, \mathrm{Gram}\begin{pmatrix} Y & U \end{pmatrix}\right)), \\ Y = MX, \\ V = M^T U. \end{cases}$$

Definition 5.4 (Set of Gram matrices associated with a class of linear operator)**.** Let classes of linear operators $\mathcal{M} \subseteq \mathcal{L}$ and square normal linear operators $\mathcal{Q} \subseteq \mathcal{S}$, the set of Gram matrices associated with $\mathcal{M}$ and $\mathcal{Q}$ are

$$\mathcal{P}^N_{\mathcal{M}} := \{(G, H) = (\mathrm{Gram}(X\ V), \mathrm{Gram}(Y\ U)) \quad : \exists M \in \mathcal{M} : Y = MX,\ V = M^T U\},$$
$$\mathcal{G}^N_{\mathcal{Q}} := \{G = \mathrm{Gram}(X\ Y\ U\ V) \quad : \exists M \in \mathcal{Q} : Y = MX,\ V = M^T U\}.$$

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

Recall a classical result on Gram matrices and semi-unitary matrices. We call a matrix semi-unitary if it has orthonormal columns.

Theorem 5.5 (Theorem 7.3.11 of [HJ12])**.** *Let $A \in \mathbb{R}^{n \times N}$ and $B \in \mathbb{R}^{m \times N}$ with $m \geq n$. We have $A^T A = B^T B$ if and only if there is a semi-unitary matrix $\bar{R} \in \mathbb{R}^{m \times n}$ such that $B = \bar{R} A$.*

By Theorem 5.5, Gram matrices associated with a linear operator (resp. normal linear operator) M are also associated with all linear operators of the form $\bar{R} M \bar{S}^T$ (resp. $\bar{R} M \bar{R}^T$).

Lemma 5.6 (Unitary invariance of Gram matrices)**.**

1. *If a Gram matrix G is associated with a normal linear operator M then it is also associated with all linear operators $\bar{R} M \bar{R}^T$ for any semi-unitary transformation $\bar{R}$.*

2. *If a pair of Gram matrices (G, H) is associated with a linear operator M then it is also associated with all linear operators $\bar{R} M \bar{S}^T$ for any semi-unitary transformations $\bar{R}$ and $\bar{S}$.*

3. *A Gram matrix G is associated with a normal linear operator M if and only if G associated with a linear operator RMR^T for any unitary transformation R.*

4. *A pair of Gram matrices (G, H) is associated with a linear operator M if and only if G is associated with a linear operator RMS^T for any unitary transformations R and S.*

Proof. We have

$$G = \text{Gram}\begin{pmatrix}X & Y & U & V\end{pmatrix} \overset{\text{Th. 5.5}}{=} \text{Gram}(\bar{R}\begin{pmatrix}X & Y & U & V\end{pmatrix}) \tag{5.5}$$

$$= \text{Gram}\begin{pmatrix}X' & Y' & U' & V'\end{pmatrix} \tag{5.6}$$

where $Y' = \bar{R}M\bar{R}^T X'$, $V' = \bar{R}^T M^T \bar{R} U'$ for all semi-unitary transformations $\bar{R}$. And we have

$$\begin{cases} G = \text{Gram}\begin{pmatrix}X & V\end{pmatrix} \overset{\text{Th. 5.5}}{=} \text{Gram}\left(\bar{S}\begin{pmatrix}X & V\end{pmatrix}\right) = \text{Gram}\begin{pmatrix}X' & V'\end{pmatrix} \\ H = \text{Gram}\begin{pmatrix}Y & U\end{pmatrix} \overset{\text{Th. 5.5}}{=} \text{Gram}\left(\bar{R}\begin{pmatrix}Y & U\end{pmatrix}\right) = \text{Gram}\begin{pmatrix}Y' & U'\end{pmatrix} \end{cases} \tag{5.7}$$

where $Y' = \bar{R}M\bar{S}^T X'$, $V' = \bar{S}M^T\bar{R}^T U'$ for all semi-unitary transformations R and S, which proves statements 1 and 2. Since the inverse of a unitary matrix is its transpose, statements 1 and 2 imply statements 3 and 4. □

Therefore, Gram matrices associated with a general (resp. normal) linear operator M are also associated with the diagonalized version $M' = RMS^T$ (resp. $M' = RMR^T$) whose diagonal elements are the singular values of M (resp. the eigenvalues of normal linear operator M) in any order. This implies that the Gram representation only allows characterizing the spectrum of linear operators M, namely, the values and multiplicity of each of their singular values (or eigenvalues for normal linear operators), in a sense that will be made clear in Corollary 5.8.

We formally define spectrums using multisets, which are sets where each element has a multiplicity. We use the usual inclusion, sum, and difference of multisets.

Definition 5.7 (Spectrum of matrices)**.** Let a matrix $M \in \mathcal{L}$ and a square normal matrix $Q \in \mathcal{S}$. The singular value spectrum of M is the following multiset of $\mathbb{R}$

$$\sigma(M) := \{\{\text{singular values of } M\}\} \tag{5.8}$$

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

$$\lambda(Q) := \{\{\text{eigenvalues of } Q\}\}. \tag{5.9}$$

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

$$\Delta(S) := \{\lambda : \forall a \in \lambda, a \in S\}. \tag{5.10}$$

λ is a multiset that considers the multiplicity of singular values and eigenvalues. We use a multiset and not a sequence since matrices with the same singular values or eigenvalues but not in the same order are the same up to a unitary transformation (and have thus the same Gram matrix by Lemma 5.6).

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

$$\mathcal{P}_\Lambda := \left\{(G, H) = (\text{Gram}(X\ V), \text{Gram}(Y\ U)) : \exists M : Y = MX, V = M^T U, \sigma(M) \in \Lambda\right\} \tag{5.11}$$

and similarly, for normal linear operators, we have

$$\mathcal{G}_\Lambda := \left\{G = \text{Gram}(X\ Y\ U\ V) : \exists M \text{ normal} : Y = MX, V = M^T U, \lambda(M) \in \Lambda\right\} \tag{5.12}$$

Due to Lemma 5.6 and the following corollary, we focus on spaces of Gram matrices associated with a set of singular or eigenvalue spectrum Λ.

Corollary 5.8 (Spectrum characterization). *Let $\mathcal{M}$ be a set of linear operators (resp. normal linear operators) of any finite dimension. We have $\mathcal{P}_\mathcal{M} = \mathcal{P}_\Lambda$ (resp. $\mathcal{G}_\mathcal{M} = \mathcal{G}_\Lambda$) where $\Lambda = \{\sigma(M) : M \in \mathcal{M}\}$ (resp. $\Lambda = \{\lambda(M) : M \in \mathcal{M}\}$).*

Proof. Thanks to Lemmas 5.6, a pair of Gram matrix (G, H) (resp. G) is also associated to its diagonalized version (and reciprocally), therefore, $(G, H) \in \mathcal{P}_\mathcal{M} \Leftrightarrow (G, H) \in \mathcal{P}_\Lambda$ (resp. $G \in \mathcal{G}_\mathcal{M} \Leftrightarrow G \in \mathcal{G}_\Lambda$). □

5.2 Main results

We can now show the two main results of the section, namely, (i) sets of Gram matrices, $\mathcal{P}_{\Delta(S)}$ and $\mathcal{G}_{\Delta(T)}$ associated with linear operators with all singular values in $S \subseteq \mathbb{R}$ and with normal linear operators with all eigenvalues in $T \subseteq \mathbb{C}$ are convex (Theorem 5.9), and, (ii) if $\mathcal{P}_{\mathcal{M}}$ or $\mathcal{G}_{\mathcal{M}}$ are convex, then they are also associated with classes of linear operators with singular values in some subset $S \subseteq \mathbb{R}$ and normal linear operators with eigenvalues in some subset $T \subseteq \mathbb{C}$ (Theorem 5.10).

Theorem 5.9. *Let $S \subseteq \mathbb{R}$ and $T \subseteq \mathbb{C}$. $\mathcal{P}_{\Delta(S)}$ and $\mathcal{G}_{\Delta(T)}$ are convex.*

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

Let $(G_i, H_i) \in \mathcal{P}_{\Delta(S)}$ for $i = 1, 2$. We have

$$(G_i, H_i) = \left(\text{Gram}\begin{pmatrix} X_i & V_i \end{pmatrix}, \text{Gram}\begin{pmatrix} Y_i & U_i \end{pmatrix}\right) \tag{5.13}$$

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

Similarly, in the square normal case, let $G_i \in \mathcal{G}_{\Delta(T)}$ for $i = 1, 2$. We have

$$G_i = \text{Gram}\begin{pmatrix} X_i & Y_i & U_i & V_i \end{pmatrix} \tag{5.14}$$

with $Y_i = M_i X_i, V_i = M_i^T U_i$ and $\lambda(M_i) \in \Delta(T)$. Let $(X\ Y\ U\ V) \triangleq \begin{pmatrix} \sqrt{c_1}(X_1 & Y_1 & U_1 & V_1) \\ \sqrt{c_2}(X_2 & Y_2 & U_2 & V_2) \end{pmatrix}$ and $M \triangleq \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}$. One can directly verify that $c_1 G_1 + c_2 G_2 = \text{Gram}\begin{pmatrix} X & Y & U & V \end{pmatrix}$, $Y = MX$, $V = M^T U$, $\lambda(M) \in \Delta(T)$, and therefore $c_1 G_1 + c_2 G_2 \in \mathcal{G}_{\Delta(T)}$. □

We introduce some remarks and technical lemmas before the proof of the following theorem.

Theorem 5.10. *Let $\mathcal{M}$ be a set of linear operators (resp. normal linear opera-*

tors). If $\mathcal{P}_{\mathcal{M}}$ (resp. $\mathcal{G}_{\mathcal{M}}$) is convex, then

$$\mathcal{P}_{\mathcal{M}} = \mathcal{P}_{\Delta(S)} \text{ (resp. } \mathcal{G}_{\mathcal{M}} = \mathcal{G}_{\Delta(T)}) \tag{5.15}$$

for some $S \subseteq \mathbb{R}$ (resp. $T \in \mathbb{C}$).

Remark 5.11. Observe that it is not possible to control the multiplicity of the singular values or eigenvalues contained in the spectrums of linear operators represented by $\mathcal{P}_{\Delta(S)}$ and $\mathcal{G}_{\Delta(T)}$.

Lemma 5.12. *Let a set of spectrum Λ. $\mathcal{P}_\Lambda$ and $\mathcal{G}_\Lambda$ are cones, namely,*

$$c(G,H) \in \mathcal{P}_\Lambda \qquad \forall c \geq 0, \forall (G,H) \in \mathcal{P}_\Lambda \tag{5.16}$$

$$cG \in \mathcal{G}_\Lambda \qquad \forall c \geq 0, \forall G \in \mathcal{G}_\Lambda. \tag{5.17}$$

Proof. For $c > 0$, we have

$$(G,H) \in \mathcal{P}_\Lambda \Rightarrow \exists X,U,M : (G,H) = \left(\text{Gram}\begin{pmatrix}X & M^T U\end{pmatrix}, \text{Gram}\begin{pmatrix}MX & U\end{pmatrix}\right), \tag{5.18}$$

$$\Rightarrow \exists X,U,M : c(G,H) = \left(\text{Gram}(\sqrt{c}X\; M^T\sqrt{c}U), \text{Gram}(\sqrt{c}MX\; \sqrt{c}U)\right), \tag{5.19}$$

$$\Rightarrow c(G,H) \in \mathcal{G}_\Lambda, \tag{5.20}$$

where $\sigma(M) \in \Lambda$. And, we have

$$\begin{aligned} G \in \mathcal{G}_\Lambda &\Rightarrow \exists X,U,\text{normal } M : G = \text{Gram}\begin{pmatrix}X & MX & U & M^T U\end{pmatrix}, \\ &\Rightarrow \exists X,U,\text{normal } M : cG = \text{Gram}\begin{pmatrix}\sqrt{c}X & M\sqrt{c}X & \sqrt{c}U & M^T\sqrt{c}U\end{pmatrix}, \\ &\Rightarrow cG \in \mathcal{G}_\Lambda, \end{aligned}$$

where $\lambda(M) \in \Lambda$. $\square$

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

Lemma 5.13 (Extension of spectrums). *Let λ_1 and λ_2 be two spectrums . If $\lambda_1 \subseteq \lambda_2$ then $\mathcal{P}_{\{\lambda_1\}} \subseteq \mathcal{P}_{\{\lambda_2\}}$ and $\mathcal{G}_{\{\lambda_1\}} \subseteq \mathcal{G}_{\{\lambda_2\}}$.*

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

$$G = \text{Gram}\begin{pmatrix} X & Y & U & V \end{pmatrix} \text{ (resp. } (G,H) = \left(\text{Gram}\begin{pmatrix} X & V \end{pmatrix}, \text{Gram}\begin{pmatrix} Y & U \end{pmatrix}\right))$$

$$\text{with } \begin{cases} Y &= M_1 X \\ V &= M_1^T U \end{cases} \text{ and } \lambda(M_1) = \lambda_1 \text{ (resp. } \sigma(M_1) = \lambda_1)$$

$$\Rightarrow G = \text{Gram}\left(\begin{pmatrix} X \\ 0 \end{pmatrix} \quad \begin{pmatrix} Y \\ 0 \end{pmatrix} \quad \begin{pmatrix} U \\ 0 \end{pmatrix} \quad \begin{pmatrix} V \\ 0 \end{pmatrix}\right)$$

$$\text{(resp. } (G,H) = \left(\text{Gram}\left(\begin{pmatrix} X \\ 0 \end{pmatrix} \quad \begin{pmatrix} V \\ 0 \end{pmatrix}\right), \text{Gram}\left(\begin{pmatrix} Y \\ 0 \end{pmatrix} \quad \begin{pmatrix} U \\ 0 \end{pmatrix}\right)\right))$$

$$\text{with } \begin{cases} \begin{pmatrix} Y \\ 0 \end{pmatrix} = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \begin{pmatrix} X \\ 0 \end{pmatrix} \\ \begin{pmatrix} V \\ 0 \end{pmatrix} = \begin{pmatrix} M_1^T & 0 \\ 0 & M_2^T \end{pmatrix} \begin{pmatrix} U \\ 0 \end{pmatrix} \end{cases}$$

$$\text{and } \begin{cases} \lambda(M_1) = \lambda_1 & \text{(resp. } \sigma(M_1) = \lambda_1) \\ \lambda(M_2) = \lambda_2 - \lambda_1 & \text{(resp. } \sigma(M_2) = \lambda_2 - \lambda_1) \end{cases}$$

$$\Rightarrow G \in \mathcal{G}_{\{\lambda_1 + (\lambda_2 - \lambda_1)\}} = \mathcal{G}_{\{\lambda_2\}} \text{ (resp. } (G,H) \in \mathcal{P}_{\{\lambda_1 + (\lambda_2 - \lambda_1)\}} = \mathcal{P}_{\{\lambda_2\}})$$

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

This property also holds for sets of spectrums.

Corollary 5.14. *Let Λ_1 and Λ_2 be sets of spectrums. If for every $\lambda_1 \in \Lambda_1$ there exists $\lambda_2 \in \Lambda_2$ such that $\lambda_1 \subseteq \lambda_2$ then $\mathcal{P}_{\Lambda_1} \subseteq \mathcal{P}_{\Lambda_2}$ and $\mathcal{G}_{\Lambda_1} \subseteq \mathcal{G}_{\Lambda_2}$.*

All symmetric matrices are normal matrices. It occurs that our proof works for the larger class of normal matrices. Therefore, we prove the result for the class of normal matrices.

Proof of Theorem 5.10. By Corollary 5.8, we have $\mathcal{P}_{\mathcal{M}} = \mathcal{P}_{\Lambda}$ (resp. $\mathcal{G}_{\mathcal{M}} = \mathcal{G}_{\Lambda}$) where $\Lambda = \{\sigma(M) : M \in \mathcal{M}\}$ (resp. $\Lambda = \{\lambda(M) : M \in \mathcal{M}\}$). We show that S (resp. T) is the set of all values appearing in the spectrums of Λ, namely, $S_\Lambda \triangleq \{\mu : \exists \lambda \in \Lambda \text{ s.t. } \mu \in \lambda\}$. Recall that

$$\Delta(S_\Lambda) = \{\lambda : \forall a \in \lambda,\ a \in S_\Lambda\} = \{\lambda : \forall a \in \lambda,\ a \in \lambda'_a,\ \lambda'_a \in \Lambda\} \quad (5.21)$$

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

(*Inclusions* $\mathcal{P}_\Lambda \subseteq \mathcal{P}_{\Delta(S_\Lambda)}$ *and* $\mathcal{G}_\Lambda \subseteq \mathcal{G}_{\Delta(S_\Lambda)}$) Let a spectrum $\lambda \in \Lambda$, then all the elements of this spectrum belong to some spectrum of Λ and therefore $\lambda \in \Delta(S_\Lambda)$. Thus, we have $\Lambda \subseteq \Delta(S_\Lambda)$ and therefore $\mathcal{P}_\Lambda \subseteq \mathcal{P}_{\Delta(S_\Lambda)}$ and $\mathcal{G}_\Lambda \subseteq \mathcal{G}_{\Delta(S_\Lambda)}$ by Corollary 5.14.

(*Inclusions* $\mathcal{P}_{\Delta(S_\Lambda)} \subseteq \mathcal{P}_\Lambda$ *and* $\mathcal{G}_{\Delta(S_\Lambda)} \subseteq \mathcal{G}_\Lambda$) First, we show that $\mathcal{G}_{\{\lambda\}} \subseteq \mathcal{G}_\Lambda$ (resp. $\mathcal{P}_{\{\lambda\}} \subseteq \mathcal{P}_\Lambda$) $\forall \lambda \in \Delta(S_\Lambda)$. Recall that $\mathcal{G}_{\{\lambda\}}$ (resp. $\mathcal{P}_{\{\lambda\}}$) is the set of Gram matrices associated with linear operators whose eigenvalue (resp. singular value) spectrum is exactly λ. Let $\lambda \in \Delta(S_\Lambda)$ and $G \in \mathcal{G}_{\{\lambda\}}$ (resp. $(G,H) \in \mathcal{P}_{\{\lambda\}}$). We have $G = \text{Gram}\begin{pmatrix}X & Y & U & V\end{pmatrix}$ (resp. $(G,H) = (\text{Gram}\begin{pmatrix}X & V\end{pmatrix}, \text{Gram}\begin{pmatrix}Y & U\end{pmatrix})$) with $Y = MX$, $V = M^TU$ and $a \in \lambda'_a$, $\lambda'_a \in \Lambda$, $\forall a \in \lambda(M)$ (resp. $\forall a \in \sigma(M)$). W.l.o.g., let M of dimension $m \times d$ with $d \leq m$ in the rectangle case and of dimension $d \times d$ in the normal case. By Lemma 5.6 and eigenvalue (resp. singular value) decomposition of M, we can choose M diagonal with its d eigenvalues (resp. singular values) m_i on its diagonal. Therefore, we have

$$(X,Y,U,V) = \left(\begin{pmatrix}X_1\\ \vdots\\ X_d\end{pmatrix}, \begin{pmatrix}V_1\\ \vdots\\ V_d\end{pmatrix}, \begin{pmatrix}Y_1\\ \vdots\\ Y_m\end{pmatrix}, \begin{pmatrix}U_1\\ \vdots\\ U_m\end{pmatrix}\right) \tag{5.22}$$

and

$$\begin{pmatrix}Y_1\\ \vdots\\ Y_m\end{pmatrix} = \begin{pmatrix}m_1 & & \\ & \ddots & \\ & & m_d\\ - & 0_{m-d\times d} & -\end{pmatrix}\begin{pmatrix}X_1\\ \vdots\\ X_d\end{pmatrix}, \quad \begin{pmatrix}V_1\\ \vdots\\ V_d\end{pmatrix} = \begin{pmatrix}m_1 & & & |\\ & \ddots & & 0_{d\times m-d}\\ & & m_d & |\end{pmatrix}\begin{pmatrix}U_1\\ \vdots\\ U_m\end{pmatrix},$$

$$\Leftrightarrow \begin{cases} Y_i = m_i X_i & \forall i = 1,\dots,d\\ V_i = m_i U_i & \forall i = 1,\dots,d\\ Y_i = 0 & \forall i = d+1,\dots,m, \end{cases}$$

$$\Leftrightarrow \begin{cases} \begin{pmatrix}Y_i\\ 0\end{pmatrix} = \overbrace{\begin{pmatrix}m_i & 0\\ 0 & \bar{M}_i\end{pmatrix}}^{M_i}\begin{pmatrix}X_i\\ 0\end{pmatrix},\\ \begin{pmatrix}V_i\\ 0\end{pmatrix} = \overbrace{\begin{pmatrix}m_i & 0\\ 0 & \bar{M}_i^T\end{pmatrix}}^{M_i^T}\begin{pmatrix}U_i\\ 0\end{pmatrix},\\ \lambda(M_i) = \lambda'_i \text{ (resp. } \sigma(M_i) = \lambda'_i), \end{cases}$$

where $m_i \in \lambda_i', \lambda_i' \in \Lambda$ and $\bar{M}_i$ completes the spectrum of M_i such that the spectrum of M_i is exactly λ_i'. Let

$$G_i = \text{Gram}\left(\begin{pmatrix} X_i \\ 0 \end{pmatrix} \quad \begin{pmatrix} Y_i \\ 0 \end{pmatrix} \quad \begin{pmatrix} U_i \\ 0 \end{pmatrix} \quad \begin{pmatrix} V_i \\ 0 \end{pmatrix}\right) \in \mathcal{G}_\Lambda$$

$$(resp.\ (G_i, H_i) = (\text{Gram}\left(\begin{pmatrix} X_i \\ 0 \end{pmatrix} \quad \begin{pmatrix} V_i \\ 0 \end{pmatrix}\right), \text{Gram}\left(\begin{pmatrix} Y_i \\ 0 \end{pmatrix} \quad \begin{pmatrix} U_i \\ 0 \end{pmatrix}\right) \in \mathcal{P}_\Lambda)$$

We have $G = \sum_{i=1}^d G_i = \sum_{i=1}^d \frac{1}{d}(dG_i)$ (resp. $(G, H) = \sum_{i=1}^d (G_i, H_i) = \sum_{i=1}^d \frac{1}{d}(d(G_i, H_i))$) since $Y_i = 0$ for $i = d+1, \ldots, m$. and $dG_i \in \mathcal{G}_\Lambda$ (resp. $d(G_i, H_i) \in \mathcal{P}_\Lambda$) by Lemma 5.12. And by convexity of $\mathcal{G}_\Lambda$ (resp. $\mathcal{P}_\Lambda$), we have $G = \sum_{i=1}^d \frac{1}{d}(dG_i) \in \mathcal{G}_\Lambda$ (resp. $(G, H) = \sum_{i=1}^d \frac{1}{d}(d(G_i, H_i)) \in \mathcal{P}_\Lambda$).
Finally, let $G \in \mathcal{G}_{\Delta(S_\Lambda)}$ (resp. $(G, H) \in \mathcal{P}_{\Delta(S_\Lambda)}$). By definition of $\Delta(S_\Lambda)$, there is a $\lambda \in \Delta(S_\Lambda)$ such that $G \in \mathcal{G}_{\{\lambda\}}$ (resp. $(G, H) \in \mathcal{P}_{\{\lambda\}}$) and therefore $G \in \mathcal{G}_\Lambda$ (resp. $(G, H) \in \mathcal{P}_\Lambda$). □

6

Quadratic functions

In Chapter 4, we obtained interpolation conditions for symmetric linear operators. We now show that it allows the development of interpolation conditions for the class of quadratic functions, and we will analyze these conditions. Afterwards, we investigate whether it is possible to estimate *a priori* the performance of a method on smooth convex functions from its performance on smooth convex quadratic functions. But before all that, we present a known technique to analyze the performance of first-order methods on quadratic functions.

6.1 Performance of first-order methods via spectral analysis

We analyze the worst-case performance of a fixed-step first-order method (FSFOM) on quadratic functions. We show that, since we can explicitly compute the gradient of a quadratic function $f(x) = \frac{1}{2}x^T Qx$, we can compute the worst-case performance as a univariate optimization problem.

Let an initial point x_0 and a general FSFOM of the form

$$x_i = x_0 - \sum_{k=0}^{i-1} h_{i,k} \nabla f(x_k), \qquad \forall i \in [N], \tag{6.1}$$

where $h_{i,k} \in \mathbb{R}$ is how much the gradient at iterate x_k is used to compute iterate x_i. When f is a quadratic function $f(x) = \frac{1}{2}x^T Qx$, we have

$$x_i = x_0 - \sum_{k=0}^{i-1} h_{i,k} Q x_k, \qquad \forall i \in [N]. \tag{6.2}$$

Developing the recurrence yields

$$\begin{aligned} x_1 =& x_0 - h_{1,0} Q x_0 = (I - h_{1,0} Q) x_0 \triangleq K_1(Q) x_0, \\ x_2 =& x_0 - h_{2,1} Q x_1 - h_{2,0} Q x_0 \\ =& (I - h_{2,1}(I - h_{1,0} Q) - h_{2,0} Q) x_0 \triangleq K_2(Q) x_0, \\ \vdots& \\ x_N =& x_0 - \sum_{k=0}^{N-1} h_{N,k} Q x_k \triangleq K_N(Q) x_0, \end{aligned} \tag{6.3}$$

where $K_i(Q)$ is a matrix polynomial "encoding" the iteration x_i. Therefore, the problem of computing the worst-case performance of the method encoded by K_N on the class of quadratic functions is

$$\begin{aligned} \max_{x_0, x_N, x^*, \mu I \preceq Q \preceq LI} \quad & \|x_N - x^*\|^2 \\ \text{s.t.} \quad & x_N = K_N(Q) x_0, \\ & \|x_0 - x^*\|^2 \leq 1, \\ & \|x^*\|^2 = 0. \end{aligned} \tag{6.4}$$

Eliminating variables x_N and x^* yields

$$\max_{\mu I \preceq Q \preceq L} \quad \max_{\|x_0\|^2 \leq 1} \quad \|K_N(Q) x_0\|^2 \tag{6.5}$$

or equivalently

$$\max_{\mu I \preceq Q \preceq L} \quad \|K_N(Q)\|^2. \tag{6.6}$$

And, using $\|K_N(Q)\| = \lambda_{\max}(K_N(Q)) = K_N(\lambda_{\max}(Q))$, the problem becomes the following optimization problem of a univariate polynomial in the interval $[\mu, L]$ that can be solved efficiently

$$\max_{\mu \leq \lambda \leq L} \quad K_N(\lambda)^2. \tag{6.7}$$

Conversely to the PEP framework, this approach does not rely on the interpolation conditions of quadratic functions but exploits the explicit form of quadratic functions. Moreover, this approach is more efficient as it requires minimizing a univariate polynomial of degree $\mathcal{O}(N)$ instead of solving a semidefinite program of size $\mathcal{O}(N^2)$. However, this approach is much less versatile as it only works for quadratic functions. This approach is not new, see [Nem95, Chapter 12], [dST⁺21, Section 2.1], and [Ped20] for recent reviews.

6.2 Interpolation conditions for quadratic functions

We now develop interpolation conditions for quadratic functions. We consider the class of homogeneous quadratic functions

$$\mathcal{Q}_{\mu,L} = \left\{ f(x) = \frac{1}{2} x^T Q x \text{ and } \mu I \preceq Q \preceq LI \right\}. \tag{6.8}$$

Theorem 4.5 allows to obtain interpolation conditions for $\mathcal{Q}_{\mu,L}$.

Theorem 6.1 ($\mathcal{Q}_{\mu,L}$-interpolation conditions)**.** *Let $-\infty < \mu \leq L < \infty$.*

The set of triplets $\{(x_i, g_i, f_i)\}_{i\in[N]}$ is $\mathcal{Q}_{\mu,L}$-interpolable if and only if

$$\begin{cases} X^T G = G^T X, \\ (G - \mu X)^T (LX - G) \succeq 0, \\ f_i = \frac{1}{2} x_i^T g_i, \quad \forall i \in [N], \end{cases} \tag{6.9}$$

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

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

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

$$\overset{\text{Th. 4.5}}{\Leftrightarrow} \begin{cases} X^T G = G^T X, \\ (G - \mu X)^T (LX - G) \succeq 0, \\ f_i = \frac{1}{2} x_i^T g_i, \quad \forall i \in [N]. \end{cases}$$

□

Since $\mathcal{Q}_{\mu,L} \subseteq \mathcal{F}_{\mu,L}$, the quadratic interpolation conditions of Theorem 6.1 must imply the general smooth strongly convex interpolation conditions of Theorem 2.3. Indeed, we can observe algebraically in the following proposition that conditions (6.9) imply conditions (2.2).

Proposition 6.2. *Let $-\infty < \mu \leq L < \infty$.*

If $\{(x_i, g_i, f_i)\}_{i\in[N]}$ is $\mathcal{Q}_{\mu,L}$-interpolable, then it is also $\mathcal{F}_{\mu,L}$-interpolable.

Proof. Conditions $(G-\mu X)^T(LX-G) \succeq 0$ and $X^TG = G^TX$ of Theorem 6.1 yield

$$\begin{aligned}
& -\mu L X^T X + (\mu+L)X^TG - G^TG && \succeq 0,\\
\Leftrightarrow\quad & z^T\big(-\mu L X^T X + (\mu+L)X^TG - G^TG\big)z && \geq 0\ \forall z\in\mathbb{R}^N,\\
\Leftrightarrow\quad & \sum_{k=1}^{N}\sum_{l=1}^{N} z_k z_l\big(-\mu L x_k^T x_l + (\mu+L)x_k^T g_l - g_k^T g_l\big) && \geq 0\ \forall z\in\mathbb{R}^N.
\end{aligned} \tag{6.10}$$

Choosing $z_i = 1, z_j = -1$ and $z_k = z_l = 0,\ \ \forall k, l \neq i, j$ yields

$$(x_i^T g_i + x_j^T g_j)(L+\mu) \geq \|g_i - g_j\|^2 + \mu L\,\|x_i - x_j\|^2 + 2(\mu+L)x_i^T g_j \tag{6.11}$$

Finally, using $f_i = \frac{x_i^T g_i}{2}$ and $x_i^T g_j = x_j^T g_i$, yields the interpolation conditions for $\mathcal{F}_{\mu,L}$ (Theorem 2.3). □

Remark 6.3. As mentioned earlier, the recent work [AdKZ23b] proposed necessary interpolation conditions (originally announced in [DS19]) 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 (2.2) in addition to the following necessary conditions for nonhomogeneous quadratic functions

$$\frac{1}{2}\left(g_i + g_j\right)^T\left(x_i - x_j\right) = f_i - f_j, \qquad \forall (i,j)\in[N]^2. \tag{6.12}$$

These conditions are not sufficient. Indeed, there exist sets of points satisfying both conditions (2.2) and (6.12) and that cannot be interpolated by a quadratic function. For example, the following set $\{(x_i, g_i, f_i)\}_{i\in[3]}$

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

satisfies conditions (2.2) and (6.12). 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).

Remark 6.4. When $\mu > 0$, the class of functions whose convex conjugates [Fen49] belong to $\mathcal{Q}_{\mu,L}$, and reciprocally, is $\mathcal{Q}_{\frac{1}{L},\frac{1}{\mu}}$. Moreover, the operation of conjugation can be seen as permuting the gradients g_i with the points x_i and we can observe the impact of this operation on the interpolation conditions (6.9). The first and third conditions remain unchanged, whereas the second one becomes

$$(X - \mu G)^T(LG - X) \succeq 0 \Leftrightarrow \left(G - \frac{X}{\mu}\right)\left(\frac{X}{L} - G\right) \succeq 0 \tag{6.13}$$

which is the second interpolation condition for the class $\mathcal{Q}_{\frac{1}{L},\frac{1}{\mu}}$.

6.2.1 Analysis of the conditions

We investigate the role of the interpolation conditions for quadratic functions of Theorem 6.1.

Condition $(G - \mu X)^T(LX - G) \succeq 0$ This condition can only be written when G^TX is symmetric. However, we can consider the following generalization of the condition (which is equivalent to $(G - \mu X)^T(LX - G) \succeq 0$ when G^TX is symmetric)

$$\frac{\mu + L}{2}(G^TX + X^TG) - G^TG - \mu LX^TX \succeq 0. \tag{6.14}$$

Theorem 4.3 on the interpolation of not necessarily symmetric linear operators allows to understand the effect of (6.14) since it can be written as

$$\left(\frac{L+\mu}{2}X - G\right)^T \left(\frac{L+\mu}{2}X - G\right) \preceq \left(\frac{L-\mu}{2}\right)^2 X^T X,$$
$$\overset{\text{Th. 4.3}}{\Leftrightarrow} \exists M \text{ such that } \|M\| \le 1 \text{ and } \frac{L+\mu}{2}X - G = M\frac{L-\mu}{2}X,$$
$$\Leftrightarrow \exists M \text{ such that } \|M\| \le 1 \text{ and } G = \frac{1}{2}\big((L+\mu)I - (L-\mu)M\big)X.$$

Therefore, if and only if a set $\{(x_i, g_i)\}_{i\in[N]}$ satisfies condition (6.14), then, there exists a linear operator T such that $g_i = Tx_i$ and

$$T = \frac{1}{2}\big((L+\mu)I - (L-\mu)M\big) \text{ with } ||M|| \le 1. \tag{6.15}$$

T is not necessarily symmetric, therefore, T is a linear operator but not the gradient of a quadratic function. Moreover, T is square since we used Theorem 4.3 with $m = n$ and, thus, has possibly complex eigenvalues $\lambda(T)$. Finally, T has the following property.

Proposition 6.5. *If $Tx = \frac{1}{2}\big((L+\mu)I - (L-\mu)M\big)x$ with $||M|| \le 1$, then,*

$$\left|\frac{L+\mu}{2} - \lambda(T)\right| \le \frac{L-\mu}{2} \tag{6.16}$$

where $|a|$ is the modulus of the complex number a.

Proof. Using $\lambda(aA + bI) = a\lambda(A) + b$, we have

$$\lambda(T) = \lambda\left(\frac{L+\mu}{2}I - \frac{L-\mu}{2}M\right) = \frac{L+\mu}{2} - \frac{L-\mu}{2}\lambda(M) \tag{6.17}$$

and we can check that the property holds

$$\left|\frac{L+\mu}{2} - \lambda(T)\right| \overset{(6.17)}{=} \left|\frac{L+\mu}{2} - \frac{L+\mu}{2} + \frac{L-\mu}{2}\lambda(M)\right| = \frac{L-\mu}{2}\,|\lambda(M)| \le \frac{L-\mu}{2}$$

which holds by $|\lambda(M)| \le ||M|| \le 1$ (since $Mv = \lambda(M)v$ with $||v|| = 1$ $\Rightarrow |\lambda(M)| = ||\lambda(M)v|| = ||Mv|| \le \max_{||v||=1} ||Mv|| = ||M||$). □

However, all operators satisfying (6.16) cannot be written under the form (6.15). Indeed, the following operator

$$T = \begin{pmatrix} 0 & 0 \\ -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} \tag{6.18}$$

satisfies $\left|\frac{1}{2} - \lambda(T)\right| \leq \frac{1}{2}$ ($\mu = 0$, $L = 1$) but the only M satisfying $Tx = \frac{1}{2}(I - M)x \ \ \forall x$ is

$$M = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \tag{6.19}$$

and is such that $||M|| = \sqrt{2} > 1$.

Condition $X^T G = G^T X$ We now investigate which operators are described by the condition $x_i^T g_j = x_j^T g_i$ alone.

If an operator T satisfy $x^T T(y) = y^T T(x)$ for every point $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, then T is linear and symmetric

Theorem 6.6. *An operator T is linear and symmetric if and only if*

$$x^T T(y) = y^T T(x) \quad \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{6.20}$$

Proof. (Necessity) If $T(x) = Qx$ with $Q = Q^T$ then $x^T Qy = y^T Qx \ \ \forall (x, y)$ (since $y^T Qx = x^T Q^T y = x^T Qy$ by symmetry of Q).

(Sufficiency) Let T satisfying $x^T T(y) = y^T T(x) \ \ \forall (x, y)$, then

$$x^T T(ay + z) = (ay + z)^T T(x) = ay^T T(x) + z^T T(x) = ax^T T(y) + x^T T(z) \tag{6.21}$$

must hold for all x, thus,

$$T(ay + z) = aT(y) + T(z) \tag{6.22}$$

meaning that T is linear. Therefore, $T(x) = Qx$ for a matrix Q and it remains to show that $Q = Q^T$. Since $y^T Qx = x^T Q^T y$ and that $x^T Qy = y^T Qx$ then $x^T Qy = x^T Q^T y \ \ \forall (x, y)$, and therefore, $Q = Q^T$. □

However, when the operator T satisfies the condition only on the discrete points $(x_i, g_i)_{i \in [N]}$, it is less clear. Condition $G^T X = X^T G$ does not describe exactly the class of symmetric linear operators without bound on the spectrum since the interpolation conditions of this class are

$$\begin{cases} G^T X = X^T G, \\ \exists L > 0 \ : \ G^T G \preceq L^2 X^T X. \end{cases} \tag{6.23}$$

Moreover, the set $\{(0,1)\}$ satisfies $G^T X = X^T G$ but there is no linear operator such that $T(0) = 1$. Maybe the constraint defines some kind of linear operator with infinite eigenvalues.

6.3 Bounding the performance on smooth strongly convex functions by the performance on quadratic ones

In this section, we investigate whether it is possible to deduce a bound on the performance of a given FSFOM on smooth strongly convex functions from its performance on quadratic functions (which is easier to compute with the technique presented in Section 6.1). More precisely, we note $w(\mathcal{E}, \mathcal{A}, \mathcal{F})$ the worst-case performance of an optimization method $\mathcal{A}$ on the function class $\mathcal{F}$ with respect to the performance criterion $\mathcal{E}$. Given a criterion $\mathcal{E}$, we would like to find an $\alpha_{\mathcal{E}}$ such that

$$1 \leq \frac{w(\mathcal{E}, \mathcal{A}, \mathcal{F}_{\mu,L})}{w(\mathcal{E}, \mathcal{A}, \mathcal{Q}_{\mu,L})} \leq \alpha_{\mathcal{E}} \ \ \forall \text{ FSFOM } \mathcal{A}. \tag{6.24}$$

We will see that there is such a finite $\alpha_{\mathcal{E}}$ for the gradient method with constant step sizes but that in general it is not possible. We prove it by constructing specific FSFOM or performance criterion $\mathcal{E}$ for which $\alpha_{\mathcal{E}}$ is unbounded.

Remark 6.7. There are examples of optimization methods in the literature that can diverge on smooth convex functions but converge on quadratic and therefore for which there is no such $\alpha_{\mathcal{E}}$, e.g., Heavy ball [Pol87, LRP16] or Barzilai-Borwein [BB88, BDH19] methods.

6.3.1 Gradient method with constant step sizes

We consider the performance criterion $\mathcal{E} = f(x_N) - f(x^*)$ and the gradient method with constant steps

$$x_{k+1} = x_k - \frac{h}{L} \nabla f(x_k). \tag{GM-h}$$

We can compute numerically the ratio $w(\mathcal{E}, \text{GM-h}, \mathcal{F}_{0,L}) / w(\mathcal{E}, \text{GM-h}, \mathcal{Q}_{0,L})$ for different step size h with PEP. Figure 6.1 shows the ratio for different step sizes h and number of iterations N. We observe that the ratio is always bounded by $\alpha_{\mathcal{E}} = e$.

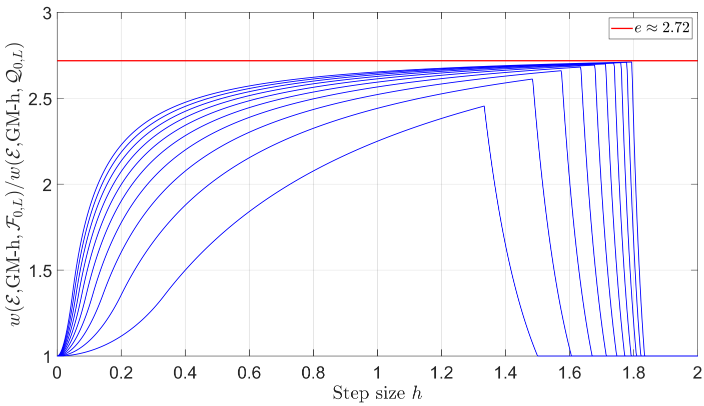


Fig. 6.1 Ratio $w(\mathcal{E}, \text{GM-h}, \mathcal{F}_{0,L})/w(\mathcal{E}, \text{GM-h}, \mathcal{Q}_{0,L})$ of the worst-case performance of GM-h on general and quadratic functions for varying step size h and number of iterations $N = 1, \ldots, 10$ (each blue curve corresponds to a given value of N, starting from $N = 1$ for the lowest to $N = 10$ for the highest on the figure).

6.3.2 Arbitrarily bad methods

Given a method $\mathcal{A}_1$, we can build another method $\mathcal{A}_2$ such that the last iterates x_N computed by both $\mathcal{A}_1$ and $\mathcal{A}_2$ are equal on all quadratic functions (but not on all general functions). Therefore, we can build a method with the same behavior as GM-1 (for example) on quadratic functions but with poor behavior on general functions.

Consider three iterations of a FSFOM (6.1) applied to a quadratic function $f(x) = \frac{1}{2}x^T Qx$, namely,

$$\begin{aligned} x_1 =& x_0 - h_{1,0}Qx_0 \\ =& (I - h_{1,0}Q)x_0, \\ x_2 =& x_0 - h_{2,0}Qx_0 - h_{2,1}Qx_1 \\ =& (I - (h_{2,0} + h_{2,1})Q + h_{2,1}h_{1,0}Q^2)x_0, \\ x_3 =& x_0 - h_{3,0}Qx_0 - h_{3,1}Qx_1 - h_{3,2}Qx_2 \\ =& (I - (h_{3,0} + h_{3,1} + h_{3,2})Q + (h_{3,1}h_{1,0} + h_{3,2}h_{2,0} \\ & + h_{3,2}h_{2,1})Q^2 - h_{3,2}h_{2,1}h_{1,0}Q^3)x_0. \end{aligned} \tag{6.25}$$

GM-1 corresponds to $h_{ik} = 1 \;\; \forall i \geq k$ therefore, any methods such that $h_{30} + h_{31} + h_{32} = h_{31}h_{10} + h_{32}h_{20} + h_{32}h_{21} = 3$ and $h_{32}h_{21}h_{10} = 1$ will have the same iterate x_3 on every quadratics. Such methods are parameterized by three parameters a, b and c and have the following iterations

$$\begin{aligned} x_1 &= x_0 - a\nabla f(x_0), \\ x_2 &= x_0 + \frac{3 - \frac{1}{a} - ab}{c}\nabla f(x_0) - \frac{1}{ac}\nabla f(x_1), \\ x_3 &= x_0 - (3 - b - c)\nabla f(x_0) - b\nabla f(x_1) - c\nabla f(x_2). \end{aligned} \tag{6.26}$$

Let us consider the methods with $a = c = 1$, namely,

$$\begin{aligned} x_1 &= x_0 - \nabla f(x_0), \\ x_2 &= x_0 + (2 - b)\nabla f(x_0) - \nabla f(x_1) \\ &= x_1 - (1 - b)\nabla f(x_0) - \nabla f(x_1), \\ x_3 &= x_0 - (2 - b)\nabla f(x_0) - b\nabla f(x_1) - \nabla f(x_2) \\ &= x_2 + (1 - b)\nabla f(x_1) - \nabla f(x_2), \end{aligned} \tag{6.27}$$

this family of methods can be written in the compact form

$$x_{k+1} = x_k + (-1)^k(1 - b)\nabla f(x_{k-1}) - \nabla f(x_k). \tag{A-b}$$

By construction, the performance of these methods A-b on any given quadratic function is equal to the performance of GM-1 on this quadratic function, therefore, a finite value.

We consider the performance criterion $\mathcal{E} = f(x_N) - f(x^*)$ and show that, when $b \geq 4$, the performance of these methods on smooth convex functions is lower bounded by

$$w(\mathcal{E}, \text{A-b}, \mathcal{F}_{0,L}) \geq \frac{(b-1)^2}{6} \tag{6.28}$$

and therefore tends to infinity when $b \to \infty$ (and so does the ratio $w(\mathcal{E},\text{A-b}, \mathcal{F}_{0,L})/w(\mathcal{E},\text{A-b}, \mathcal{Q}_{0,L})$). We compute a lower bound on the worst-case performance by exhibiting a function and computing the iterations of the method on this function.

The functions belong to a family of Huber loss of the form

$$f_b(x) = \begin{cases} \tau_b|x| - \frac{\tau_b^2}{2} & \text{if } |x| \geq \tau_b, \\ \frac{x^2}{2} & \text{else,} \end{cases} \tag{6.29}$$

with $\tau_b = \frac{b-1}{3}$. Let $b \geq 4$ and $x_0 = -1$. Since $b \geq 4$, we have $\tau_b \geq 1$ and $|x_0| \leq \tau_b$, $\nabla f(x_0) = -1$, $x_1 = x_0 - \nabla f(x_0) = 0$, $\nabla f(x_1) = 0$ and $x_2 = x_1 - (1-b)\nabla f(x_0) - \nabla f(x_1) = (1-b) = -3\tau_b$. Therefore, $|x_2| = 3\tau_b \geq \tau_b$, $\nabla f(x_2) = -\tau_b$ and $x_3 = x_2 - (1-b)\nabla f(x_1) - \nabla f(x_2) = -2\tau_b$. Therefore, $|x_3| = 2\tau_b \geq \tau_b$ and $f_b(x_3) = \frac{3}{2}\tau_b^2 = \frac{(b-1)^2}{6}$.

Besides, the worst-case performance of A-b and GM-1 on quadratic functions is

$$w(\mathcal{E}, \text{A-b}, \mathcal{Q}_{0,L}) = w(\mathcal{E}, \text{GM-1}, \mathcal{Q}_{0,L}) = \max_{\rho \in [0,L]} \frac{1}{2}\rho(1-\rho)^6. \tag{6.30}$$

By construction, it does not depend on the parameter b.

Remark 6.8. Note that the strongly convex version of the Huber loss,

$$f_{b,\mu} = \begin{cases} \frac{\mu x^2}{2} + (1-\mu)\tau_b|x| - \frac{1-\mu}{2}\tau_b^2 & \text{if } |x| \geq \tau_b, \\ \frac{x^2}{2} & \text{else,} \end{cases} \tag{6.31}$$

allows to obtain the same observations for the strongly convex case $\mu > 0$.

We summarize our reasoning in the following theorem.

Theorem 6.9. *Method A-b has the same performance as GM-1 on smooth convex quadratic functions, i.e.,* (6.30)*, but can have arbitrarily bad performance on smooth convex functions when b tends to infinity, i.e.,* (6.28)*. Therefore, it is not always possible to bound the performance of a method on general smooth convex functions by a multiple of its performance on quadratic functions.*

6.3.3 A particular choice of criterion

Another reason why it will not be possible to bound the performance of methods on general functions by their performance on quadratics in general is the following performance criterion

$$\mathcal{E} = f_N - \frac{1}{2}x_N^T g_N. \tag{6.32}$$

This quantity is always zero for homogeneous quadratic functions and may not be zero for general functions. For example, after 0 iteration, on all quadratics we have $f_0 - \frac{1}{2}x_0^T g_0 = 0$ whereas on the function $f(x) = ax$, we have $f_0 - \frac{1}{2}x_0^T g_0 = \frac{1}{2}ax_0$ which can be arbitrarily large when increasing a (when $x_0 > 0$).

7

Performance estimation of methods involving linear operators

Discussion with Adrien Taylor, Moslem Zamani, and Teodor Rotaru inspired some of the results of this chapter.

THE new interpolation conditions of Theorems 4.3 and 4.5 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` [THG17b] to which we added our new interpolation conditions. The toolbox and our extension are available in *Python* via the library `PEPit` [GMG+24], and the numerical experiments performed in the two first sections can be found at

`https://github.com/NizarBousselmi/PEP-Linear-Operator-code/tree/main`.

Semidefinite programs are solved by the *MOSEK* [ApS24] interior-point semidefinite optimization solver.

In this chapter, 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 in Chapter 3, $\min_x g(Mx)$ (which covers problem $\min_x \frac{1}{2}x^TQx$ as a special case). Then, we analyze the more recent Chambolle-Pock algorithm [CP11a], which was our third 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. Afterwards, we analyze the Barzilai-Borwein method applied to quadratic functions thanks to the new interpolation conditions for quadratic functions. Finally, we show that the worst-case linear operators are not always a scaling operator of the form $M = \alpha I$ even though it seems to be often the case on the settings studied.

7.1 Gradient method on $g \circ \mathcal{M}$

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

$$F = g \circ \mathcal{M} \tag{7.1}$$

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 μ_M and L_M. We also define the class $\mathcal{D}_{\mu_g,L_g}^{\mu_M,L_M}$ 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 $\mathcal{C}_{\mu_g,L_g}^{\mu_M,L_M} = \mathcal{C}_{\mu_g L_M^2, L_g L_M^2}^{\mu_M/L_M,1}$ and $\mathcal{D}_{\mu_g,L_g}^{\mu_M,L_M} = \mathcal{D}_{\mu_g L_M^2, L_g L_M^2}^{\mu_M/L_M,1}$, 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

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

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

$$F(x_N) - F(x^*) \leq w\left(\mathcal{F}, R, N, \frac{h}{L_F}\right) \quad \forall F \in \mathcal{F} \tag{7.2}$$

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

The worst-cases guarantees for classes $\mathcal{F}_{\mu_f, L_f}$, $\mathcal{C}^{\mu_M, 1}_{\mu_g, L_g}$, and $\mathcal{C}^{\mu_M, 1}_{\mu_g, L_g}$ reduce to simpler cases with the following homogeneity relations, see [THG17c, Section 3.5] for a proof (semicolons are introduced for readability),

$$\begin{aligned} w\left(\mathcal{F}_{\mu_f, L_f}; R, N, \frac{h}{L_f}\right) &= L_f R^2 w\left(\mathcal{F}_{\frac{\mu_f}{L_f}, 1}; 1, N, h\right), \\ w\left(\mathcal{C}^{\mu_M, 1}_{\mu_g, L_g}; R, N, \frac{h}{L_g}\right) &= L_g R^2 w\left(\mathcal{C}^{\mu_M, 1}_{\frac{\mu_g}{L_g}, 1}; 1, N, h\right), \\ w\left(\mathcal{D}^{\mu_M, 1}_{\mu_g, L_g}; R, N, \frac{h}{L_g}\right) &= L_g R^2 w\left(\mathcal{D}^{\mu_M, 1}_{\frac{\mu_g}{L_g}, 1}; 1, N, h\right). \end{aligned} \tag{7.3}$$

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

$$\begin{aligned} w\left(\mathcal{F}_{\mu_f, 1}; 1, N, h\right) &= w\left(\mathcal{F}_{\mu_f}; h\right), \\ w\left(\mathcal{C}^{\mu_M, 1}_{\mu_g, 1}; 1, N, h\right) &= w\left(\mathcal{C}^{\mu_M}_{\mu_g}; h\right), \\ w\left(\mathcal{D}^{\mu_M, 1}_{\mu_g, 1}; 1, N, h\right) &= w\left(\mathcal{D}^{\mu_M}_{\mu_g}; h\right), \end{aligned} \tag{7.4}$$

Note that $\mathcal{F}_{\mu_g} \subseteq \mathcal{C}^{\mu_M}_{\mu_g} \subseteq \mathcal{F}_{\mu_g \mu_M^2}$ and $\mathcal{F}_{\mu_g} \subseteq \mathcal{D}^{\mu_M}_{\mu_g} \subseteq \mathcal{F}_{\mu_g \mu_M^2}$, therefore, $w(\mathcal{F}_{\mu_g}; h) \leq w(\mathcal{C}^{\mu_M}_{\mu_g}; h) \leq w(\mathcal{F}_{\mu_g \mu_M^2}; h)$ 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 (except [NNG19] that provided a bound on the performance of the gradient method with unit step size on functions $F = g \circ M$ when M has a non-zero lower bound on its singular values).

Solving (PEP) for the new class $\mathcal{C}^0_{0.1}$ yields numerical results in Figure 7.1, namely, the worst-case performance of the gradient method (GM) 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 $\mathcal{F}_{0.1}$. 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.

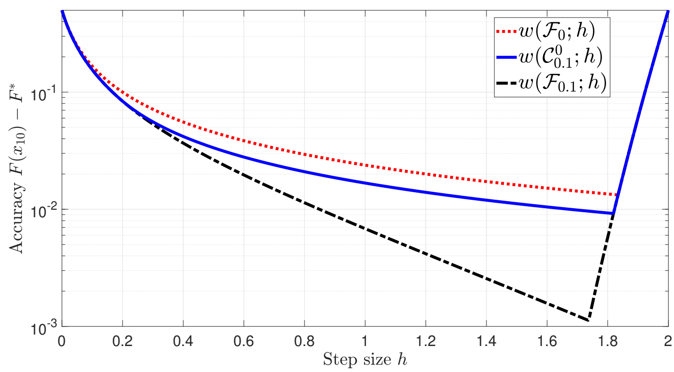


Fig. 7.1 Worst-case performance of 10 iterations of (GM) for varying step size $h \in [0, 2]$ on classes $\mathcal{F}_0$ of 1-smooth convex functions f (dotted red line), $\mathcal{C}^0_{0.1}$ of 1-smooth 0.1-strongly convex functions $g \circ M$ (solid blue line) and $\mathcal{F}_{0.1}$ of 1-smooth 0.1-strongly convex functions f (broken black line).

After extensive numerical computations with `PESTO` and analysis of results such that depicted in Figure 7.1, we were able to identify the exact worst-case performances $w(\mathcal{C}^{\mu_M}_{\mu_g}; h)$ 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 7.1.2, we review the results of [THG17c] on the worst-case performance of (GM) on the class $\mathcal{F}_{\mu_f}$.

7.1.1 Performance on $\mathcal{F}_\mu$

Let $N \geq 0$, $h \in [0, 2]$, functions $\ell, q \in \mathcal{F}_\mu$ as

$$\ell_{\mu,h}(x) = \begin{cases} \frac{\mu}{2}x^2 + (1-\mu)\tau_{\mu,h}|x| - \left(\frac{1-\mu}{2}\right)\tau^2_{\mu,h} & \text{if } |x| \geq \tau_{\mu,h}, \\ \frac{1}{2}x^2 & \text{else}, \end{cases} \tag{7.5}$$
$$q(x) = \frac{1}{2}x^2,$$

and $\tau_{\mu,h} = \frac{\mu}{\mu-1+(1-\mu h)^{-2N}}$. It is conjectured [THG17c, Conjecture 2] with very strong numerical evidence, that the worst-case performance of (GM) on $\mathcal{F}_{\mu_f}$ is given by

$$w\left(\mathcal{F}_{\mu_f};h\right) = \frac{1}{2}\max\left\{\frac{\mu_f}{\mu_f - 1 + (1-\mu_f h)^{-2N}}, (1-h)^{2N}\right\}. \tag{7.6}$$

Moreover, as one can check that this worst-case performance corresponds to that of the univariate 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 univariate functions.

Remark 7.1. A proof of [THG17c, Conjecture 2] has recently been proposed in [Kim24] but has not yet been published.

7.1.2 Performance on $\mathcal{C}_{\mu_g}^{\mu_M}$ and $\mathcal{D}_{\mu_g}^{\mu_M}$

Through numerous numerical experiments, we also observed that the worst-case functions are univariate 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 $\tilde{\mathcal{C}}_{\mu_g}^{\mu_M}$ (resp. $\tilde{\mathcal{D}}_{\mu_g}^{\mu_M}$) the sub-class of functions $g \circ M$ where $M = \alpha I$ is a scaling operator and propose the following conjecture, supported by our numerical evidence.

Conjecture 7.2. *Worst-case performances of $\mathcal{C}_{\mu_g}^{\mu_M}$ and $\mathcal{D}_{\mu_g}^{\mu_M}$ are attained by scaling operator $M = \alpha I$, i.e., $w(\mathcal{C}_{\mu_g}^{\mu_M};h) = w(\tilde{\mathcal{C}}_{\mu_g}^{\mu_M};h)$ and $w(\mathcal{D}_{\mu_g}^{\mu_M};h) = w(\tilde{\mathcal{D}}_{\mu_g}^{\mu_M};h)$.*

From now on, given Conjecture 7.2 and the fact that $\tilde{\mathcal{C}}_{\mu_g}^{\mu_M} = \tilde{\mathcal{D}}_{\mu_g}^{\mu_M}$, we will only present the analysis for the symmetric case, as the general case shares the same analysis. The class of functions $g \circ \alpha I$ can be written as a union of classes of functions f,

$$\tilde{\mathcal{D}}_{\mu_g}^{\mu_M} = \bigcup_{\alpha \in [\mu_M, 1]} \mathcal{F}_{\mu_g \alpha^2, \alpha^2}, \tag{7.7}$$

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

$w(\mathcal{F}_{\mu_f};h)$ with

$$\begin{aligned} w\left(\mathcal{D}_{\mu_g}^{\mu_M};h\right) &\overset{\text{Conj. 7.2}}{=} w\left(\tilde{\mathcal{D}}_{\mu_g}^{\mu_M};h\right) \\ &\overset{(7.7)}{=} w\left(\bigcup_{\alpha\in[\mu_M,1]} \mathcal{F}_{\mu_g\alpha^2,\alpha^2};h\right) \\ &= \max_{\alpha\in[\mu_M,1]} w\left(\mathcal{F}_{\mu_g\alpha^2,\alpha^2};h\right) \\ &\overset{(7.3)}{=} \max_{\alpha\in[\mu_M,1]} \alpha^2 w\left(\mathcal{F}_{\mu_g};\alpha^2 h\right). \end{aligned} \tag{7.8}$$

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 (see Section 7.4). 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 their performance on general linear operators.

Given expression (7.6) of $w(\mathcal{F}_{\mu_f};h)$, it is possible to solve the last maximization problem of (7.8) and end up with the following conjecture featuring an explicit convergence rate.

Conjecture 7.3 (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*

$$w\left(\mathcal{D}_{\mu_g}^{\mu_M};h\right) = \frac{1}{2}\max\left\{\frac{\mu_g {\alpha^*}^2}{\mu_g - 1 + \left(1-\mu_g{\alpha^*}^2 h\right)^{-2N}}, (1-h)^{2N}\right\} \tag{7.9}$$

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

$$(1-\mu_g)(1-\mu_g h_0)^{2N+1} = 1-(2N+1)\mu_g h_0. \tag{7.10}$$

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

Theorem 7.4 (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*

$$w\left(\mathcal{D}_{\mu_g}^{\mu_M}; h\right) \geq \frac{1}{2} \max \left\{ \frac{\mu_g {\alpha^*}^2}{\mu_g - 1 + \left(1 - \mu_g {\alpha^*}^2 h\right)^{-2N}}, (1-h)^{2N} \right\} \quad (7.11)$$

with α^* *defined in Conjecture 7.3.*

Proof. Functions ${\alpha^*}^2 \ell_{\mu_g, {\alpha^*}^2 h}$ and q, see (7.5), belong to $\mathcal{D}_{\mu_g}^{\mu_M}$ and attain (7.11). □

The worst-case performance established in Conjecture 7.3 matches exactly the large number of numerical experiments we performed for many different values of parameters μ_g, μ_M, and h. Moreover, development (7.8) shows that Conjecture 7.3 relies only on the weaker Conjecture 7.2 and on the previous conjectures of [THG17c]. We summarize this observation in a corollary.

Corollary 7.5. *If Conjecture 7.2 holds and* $w(\mathcal{F}_{\mu_f}; h)$ *is given by* (7.6) *as conjectured in [THG17c], then Conjecture 7.3 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 have proposed several partial justifications for Conjecture 7.3 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 7.3 when $N = 1, h = 1, \mu_M = 0$ The optimal primal solution of (PEP) yields a function attaining 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 [GDT23b] 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 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 7.3 in the case $N = 1$, $h = 1$, $\mu_M = 0$, and

$$\mu_g \leq \frac{1}{6}\left(7 - \frac{7}{\sqrt[3]{44 - 3\sqrt{177}}} - \sqrt[3]{44 - 3\sqrt{177}}\right) \approx 0.17. \tag{7.12}$$

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 $F(x_1) - F(x^*) \leq \tau||x_0 - x^*||^2$, was identified from the numerical dual solution of the PEP.

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

where $P = \begin{pmatrix} Mx_0 & Mx_1 & Mx^* & \nabla g(Mx_0) & \nabla g(Mx_1) & \nabla g(Mx^*) \end{pmatrix}$, $\gamma = \frac{1-\mu_g}{2-\mu_g}$, $\tau = \frac{\mu_g}{2(\mu_g - 1 + (1-\mu_g)^{-2})}$, $x_1 = x_0 - M^T \nabla g(Mx_0)$,

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

$$b_{\mu_g} = \begin{pmatrix} 0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix} = \begin{pmatrix} 0 \\ -\sqrt{\frac{\mu_g}{2(1-\mu_g)} + \frac{1}{2-\mu_g} + (1+\mu_g)^2 a_2^2} \\ -b_1 \\ \frac{2(1+\mu_g)^2(2-\mu_g)a_2^2-(1+\mu_g)}{-2(2-\mu_g)b_1} \\ \frac{-2(1-\mu_g^2)a_2^2+1}{-2(1-\mu_g)b_1} \\ \frac{-2\mu_g(1-\mu_g^2)(2-\mu_g)a_2^2+(1-\mu_g)^2+\mu_g}{2(1-\mu_g)(2-\mu_g)b_1} \end{pmatrix}, c_{\mu_g} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix},$$

$$a_1 = -\sqrt{\frac{\mu_g}{2-\mu_g} + \tau}, \quad c_3 = \sqrt{\frac{1}{2-\mu_g} - (1+\mu_g)^2 a_2^2 - b_2^2},$$

$$a_2 = \frac{1}{2(2-\mu_g)a_1}, \qquad c_4 = -\sqrt{\frac{1}{2(1-\mu_g)} - a_2^2 - b_3^2},$$

$$a_3 = \sqrt{\tau - \frac{1}{2-\mu_g} + (1+\mu_g)^2 a_2^2}, \quad c_5 = \sqrt{\frac{1}{2(1-\mu_g)} - (\mu_g a_2)^2 - b_4^2}.$$

The expression of the rate τ matches Conjecture 7.3 for this set of parameters. The equality is an algebraic reformulation that is always satisfied. 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 $\mathcal{L}_{0,L_M}$-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 7.1 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,

$$\begin{aligned} p_1(\mu, h) &\triangleq \ell_{\mu,h}(x_N) - \ell^* = \frac{1}{2}\frac{\mu}{\mu - 1 + (1-\mu h)^{-2N}}, \\ p_2(h) &\triangleq q(x_N) - q^* \quad = \frac{1}{2}(1-h)^{2N}. \end{aligned} \tag{7.13}$$

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.

Table 7.1 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.	$\begin{cases}\ell_{\mu_f,h}(x)\\ q(x)\end{cases}$	$\begin{cases}\alpha^{*2}\ell_{\mu_g,M^{*2}h}(x)\\ q(x)\end{cases}$

Optimal step sizes Understanding the exact worst-case performance of (GM) on $\mathcal{D}_{\mu_g}^{\mu_M}$ allows 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 7.3. Optimal steps $h^*(\mu_f)$ for $\mathcal{F}_{\mu_f}$ (see [THG17c]) and $h^*(\mu_g, \mu_M)$ for $\mathcal{D}_{\mu_g}^{\mu_M}$ satisfy

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

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

where $\tilde{\mu}(h) = \mu_g \text{proj}_{[\mu_M^2, 1]}\left(\frac{h_0}{h}\right)$. Both $h^*(\mu_f)$ and $h^*(\mu_g, \mu_M)$ can be easily computed numerically and 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 $h^*\left(\mu_g, \mu_M\right)$.

Performance on quadratic functions The class of L-smooth, μ-strongly convex, not necessarily homogeneous quadratic functions is $\mathcal{D}_{1,1}^{\sqrt{\mu},\sqrt{L}}$. 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 7.3, the worst-case performance of the gradient method over the class of quadratic functions is

$$w\left(\mathcal{Q}_{\mu,L}; h\right) = w\left(\mathcal{D}_{1,1}^{\sqrt{\mu},\sqrt{L}}; h\right) = \frac{LR^2}{2}\max\left\{q\left(1-qh\right)^{2N}, (1-h)^{2N}\right\} \tag{7.14}$$

where $q = \text{proj}_{[\frac{\mu}{L},1]}(\frac{1}{h(2N+1)})$. Moreover, the worst-case over the class of quadratic functions is univariate, a consequence of the polynomial-based analysis of Section 6.1. Therefore, Conjecture 7.2 holds in the case $L_g = \mu_g$.

Performance on linear operators with eigenvalues on a union of intervals

The results of this paragraph were obtained in collaboration with Zhicheng Deng.

We compute the worst-case performance of GM when the linear operator M has eigenvalues in a union of intervals. As an example, we compare the worst-case behavior of GM when M has eigenvalues in $[0.6, 1]$ or in $[0.6, 0.65] \cup [0.95, 1]$. It appears that in both cases the worst-case linear operator is scalar. Therefore, Figure 7.2 compares the scalar worst-case linear operator of GM when varying the step size h when $M \in \mathcal{S}_{[0.6,1]}$ (red dots) and $M \in \mathcal{S}_{[0.6,0.65]\cup[0.95,1]}$ (blue dots). In other words, either we include the whole interval $[0.6, 1]$, or we exclude the sub-interval $[0.65, 0.95]$.

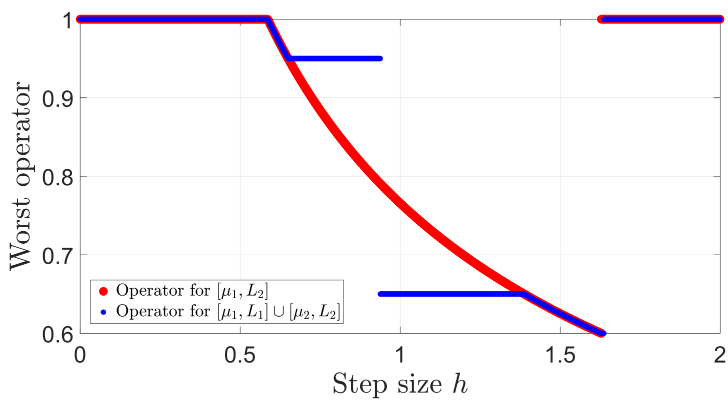


Fig. 7.2 Comparison of the scalar worst-case linear operators M when allowing the whole interval (red curve) or only the union of two sub-intervals (blue curve).

Interestingly, the worst-case linear operator remains a scalar. This could not be observed without the interpolation conditions of Section 4.3.

7.2 Chambolle-Pock method

We show how to tackle with PEP the analysis of a more sophisticated algorithm, namely the Chambolle-Pock algorithm [CP11a].

The Chambolle-Pock algorithm solves problems of the form

$$\min_x f(x) + g(Mx) \tag{7.15}$$

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

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

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

Remark 7.6. Note that we could also analyze the accelerated version of Chambolle-Pock method [CP11a, Algorithm 2] with our extension of PEP.

7.2.1 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 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 [CP11a], reproduced below, requires the existence of sets B_1 and B_2 "large enough". We note $\mathcal{L}(x,u) = u^T Mx + f(x) - g^*(u)$ the Lagrangian of problem (7.15) 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 7.7 ([CP11a], 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*

$$\mathcal{G}_{B_1 \times B_2}(\bar{x}_N, \bar{u}_N) \leq \frac{D(B_1, B_2)}{N} \tag{7.16}$$

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

$$\begin{aligned} \mathcal{G}_{B_1 \times B_2}(x,u) &= \max_{u' \in B_2} \mathcal{L}(x,u') - \min_{x' \in B_1} \mathcal{L}(x',u), \\ D(B_1,B_2) &= \sup_{(x,u) \in B_1 \times B_2} \frac{||x-x_0||^2}{2\tau} + \frac{||u-u_0||^2}{2\sigma}. \end{aligned} \tag{7.17}$$

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 [CP16b] adapts the result to remove the dependency in M at the cost of an additional factor).

Theorem 7.8 ([CP16b], Theorem 1). *Let f and g 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*

$$\mathcal{L}(\bar{x}_N, u) - \mathcal{L}(x, \bar{u}_N) \leq \frac{1}{2N}\left(\frac{\|x - x_0\|^2}{\tau} + \frac{\|u - u_0\|^2}{\sigma} - 2(u - u_0)^T M(x - x_0)\right). \tag{7.18}$$

Note that Theorem 2 of [Yan18] 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 Prof. Beck (slide 29 of [Bec22]), 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.

Theorem 7.9. *Let f and g 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*

$$F(\bar{x}_N) - F(x^*) \leq \frac{1}{2N}\left(\frac{||x^* - x_0||^2}{\tau} + \frac{||\tilde{u}_N - u_0||^2}{\sigma} - 2(\tilde{u}_N - u_0)^T M(x^* - x_0)\right) \tag{7.19}$$

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

Proof. Since the assumptions of Theorem 7.8 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 Conjugate Subgradient Theorem (CST) and Fenchel's Inequality (FI) yield

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

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

and therefore

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

□

7.2.2 Convergence results obtained with the new PEP framework

Theorems 7.8 and 7.9 can be analyzed using our extension of the PEP framework. First, to study the setting of Theorem 7.8, 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 (7.18) 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 7.3. 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 (7.18) but with $N+1$ instead of N (shown as a blue line on Figure 7.3). Based on the hint, provided by PEP, that this better bound actually holds, we were able to adapt the proof of Theorem 7.8 in [CP16b] to obtain the following theorem, that gives the exact convergence rate. More precisely, PEP allowed us to identify the inequality where the tighness was lost in the original proof.

Theorem 7.10. *Let f and g 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*

$$\mathcal{L}(\bar{x}_N, u) - \mathcal{L}(x, \bar{u}_N) \leq \frac{1}{2(N+1)}\left(\frac{\|x - x_0\|^2}{\tau} + \frac{\|u - u_0\|^2}{\sigma} - 2(u - u_0)^T M(x - x_0)\right). \tag{7.20}$$

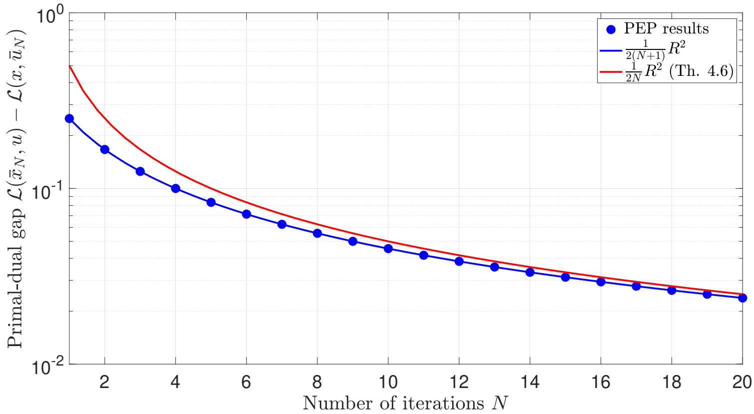


Fig. 7.3 Worst-case performance obtained by our extension of PEP for N iterations of the Chambolle-Pock algorithm (CP) with any step size parameters σ and τ satisfying $\tau\sigma L_M^2 \leq 1$ on the problem $\min_x F(x)$ where $F = f + g \circ M$, f and g are convex proximable and M is such that $||M|| \leq 1$. 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 (7.18) of Theorem 7.8 (red line).

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 [CP16b])

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

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

$s_{N+1} = z$:

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

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

Remark 7.11. Since Theorem 7.9 was based on Theorem 7.8, Theorem 7.10 that improves the convergence of Theorem 7.8, also improves the one of Theorem 7.9.

Instead of bounding the quantity

$$R^2 = \frac{\|x - x_0\|^2}{\tau} + \frac{\|u - u_0\|^2}{\sigma} - 2(u - u_0)^T M(x - x_0), \tag{7.21}$$

it is also possible with PEP to compute rates that depend on the simpler squared distance

$$R_0^2 = \|x - x_0\|^2 + \|u - u_0\|^2. \tag{7.22}$$

In this case, our numerical results indicate that the primal-dual gap depends on the step sizes τ and σ. Figure 7.4 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 [CP16b]. 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.

7.2.3 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.

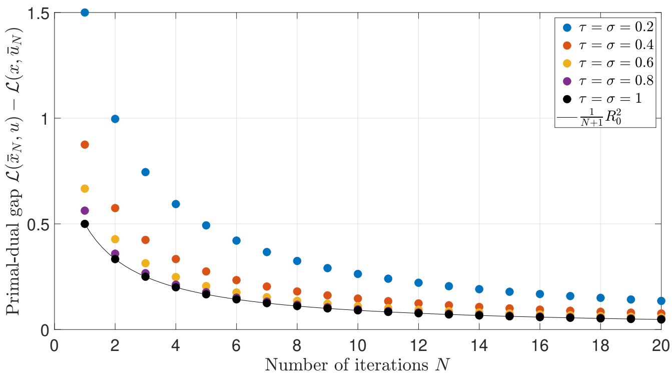


Fig. 7.4 Worst-case performance obtained by our extension of PEP for N iterations of the Chambolle-Pock algorithm (CP) with different step sizes $\tau = \sigma$ on the problem $\min_x F(x)$ where $F = f + g \circ M$, f and g are convex proximable and M is such that $||M|| \leq 1$. The performance criterion is the primal-dual gap $\mathcal{L}(\bar{x}_N, u) - \mathcal{L}(x, \bar{u}_N)$ and the initial distance is $R_0^2 = \|x - x_0\|^2 + \|u - u_0\|^2 \leq 1$.

Theorems 7.7 and 7.8 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 7.5 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.

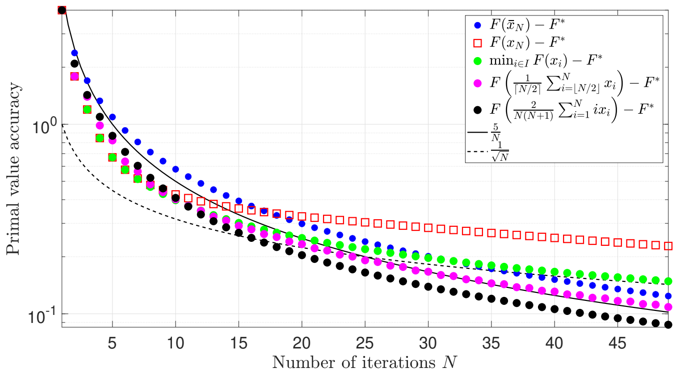


Fig. 7.5 Worst-case performance obtained by our extension of PEP for N iterations of Chambolle-Pock algorithm (CP) with step size parameters $\tau = \sigma = 1$ on the problem $\min_x F(x)$ where $F = f + g \circ M$, f and g are 1-Lipschitz convex proximable, and M is such that $||M|| \leq 1$. The performance 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.

Numerical guarantees such as those depicted in Figure 7.5 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 manner will be exact, i.e., unimprovable 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 re-developed when changing the framework of evaluation. For example, the convergence of a weighted sum of the iterates as done in Figure 7.5 would in all likelihood be quite difficult to analyze with standard techniques.

7.3 Barzilai-Borwein method on quadratic functions

Discussions with Adrien Taylor and Moslem Zamani strongly inspired the content of this section.

The goal of this section is to exploit the idea of [BTd20] to use the Performance Estimation Problem framework to analyze an adaptive first-order method, namely, the Barzilai-Borwein method. As the method is known to diverge on general functions [BDH19], we analyze it on the class of quadratic functions. We only analyze a single iteration.

The Barzilai-Borwein method (BB) is an adaptive first-order method with the following iteration

$$x_{i+1} = x_i - \alpha_i \nabla f(x_i) \tag{7.23}$$

where α_i is either a long step or a short step

$$\begin{aligned}\alpha_i^{(long)} &= \frac{\|x_i - x_{i-1}\|^2}{(x_i - x_{i-1})^T(\nabla f(x_i) - \nabla f(x_{i-1}))},\\ \alpha_i^{(short)} &= \frac{(\nabla f(x_i) - \nabla f(x_{i-1}))^T(x_i - x_{i-1})}{\|\nabla f(x_i) - \nabla f(x_{i-1})\|^2}.\end{aligned}$$

The method works well in practice [BB88] but is not entirely understood theoretically. It converges on quadratic functions [DL02] but not on general smooth convex functions [BDH19].

Two recent ideas make it possible to analyze one iteration of the method on the class of quadratic functions with PEP:

- Development of interpolation conditions for linear operators and quadratic functions (Theorem 6.1);
- Formulating an adaptive step size via a grid search [BTd20]. In principle, we can only analyze fixed-step first-order methods with PEP and not adaptive methods. However, if the adaptive method has the form $x_{i+1} = x_i - \alpha_i \nabla f(x_i)$ and that the adaptive step α_i can be written under the form $\alpha_i = \frac{u(G)}{v(G)}$ (where G is the Gram matrix formed by the scalar products between iterates x_i and their gradients $\nabla f(x_i)$), then

we can add the constraint

$$\alpha_i v(G) = u(G) \tag{7.24}$$

and search for the worst value of α_i. It works only if the constraint $\alpha_i v(G) = u(G)$ is Gram-representable (which is the case for the long and short BB steps). See [BTd20, Section 4] for an example of this technique applied to the Polyak steps.

PEP formulation In practice, we solve the convex reformulation of the following conceptual PEP

$$\begin{aligned}
\rho(\alpha) = \max_{x_0, f} \quad & ||x_2 - x^*||^2 \\
\text{s.t.} \quad & f \in \mathcal{Q}_{\mu,L}, \\
& x_1 = x_0 - \frac{1}{L}\nabla f(x_0), \\
& x_2 = x_1 - \alpha \nabla f(x_1), \\
& \alpha(x_1 - x_0)^T(\nabla f(x_1) - \nabla f(x_0)) = \|x_1 - x_0\|^2, \\
& ||x_0 - x^*||^2 \le R^2, \\
& ||\nabla f(x^*)||^2 = 0.
\end{aligned} \tag{7.25}$$

And then, we solve $\max_{\alpha \in \left[\frac{1}{L}, \frac{1}{\mu}\right]} \rho(\alpha)$ with a grid search to obtain the worst-case performance. We set x_1 to be a classical gradient iteration.

Numerical results First, we fix $L = 1$ and $\mu = 0.5$ and we solve (7.25) for different values of $\alpha \in [1/L, 1/\mu]$ in Figure 7.6. This provide the function $\rho(\alpha)$ for $\alpha \in [1/L, 1/\mu]$. The maximum of $\rho(\alpha)$ (red dot) is the worst-case performance for the parameters $L = 1$ and $\mu = 0.5$. We repeat this for all $\mu \in [0, 1]$ to produce Figure 7.7 and obtain the worst-case performance with respect to μ.

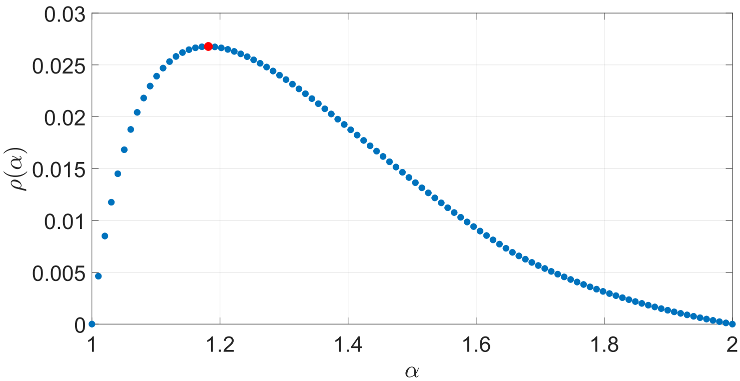


Fig. 7.6 Performance $\rho(\alpha)$ of one gradient iteration with step size $\frac{1}{L}$ and one BB iteration with long step on L-smooth μ-strongly convex quadratic functions with $L = 1$ and $\mu = 0.5$. The performance measure is $\|x_2 - x^*\|^2$.

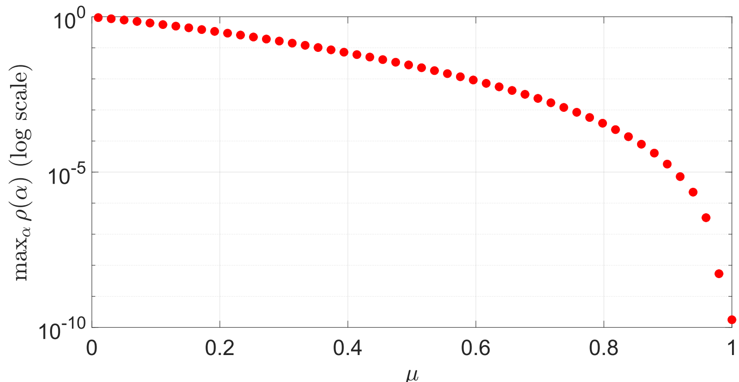


Fig. 7.7 Worst-case performance $\max_{\alpha \in \left[\frac{1}{L}, \frac{1}{\mu}\right]} \rho(\alpha)$ of one gradient iteration with step size $\frac{1}{L}$ and one BB iteration with long step on L-smooth μ-strongly convex quadratic functions with $L = 1$ for varying value of μ. The performance measure is $\|x_2 - x^*\|^2$.

In Figure 7.6, we observe $\rho(1/L = 1) = \rho(1/\mu = 2) = 0$, since the values of α beyond the interval $[1/L, 1/\mu]$ do not correspond to any quadratic function and the worst-case is thus zero.

These figures are inspired by [BTd20, Figure 2] (applied to a different setting).

Remark 7.12. Such analysis could not be performed using the spectral technique described in Section 6.1 because BB is not an FSFOM. We had to rely on PEP and the interpolation conditions for quadratic functions.

7.4 Worst-case linear operator is not always a scaling

The worst-case behavior of optimization methods on problems involving linear operators is often attained by a scalar operator of the form $M = \alpha I$ where $\alpha \in \mathbb{R}$ (see, e.g., Conjecture 7.2). However, it is not always the case, and we propose two examples where it is not the case.

Example 7.13 (Gradient method with exact line search for quadratic functions). We consider the gradient method with exact line search

$$\gamma = \arg\min_{\gamma} f(x_i - \gamma \nabla f(x_i)) \tag{7.26}$$

$$x_{i+1} = x_i - \gamma \nabla f(x_i) \tag{7.27}$$

on the class of L-smooth μ-strongly convex quadratic functions $f(x) = g(Ax) = \frac{1}{2}\|Ax\|^2 = \frac{1}{2}x^T A^T A x$ where $\mu \preceq A^T A \preceq L$, the performance measure $f(x_{i+1}) - f(x^*)$, and the initial condition $f(x_i) - f(x^*)$ where x^* is an optimal point of f. It has been shown in [dKGT17] that the worst-case performance of this method on the general quadratics is

$$f(x_{i+1}) - f(x^*) = \left(\frac{L-\mu}{L+\mu}\right)^2 (f(x_i) - f(x^*)). \tag{7.28}$$

We show that the method is strictly better on quadratic functions of the form $f(x) = \frac{\alpha}{2}\|x\|^2$ than on general quadratics. The optimality condition of the iteration yields

$$\nabla f(x_i - \gamma \nabla f(x_i))^T \nabla f(x_i) = (\alpha \|x_i\|)^2 (1 - \gamma\alpha) = 0 \tag{7.29}$$

therefore $\gamma = \frac{1}{\alpha}$, the method converges in one iteration, and $f(x_{i+1}) - f(x^*) = 0$. Therefore, the performance when the operator is a scaling operator is strictly lower (i.e., 0) than the performance among linear operators (i.e, $\left(\frac{L-\mu}{L+\mu}\right)^2 (f(x_i) - f(x^*))$).

Example 7.14 (Artificial example). We propose an artificial example that is not directly interpretable as the analysis of an optimization method but for which the worst-case cannot be a scaling.

$$\max_{\|x\|\leq 1, \|A\| \leq 1} x^T A^T A x \text{ s.t. } x^T A x = 0. \tag{7.30}$$

An optimal solution is $A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ with value 1 whereas a scaling operator $A = \alpha I$ can only yield a value of 0.

7.5 Summary

Our extension of the PEP framework 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. Then, we analyzed the adaptive Barzilai-Borwein methods on quadratic functions and obtained the numerical convergence rate. Finally, we showed that there exist settings where the worst-case linear operator is not a scaling operator.

Throughout this chapter, 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 7.1.2).
2. identification of the analytical expression of an exact convergence rate, later proved using an improvement of the classic argument (see Theorem 7.10 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 7.3 on the gradient method).
4. identification of a new analytical expression for a performance guarantee (see Figure 7.4 for the case $\sigma = \tau = 1$ on the primal-dual gap of CP with another initial distance).
5. purely numerical performance guarantee (see Figure 7.5 for CP on Lipschitz convex functions and Figure 7.7 for BB on quadratic functions).

8

Application to image processing

The results of this chapter were obtained in collaboration with Nelly Pustelnik.

MANY applications in signal processing and machine learning [CP11b, BJM+12, CP16a, PBBZP16] require to solve composite problems of the form

$$\min_{x \in \mathcal{H}} f(x) + g(x) \tag{8.1}$$

or

$$\min_{x \in \mathcal{H}} f(x) + h(Mx) \tag{8.2}$$

for a finite-dimensional Hilbert space $\mathcal{H}$ and some properties of smoothness and convexity on functions f, g, or h and a linear operator M. A large number of gradient and proximal-based algorithms have been developed to solve such minimization problems [Nem04, CP11b, Con13, CP16a, Bec17, CKCH23]. However, selecting the fastest method according to a target class of functions remains a tedious task, especially when no tight theoretical convergence rate is available.

In this chapter, we propose to apply the Performance Estimation Problem (PEP) framework [DT14, THG17c, RTBG20] to (8.1) and (8.2). This analysis allows us to obtain the exact worst-case performance for these optimization schemes, and therefore to identify the fastest strategy for a given class of functions.

8.1 PEP analysis for the composite problem $f + g$

We consider problem (8.1) when f is α^{-1}-smooth (i.e., differentiable with α^{-1}-Lipschitz gradient) ρ-strongly convex and g is β^{-1}-smooth μ-strongly convex with $\alpha, \beta \in]0, \infty[$ and $\rho, \mu \in [0, \infty[$, and we seek the largest (i.e., worst case) contraction factor r such that, for every $(x, y) \in \mathcal{H} \times \mathcal{H}$,

$$||\Phi_{f,g}x - \Phi_{f,g}y|| \leq r||x - y|| \tag{8.3}$$

where $\Phi_{f,g} : \mathcal{H} \to \mathcal{H}$ denotes an algorithmic iteration to solve (8.1). The corresponding conceptual PEP consists in

$$\begin{aligned} \max_{x,y,f,g} \quad & ||\Phi_{f,g}x - \Phi_{f,g}y||^2 \\ \text{s.t.} \quad & f \in \mathcal{F}_{\rho,\alpha^{-1}}, g \in \mathcal{F}_{\mu,\beta^{-1}}, \\ & ||x - y||^2 \leq 1, \end{aligned} \tag{8.4}$$

where $\mathcal{F}_{\rho,\alpha^{-1}}$ denotes the class of α^{-1}-smooth ρ-strongly convex functions. The optimal value of (8.4) is then equal to the square worst-case contraction factor r^2.

8.2 First-order proximal methods

PEP to validate the tightness of theoretical analysis First, we consider the cases of convex g (i.e., $\mu = 0$). We compare the following five optimization methods and recall for each the upper bound on the worst-case contraction factor r proved in [BAP23]:

- **Gradient method (GM):** Let $x \in \mathcal{H}$, $\tau \in]0, \frac{2}{\alpha^{-1}+\beta^{-1}}[$,

$$\Phi_{f,g}x = x - \tau\left(\nabla f(x) + \nabla g(x)\right), \tag{GM}$$

$$r_{\text{GM}}(\tau) \leq \max\{|1 - \tau\rho|, |1 - \tau(\beta^{-1} + \alpha^{-1})|\}. \tag{8.5}$$

- **Forward-backward splitting (FBS):** Let $x \in \mathcal{H}$, $\tau \in]0, 2\alpha[$,

$$\Phi_{f,g}x = \operatorname{prox}_{\tau g}\left(x - \tau\nabla f(x)\right), \tag{FBS1}$$

$$r_{\text{FBS1}}(\tau) \le \max\{|1-\tau\rho|, |1-\tau\alpha^{-1}|\}. \tag{8.6}$$

- **Forward-backward splitting (FBS2):** Let $x \in \mathcal{H}$, $\tau \in]0, 2\beta]$,

$$\Phi_{f,g}x = \operatorname{prox}_{\tau f}\left(x - \tau\nabla g(x)\right), \tag{FBS2}$$

$$r_{\text{FBS2}}(\tau) \le \frac{1}{1+\tau\rho}. \tag{8.7}$$

- **Peaceman-Rachford splitting (PRS):** Let $x \in \mathcal{H}$, $\tau > 0$,

$$\begin{cases} y & = \operatorname{prox}_{\tau f}(x), \\ \Phi_{f,g}x & = 2\operatorname{prox}_{\tau g}(2y - x) - 2y + x, \end{cases} \tag{PRS}$$

$$r_{\text{PRS}}(\tau) = \max_{a \in \{\rho, \alpha^{-1}\}} \frac{|1-\tau a|}{1+\tau a}. \tag{8.8}$$

- **Douglas-Rachford splitting (DRS):** Let $x \in \mathcal{H}$, $\tau > 0$,

$$\begin{cases} y & = \operatorname{prox}_{\tau f}(x), \\ \Phi_{f,g}x & = \operatorname{prox}_{\tau g}(2y - x) - y + x, \end{cases} \tag{DRS}$$

$$r_{\text{DRS}}(\tau) \le \min\left\{\frac{1+r_{\text{PRS}}(\tau)}{2}, \frac{1+\tau^2\rho\beta^{-1}}{(1+\tau\rho)(1+\tau\beta^{-1})}\right\}. \tag{8.9}$$

Figure 8.1 compares the above theoretical rates and those obtained with the PEP analysis described in (8.4). The results are displayed for three selected configurations of parameters (α, β, ρ), but are representative of a large number of additional experiments, which are not displayed here.

The numerical results obtained by PEP exactly match the above upper bounds for the first four methods (i.e., GM, FBS1, FBS2, PRS) showing that these upper bounds are tight, i.e., unimprovable. Moreover, inspection of the PEP results suggested that the worst-case functions are univariate and quadratic. Searching for the coefficients of those quadratic functions can then be formulated as a simple optimization problem (see Section

8.3), whose solution allows us to formally establish tightness of those four bounds.

Proposition 8.1. *Each upper bound* (8.5), (8.6), (8.7), *and* (8.8) *is attained by a quadratic function, and hence describes the exact (unimprovable) worst-case contraction factors of (GM), (FBS1), (FBS2), and (PRS).*

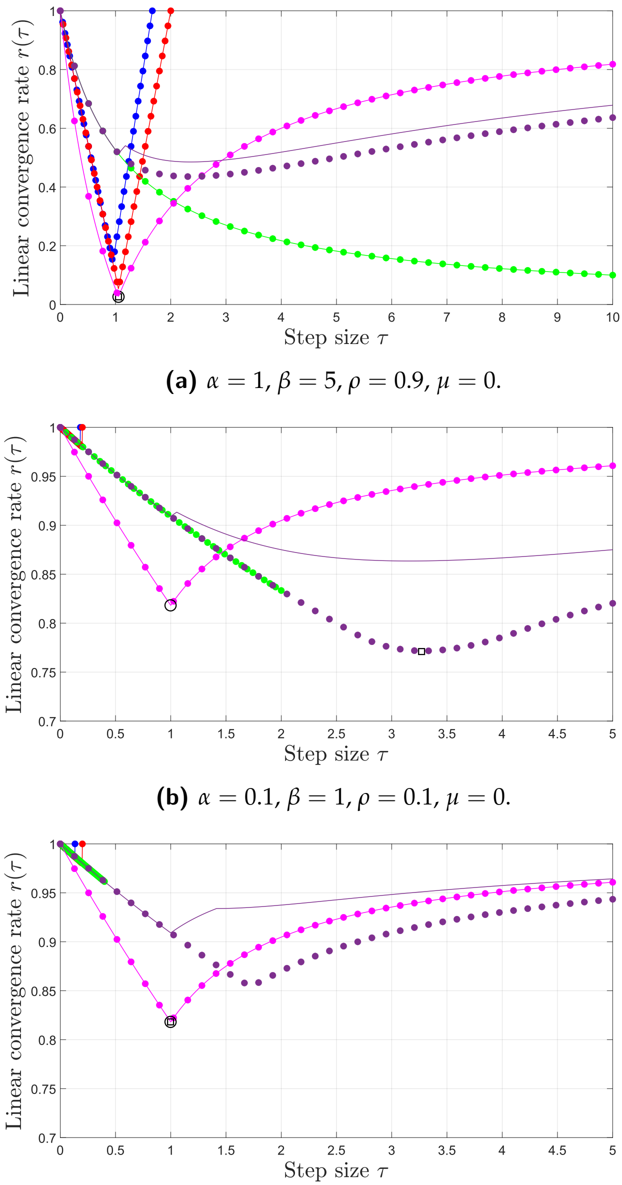
Proof. The quadratic functions are $f(x) = a\frac{x^2}{2}$ and $g(x) = b\frac{x^2}{2}$ with $a \in \{\rho, \alpha^{-1}\}$ and $b \in \{0, \beta^{-1}\}$. □

Remark 8.2. The bounds of Proposition 8.1 are tight for any number of iterations since they are attained by quadratic functions that attain these bounds for any number of iterations. In general, a bound can be tight for a single iteration and overconservative when used to bound the performance after several iterations.

For the fifth method, DRS, Figure 8.1 shows that the theoretical bound (8.9) is conservative and can thus be improved. A direct practical consequence of this conservatism is that it may lead us to make sub-optimal choices of methods. For instance, in Figure 8.1b, the upper bounds obtained by [BAP23] suggested to use PRS with the step-size $\tau = 1$ (black circle). However, DRS behaves much better than predicted by the existing bound in this case. Indeed, given the exact results provided by PEP among all methods, DRS achieves a better convergence rate when considering a step-size $\tau = 3.3$ (black square in Figure 8.1b). The configuration displayed in Figure 8.1b illustrates the drawback of relying on a non-tight analysis of the methods.

PEP to evaluate the impact of changing the regularization parameter in the objective function Knowing the exact worst-case contraction factor of the methods also allows us to understand what is happening when changing the regularizing hyperparameter in $f + \lambda h$, since it can be seen as an instance of our problem with $g = \lambda h$. The value of λ directly impacts the smoothness parameter of g. For instance, between Figure 8.1b and 8.1c, parameter β is divided by 5, which happens when hyperparameter λ is multiplied by 5. We observe that, as expected, the behaviors of FBS2, PRS, and DRS are impacted, and, considering the PEP results, the optimal method is DRS when $\beta = 1$ and PRS when $\beta = 5$.

Focus on Douglas-Rachford Following our observation that the DRS bound is not tight, we now investigate this method in more detail. Figure



(a) $\alpha = 1$, $\beta = 5$, $\rho = 0.9$, $\mu = 0$.

(b) $\alpha = 0.1$, $\beta = 1$, $\rho = 0.1$, $\mu = 0$.

(c) $\alpha = 0.1$, $\beta = 0.2$, $\rho = 0.1$, $\mu = 0$.

Fig. 8.1 Comparison of the contraction factors from [BAP23] (solid lines) and PEP (dots) of GM (blue), FBS1 (red), FBS2 (green), PRS (magenta), DRS (purple), and the optimal rate predicted by [BAP23] (black circle) and PEP (black square).

8.2 shows that the rate proved in [BAP23], namely $\frac{1}{1+\tau\rho}$ (displayed in blue), is exact for short step sizes. Looking now at larger step sizes, we find by matching the PEP numerical results and the known analytical performance of the method on simple functions (see Section 8.3) that the worst-case contraction factor is given by $\frac{1+\tau^2\alpha^{-1}\beta^{-1}}{(1+\tau\alpha^{-1})(1+\tau\beta^{-1})}$ (displayed in red). However, we could not identify the precise behavior of the contraction factor in the middle range (cf. Figure 8.2 around $\tau = 7$).

Proposition 8.3. *The exact worst-case contraction factor r_{DRS} of DRS satisfies*

$$r_{DRS}(\tau) \geq \max\left\{ \max_{\substack{a\in\{\rho,\alpha^{-1}\}\\ b\in\{0,\beta^{-1}\}}} \frac{1+\tau^2 ab}{(1+\tau a)(1+\tau b)}, T_0(\tau) \right\} \tag{8.10}$$

where $T_0(\tau)$ is unidentified.

Proof. The bound is attained by quadratic functions $f(x) = a\frac{x^2}{2}, g(x) = b\frac{x^2}{2}$ with $a \in \{\rho, \alpha^{-1}\}$, $b \in \{0, \beta^{-1}\}$ and unidentified functions $f(x) = s_0(x)$ and $g(x) = t_0(x)$ such that $||\phi_{s_0,t_0}x - \phi_{s_0,t_0}y|| = T_0(\tau)||x-y||$ for some x, y. □

Generalization to strongly convex f and g Based on PEP analysis, we extend the algorithmic comparison proposed in [BAP23] to the case where g is also strongly convex (see Figure 8.3), which was not covered in [BAP23]. Similarly to the convex case $\mu = 0$, we prove in this setting a lower bound on the worst-case contraction factor for the five methods. We summarize the identified analytical expressions in the following proposition.

Proposition 8.4. *The exact worst-case contraction factor of GM, FBS1, FBS2, PRS, and DRS respectively satisfy*

$$r_{GM}(\tau) \geq \max\{|1-\tau(\rho+\mu)|, |1-\tau(\beta^{-1}+\alpha^{-1})|\}, \tag{8.11}$$

$$r_{FBS1}(\tau) \geq \frac{1}{1+\tau\mu}\max\{|1-\tau\rho|, |1-\tau\alpha^{-1}|\}, \tag{8.12}$$

$$r_{FBS2}(\tau) \geq \frac{1}{1+\tau\rho}\max\{|1-\tau\mu|, |1-\tau\beta^{-1}|\}, \tag{8.13}$$

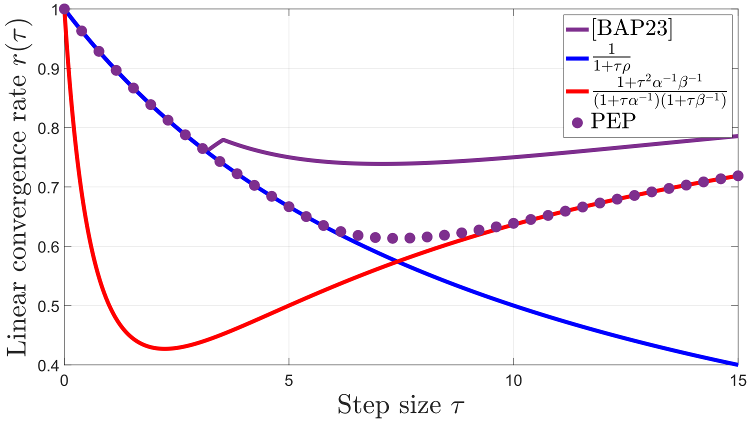


Fig. 8.2 Worst-case contraction factor of DRS for $\alpha = 1$, $\beta = 5$, $\rho = 0.1$, $\mu = 0$ from [BAP23] (purple solid line) and PEP (purple dots). We identified two regimes (blue and red solid lines).

$$r_{PRS}(\tau) \geq \max_{\substack{a\in\{\rho,\alpha^{-1}\}\\ b\in\{\mu,\beta^{-1}\}}} \frac{|1-\tau a|}{1+\tau a}\frac{|1-\tau b|}{1+\tau b}, \tag{8.14}$$

$$r_{DRS}(\tau) \geq \max\left\{\max_{\substack{a\in\{\rho,\alpha^{-1}\}\\ b\in\{\mu,\beta^{-1}\}}} \frac{1+\tau^2 ab}{(1+\tau a)(1+\tau b)}, T_\mu(\tau)\right\}, \tag{8.15}$$

where T_μ is unidentified.

Proof. All the bounds are attained by quadratic functions $f(x) = a\frac{x^2}{2}$, $g(x) = b\frac{x^2}{2}$ with $a \in \{\rho, \alpha^{-1}\}$, $b \in \{\mu, \beta^{-1}\}$ and, for DRS only, unidentified functions $f(x) = s_\mu(x)$, $g(x) = t_\mu(x)$ such that $||\phi_{s_\mu,t_\mu}x - \phi_{s_\mu,t_\mu}y|| = T_\mu(\tau)||x-y||$ for some x, y. □

The numerical results of all our experiments were consistent with the contraction factors used as lower bounds in Proposition 8.4. Therefore, we conjecture that they are exact. As for the case $\mu = 0$, we were not able to identify the regimes T_μ of DRS nor the associated worst-case functions s_μ and t_μ.

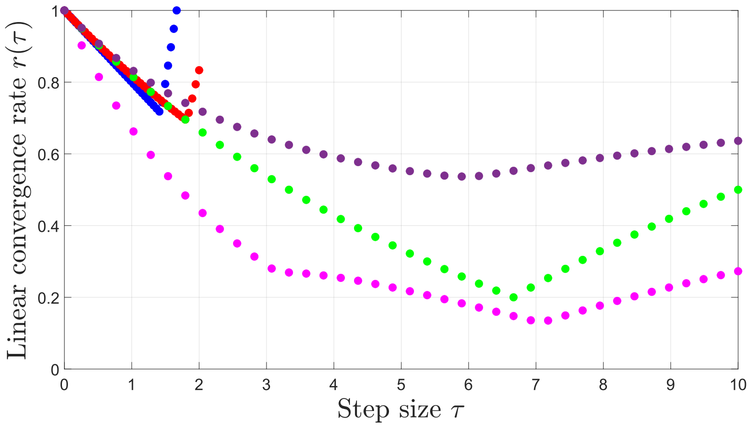


Fig. 8.3 Contraction factors obtained by PEP of GM (blue), FBS1 (red), FBS2 (green), PRS (magenta), DRS (purple) in the doubly strongly convex case $\alpha = 1$, $\beta = 5$, $\rho = \mu = 0.1$.

8.3 Linear rate on quadratic functions

Interestingly, all worst cases identified so far are attained by univariate quadratic functions. The contraction factor of a method on the class of quadratic functions can be explicitly computed by solving a simple maximization problem (using the same idea as the spectral analysis presented in Section 6.1).

For example, let us consider PRS applied to the problem $f(x) = a\frac{x^2}{2}$ and $g(x) = b\frac{x^2}{2}$. The proximal step on a quadratic function $f(x) = a\frac{x^2}{2}$ can be explicitly computed as

$$\mathrm{prox}_{\tau f}(x) = \frac{x}{1 + \tau a}. \tag{8.16}$$

Therefore, after simplifications, PRS becomes

$$\Phi_{f,g} x = \frac{1 - \tau a}{1 + \tau a} \frac{1 - \tau b}{1 + \tau b} x. \tag{8.17}$$

In this case, problem (8.4) of computing the worst-case contraction factor reduces to

$$r^2(\tau) = \max_{\substack{a \in [\rho, \alpha^{-1}] \\ b \in [\mu, \beta^{-1}]}} \left| \frac{1 - \tau a}{1 + \tau a} \frac{1 - \tau b}{1 + \tau b} \right|^2, \tag{8.18}$$

namely, a simple optimization problem involving two variables a and b that can often be solved analytically. The same methodology applies to the other methods and also to general quadratic functions (not necessarily univariate).

This methodology allowed us to find the analytical expressions of all the contraction factors gathered in Proposition 8.4 since the worst-case is achieved with a quadratic function for each of these methods (except for an unidentified non-quadratic worst-case of DRS). Nonetheless, it is important to note that it was not known a priori that the worst-case contraction factors of the methods over general smooth convex functions are attained by quadratic functions. The analysis and numerical results provided by PEP are critical to reach this conclusion.

Conjecture 8.5. *The worst-case contraction factor of GM, FBS1, FBS2, and PRS on the problem* $\min_x f(x) + g(x)$ *where* $f \in \mathcal{F}_{\rho,\alpha^{-1}}$ *and* $g \in \mathcal{F}_{\mu,\beta^{-1}}$ *are attained by univariate quadratic functions* f *and* g*.*

8.4 Case $f + h \circ M$

We now move to composite functions of the form (8.2) involving a linear operator, whose inclusion in the PEP framework was made possible by the very recent work [BHG24]. Specifically, we assume that f is α^{-1}-smooth ρ-strongly convex, h is γ^{-1}-smooth δ-strongly convex, and matrix M such that $||M|| \triangleq \sigma_{\max}(M) \leq L$ with $\alpha, \rho, L \in]0, \infty[$ and $\rho, \delta \in [0, \infty[$. This is a particular case of problem (8.1), therefore, the previous results still apply. However, the composite structure of $g = h \circ M$ can in principle lead to stronger results. We still consider the five same methods and want to quantify the potential gain of performance when applying a method on (8.2).

The only setting where an improvement could be expected is for DRS. Indeed, for the four other methods, the worst-case functions on the class of functions of the form $f + g$ are quadratic. Since all quadratic functions g can be expressed as a composite function of the form $h \circ M$, the worst-case performance cannot be improved. However, since the worst-case functions of DRS are not identified, this leaves a possibility for an improved rate in the composite case. Still, the PEP numerical results are the same as in Section 8.2, suggesting that there is actually no improvement. It indicates that there exists a worst-case function in the case $f + g$ for which g is of the form $h \circ M$ (and that is not a quadratic function).

8.5 Primal-dual methods

Unlike the results of [BAP23] for primal methods, no theoretical comparison for proximal primal-dual algorithms seems to be available in the literature so far. We now show how the recent extension of PEP allows us to numerically compute these rates and, therefore, to compare methods. We focus on two primal-dual methods that exploit access to linear operators in their iterations:

- **Chambolle-Pock method (CP)** [CP11a]: Let x, u, and τ, σ such that $\sigma\tau||M||^2 \leq 1$:

$$\begin{cases} x_+ & = \text{prox}_{\tau f}(x - \tau M^T u), \\ u_+ & = \text{prox}_{\sigma h^*}\left(u + \sigma M(2x_+ - x)\right). \end{cases} \tag{8.19}$$

- **Condat-Vũ method (CV)** [Con13, Vũ13]: Let x, u and τ, σ such that $\frac{1}{\tau} - \sigma||M||^2 \geq \frac{1}{2\alpha}$ (where f is assumed to be α^{-1}-smooth):

$$\begin{cases} x_+ & = x - \tau\nabla f(x) - \tau M^T u, \\ u_+ & = \text{prox}_{\sigma h^*}\left(u + \sigma M(2x_+ - x)\right). \end{cases} \tag{8.20}$$

For such primal-dual methods, we are interested in the contraction factor r of the primal-dual iterates (x, u), defined as

$$||\phi_{f,g}(x,u) - \phi_{f,g}(x',u')|| \leq r||(x,u) - (x',u')||, \tag{8.21}$$

which cannot be directly compared to the primal contraction factor (8.3). Similarly to Section 8.4, such worst-case analysis can be formulated and obtained with PEP. Figure 8.4 below depicts the worst-case contraction factor of CP (with $\sigma(\tau) = \frac{1}{\tau||M||^2}$) and CV (with $\sigma(\tau) = \frac{1}{\tau||M||^2} - \frac{1}{2\alpha||M||^2}$) on problem (8.2) for $\delta = 0$ and $\delta = 0.1$.

We observe the impact of the presence of strong convexity in h. In these two specific settings, CP performs better.

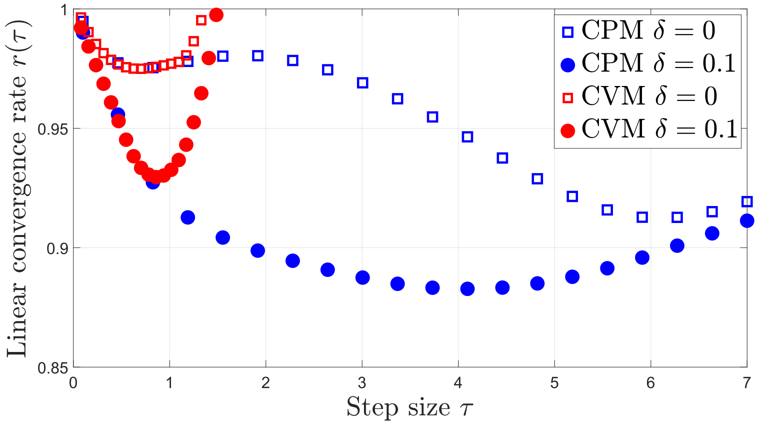


Fig. 8.4 Contraction factors of CP (blue) and CV (red) obtained by PEP for $\alpha = 1$, $\beta = 5$, $\rho = 0.1$, $\delta = 0$ (squares) and $\delta = 0.1$ (dots) and $||M|| \leq 1$.

8.6 Summary

We have shown how the PEP framework allows to easily certify the tightness of (or, in one case, to improve) recent bounds on the contraction factor of the gradient method and the forward-backward, Peaceman-Rachford, and Douglas-Rachford splitting algorithms, applied to smooth convex functions. We also perform the same analysis in the strongly convex case, obtaining again tight numerical rates. The methodology also highlights interesting and a priori unexpected phenomena, such as the existence of quadratic worst-case functions.

This type of analysis can be extended to any other method, function class, and performance criterion. We demonstrate this possibility on Chambolle-Pock and Condat-Vũ methods opening the way for more extended analysis in further works.

A

Appendix

A.1 Properties of matrices

We review some results of linear algebra used in our proofs. Recall that we note I_d and $0_{m,n}$ for identity and zero matrices (dimension may be omitted), $M^\dagger$ and $M^{\frac{1}{2}}$ for the pseudo-inverse and square root of M, and $||M|| = \sigma_{\max}(M)$ for the norm of M.

Proposition A.1 ([BV04]; [Gal10], Theorem 4.3). *Let $G = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$ symmetric. We have the following three equivalences*

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

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 pseudo inverse and square root matrix. □

Proposition A.3. *Let two matrices 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 $A \succeq 0$ 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 $A \succeq 0$ 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 $\exists L_1 > 0 \;:\; C \preceq L_1^2 A \Leftrightarrow AA^\dagger C = C \Leftrightarrow AA^\dagger C^2 = C^2 \Leftrightarrow \exists L_2 > 0 \;:\; C^2 \preceq L_2^2 A$. □

Proposition A.6. *Let two symmetric matrices $B \succeq 0$ and $X \succeq 0$ and $A = B + X$. We have*

$$A^\dagger A B^\dagger = B^\dagger. \tag{A.4}$$

Proof. Let $\mathcal{C}(M)$ and $\mathcal{N}(M)$ denote the range and nullspace of the matrix M. We want to show that $(I - A^\dagger A)B^\dagger = 0$, in other words, that $\mathcal{C}(B^\dagger) \subseteq \mathcal{N}(I - A^\dagger A)$.

Firstly, by definition of the pseudo-inverse and symmetry of B, we have $\mathcal{C}(B^\dagger) = \mathcal{C}(B^T) = \mathcal{C}(B)$. Secondly, since $A(I - A^\dagger A) = 0$, we have $\mathcal{N}(I - A^\dagger A) = \mathcal{C}(A)$. Therefore, we have

$$\begin{aligned} A^\dagger A B^\dagger = B^\dagger &\Leftrightarrow \mathcal{C}(B^\dagger) \subseteq \mathcal{N}(I - A^\dagger A) \\ &\Leftrightarrow \mathcal{C}(B) \subseteq \mathcal{C}(A) \\ &\Leftrightarrow \mathcal{N}(A) \subseteq \mathcal{N}(B). \end{aligned} \tag{A.5}$$

Therefore, it remains to show that $\mathcal{N}(A) \subseteq \mathcal{N}(B)$.

$$\begin{aligned} x \in \mathcal{N}(A) \Leftrightarrow Ax = 0 & \\ & \Rightarrow x^T A x = \overbrace{x^T B x}^{\geq 0} + \overbrace{x^T X x}^{\geq 0} = 0 \\ & \Rightarrow x^T B x = 0 \\ & \Leftrightarrow x \in \mathcal{N}(B). \end{aligned} \tag{A.6}$$

□

The following result allows to extend a block matrix without increasing its maximal singular value.

Theorem A.7 ([DKW82], 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*

$$\left\| \begin{pmatrix} M_1 & M_2 \end{pmatrix} \right\| \leq L \text{ and } \left\| \begin{pmatrix} M_1 \\ M_3 \end{pmatrix} \right\| \leq L. \tag{A.7}$$

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, $\left\| \begin{pmatrix} 1 & 1 \end{pmatrix} \right\| = \sqrt{2}$, $\left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\| = \sqrt{2}$, $\left\| \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \right\| = \frac{1+\sqrt{5}}{2} \approx 1.618 > \sqrt{2}$ 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.

Theorem A.8 ([HJ12], 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*

$$\exists V \in \mathbb{R}^{d \times d} \text{ unitary } : B = VA. \tag{A.8}$$

A.2 Extension of matrix

We propose an alternative proof of Theorem A.7.

Proof of Theorem A.7.
(Necessity) $\left\| \begin{pmatrix} M_1 & M_2 \\ M_3 & W \end{pmatrix} \right\| \leq L \Rightarrow || \begin{pmatrix} M_1 & M_2 \end{pmatrix} || \leq L$ and $\left|\left| \begin{pmatrix} M_1 \\ M_3 \end{pmatrix} \right|\right| \leq L$

If there is no unitary vector x such that $\left|\left| \begin{pmatrix} M_1 & M_2 \\ M_3 & W \end{pmatrix} x \right|\right| > L$, then, there is no unitary vectors y_1 or y_2 such that $|| \begin{pmatrix} M_1 & M_2 \end{pmatrix} y_1 || > L$ nor $\left|\left| \begin{pmatrix} M_1 \\ M_3 \end{pmatrix} y_2 \right|\right| > L$.

(Sufficiency) $|| \begin{pmatrix} M_1 & M_2 \end{pmatrix} || \leq L$ and $\left|\left| \begin{pmatrix} M_1 \\ M_3 \end{pmatrix} \right|\right| \leq L \Rightarrow \left\| \begin{pmatrix} M_1 & M_2 \\ M_3 & W \end{pmatrix} \right\| \leq L$

First of all, we observe that

$$|| \begin{pmatrix} M_1 & M_2 \end{pmatrix} || \leq L \Leftrightarrow \begin{pmatrix} M_1 & M_2 \end{pmatrix} \begin{pmatrix} M_1^T \\ M_2^T \end{pmatrix} = M_1 M_1^T + M_2 M_2^T \preceq L^2 I \quad \text{(A.9)}$$

and

$$\begin{aligned} \left|\left| \begin{pmatrix} M_1 \\ M_3 \end{pmatrix} \right|\right| \leq L \Leftrightarrow \begin{pmatrix} M_1 \\ M_3 \end{pmatrix} \begin{pmatrix} M_1^T & M_3^T \end{pmatrix} &= \begin{pmatrix} M_1 M_1^T & M_1 M_3^T \\ M_3 M_1^T & M_3 M_3^T \end{pmatrix} \preceq L^2 I \\ &\overset{\text{(A.1)}}{\Leftrightarrow} \begin{cases} L^2 I - M_1 M_1^T & \succeq 0 \\ L^2 I - M_3 M_3^T - M_3 M_1^T (L^2 I - M_1 M_1^T)^\dagger M_1 M_3^T & \succeq 0 \\ (I - (L^2 I - M_1 M_1^T)(L^2 I - M_1 M_1^T)^\dagger) M_1 M_3^T & = 0 \end{cases} \end{aligned} \quad \text{(A.10)}$$

We want a W such that the singular values of $M = \begin{pmatrix} M_1 & M_2 \\ M_3 & W \end{pmatrix}$ (i.e., square roots of the eigenvalues of MM^T) are lower than L, namely,

$$W \; : \; MM^T \preceq L^2 I. \quad \text{(A.11)}$$

We propose a W such that the following matrix is positive semidefinite

$$T = \begin{pmatrix} T_1 & T_2(W) \\ T_2^T(W) & T_3(W) \end{pmatrix} \triangleq L^2 I - MM^T$$
$$= \begin{pmatrix} L^2 I - M_1 M_1^T - M_2 M_2^T & -M_1 M_3^T - M_2 W^T \\ -M_3 M_1^T - W M_2^T & L^2 I - M_3 M_3^T - WW^T \end{pmatrix} \tag{A.12}$$

or equivalently by Proposition A.1

$$\begin{cases} T_1 & \succeq 0, \\ T_3(W) - T_2^T(W) T_1^\dagger T_2(W) & \succeq 0, \\ (I - T_1 T_1^\dagger) T_2(W) & = 0, \end{cases} \tag{A.13}$$

where the first equation is satisfied by (A.9). We note $T_0 \triangleq L^2 I - M_1 M_1^T$ and show that the solution $W = -M_3 M_1^T T_0^\dagger M_2$ satisfies (A.13). Note the identity $T_0 - T_1 = M_2 M_2^T$.

(i) Verifying $T_3(W) - T_2^T(W) T_1^\dagger T_2(W) \succeq 0$:

We show that $T_3(W) \succeq T_2^T(W) T_1^\dagger T_2(W)$. Using (A.12) and $T_0 - T_1 = M_2 M_2^T$ it is equivalent to

$$\overbrace{L^2 I - M_3 M_3^T - M_3 M_1^T \underbrace{T_0^\dagger T_0 T_0^\dagger}_{T_0^\dagger} M_1 M_3^T}^{\succeq 0 \text{ by (A.10)}} + M_3 M_1^T T_0^\dagger T_1 T_0^\dagger M_1 M_3^T \succeq T_2^T(W) T_1^\dagger T_2(W).$$

therefore, it remains to show that $M_3 M_1^T T_0^\dagger T_1 T_0^\dagger M_1 M_3^T = T_2^T(W) T_1^\dagger T_2(W)$. Using (A.12) yields

$$M_3 M_1^T T_0^\dagger T_1 T_0^\dagger M_1 M_3^T = M_3 M_1^T \left(I - T_0^\dagger M_2 M_2^T\right) T_1^\dagger \left(I - M_2 M_2^T T_0^\dagger\right) M_1 M_3^T \tag{A.14}$$

which holds thanks to the identities $T_0^\dagger T_1 T_0^\dagger = (I - T_0^\dagger M_2 M_2^T) T_1^\dagger (I - M_2 M_2^T T_0^\dagger)$, $M_2 M_2^T = T_0 - T_1$ and $T_1^\dagger = T_0^\dagger T_0 T_1^\dagger$ (which holds by Proposition A.6 since $T_0 = T_1 + M_2 M_2^T$ where T_1 and $M_2 M_2^T$ are both positive semidefinite).

(ii) Verifying $(I - T_1 T_1^\dagger) T_2(W) = 0$:

We show that $T_1 T_1^\dagger T_2(W) = T_2(W)$

$$
\begin{aligned}
T_1 T_1^\dagger T_2(W) &\overset{(A.12)}{=} T_1 T_1^\dagger (-M_1 M_3^T + \overbrace{M_2 M_2^T}^{T_0 - T_1} T_0^\dagger M_1 M_3^T) \\
&= T_1 T_1^\dagger (\overbrace{-M_1 M_3^T + T_0 T_0^\dagger M_1 M_3^T}^{=0 \text{ by (A.10)}} - T_1 T_0^\dagger M_1 M_3^T) \\
&= -\overbrace{T_1 T_1^\dagger T_1}^{T_1} T_0^\dagger M_1 M_3^T \\
&= -\overbrace{(T_0 - M_2 M_2^T)}^{T_1} T_0^\dagger M_1 M_3^T \\
&= -T_0 T_0^\dagger M_1 M_3^T + M_2 M_2^T T_0^\dagger M_1 M_3^T \\
&= -M_1 M_3^T + M_2 M_2^T T_0^\dagger M_1 M_3^T \\
&\overset{(A.12)}{=} T_2(W)
\end{aligned}
\tag{A.15}
$$

where we used $M_1 = T_0 T_0^\dagger M_1$ by Proposition A.6.

(iii) Expression, symmetry and skew-symmetry of W:

We conclude the proof by exhibiting an alternative expression of W and showing its symmetry and skew-symmetry in the required conditions. Computing the singular value decomposition of M_1 yields

$$
\begin{aligned}
M_1 &= U \Sigma V^T \\
M_1 M_1^T &= U \Sigma \Sigma^T U^T \\
L^2 I - M_1 M_1^T &= U (L^2 I - \Sigma \Sigma^T) U^T \\
(L^2 I - M_1 M_1^T)^\dagger &= U (L^2 I - \Sigma \Sigma^T)^\dagger U^T
\end{aligned}
\tag{A.16}
$$

$$
M_1^T T_0^\dagger = M_1^T (L^2 I - M_1 M_1^T)^\dagger = V \overbrace{\Sigma^T (L^2 I - \Sigma \Sigma^T)^\dagger}^{D} U^T
$$

where D is a diagonal matrix. Therefore, W can be written as

$$
W = -M_3 M_1^T T_0^\dagger M_2 = -M_3 V D U^T M_2 \tag{A.17}
$$

If M_1 is symmetric and $M_2 = M_3^T$ then W is symmetric, namely,

$$
W = -M_2^T U D U^T M_2 = W^T \tag{A.18}
$$

and if M_1 is skew-symmetric and $M_2 = -M_3^T$ then W is skew-symmetric, namely,

$$W = M_2^T U D U^T M_2 = -W^T. \tag{A.19}$$

□

Remark A.9. Using $W = -M_3 M_1^T (L^2 I - M_1 M_1^T)^\dagger M_2 = -M_3 V D U^T M_2$ where $M_1 = U \Sigma V^T$ and $D = \Sigma^T (L^2 I - \Sigma \Sigma^T)^\dagger$, we can express the whole matrix as

$$\begin{aligned} \begin{pmatrix} M_1 & M_2 \\ M_3 & W \end{pmatrix} &= \begin{pmatrix} U\Sigma V^T & UU^T M_2 \\ M_3 V V^T & -M_3 V D U^T M_2 \end{pmatrix} \\ &= \begin{pmatrix} U & 0 \\ 0 & M_3 V \end{pmatrix} \begin{pmatrix} \Sigma & I \\ I & -D \end{pmatrix} \begin{pmatrix} V^T & 0 \\ 0 & U^T M_2 \end{pmatrix}. \end{aligned} \tag{A.20}$$

PART II
Non-convex Performance Estimation

9

Motivation and limitations of convex PEP framework

THE Performance Estimation Problem (PEP) framework formulates the problem of finding the worst-case performance of an optimization method on a function class as an optimization problem itself. For a lot of first-order optimization methods, this formulation is a convex semidefinite program [THG17c, GMG+24]. However, when the problem is not convex, it can still be formulated as a non-convex quadratically constrained quadratic program. Recently, it has been made possible to solve globally non-convex problems (e.g., Gurobi's branch and bound [Gur24]).

This offers a new path for the utilization of the PEP framework [BTd20, DGFST24, DGVPR24b, DG24, RS25, LG24, RTBG20]. If we accept the possibly expensive computational cost of solving non-convex problems, we can analyze new problems that were inaccessible through the classical convex PEP framework. Note that we are still limited by the existence and knowledge of necessary and sufficient interpolation conditions of the classes of interest [RBCH23].

We present what kind of problems can be solved and propose a list of examples of new methods that could be analyzed with the non-convex PEP framework and the current available solvers.

9.1 Formulation of a non-convex PEP

In the convex PEP framework, the variables are the function values, f_k, and the scalar products between gradients and iterates, i.e., $x_i^T x_k$, $x_i^T g_k$ and $g_i^T g_k$. The objective function of the PEP must be linear on these variables (e.g., $f_N - f^*$, $||x_N - x^*||^2$, etc). The constraints of the PEP must be linear (or at least convex) on the variables or semidefinite constraints on (the blocks of) the Gram matrix containing all the scalar products. Importantly, in the convex PEP formulation, we do not specify the dimension of the function and search for the worst-case function f of any dimension.

In the non-convex PEP, the variables are the function values f_k, the gradients g_k, and the iterates x_k themselves (and possibly auxiliary variables). Therefore, their dimensions and the dimension of the function f are specified. The objective function and the constraints can be non-homogeneous (non-convex) quadratic in the variables. Thanks to the introduction of auxiliary variables and constraints, we can also represent products, quotients, n^{th} root of variables, etc. Here are some examples of reformulation ending up in non-convex quadratic constraints:

- Product xyz can be replaced by product uz and adding constraint $u = xy$;
- Ratio $\frac{x}{y}$ can be replaced by variable u and adding constraint $yu = x$;
- Square root $\sqrt{x}$ can be replaced by u and adding constraints $u^2 = x$ and $u \geq 0$.

Remark 9.1. The current version of Gurobi handles integer variables, thus allowing modeling *if conditions*.

The two main limitations of the convex formulation of a convex PEP are (i) each iteration of the method must lead to a linear combination of the iterates and their gradients and (ii) the access to explicit and convex interpolation conditions of the class of interest. In the non-convex PEP framework, we can formulate the iterations and interpolation conditions as any non-convex quadratic constraints. We will show that it allows to analyze different methods.

Remark 9.2. In the convex PEP framework, we work with dimension-free functions, whereas in the non-convex PEP framework, we have to explicitly choose *a priori* the dimension. Of course, increasing the dimension of the functions increases the dimension of the problem solved.

9.2 Applications of the non-convex PEP framework

We now show how we can exploit this tool to analyze new problems.

- **Second-order optimization**: One of the reasons why it is difficult to analyze second-order methods with the convex PEP framework is the representation of the Hessian. In the non-convex PEP framework, we can just add a variable h_k to represent the Hessian at the point x_k. However, the interpolation conditions for relevant classes of second-order optimization are challenging. An extensive analysis of second-order methods with non-convex PEP is performed in Chapter 10.

- **High-order methods**: We could also analyze third- (and higher-) order methods [Doi21] by introducing the needed derivatives as variable of the non-convex PEP. However, it will require interpolation conditions on all the derivatives, which is already challenging for the second-order case.

- **Zeroth-order methods**: We can also consider zeroth-order methods [CSV09] with the non-convex PEP framework. An example of a zeroth-order method for a univariate function is

$$x_{k+1} = x_k - \frac{f(x_k + h) - f(x_k)}{h}. \tag{9.1}$$

  The limitation of the convex PEP framework is that it works in a dimension-free fashion whereas here we could fix the dimension of f to $d = 1$.

- **Subgradient method with constant step lengths**: The subgradient method [Sho12]

$$x_{k+1} = x_k - t\frac{1}{||g_k||}g_k \tag{9.2}$$

  with fixed step size t can be formulated with quadratic constraints as

$$\begin{cases} x_{k+1} &= x_k - tp_k, \\ u_k^2 &= ||g_k||^2, \\ u_k &\geq 0, \\ u_k p_k &= g_k. \end{cases} \tag{9.3}$$

- **Quasi-Newton and adaptive methods**: We consider the family of quasi-Newton methods [Rod22] satisfying

$$\begin{cases} x_{k+1} = x_k - H_k \nabla f(x_k), \\ H_k(\nabla f(x_{k+1}) - \nabla f(x_k)) = x_{k+1} - x_k, \end{cases} \tag{9.4}$$

which consists of requiring the secant equation. In this case, it may be more interesting to consider functions f with at least two variables since every quasi-Newton method reduces to the same method in the univariate case (i.e., the secant method).

We can analyze (almost) any adaptive methods. For example, we can consider the adaptive gradient descent (with approximation of the local Lipschitz constant) [MM20, Algorithm 1]

$$\begin{cases} h_k & = \min\left\{ \sqrt{1+\theta_{k-1}} h_{k-1}, \frac{||x_k - x_{k-1}||}{2||\nabla f(x_k) - \nabla f(x_{k-1})||} \right\}, \\ x_{k+1} & = x_k - h_k \nabla f(x_k), \\ \theta_k & = \frac{h_k}{h_{k-1}}. \end{cases} \tag{9.5}$$

Remark 9.3. This kind of adaptive method could still be analyzed with the convex PEP framework and the ideas of [BTd20] (see, e.g., Section 7.3 on Barzilai-Borwein method).

- **Penalty method**: The family of penalty methods [NW06] consists of solving a sequence of unconstrained problems to solve a constrained problem. For example, the quadratic penalty method solves the problem

$$\begin{aligned} \min_x \ & f(x) \\ & c(x) = 0, \end{aligned} \tag{9.6}$$

by solving

$$x_{k+1} = \arg\min_x f(x) + \frac{\sigma_k}{2}\|c(x)\|^2, \tag{9.7}$$

and increasing σ_k until $\|c(x_{k+1})\|^2 = 0$.

- **Methods with linear operators**: We can introduce the variable M to model a linear operator and analyze methods applied to problems involving linear operators. For example, we can consider the gradient method on functions $g(Mx)$ with the iteration

$$x_{k+1} = x_k - M^T \nabla g(Mx) \tag{9.8}$$

or any primal-dual methods.

Remark 9.4. This approach would be complementary to the one presented in Part I. It would allow more flexibility on the representation of linear operators, but at a possibly heavy computational cost.

- **Finding the minimal dimension of a worst-case function**: Finally, the last application is a complementary tool to the convex PEP framework. When we have a method and a class that we can analyze with the convex PEP framework, we obtain the worst-case performance together with an actual function reaching this worst-case. The function can be any worst-case function, in particular, there is no guarantee (yet very good heuristics) that this function has the minimal dimension among the worst ones. If we know the worst-case performance, we can solve the non-convex PEP by increasing the dimension until reaching this worst-case performance, meaning that we have found the smallest dimension for the worst-case.

  Moreover, if the worst-case performance is not attained by a univariate function, it can still provide the worst univariate function of the class, which can be of interest. In Section B.1, we propose a list of settings where the worst-case is attained by a univariate function and a list of settings where there is no univariate function attaining the worst-case performance.

10

Analysis of second-order optimization

The results of this chapter were obtained in collaboration with Anne Rubbens. Discussions with Adrien Taylor, Moslem Zamani, Nikita Doikov, Shuvomoy Das Gupta, and Anton Rodomanov inspired some of the results of this chapter.

10.1 Introduction

Consider the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^d} f(x), \tag{10.1}$$

where f belongs to a given function class $\mathcal{F}$, e.g, the class of twice-differentiable functions $f : \mathbb{R}^d \to \mathbb{R}$ whose Hessian is M-Lipschitz continuous, i.e,

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \le M\|x - y\|, \quad \forall x, y \in \mathbb{R}^d. \tag{10.2}$$

To solve (10.1), consider N iterations of a black-box (oracle-based) second-order optimization method $\mathcal{M}$, e.g., Newton's method, i.e,

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k).$$

Our goal is to answer the following question.

What is the worst-case convergence rate after N iterations of method $\mathcal{M}$ on the class $\mathcal{F}$?

A worst-case convergence rate of a method on a function class is an upper bound on how quickly the method converges to a (local or global) minimum for any function in the class. This bound is said to be *tight* or *exact* when it is attained by at least one function in the class under consideration. Obtaining an exact rate allows understanding the possible sources of inefficiency and to tune a given method, by choosing its parameters as to minimize the rate. In addition, obtaining the exact convergence rate of different methods provides a basis for accurately comparing these methods. In this chapter, we extend the Performance Estimation framework [DT14, THG17c] that allows to automatically compute tight convergence rates of first-order methods, to the tight analysis of second-order methods applied to *univariate* functions.

10.1.1 Second-order optimization

Second-order methods such as Newton's method or Cubic Regularized Newton method [NP06], exploit evaluations of the Hessian of the objective function. They have attracted a lot of attention due to their local quadratic convergence and importance for interior-point methods [CGT00, NN94] based on self-concordant functions. Several globally convergent variants of Newton's method were then also proposed on different classes of functions [CGT11, DMN24, HKP$^+$22, Hil21, IH24, Mis23, NP06, Pol09].

The potential tightness of existing convergence rates is an open question in general. In practice, the methods often behave significantly better than predicted by the theory, either because worst-case instances are very unlikely to appear in practical problems, or because the current convergence rates are too pessimistic and can be improved. In addition, a fair comparison between the different variants of Newton's method is difficult to perform, in the sense that the performance measures and hypotheses vary for each analysis. A tight and systematic analysis of these variants on a few well-chosen performance measures would thus allow improving on the existing bounds, provide insight into their level of pessimism, and allow for a fairer comparison. A more detailed state-of-the-art about the methods we analyze is provided in Section 10.5.

10.1.2 Obtaining convergence rates: the Performance Estimation Framework

For a wide range of methods, the convergence rate after a fixed number of iterations can be computed via the Performance Estimation Problem (PEP) approach [DT14, THG17c], which builds on the following observation: in principle, our main question can be reformulated as the following type of problem, where we select the distance to the minimizer as the performance measure:

$$
\begin{aligned}
& \max_{x_0, x_\star \in \mathbb{R}^d, f \in \mathcal{F}} ||x_N - x_\star|| \\
\text{Method's definition:} \quad & \text{s.t. } x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k), \quad k = 0, ..., N-1, \\
\text{Initial condition:} \quad & ||x_0 - x_\star|| \leq R, \\
\text{Local optimality:} \quad & \nabla f(x_\star) = 0, \nabla^2 f(x_\star) \succeq 0.
\end{aligned}
\tag{10.3}
$$

That is, we are looking for the function f, initial point x_0, and local optimizer $x_\star$ leading to the worst performance of the method $\mathcal{M}$, and hence ensure that $\mathcal{M}$ behaves better or equivalently on all functions of the class $\mathcal{F}$. Problem (10.3) is an infinite-dimensional problem, but [DT14, THG17c] have shown how it admits a finite reformulation. The idea is to replace optimization over $f \in \mathcal{F}$ by optimization over the set $S = \{(x_k, g_k, h_k, f_k)\}_{k=0,...,N,\star}$ containing all iterates x_k (and the minimizer $x_\star$) along with the corresponding gradients g_k, Hessians h_k, and function values f_k, under the constraint that S is consistent with an actual function in $\mathcal{F}$. Indeed, (10.3) is equivalent to

$$
\begin{aligned}
& \max_{S = \{(x_k, f_k, g_k, h_k)\}_{k=0,...,N,\star}} ||x_N - x_\star|| \\
\text{Method's definition:} \quad & \text{s.t. } x_{k+1} = x_k - h_k^{-1} g_k, \quad k = 0, ..., N-1, \\
\text{Initial condition:} \quad & ||x_0 - x_\star|| \leq R, \\
\text{Local optimality:} \quad & g_\star = 0, h_\star \succeq 0, \\
\text{Consistency with } \mathcal{F}: \quad & \exists f \in \mathcal{F} : f(x_k) = f_k, \nabla f(x_k) = g_k, \nabla^2 f(x_k) = h_k, \\
& \qquad k = 0, \ldots, N, \star.
\end{aligned}
\tag{10.4}
$$

Making the last constraint in (10.4) explicit requires having access to *interpolation conditions* for $\mathcal{F}$, that is necessary and sufficient conditions to impose on a data set $S = \{(x_k, f_k, g_k, h_k)\}_{k=0,...,N,\star}$ to ensure its consistency with a function in $\mathcal{F}$. Solving a PEP as in (10.4) with interpolation conditions ensures tightness of the resulting bound. If instead we impose necessary but not sufficient conditions on the set of iterates, we obtain a relaxation of (10.3), potentially leading to bounds that are not tight (see [RH24, THG17c] for examples). Figure 10.1 illustrates the difference between relying on exact interpolation conditions and necessary conditions when solving a PEP.

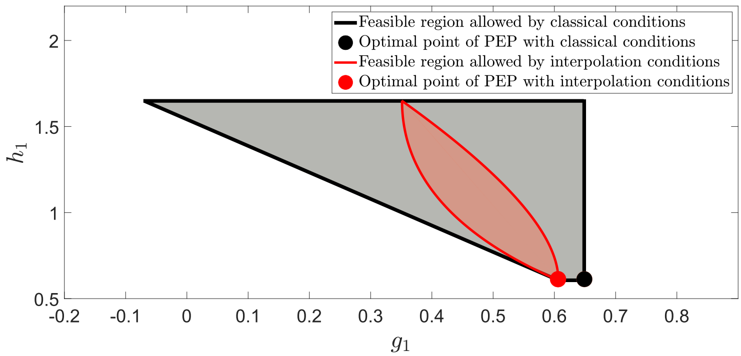


Fig. 10.1 Given $(x_0, g_0, h_0) = (0, 1, 1)$, $M = 1$ and $x_1 = \frac{1}{2}$, the plot shows the admissible region for (g_1, h_1) such that $\{(x_i, g_i, h_i)\}_{i=0,1}$ satisfies (i) classical conditions defining quasi-self-concordant functions, i.e., (10.52), (black curves), or (ii) interpolation conditions for this function class, i.e., Corollary 10.35 (red curves). It also shows the points (g_1, h_1) corresponding to the worst-case of a variant of Newton's method, whose iterations are given by (GNM1), according to both conditions. Using interpolation conditions instead of necessary conditions significantly restricts the domain (g_1, h_1) of the PEP, and the optimal point obtained by solving a PEP with necessary conditions (black dot) is not consistent with any actual function of the class, hence the result is not tight.

Obtaining tight bounds thus requires (i) obtaining interpolation constraints for the function class $\mathcal{F}$ of interest, and (ii) solving the PEP (10.4), either analytically or numerically.

10.1.3 Organization and contributions

We first propose a generic approach to derive exact interpolation conditions for classes of univariate functions, namely, we

(i) Propose a technique to lift interpolation conditions from a given univariate function class $\mathcal{F}$ to an associated univariate function class $\int \mathcal{F}$, defined as the set of functions whose derivative belongs to $\mathcal{F}$ (Section 10.2 and Theorem 10.18).

(ii) Present a range of function classes on which the technique applies, that is functions satisfying $|f^{(k+1)}(x)| \leq Af^{(k)}(x)^{\alpha}$ for some $A, \alpha \geq 0$, where $f^{(k)}$ denotes the k^{th} derivative of f (Section 10.3). This definition includes generalized smoothness [LQT+24] and generalized self-concordance [STD19], and in particular adapts to (strongly convex) functions with M-Lipschitz continuous Hessian and (quasi)-self-concordant functions [STD19]. We propose a unified interpretation of these classes by showing they satisfy a Lipschitz condition on a specific quantity depending on α (Proposition 10.21).

(iii) Exploit the technique to derive explicit interpolation conditions for these function classes, summarized in Table 10.1 (Section 10.4).

Table 10.1 Summary of interpolation conditions of classes of univariate functions.

Class	$\{(x_i, h_i)\}$	$\{(x_i, g_i, h_i)\}$	$\{(x_i, f_i, g_i, h_i)\}$
Hessian Lipschitz	Th. 10.23	Th. 10.27 or Cor. 10.30	Th. 10.38 (new)
Hessian Lipschitz & strongly convex	Th. 10.23 (new)	Th. 10.27 or Cor. 10.31 (new)	
Self-concordant	Th. 10.23 (new)	Th. 10.27 or Cor. 10.33 (new)	
Quasi-self-concordant	Th. 10.23 (new)	Th. 10.26 or Cor. 10.35 (new)	
Generalized self-concordant	Th. 10.23 (new)	Th. 10.27 (new)	

We then use these conditions to analytically analyze several second-order methods (Section 10.5), thereby

(i) Obtaining a tight bound for one iteration of the Cubic Regularized Newton method on Hessian Lipschitz univariate functions [NP06] (Theorem 10.43) and Gradient Regularized Newton method on quasi-self-concordant univariate functions [Doi23] (Lemma 10.59). These automatically serve as multivariate lower bounds, that improve on existing ones.

(ii) Proving tightness of existing results in the *multivariate case*: Newton's method and Gradient method on Hessian Lipschitz functions [N+18] (Theorems 10.52 and 10.53) and Newton's method on quasi-self-concordant and strongly convex functions (Lemma 10.61).

(iii) Proposing an alternative proof of an existing result on the convergence of Newton's method on self-concordant functions (Section 10.5.4).

Finally, we leverage on the advances in solving non-convex PEPs to solve formulations associated with several second-order methods (Section 10.5), thereby

(i) Numerically improving convergence rates in the univariate case: Cubic Regularized Newton method on Hessian Lipschitz functions (Figure 10.3, for more than one iteration), Gradient Regularized Newton method on (strongly) convex Hessian Lipschitz functions (Conjecture 10.49 and Figure 10.5), Newton and damped Newton methods on self-concordant functions (Theorem 10.57), Gradient Regularized Newton method on quasi-self-concordant functions (Figure 10.11, for more than one iteration).

(ii) Tuning methods in the univariate case: Cubic Regularized Newton method with a stepsize on Hessian Lipschitz functions (Figure 10.4), and Fixed damped Newton method on Hessian Lipschitz functions (Figure 10.8).

(iii) Comparing different variants of Newton's method in the univariate case (Figure 10.12).

Table 10.2 summarizes the state-of-the-art bounds on the performance of the second-order methods we analyzed.

10.1.4 Restriction to the univariate setting

Our method and analysis are restricted to classes of univariate functions ($d = 1$) due to two main difficulties, preventing a straightforward generalization to higher dimensions. First, it is in general not straightforward to obtain interpolation conditions for given function classes. Nevertheless, we propose a technique to automatically obtain such conditions in the univariate case, which heavily relies on properties specific to univariate functions. Second-order interpolation conditions for classes of multivariate functions remain an open question. Second, there is no known efficient way to solve PEPs involving second-order quantities in the multivariate case: these are a priori highly non-convex and of large dimensions. On the contrary, we manage to solve PEPs in various univariate second-order settings due to their smaller size.

This work thus only takes a first step into the problem of tightly analyzing second-order methods. However, we believe it is important for several reasons. First, univariate interpolation conditions can provide intuition

Table 10.2 Summary of existing convergence results of the literature together with our analytical or numerical contributions (in the univariate case).

Method[1]	Class[2]	Perf. meas.	Initial condition	References	Contributions[3] (univariate case)
CNM	HL	$\min_k \lvert f'(x_k)\rvert$	$f(x_0) - f_\star$	[NP06, Th. 1]	Improved guarantee (Th. 10.43 and Fig. 10.3)
CNM	HL	$\lvert f'(x_N)\rvert$	$f(x_0) - f_\star$		First (numerical) guarantee (Fig. 10.3)
CNM	HL ∩ C	$f(x_N) - f_\star$	$f(x_0) - f_\star$	[NP06, Th. 6]	Improved guarantee (Rem. 10.45)
CNM	HL ∩ sC	$f(x_N) - f_\star$	$f(x_0) - f_\star$	[NP06, Th. 7]	Improved guarantee (Rem. 10.45)
GNM2	HL ∩ C	$\lvert f'(x_k)\rvert$	$f(x_k) - f(x_{k+1})$	[Mis23, Th. 2.6]	Improved guarantee (Conj. 10.49)
GNM2	HL ∩ sC	$\lvert f'(x_{k+1})\rvert$	$\lvert f'(x_k)\rvert$	[Mis23, Th. 2.7]	Improved (Fig. 10.5)
NM	HL	$\lvert x_{k+1} - x_\star\rvert$	$\lvert x_k - x_\star\rvert$	[N+18, Th. 1.2.5]	Tightness proved (Th. 10.52)
GM	HL	$\lvert x_{k+1} - x_\star\rvert$	$\lvert x_k - x_\star\rvert$	[N+18, Th. 1.2.4]	Tightness proved (Th. 10.54)
DNM	HL	$\lvert x_{k+1} - x_\star\rvert$	$\lvert x_k - x_\star\rvert$		First (numerical) guarantee (Figs. 10.7 and 10.8)
NM	SC	$\lambda_f(x_{k+1})$	$\lambda_f(x_k)$	[Hil21, Eq. (11)]	Worst-case function (Sect. 10.5.4)
DNM	SC	$\lambda_f(x_N)$	$\lambda_f(x_0)$	[Hil21, Eq. (11)]	Tightness proved (Th. 10.57)
GNM1	QSC	$\lvert f'(x_N)\rvert / f''(x_N)$	$\lvert f'(x_0)\rvert / f''(x_0)$	[Doi23, Eq. (49)]	Improved guarantee (Lem. 10.59 and Fig. 10.11)
NM	QSC ∩ sC	$\lvert f'(x_{k+1})\rvert$	$\lvert f'(x_k)\rvert$	[Doi23, Eq. (53)]	Tightness proved (Lem. 10.61)
all	HL ∩ sC	$\lvert f'(x_{k+1})\rvert$	$\lvert f'(x_k)\rvert$		First (numerical) comparison (Fig. 10.12)

[1] NM: Newton's method, CNM: Cubic Regularized Newton method, GNM: Gradient Regularized Newton method, GM: Gradient method, DNM: fixed damped Newton method.
[2] HL: Hessian Lipschitz, C: Convex, sC: Strongly convex, SC: Self-Concordant, QSC: Quasi-Self-Concordant functions.
[3] Our contributions are proved when referred to as Lem., Rem., Sect., or Th. and they are only conjectured based on numerical experiments when referred to as Fig. or Conj.

about candidate multivariate conditions. Second, exact univariate bounds provide lower bounds on the multivariate worst-case performance of the method. In fact, in many settings (first and second-order methods), univariate functions are observed to provide the general multivariate worst-case [CGT10, Doi21, THG17c, Toi24, RGP24] (see also Appendix B). It is even more often the case when considering a single iteration since the method only "sees" a univariate segment between the initial point and the next iterate. Finally, in case the obtained univariate lower bound matches a known analytical multivariate upper bound, the associated worst-case instance proves tightness of this upper bound.

10.1.5 Notation

In this chapter, we let $[N] = \{0, \cdots, N\}$. We denote by $\mathcal{C}^m$ the class of univariate functions $f : \mathbb{R} \to \mathbb{R}$ that are at least m times everywhere differentiable, with derivative up to order m everywhere continuous. To deal with functions taking infinite values, we let $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$, the projectively extended real line, see, e.g., [AB98, p. 29] for details on this topology. We denote by $\bar{\mathcal{C}}^0$ the class of functions $f : \mathbb{R} \to \bar{\mathbb{R}}$ that are everywhere continuous, i.e., $\forall c \in \mathbb{R}, \lim_{x\to c} f(x) = f(c)$, in the following sense:

$$\begin{cases} \forall \epsilon > 0,\ \exists \delta > 0, \forall x \in \mathbb{R}:\ |x-c| < \delta \Rightarrow |f(x) - f(c)| < \epsilon & \text{for finite } f(c), \\ \forall N > 0,\ \exists \delta > 0, \forall x \in \mathbb{R}:\ |x-c| < \delta \Rightarrow |f(x)| > N & \text{for infinite } f(c). \end{cases}$$

For instance, the function $f(x) = \frac{1}{x}$ if $x \neq 0,\ f(0) = \infty$ belongs to $\bar{\mathcal{C}}^0$. We denote by $\text{dom} f = \{x \in \mathbb{R}:\ f(x) \neq +\infty\}$ the *effective domain of* $f \in \bar{\mathcal{C}}^0$. Further, we say a function $f : \mathbb{R} \to \bar{\mathbb{R}} \in \bar{\mathcal{C}}^m$ if (i) f is at least m times everywhere differentiable on its effective domain, and (ii) its derivative up to order m, i.e., it differential on the effective domain extended to $\mathbb{R}$ with value ∞, belongs to $\bar{\mathcal{C}}^0$. We have $\mathcal{C}^m \subseteq \bar{\mathcal{C}}^m$.

10.2 Generic technique to lift interpolation conditions from known classes to higher-order classes

We seek for *interpolation conditions* for classes of univariate functions $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$, i.e., necessary and sufficient conditions ensuring $\mathcal{F}$*-interpolability* of a discrete data set:

Definition 10.1 (Interpolation with function values)**.** Given $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$, a set $S = \{(x_i, f_i^0, f_i^1, ..., f_i^m)\}_{i\in[N]} \in (\mathbb{R} \times \ldots \times \mathbb{R})^N$ is $\mathcal{F}$-interpolable with function values if and only if

$$\exists f \in \mathcal{F} : f^{(k)}(x_i) = f_i^k,\ \forall k \in [m], \quad \forall i \in [N]. \tag{10.5}$$

In some circumstances, given a dataset S, one seeks to ensure the existence of a function in $\mathcal{F}$ consistent with S except for its function values f_i^0. Definition 10.1 can be straightforwardly extended to this case, called interpolation without function values:

Definition 10.2 (Interpolation without function values)**.** Given $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$, a set $S = \{(x_i, f_i^0, f_i^1, ..., f_i^m)\}_{i\in[N]} \in (\mathbb{R} \times \ldots \times \mathbb{R})^N$ is $\mathcal{F}$-interpolable without function values if and only if

$$\exists f \in \mathcal{F} : f^{(k)}(x_i) = f_i^k, \quad \forall k = 1, \cdots, m,\ \forall i \in [N]. \tag{10.6}$$

The quantities f_i^0 are thus ignored in interpolation without function values and are not necessarily equal to function values $f^{(0)}(x_i)$.

Remark 10.3. In the univariate case, interpolation conditions without function values are often more concise than interpolation conditions with function values. Hence, we resort to this formulation whenever possible, for instance when analyzing methods that do not involve function values in their settings (see, e.g., [Tay17, Section 3.4]).

Whenever considering the cases $m = 1, 2$, we denote $f_i^0 := f_i$, $f_i^1 := g_i$, $f_i^2 := h_i$, $f^{(1)}(x) := f'(x)$ and $f^{(2)}(x) := f''(x)$.

10.2.1 Overview of the technique

Suppose interpolation conditions are known for a class $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$. We propose a technique, summarized in Theorem 10.18, to obtain interpolation conditions for the associated class $\int \mathcal{F} \subseteq \bar{\mathcal{C}}^{m+1}$ of functions whose derivative belongs to $\mathcal{F}$:

$$\int \mathcal{F} := \{f : \mathbb{R} \to \bar{\mathbb{R}} : \; f' \in \mathcal{F}\}. \tag{10.7}$$

We thus lift interpolation conditions from the function class $\mathcal{F}$ to the higher-order function class $\int \mathcal{F}$.

The proposed method consists of two main steps. First, we tackle the question of $\int \mathcal{F}$-interpolation without function values of $S = \{(x_i, f_i^0, f_i^1, ..., f_i^m)\}_{i\in[N]}$. For univariate function classes, these conditions are the interpolation conditions with function values for $\mathcal{F}$ (Lemma 10.4), and the question is straightforwardly answered, by imposing these known interpolation conditions on $\tilde{S} = \{(x_i, f_i^1, ..., f_i^m)\}_{i\in[N]}$. Second, we build on these interpolation conditions without function values to tackle the case of interpolation with function values of $\int \mathcal{F}$. Informally, we impose further conditions on f_i^0, $i \in [N]$, ensuring that these belong to the interval defined by the integral of all functions in $\mathcal{F}$ interpolating S (Theorem 10.18).

The method thus heavily relies on univariate properties and cannot be directly generalized to other dimensions. In addition, its second step applies to function classes satisfying Assumptions 10.9, 10.11, and 10.14. Before stating Theorem 10.18, we explicit both steps and alongside present these assumptions and why they are required for validity of the method. In Section 10.3, we then show that Theorem 10.18 can be applied to a large set of function classes, including (convex) functions with Lipschitz Hessian, or (quasi)-self-concordant functions. For an illustration of the technique on these classes, among others, see Section 10.4.1.

10.2.2 Interpolation without function values

Unlike the multidimensional case, any univariate function is the derivative of another univariate function. Thereby, in the univariate case, interpolation without function values of $\int \mathcal{F}$ amounts exactly to interpolation with function values of $\mathcal{F}$:

Lemma 10.4. *Consider $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ and $\int \mathcal{F} \subseteq \bar{\mathcal{C}}^{m+1}$ (defined in (10.7)). A set $S = \{(x_i, f_i^0, f_i^1, ..., f_i^m)\}_{i\in[N]}$ is $\int \mathcal{F}$-interpolable (without function values) if and only if $\tilde{S} := \{(x_i, f_i^1, ..., f_i^m)\}_{i\in[N]}$ is $\mathcal{F}$-interpolable (with function values).*

Proof. By Definition 10.2, S is $\int \mathcal{F}$-interpolable without function values if

$$\begin{aligned}
&\exists\, f \in \textstyle\int \mathcal{F} &&: f^{(k)}(x_i) = f_i^k, \quad \forall k = 1, \dots, m,\ \forall i \in [N],\\
\overset{(10.7)}{\Leftrightarrow}\ &\exists\, f : f' = g \in \mathcal{F},\ && f^{(k)}(x_i) = f_i^k, \quad \forall k = 1, \dots, m,\ \forall i \in [N],\\
\Leftrightarrow\ &\exists\, g \in \mathcal{F} &&: g^{(k)}(x_i) = f_i^{k+1}, \forall k = 0, \dots, m-1,\ \forall i \in [N],\\
\overset{\text{Def. 10.1}}{\Leftrightarrow}\ &\tilde{S} \text{ is } \mathcal{F}\text{-interpolable with function values.}
\end{aligned}$$

□

10.2.3 A sufficient condition for interpolation with function values

We now show how to go from interpolation without function values to interpolation with function values. Informally, given a set $S = \{(x_i, f_i^0, f_i^1, \dots, f_i^{m+1})\}_{i\in[N]}$, which is $\int \mathcal{F}$-interpolable without function values, we ensure that one of the functions in $\mathcal{F}$ interpolating S (except possibly for f_i^0, $i \in [N]$) is the derivative of a function f satisfying $f(x_i) = f_i^0$, $i \in [N]$.

We take the simplest situation possible and restrict our attention to the interpolation of a single pair of data points $x_1 < x_2$.

Consider a set $S = \{(x_i, f_i^0, f_i^1, \dots, f_i^{m+1})\}_{i=1,2}$, which is $\int \mathcal{F}$-interpolable without function values and the associated set $\tilde{S} = \{(x_i, f_i^1, \dots, f_i^{m+1})\}_{i=1,2}$. By Lemma 10.4, $\tilde{S}$ is $\mathcal{F}$-interpolable, hence there exists at least one function $g \in \mathcal{F}$ interpolating $\tilde{S}$. For each of such g, we can construct an associated function $f \in \int \mathcal{F}$, interpolating S with the exception of $f(x_2)$ which could differ from f_2^0:

Lemma 10.5. *Consider a class $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ of univariate functions. Let $S = \{(x_i, f_i^0, f_i^1, \dots, f_i^{m+1})\}_{i=1,2}$ be $\int \mathcal{F}$-interpolable without function values, $\tilde{S} = \{(x_i, f_i^1, \dots, f_i^{m+1})\}_{i=1,2}$, and $g \in \mathcal{F}$ be a function interpolating $\tilde{S}$. Then,*

$$f(x) := f_1^0 - \int_{-\infty}^{x_1} g(z)\mathrm{d}z + \int_{-\infty}^{x} g(z)\mathrm{d}z, \tag{10.8}$$

belongs to $\int \mathcal{F}$ *and interpolates* S *except* $f(x_2)$ *which could differ from* f_2^0.

Proof. It holds that $f'(x) = g(x)$, hence by Definition 10.7, $f \in \int \mathcal{F}$. In addition:

$$\begin{cases} f'(x_1) = g(x_1) = f_1^1, \\ f(x_1) = f_1^0, \\ f'(x_2) = g(x_2) = f_2^1, \\ f(x_2) = f_1^0 + \int_{x_1}^{x_2} g(z)\mathrm{d}z, \end{cases} \tag{10.9}$$

hence f interpolates S except possibly for $f(x_2)$.

□

Suppose now there exist two functions g_{lo}, $g_{\text{hi}} \in \mathcal{F}$ interpolating $\tilde{S}$, and such that the associated functions $f_{\text{lo}}, f_{\text{hi}} \in \int \mathcal{F}$ as defined in (10.8) satisfy

$$f_{\text{lo}}(x_2) \le f_2^0 \le f_{\text{hi}}(x_2) \Leftrightarrow f_2^0 = \lambda f_{\text{lo}}(x_2) + (1-\lambda) f_{\text{hi}}(x_2), \tag{10.10}$$

for some $\lambda \in [0,1]$. Then, there exists a function $f \in \int \mathcal{F}$ interpolating S (including $f(x_2) = f_2^0$), provided $\int \mathcal{F}$ is a *convex* function class, namely, any convex combination of functions in $\int \mathcal{F}$ is itself in $\int \mathcal{F}$.

Assumption 10.6 (Convex function class)**.** *We say a class* $\mathcal{F}$ *of univariate functions is* convex *if, given any* $f_a, f_b \in \mathcal{F}$ *and any* $\lambda \in [0,1]$,

$$\lambda f_a + (1-\lambda) f_b \in \mathcal{F}. \tag{10.11}$$

We show in Proposition B.2 that a wide range of function classes of interest are convex, e.g., functions with Lipschitz Hessian or quasi-self-concordant functions.

For convex classes $\int \mathcal{F}$, existence of interpolants g_{lo} and g_{hi} whose integrals evaluated at x_2 are respectively smaller and larger than f_2 is a sufficient condition for $\int \mathcal{F}$-interpolability of S, in addition to S being $\mathcal{F}$-interpolable:

Lemma 10.7. *Let* $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ *be a convex (Assumption 10.6) class of univariate functions, and let* $\int \mathcal{F} \subseteq \bar{\mathcal{C}}^{m+1}$ *(defined in* (10.7)*). Let* $S = \{(x_i, f_i^0, f_i^1, \ldots, f_i^{m+1})\}_{i=1,2}$ *and* $\tilde{S} = \{(x_i, f_i^1, \ldots, f_i^{m+1})\}_{i=1,2}$, *where* $x_1 < x_2$.

If $\tilde{S}$ *is* $\int \mathcal{F}$*-interpolable without function values, and there exists two functions*

$g_{lo}, g_{hi} \in \mathcal{F}$ interpolating $\tilde{S}$ such that

$$f_1^0 + \int_{x_1}^{x_2} g_{lo}(z)\mathrm{d}z \le f_2^0 \le f_1^0 + \int_{x_1}^{x_2} g_{hi}(z)\mathrm{d}z, \tag{10.12}$$

then S is $\int \mathcal{F}$-interpolable with function values.

Proof. Let

$$f_{\mathrm{lo}}(x) := f_1^0 - \int_{-\infty}^{x_1} g_{\mathrm{lo}}(z)\mathrm{d}z + \int_{-\infty}^{x} g_{\mathrm{lo}}(z)\mathrm{d}z, \tag{10.13}$$

$$f_{\mathrm{hi}}(x) := f_1^0 - \int_{-\infty}^{x_1} g_{\mathrm{hi}}(z)\mathrm{d}z + \int_{-\infty}^{x} g_{\mathrm{hi}}(z)\mathrm{d}z. \tag{10.14}$$

By Lemma 10.5, f_{lo} and f_{hi} belong to $\int \mathcal{F}$. In addition, the two functions interpolate S with the exception of f_2^0 but there exists $\lambda \in (0,1)$ such that

$$f_2^0 = \lambda f_{\mathrm{lo}}(x_2) + (1-\lambda) f_{\mathrm{hi}}(x_2). \tag{10.15}$$

Therefore, $f := \lambda f_{\mathrm{lo}} + (1-\lambda) f_{\mathrm{hi}}$ interpolates S (including $f(x_2) = f_2^0$) and $f \in \mathcal{F}$ by convexity of $\mathcal{F}$.

□

10.2.4 Interpolation conditions with function values (2 points)

The sufficient condition in Lemma 10.7 becomes necessary if g_{lo}, g_{hi} are chosen to be the *extremal interpolants* of S, that is the lowest and highest functions in $\mathcal{F}$ interpolating S without function values (see Figure 10.2 for an illustration).

As such, extremal interpolants are not well-defined and might not exist or be non-unique. We thus introduce the notion of *lower and upper interpolating envelopes*, that is pointwise minimum and maximum of all functions in $\mathcal{F}$ interpolating S:

Definition 10.8 (Extremal interpolating envelopes)**.** Consider a function class $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ and an $\mathcal{F}$-interpolable set $S = \{(x_i, g_i^0, g_i^1, \ldots, g_i^m)\}_{i=1,2}$, where $x_1 < x_2$. The *extremal interpolating envelopes* $g_{\min}$ and $g_{\max}$ of S are defined

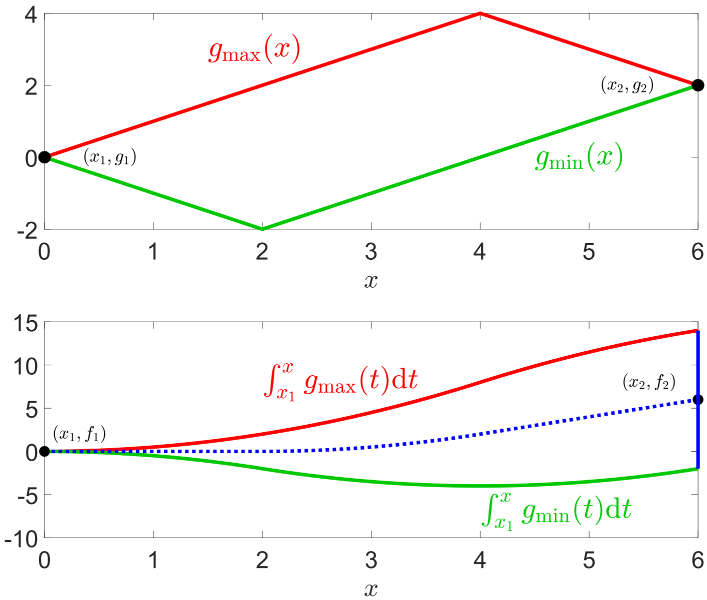


Fig. 10.2 Illustration of the method to obtain interpolation conditions ensuring $S = \{(x_i, f_i, g_i)\}_{i=1,2}$ to be $\mathcal{F}_{M,0}$-interpolable: extremal M-Lipschitz interpolants $g_{\min}$ (above, green curve) and $g_{\max}$ (above, red curve) of $\{(x_i, g_i)\}_{i=1,2}$ (see Proposition B.2 for their derivation). These extremal interpolants are integrated (below, red and green curves), and define an interval including all possible smooth functions interpolating S, except possibly f_2. The interval $[f_1 + \int_{x_1}^{x_2} g_{\min}(x)\mathrm{d}x, f_1 + \int_{x_1}^{x_2} g_{\max}(x)\mathrm{d}x]$ (below, blue line) is exactly the interval of admissible values for f_2. Whenever f_2 belongs to this interval, there exists a convex combination of the integrals of $g_{\min}$ and $g_{\max}$ (below, blue dotted curve), which is in $\int \mathcal{F}_{M,0}$ and interpolates S.

as, $\forall x \in \mathbb{R}$,

$$\begin{aligned} g_{\min}(x) = \inf_{g:\mathbb{R}\to\mathbb{R}} \quad & g(x) \\ \text{s.t.} \quad & g \in \mathcal{F} \text{ and } g^{(k)}(x_i) = g_i^k,\ i = 1,2,\ k \in [m], \\ g_{\max}(x) = \sup_{g:\mathbb{R}\to\mathbb{R}} \quad & g(x) \\ \text{s.t.} \quad & g \in \mathcal{F} \text{ and } g^{(k)}(x_i) = g_i^k,\ i = 1,2,\ k \in [m]. \end{aligned} \tag{10.16}$$

Extremal interpolating envelopes always exist, but nothing ensures they belong to $\mathcal{F}$. To ensure it is the case, we require $\mathcal{F}$ to be *extremally interpolable*:

Assumption 10.9 (Extremally interpolable function class). *We say a function class $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ of univariate functions is* extremally interpolable *if any $\mathcal{F}$-interpolable set $S = \{(x_i, g_i^0, g_i^1, ..., g_i^m)\}_{i=1,2}$ satisfies*

$$g_{\min}, g_{\max} \in \mathcal{F} \tag{10.17}$$

where $g_{\min}$ and $g_{\max}$ are defined in (10.16).

Again, we will see in Section 10.4.1 that a wide range of classes of interest are extremally interpolable, e.g., classes $\mathcal{F}$ such that $\int \mathcal{F}$ describes functions with Lipschitz Hessian, or (quasi)-self-concordant functions.

Whenever $\mathcal{F}$ is extremally interpolable, we refer to extremal interpolating envelopes of S as extremal interpolants of S. We are now ready to summarize the interpolation conditions for $\int \mathcal{F}$:

Lemma 10.10. *Let $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ be an extremally interpolable (Assumption 10.9) class of univariate functions and let $\int \mathcal{F} \subseteq \bar{\mathcal{C}}^{m+1}$ (defined in (10.7)) be convex (Assumption 10.6).*

A set $S = \{(x_i, f_i^0, f_i^1, ..., f_i^{m+1})\}_{i=1,2}$, where $x_1 < x_2$ is $\int \mathcal{F}$-interpolable with function values if and only if S is $\int \mathcal{F}$-interpolable without function values, and f_1^0, f_2^0 satisfy

$$\int_{x_1}^{x_2} g_{\min}(x)\mathrm{d}x \le f_2^0 - f_1^0 \le \int_{x_1}^{x_2} g_{\max}(x)\mathrm{d}x, \tag{10.18}$$

where $g_{\min}$ and $g_{\max}$ are defined as in (10.16).

Proof. Sufficiency follows from Lemma 10.7. It remains to prove necessity of (10.18) for $\int \mathcal{F}$-interpolability. Suppose $S = \{(x_i, f_i^0, f_i^1, ..., f_i^{m+1})\}_{i=1,2}$ is $\int \mathcal{F}$-interpolable (with function values). Then, by Lemma 10.4, S is $\mathcal{F}$-interpolable (without function values). In addition, by Definition 10.8 of extremal interpolants, any function $g : \mathbb{R} \to \mathbb{R}$ such that (i) $\exists f \in \int \mathcal{F} : g(x) = f'(x), \forall x \in \mathbb{R}$ and (ii) g interpolates S satisfies

$$g_{\min}(x) \le g(x) \le g_{\max}(x), \ \forall x \in [x_1, x_2]. \tag{10.19}$$

Hence, (10.18) is necessarily satisfied.

□

Convexity of $\int \mathcal{F}$, though natural to introduce and sufficient for validity of Lemma 10.10, proves to be too restrictive, since it is not satisfied by many classes for which we would like to obtain interpolation conditions, e.g., the class of self-concordant functions. We therefore introduce a weaker condition, sufficient for validity of Lemma 10.10 and satisfied, e.g., by self-concordant functions.

Assumption 10.11 (Extremally completable function class). *Let $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ be an extremally interpolable (Assumption 10.9) class of univariate functions, and consider $\int \mathcal{F} \in \bar{\mathcal{F}}^{m+1}$.*

Let $S = \{(x_i, f_i^0, f_i^1, \dots, f_i^{m+1})\}_{i=1,2}$ be $\int \mathcal{F}$-interpolable without function values, $\tilde{S} = \{(x_i, f_i^1, \dots, f_i^{m+1})\}_{i=1,2}$, and $g_{\min}, g_{\max} \in \mathcal{F}$ be the extremal interpolants of $\tilde{S}$, as defined in (10.16)*. Let, in addition,*

$$f_1^0 + \int_{x_1}^{x_2} g_{\min}(z)\mathrm{d}z \leq f_2^0 \leq f_1^0 + \int_{x_1}^{x_2} g_{\max}(z)\mathrm{d}z. \tag{10.20}$$

We then say $\int \mathcal{F}$ is extremally completable *if*

$$\exists f \in \int \mathcal{F} : \; f^{(k)}(x_i) = f_i^k, \quad \forall i \in [N], \quad \forall k \in [m]. \tag{10.21}$$

Convex and extremally completable classes are related as follows.

Lemma 10.12. *Let $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ be an extremally interpolable (Assumption 10.9) class of univariate functions. If $\int \mathcal{F}$ is convex (Assumption 10.6), then $\int \mathcal{F}$ is extremally completable (Assumption 10.11).*

Proof. With the same notation as in Assumption 10.11, let

$$f_{\min}(x) := f_1^0 - \int_{-\infty}^{x_1} g_{\min}(z)\mathrm{d}z + \int_{-\infty}^{x} g_{\min}(z)\mathrm{d}z. \tag{10.22}$$

$$f_{\max}(x) := f_1^0 - \int_{-\infty}^{x_1} g_{\max}(z)\mathrm{d}z + \int_{-\infty}^{x} g_{\max}(z)\mathrm{d}z. \tag{10.23}$$

By Lemma 10.5, $f_{\min}$ and $f_{\max}$ belong to $\int \mathcal{F}$. In addition, the two functions interpolate S with the exception of $f_2^0 = \lambda f_{\mathrm{lo}}(x_2) + (1-\lambda) f_{\mathrm{hi}}(x_2)$ for some $\lambda \in (0,1)$. By convexity of $\int \mathcal{F}$, $f := \lambda f_{\mathrm{lo}} + (1-\lambda) f_{\mathrm{hi}}$ interpolates S (including $f(x_2) = f_2$), and $f \in \mathcal{F}$.

□

We show that extremal completability suffices for validity of Lemma 10.10.

Lemma 10.13. *Let $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ be an extremally interpolable (Assumption 10.9) class of univariate functions and let $\int \mathcal{F} \subseteq \bar{\mathcal{C}}^m$ (defined in (10.7)) be extremally completable (Assumption 10.11).*

A set $S = \{(x_i, f_i^0, f_i^1, ..., f_i^{m+1})\}_{i=1,2}$, where $x_1 < x_2$ is $\int \mathcal{F}$-interpolable with function values if and only if S is $\int \mathcal{F}$-interpolable without function values, and f_1^0, f_2^0 satisfy

$$\int_{x_1}^{x_2} g_{\min}(x)\mathrm{d}x \leq f_2^0 - f_1^0 \leq \int_{x_1}^{x_2} g_{\max}(x)\mathrm{d}x, \tag{10.24}$$

where $g_{\min}$ and $g_{\max}$ are defined as in (10.16).

Proof. Sufficiency follows the definition of extremal completability, and necessity follows the same argument as in the proof of Lemma 10.10. □

To obtain full interpolation conditions for $\int \mathcal{F}$, it remains to consider the case of an arbitrary number of points to interpolate.

10.2.5 Interpolation conditions with function values (N points)

In the univariate case, interpolation of an arbitrary number of points is equivalent to that of a single pair, provided the class $\int \mathcal{F}$ is *order $m+1$ connectable*, in the sense that the juxtaposition of different functions in $\int \mathcal{F}$, where the functions coincide up to order $m+1$ on the boundaries of some intervals, is itself a function in $\int \mathcal{F}$. Formally:

Assumption 10.14 (Order m connectable)**.** *A class $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ of univariate functions is* order m connectable *if, given any $x_0 \leq ... \leq x_{K+1} \in \mathbb{R} \cup \{-\infty, \infty\}$, where $x_0 = -\infty$ and $x_{K+1} = \infty$, and any K functions $f_j \in \mathcal{F}$ such that*

$$f_j^{(l)}(x_{j+1}) = f_{j+1}^{(l)}(x_{j+1}), \quad \forall l \in [m], \quad j \in [K-1],$$

then the piecewise function

$$f : \mathbb{R} \to \mathbb{R} : \ f(x) := f_j(x) \quad \forall x \in [x_j, x_{j+1}], \forall j \in [K] \text{ belongs to } \mathcal{F}.$$

Remark 10.15. Linear functions are order 1 connectable, since the juxtaposition of several linear functions whose slope is identical is linear, but not order 0 connectable, since the juxtaposition of linear functions with different slopes is not linear. Hence, an appropriate choice of order is essential when determining whether a function class is connectable.

Once more, we will see in Proposition B.2 that a wide range of function classes of interest are order 2-connectable, e.g., functions with Lipschitz Hessian or (quasi)-self-concordant functions.

Order $m+1$ connectivity of $\int \mathcal{F}$ allows extending Lemma 10.10 to any arbitrary number of points directly:

Lemma 10.16 (Interpolation on a single interval). *Let $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ be an order m connectable class of univariate functions (Assumption 10.14). A set $S = \{(x_i, f_i^0, f_i^1, ..., f_i^m)\}_{i\in[N]}$ is $\mathcal{F}$-interpolable with function values if and only if,*

$$\forall i \in [N-1] : \left\{(x_i, f_i^0, f_i^1, ..., f_i^m), (x_{i+1}, f_{i+1}^0, f_{i+1}^1, ..., f_{i+1}^m)\right\} \text{ is } \mathcal{F}\text{-interpolable.} \tag{10.25}$$

Proof. (Necessity) Suppose $S = \{(x_i, f_i^0, f_i^1, ..., f_i^m)\}_{i\in[N]}$ is $\mathcal{F}$-interpolable, and let $f \in \mathcal{F}$ be a function interpolating S, and in particular all pairs (i, j). For each pair (i, j), the restriction of f to the interval $[x_i, x_j]$ is a function in $\mathcal{F}$ interpolating $\left\{(x_i, f_i^0, f_i^1, ..., f_i^m), (x_j, f_j^0, f_j^1, ..., f_j^m)\right\}$, and (10.25) is satisfied.

(Sufficiency) Suppose (10.25) is satisfied, let $x_1 \leq \ldots \leq x_N$ and let $f^{(i)} \in \mathcal{F}$ be a function interpolating the pair $(i, i+1)$ $\forall i \in [N-1]$. Define

$$\mathcal{I}_i = \begin{cases}]-\infty, x_2] & i = 1, \\ [x_i, x_{i+1}] & i = 2, \cdots, N-2, \\ [x_{N-1}, \infty[& i = N-1, \end{cases} \tag{10.26}$$

where $\mathcal{I}_1$ and $\mathcal{I}_{N-1}$ are defined such that $\bigcup_{i=1}^{N-1} \mathcal{I}_i = \mathbb{R}$ and not just $[x_1, x_N]$. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by:

$$f(x) := f^{(i)}(x) \quad \forall x \in \mathcal{I}_i,\ \forall i \in [N-1]. \tag{10.27}$$

By construction, f interpolates $S = \{(x_i, f_i^0, f_i^1, ..., f_i^m)\}_{i\in[N]}$. Furthermore, by order m piecewise invariance of $\mathcal{F}$, it holds that $f \in \mathcal{F}$, which completes the proof. □

Remark 10.17. This differs from the multivariate case, where interpolation of a single pair can significantly differ from interpolation of an arbitrary number of points (see, e.g., [RTBG20, Proposition 4]). For example, some inequalities are interpolation conditions whenever applied to a single pair but are only necessary otherwise.

10.2.6 Main Theorem

We are now ready to present our main Theorem, allowing to lift interpolation conditions from $\mathcal{F}$ to $\int \mathcal{F}$:

Theorem 10.18. *Let $\mathcal{F} \subseteq \bar{\mathcal{C}}^m$ be an extremally interpolable (Assumption 10.9) class of univariate functions, and let $\int \mathcal{F} \subseteq \bar{\mathcal{C}}^{m+1}$ (defined in (10.7)) be extremally completable (Assumption 10.11) and order $m+1$ connectable (Assumption 10.14).*

A set $S = \{(x_i, f_i^0, f_i^1, ..., f_i^m)\}_{i\in[N]}$, where $x_0 < x_1 < ... < x_N$ is $\int \mathcal{F}$-interpolable if and only if S is $\int \mathcal{F}$-interpolable without function values, and $\forall i \in [N]$

$$\frac{1}{x_{i+1}-x_i}\int_{x_i}^{x_{i+1}} g_{\min}(x)\mathrm{d}x \leq \frac{f_{i+1}-f_i}{x_{i+1}-x_i} \leq \frac{1}{x_{i+1}-x_i}\int_{x_i}^{x_{i+1}} g_{\max}(x)\mathrm{d}x,$$

where $g_{\min}$ and $g_{\max}$ are defined as in (10.16).

Proof. Combining Lemma 10.10 and Lemma 10.16 provides the result. □

Remark 10.19. While we focus in the sequel on function classes with second-order properties, the technique can provide interpolation conditions for function classes with higher-order or first-order properties.

10.3 Generalized Lipschitz functions

This section presents a unified characterization of classes of interest in second-order optimization, including, e.g., (quasi)-self-concordant functions and (convex) functions with Lipschitz Hessian, and on which the technique presented in Section 10.2 applies. We show that these functions can all be defined as functions whose second derivative satisfies some kind of generalized Lipschitz condition. We call such classes $\mathcal{F}$ *basic function classes*, and will be interested in obtaining interpolation conditions for $\int^{(2)} \mathcal{F}$.

10.3.1 Basic function classes

Basic classes $\mathcal{F}$ are defined as satisfying a generalized Lipschitz property, in the spirit of generalized smoothness introduced in [LQT+24] and generalized self-concordance introduced in [STD19]. Their interest lies in the fact that the associated $\int^{(2)} \mathcal{F}$ include all classes of generalized self-concordant functions [STD19, Equation (1)] of the form

$$|f'''(x)| \leq Af''(x)^\alpha, \; f''(x) \geq 0, \tag{10.28}$$

for some $A, \alpha \geq 0$, as well as the class of functions with Lipschitz Hessian. Throughout, given $\alpha \geq 0$, we define

$$\beta(\alpha) = \begin{cases} \frac{1}{1-\alpha} & \text{if } \alpha \neq 1, \\ 1 & \text{if } \alpha = 1. \end{cases} \tag{10.29}$$

Definition 10.20 (Generalized Lipschitz function)**.** Let $M, \alpha \geq 0$, $f \in \bar{\mathcal{C}}^0$, and $\beta(\alpha)$ be defined as in (10.29). We say that f is $(M, \alpha, +)$-generalized Lipschitz if f (i) belongs to $\mathcal{C}^0$ if $\alpha \leq 1$, (ii) is non-negative everywhere, (iii) is piecewise $\bar{\mathcal{C}}^1$ (or $\mathcal{C}^1$ if $\alpha \leq 1$), hence non-differentiable at a finite number of points only, (iv) satisfies, whenever differentiable,

$$|f'(x)| \leq |\beta(\alpha)| M f(x)^\alpha. \tag{10.30}$$

We denote by $\mathcal{F}_{M,\alpha,+} \subseteq \bar{\mathcal{C}}^0$ (and $\mathcal{C}^0$ when $\alpha \leq 1$) the class of $(M, \alpha, +)$-generalized Lipschitz functions. In addition, we say f is $(M, 0)$-generalized Lipschitz if $(M, 0, +)$-generalized Lipschitz without the non-negativity constraint. In this case, we recover the class of Lipschitz functions $\mathcal{F}_{M,0} \triangleq \mathcal{F}_M$.

We use the notation $\mathcal{F}_{M,\alpha,(+)}$ to refer to both $\mathcal{F}_{M,\alpha,+}$ and $\mathcal{F}_{M,0}$, and propose an alternative definition of $\mathcal{F}_{M,\alpha,(+)}$ that does not involve derivatives.

Proposition 10.21 (2-points definition)**.** *Let $M, \alpha \geq 0$, $f \in \bar{\mathcal{C}}^0$, and f piecewise $\bar{\mathcal{C}}^1$. In addition, if $\alpha \leq 1$, let $f \in \mathcal{C}^0$, and f piecewise $\mathcal{C}^1$. Then, $f \in \mathcal{F}_{M,\alpha,+}$, if and only if $f = 0$, or $\forall x, y \in \mathbb{R}$:*

$$|\tilde{f}(x) - \tilde{f}(y)| \leq M|x - y|, \tag{10.31}$$

$$\begin{cases} f(x) \geq 0, & \text{if } \alpha \leq 1, \\ f(x) > 0, & \text{if } \alpha > 1, \end{cases} \tag{10.32}$$

where $\tilde{f}(x) = \begin{cases} f(x)^{1-\alpha}, & \text{if } \alpha \neq 1, \\ \log(f(x)), & \text{if } \alpha = 1. \end{cases}$

In addition, $f \in \mathcal{F}_{M,0}$ *if and only if it satisfies* (10.31).

Proof. When $f'(x)$ exists, we have $\tilde{f}'(x) = \frac{f'(x)}{|\beta(\alpha)|f(x)^\alpha}$ $\forall \alpha \geq 0$.
(Necessity of (10.31)*)* Let $f \in \mathcal{F}_{M,\alpha,+}$. By assumptions on f and definition of $\tilde{f}$, $\tilde{f}$ is continuous and piecewise $\mathcal{C}^1$. Indeed, if $\alpha \leq 1$, $f \in \mathcal{C}^0$ and is piecewise $\mathcal{C}^1$, and $\tilde{f}$ preserves these properties. On the other hand, if $\alpha > 1$, then $\tilde{f} = 0$ when $f = \infty$, hence, $\tilde{f} \in \mathcal{C}^0$ and is piecewise $\mathcal{C}^1$ despite $f \in \bar{\mathcal{C}}^0$, and piecewise $\bar{\mathcal{C}}^1$.

Let $x_1 < x_2 < \ldots < x_K$ be the points at which $\tilde{f}$ is non-differentiable. Consider any interval $\mathcal{I}_i$ on which $\tilde{f}$ is differentiable, i.e., $[x_i, x_{i+1}]$ for $i \in \{1, 2, \ldots, K-1\}$, $(-\infty, x_1]$ or $[x_K, \infty)$. By the mean value theorem, it holds that $x < y \in \mathcal{I}_i$, $\exists c \in (x, y)$ such that

$$\frac{|\tilde{f}(x) - \tilde{f}(y)|}{|x-y|} = |\tilde{f}'(c)| = \frac{|f'(c)|}{|\beta(\alpha)||f(c)^\alpha|} \overset{(10.30)}{\leq} \frac{|\beta(\alpha)|M|f(c)^\alpha|}{|\beta(\alpha)||f(c)^\alpha|} = M$$

hence $\tilde{f}$ is piecewise M-Lipschitz. Consider now $x < x_i < \ldots < x_j < y$ where $1 \leq i \leq j \leq K$, i.e., x, y may belong to distinct intervals. Then,

$$\begin{aligned} \frac{|\tilde{f}(x) - \tilde{f}(y)|}{|x-y|} &\leq \frac{|\tilde{f}(x) - \tilde{f}(x_i)| + |\tilde{f}(x_i) - \tilde{f}(x_{i+1})| + \ldots + |\tilde{f}(x_j) - \tilde{f}(y)|}{|x-y|} \\ &\leq M \frac{x_i - x + x_{i+1} - x_i + \ldots + y - x_j}{y - x} = M. \end{aligned}$$

(Sufficiency of (10.31)*)* Let $f \in \mathcal{C}^0$, piecewise $\mathcal{C}^1$ and satisfying (10.31). At all $x \in \mathbb{R}$ where f is differentiable, it holds:

$$|\tilde{f}'(x)| = \frac{|f'(x)|}{|\beta(\alpha)||f(x)|^\alpha} = \lim_{h \to 0} \frac{|\tilde{f}(x+h) - \tilde{f}(x)|}{h} \overset{(10.31)}{\leq} M, \tag{10.33}$$

$$\Rightarrow |f'(x)| \leq |\beta(\alpha)| M f(x)^\alpha. \tag{10.34}$$

To conclude, nonnegativity is due to Definition 10.20 and positivity in the case $\alpha \geq 1$ arises from the limit case of (10.31): if $f \in \mathcal{F}_{M,\alpha,+}$, and $f(x) = 0$ at some $x \in \mathbb{R}$, then $f = 0$.

□

Remark 10.22. Proposition 10.21 can be extended to be valid in $\mathbb{R}^d$. It suffices to consider the function $\phi(t) = x + t(y - x)$, and to conduct the proof on $\tilde{f}(\phi(t))$.

Proposition 10.21 thus reduces generalized self-concordant functions to functions f whose associated quantity $\tilde{f}''$ is simply Lipschitz continuous. This allows conducting almost the same reasoning on these functions as on Lipschitz functions.

10.3.2 Examples of function classes $\mathcal{F}_{M,\alpha,(+)}$ and $\int^{(k)} \mathcal{F}_{M,\alpha,(+)}$

We propose a non-exhaustive list of function classes that fall under Definition 10.20.

Zeroth and first-order classes Classical classes of functions whose description involves zeroth and first-order derivatives include the class of M-Lipschitz functions $\mathcal{F}_{M,0} \subseteq \mathcal{C}^0$, i.e., functions satisfying $|f(x) - f(y)| \leq M|x-y|$, $\forall x, y \in \mathbb{R}$ and the class of M-smooth functions $\mathcal{G}_M \triangleq \int \mathcal{F}_{M,0} \subseteq \mathcal{C}^1$, i.e., functions satisfying $|f'(x) - f'(y)| \leq M|x-y|$, $\forall x, y \in \mathbb{R}$.

Second-order classes The class of Hessian Lipschitz functions is $\mathcal{H}_M \triangleq \int^{(2)} \mathcal{F}_{M,0} \subseteq \mathcal{C}^2$, and the class of convex generalized self-concordant functions (10.28) is $\mathcal{H}_{M,\alpha,+} \triangleq \int^{(2)} \mathcal{F}_{M,\alpha,+} \subseteq \bar{\mathcal{C}}^2$ if $\alpha > 1$, and $\mathcal{H}_{M,\alpha,+} \triangleq \int^{(2)} \mathcal{F}_{M,\alpha,+} \subseteq \mathcal{C}^2$ if $\alpha \leq 1$. This class includes, e.g., convex functions with Lipschitz Hessian, $\mathcal{H}_{M,+}$, self-concordant functions, $\mathcal{S}_{M,+} \triangleq \mathcal{H}_{M,3/2,+}$, and quasi-self-concordant functions, $\mathcal{T}_{M,+} \triangleq \mathcal{H}_{M,1,+}$.

10.4 Interpolation conditions for classes of univariate functions

We now exploit the technique proposed in Section 10.2 to obtain interpolation conditions for the basic function classes of Section 10.3. First, we obtain interpolation conditions for $\mathcal{F}_{M,\alpha,(+)}$ by exploiting Proposition 10.21. Then, we rely on our generic technique and Theorem 10.18 to obtain interpolation conditions for $\int \mathcal{F}_{M,\alpha,(+)}$. We then proceed iteratively to obtain interpolation conditions for the specific class $\int^{(2)} \mathcal{F}_{M,0}$.

We show that the classes considered satisfy the assumptions required by Theorem 10.18 in Appendix B.3.1. Moreover, we only present the new interpolation conditions. The proof of their obtention from Theorem 10.18 is deferred to Appendix B.3.2.

10.4.1 Interpolation conditions for $\mathcal{F}_{M,\alpha,(+)}$

Interpolation conditions for classes $\mathcal{F}_{M,\alpha,(+)}$ happen to be discretized versions of Definition (10.31).

Theorem 10.23. *Let $M, \alpha \geq 0$. A set $S = \{(x_i, f_i)\}_{i\in[N]}$ is $\mathcal{F}_{M,\alpha,+}$-interpolable if and only if $f_i = 0$, $\forall i \in [N]$, or $\forall i, j \in [N]$:*

$$|\tilde{f}_i - \tilde{f}_j| \leq M|x_i - x_j|, \tag{10.35}$$

$$\begin{cases} f_i \geq 0 & \text{if } \alpha \leq 1, \\ f_i > 0 & \text{if } \alpha > 1, \end{cases} \tag{10.36}$$

where $\tilde{f}_i = \begin{cases} f_i^{1-\alpha}, & \text{if } \alpha \neq 1, \\ \log(f_i), & \text{if } \alpha = 1. \end{cases}$
In addition, S is $\mathcal{F}_{M,0}$-interpolable if and only if it satisfies (10.35).

Proof. By Proposition 10.21, these conditions are necessary for interpolation. Let us now prove they are also sufficient.

Suppose S satisfies (10.35). We construct a function $f \in \mathcal{F}_{M,\alpha,(+)}$ interpolating S. Let $\tilde{f}(x) = \min_k \tilde{f}_k + M|x - x_k|$ and

$$f(x) = \begin{cases} \tilde{f}(x)^{\alpha-1} & \text{if } \alpha \neq 1, \\ e^{\tilde{f}(x)} & \text{if } \alpha = 1. \end{cases} \tag{10.37}$$

This function interpolates S, i.e.,

$$\text{when } \alpha \neq 1 \ : f(x_i) = (\min_k \tilde{f}_k + M|x_i - x_k|)^{\alpha-1} = \tilde{f}_i^{\alpha-1} = f_i, \tag{10.38}$$

$$\text{when } \alpha = 1 \ : f(x_i) = e^{\min_k \tilde{f}_k + M|x_i - x_k|} = e^{\tilde{f}_i} = f_i, \tag{10.39}$$

since by (10.35)

$$\tilde{f}_i \leq \tilde{f}_j + M|x_j - x_i|, \ \forall j \in [N].$$

Moreover, we can show that $f \in \mathcal{F}_{M,\alpha,+}$ since $\forall x, y \in \mathbb{R}$, we have

$$\begin{aligned} |\tilde{f}(x) - \tilde{f}(y)| &= |\tilde{f}_{k_\star(x)} + M|x - x_{k_\star(x)}| - \tilde{f}_{j_\star(y)} - M|y - x_{j_\star(y)}|| \\ &\leq |\tilde{f}_{j_\star(y)} + M|x - x_{j_\star(y)}| - \tilde{f}_{j_\star(y)} - M|y - x_{j_\star(y)}|| \\ &= M||x - x_{j_\star(y)}| - |y - x_{j_\star(y)}|| \leq M|x - y|, \end{aligned}$$

where $k_\star(x)$ and $j_\star(y)$ are the optimal indices in the definition of $\tilde{f}(x)$ and $\tilde{f}(y)$. The last inequality follows from the reverse triangle inequality. In addition, if $f_i \geq 0\ \forall i \in [N]$ then $f(x) \geq 0\ \forall x \in \mathbb{R}$, and the same holds if $f_i > 0$. Finally, $\tilde{f} \in \mathcal{C}^0$ and piecewise $\mathcal{C}^1$. □

Theorem 10.23 covers the particular case of interpolation conditions for Lipschitz continuous functions $\mathcal{F}_{M,0}$ [Val43].

Remark 10.24. Theorem 10.23 can be extended to be valid in $\mathbb{R}^d$.

Remark 10.25. By Lemma 10.4, Theorem 10.23 furnishes interpolation conditions without function values for $\int \mathcal{F}_{M,\alpha,(+)}$, and interpolation conditions without function values and first derivatives for $\int^{(2)} \mathcal{F}_{M,\alpha,(+)}$.

10.4.2 Interpolation conditions for $\int \mathcal{F}_{M,\alpha,(+)}$

Exploiting Theorem 10.18, we obtain interpolation conditions for $\int \mathcal{F}_{M,\alpha,(+)}$. For the sake of clarity, we separate the case $\alpha = 1$.

Theorem 10.26 (Interpolation conditions for $\int \mathcal{F}_{M,1,+}$)**.** *Let $M \geq 0$. A set $S = \{(x_i, f_i, g_i)\}_{i\in[N]}$ is $\int \mathcal{F}_{M,1,(+)}$-interpolable if and only if, $\forall i, j \in [N]$, $g_i \geq 0$ and*

$$f_j - f_i \geq \frac{1}{M}(g_i + g_j) - \frac{2}{M}\sqrt{g_i g_j} e^{-\frac{M}{2}(x_j - x_i)}. \tag{10.40}$$

Theorem 10.27 (Interpolation conditions for $\int \mathcal{F}_{M,\alpha,(+)}$ $(\alpha \neq 1)$)**.** *Let $\alpha \neq 1, M \geq 0$. A set $S = \{(x_i, f_i, g_i)\}_{i\in[N]}$ is $\int \mathcal{F}_{M,\alpha,+}$-interpolable if and only if, $\forall i, j \in [N]$, $g_i = 0$ and $f_i = f_j$, or $\forall i, j \in [N]$:*

$$|\tilde{g}_i - \tilde{g}_j| \leq M|x_i - x_j|, \tag{10.41}$$

$$\begin{cases} g_i \geq 0 & \text{if } \alpha < 1, \\ g_i > 0 & \text{if } \alpha \geq 1, \end{cases} \tag{10.42}$$

$$\text{If } \tilde{g}_i + \tilde{g}_j \geq \frac{\beta(\alpha)}{|\beta(\alpha)|} M(x_j - x_i) :$$

$$f_j - f_i \geq \frac{\beta(\alpha)}{|\beta(\alpha)| M(\beta(\alpha)+1)} \Big(\tilde{g}_i^{\beta(\alpha)+1} + \tilde{g}_j^{\beta(\alpha)+1} - \frac{1}{2^{\beta(\alpha)}} (\tilde{g}_i + \tilde{g}_j - \frac{\beta(\alpha)}{|\beta(\alpha)|} M(x_j - x_i))^{\beta(\alpha)+1} \Big), \tag{10.43}$$

$$\text{If } \alpha < 1 \text{ and } \tilde{g}_i + \tilde{g}_j \leq M(x_j - x_i) :$$

$$f_j - f_i \geq \frac{1}{M(\beta(\alpha)+1)} \left(\tilde{g}_i^{\beta(\alpha)+1} + \tilde{g}_j^{\beta(\alpha)+1} \right), \tag{10.44}$$

where $\tilde{g}_i = g_i^{1/\beta(\alpha)}$*. In addition, S is* $\int \mathcal{F}_{M,0}$ *if and only if it satisfies* (10.43).

The proofs of Theorem 10.26 and 10.27 are deferred to Appendix B.3.2.

Remark 10.28. By Proposition 10.4, Theorems 10.26 and 10.27 furnish interpolation conditions without function values for $\int^{(2)} \mathcal{F}_{M,\alpha,(+)}$.

Remark 10.29. Necessarily, the interpolation conditions of Theorems 10.26 and 10.27 are satisfied everywhere by any function in $\int \mathcal{F}_{M,\alpha,(+)}$. They are thus completely equivalent to the initial definition of $\int \mathcal{F}_{M,\alpha,(+)}$, i.e., $|f''(x)| \leq |\beta(\alpha)| M f'(x)^\alpha$, when imposed everywhere, even though it is difficult to prove it straightforwardly.

10.4.3 Applications of Theorems 10.26 and 10.27

We propose a list of corollaries to Theorems 10.26 and 10.27, furnishing interpolation conditions without function values to several specific classes $\int^{(2)} \mathcal{F}_{M,\alpha,(+)}$. Whenever possible, we compare these conditions with existing results in the literature.

Class $\mathcal{H}_M$ of functions with M-Lipschitz Hessian We recover the interpolation conditions for M-smooth functions derived in [THG17c, Theorem 4], or equivalently, the interpolation conditions without function values for $\mathcal{H}_M$.

Corollary 10.30. *A set* $\{(x_i, f_i, g_i, h_i)\}_{i\in[N]}$ *is* $\mathcal{H}_M$*-interpolable without functions values if and only if,* $\forall i,j \in [N]$*:*

$$g_j - g_i - h_i(x_j - x_i) \geq -\frac{M}{2}(x_j - x_i)^2 + \frac{1}{4M}(h_j - h_i + M(x_j - x_i))^2. \tag{10.45}$$

Proof. We apply Theorem 10.27 to the set $\{(x_i, g_i, h_i)\}_{i\in[N]}$. □

Class $\mathcal{H}_{M,+}$ of μ-strongly convex functions with M-Lipschitz Hessian

Corollary 10.31. *A set* $\{(x_i, f_i, g_i, h_i)\}_{i\in[N]}$ *is* $\mathcal{H}_{M,+}$*-interpolable without functions values if and only if* $\forall i,j \in [N]$*:*

$$g_j - g_i \geq h_i(x_j - x_i) - \frac{M}{2}(x_j - x_i)^2 + \frac{1}{4M}(h_j - h_i + M(x_j - x_i))^2 \tag{10.46}$$

$$h_i \geq \mu \tag{10.47}$$

$$\textit{If } x_j - x_i \geq \frac{h_i + h_j - 2\mu}{M}, \textit{ then } g_j - g_i \geq \mu(x_j - x_i) + \frac{(h_i - \mu)^2 + (h_j - \mu)^2}{2M}. \tag{10.48}$$

Proof. When $\mu = 0$, $\mathcal{H}_{M,+} = \int^{(2)} \mathcal{F}_{M,0,+}$, hence Corollary 10.31 is an application of Theorem 10.27 as applied to $\mathcal{F}_{M,0,+}$. The case $\mu \geq 0$ follows from the observation that $f \in \mathcal{F}_{M,0,+} \Leftrightarrow f + \frac{\mu}{2}x^2 \in \mathcal{H}_{M,+}$, since the Lipschitzness condition remains valid. □

Remark 10.32. Condition (10.46) is exactly (10.45) and hence is an interpolation condition for the M-smoothness of S, while (10.47) (lower bound on h) ensures S to be consistent with an increasing function, of slope at least μ. However, juxtaposing these conditions alone is not an interpolation condition for the class of M-smooth strongly monotone functions. One must add (10.48), which links both properties, and ensures the existence of some M-Lipschitz continuous function $h(x)$ interpolating h_i and h_j, and whose integral is smaller than or equal to $g_j - g_i$.

Class $\mathcal{S}_{M,+} = \int^{(2)} \mathcal{F}_{M,3/2,+}$ of self-concordant functions

Corollary 10.33. *A set $S = \{(x_i, g_i, h_i)\}_{i\in[N]}$ is $\mathcal{S}_{M,+}$-interpolable if and only if, $\forall i, j \in [N]$, $h_i = 0$ and $g_i = g_j$, or $\forall i, j \in [N]$,*

$$|\tilde{h}_j - \tilde{h}_i| \leq M|x_j - x_i| \textit{ and } h_i > 0 \tag{10.49}$$

$$\textit{If } \tilde{h}_i + \tilde{h}_j > -M(x_j - x_i), \textit{ then } g_j - g_i \geq \frac{1}{M\tilde{h}_i} + \frac{1}{M\tilde{h}_j} - \frac{4}{M(\tilde{h}_i + \tilde{h}_j + M(x_j - x_i))}, \tag{10.50}$$

where $\tilde{h}_i = h_i^{-1/2}$.

Proof. We apply Theorem 10.27 to the set $\{(x_i, g_i, h_i)\}_{i\in[N]}$.. □

Remark 10.34. The class of M-self-concordant functions is affine-invariant, namely, if $f(x) \in \mathcal{S}_{M,+}$ then $g(x) = f(ax + b) \in \mathcal{S}_{M,+}$ $\forall a, b$. This property appears in the interpolation conditions since replacing (x_i, x_j) by $(\frac{x_i - b}{a}, \frac{x_j - b}{a})$, (g_i, g_j) by $a(g_i, g_j)$ and (h_i, h_j) by $a^2(h_i, h_j)$ does not modify them.

Class $\mathcal{T}_{M,+} = \int^{(2)} \mathcal{F}_{M,1,+}$ of quasi-self-concordant functions

Corollary 10.35. *A set $S = \{(x_i, g_i, h_i)\}_{i\in[N]}$ is $\mathcal{T}_{M,+}$-interpolable if and only if, $\forall i, j \in [N]$, $h_i \geq 0$ and*

$$g_j - g_i \geq \frac{h_i + h_j}{M} - \frac{2}{M}\sqrt{h_i h_j e^{-\frac{M}{2}(x_j - x_i)}}. \tag{10.51}$$

Proof. We apply Theorem 10.26 to the set $\{(x_i, g_i, h_i)\}_{i\in[N]}$.. □

Remark 10.36. The class of M-quasi-self-concordant is scale-invariant, namely, if $f \in \mathcal{T}_{M,+}$ then $cf \in \mathcal{T}_{M,+}$ $\forall c > 0$. This property appears in the interpolation conditions since replacing (g_i, g_j, h_i, h_j) by $c(g_i, g_j, h_i, h_j)$ does not modify them.

It is known [Doi23, Lemma 2.7] that quasi-self-concordant functions satisfy:

$$f'(y) - f'(x) - f''(x)(y - x) \leq \frac{1}{M} f''(x)(e^{M|y-x|} - M|y - x| - 1). \tag{10.52}$$

We rely on (10.51) to strengthen this condition:

Lemma 10.37. *If $f \in \mathcal{T}_{M,+}$, then $\forall x, y \in \mathbb{R}$,*

$$\begin{aligned} f'(y) - f'(x) - f''(x)(y - x) \leq &\frac{1}{M} f''(x)\left(e^{M|y-x|} - M|y - x| - 1\right) \\ &- \frac{1}{M}\left(\sqrt{f''(y)} - \sqrt{f''(x)e^{M(y-x)}}\right)^2. \end{aligned} \tag{10.53}$$

Proof. By Corollary 10.35 and Remark 10.29, f satisfies, $\forall x, y \in \mathbb{R}$,

$$\begin{aligned} f'(y) - f'(x) &\leq -\frac{f''(x)}{M} - \frac{1}{M}\left(f''(y) - 2\sqrt{f''(x)f''(y)e^{M(y-x)}}\right) \\ &= -\frac{f''(x)}{M} - \frac{1}{M}\left(-f''(x)e^{M(y-x)} + \left(\sqrt{f''(y)} - \sqrt{f''(x)e^{M(y-x)}}\right)^2\right) \\ &\leq -\frac{f''(x)}{M} - \frac{1}{M}(-f''(x)e^{M|y-x|} + f''(x)M(|y - x| - (y - x))) \\ &\quad - \frac{1}{M}\left(\sqrt{f''(y)} - \sqrt{f''(x)e^{M(y-x)}}\right)^2, \end{aligned}$$

where we used the identity $e^{|t|} - e^t - |t| + t \geq 0$. □

10.4.4 Interpolation conditions for $\int^{(2)} \mathcal{F}_{M,0,(+)}$ with function values

We rely on Theorems 10.27 and 10.18 to obtain interpolation conditions with function values for $\mathcal{H}_M$.

Interpolation conditions with function values for $\mathcal{H}_M$

Theorem 10.38. *A set $S = \{(x_i, f_i, g_i, h_i)\}_{i\in[N]}$ is $\mathcal{H}_M$-interpolable, if and only if, $\forall i, j \in [N]$*

$$|h_j - h_i| \leq M|x_j - x_i|, \tag{10.54}$$

if $h_j - h_i + M|x_j - x_i| \neq 0$, then

$$\begin{aligned} f_j - f_i - g_i(x_j - x_i) - \frac{h_i}{2}(x_j - x_i)^2 \geq &- \frac{M}{6}|x_j - x_i|^3 \\ &+ \frac{\left(g_j - g_i - h_i(x_j - x_i) + \frac{M}{2}|x_j - x_i|(x_j - x_i)\right)^2}{2\left(h_j - h_i + M|x_j - x_i|\right)} \\ &+ \frac{\left(h_j - h_i + M|x_j - x_i|\right)^3}{96M^2}, \end{aligned} \tag{10.55}$$

if $h_j - h_i + M|x_j - x_i| = 0$, then

$$g_j - g_i - h_i(x_j - x_i) = -\frac{M}{2}|x_j - x_i|(x_j - x_i), \tag{10.56}$$

$$f_j - f_i - g_i(x_j - x_i) - \frac{h_i}{2}(x_j - x_i)^2 = -\frac{M}{6}|x_j - x_i|^3. \tag{10.57}$$

The proof of Theorem 10.38 is deferred to Appendix B.4.

Remark 10.39. Thanks to Lemma 10.4, Theorem 10.38 provides interpolation conditions without function values for the class of univariate functions with Lipschitz third derivatives.

It is known [N+18, Lemma 1.2.4] that functions in $\mathcal{H}_M$ satisfy the cubic lower and upper bounds:

$$-\frac{M}{6}|y-x|^3 \leq f(y) - f(x) - f'(x)(y-x) - \frac{f''(x)}{2}(y-x)^2 \leq \frac{M}{6}|y-x|^3 \;\forall x, y \in \mathbb{R}. \tag{10.58}$$

We now show that the conditions of Theorem 10.38 strengthen (10.58), since the additional terms in (10.55) are non-positive.

Lemma 10.40. *Consider a set* $S = \{(x_i, g_i, h_i, f_i)\}_{i\in[N]}$. *If* S *satisfies* (10.54) *and* (10.55) *for both pairs, then for both pairs* S *satisfies* (10.45), *the discretized version of* (10.58), *and*

$$f_j - f_i - g_i(x_j - x_i) - \frac{h_i}{2}(x_j - x_i)^2 \le \frac{M}{6}|x_j - x_i|^3 - \frac{\left(g_j - g_i - h_i(x_j - x_i) - M|x_j - x_i|(x_j - x_i)\right)^2}{2\left(M|x_j - x_i| - h_j - h_i\right)} - \frac{\left(M|x_j - x_i| - (h_j - h_i)\right)^3}{96M^2}. \tag{10.59}$$

The proof of Lemma 10.40 is deferred to Appendix B.5.

We now propose a relaxation of (10.55), satisfied by all functions in $\mathcal{H}_M$:

Corollary 10.41. *If* f *is a univariate twice-differentiable function in* $\mathcal{H}_M$, *then* $\forall x, y \in \mathbb{R}$ *such that* $|f''(x) - f''(y)| \neq M|x - y|$, *we have:*

$$f(y) - f(x) - f'(x)(y - x) - \frac{f''(x)}{2}(y - x)^2 \le \frac{M}{6}|y - x|^3 - \frac{M}{3}\left(\frac{|f'(y) - f'(x) - f''(x)(y - x) - M|y - x|(y - x)|}{M}\right)^{\frac{3}{2}} \tag{10.60}$$

Proof. By Remark 10.29, $f \in \mathcal{H}_M$ satisfies (10.55) everywhere. In addition, for any $a, b, c \ge 0$ it holds:

$$3\frac{a}{c} + \frac{c^3}{b} \ge 4\left(\frac{a^3}{c^3}\frac{c^3}{b}\right)^{\frac{1}{4}} = 4\left(\frac{a^3}{b}\right)^{\frac{1}{4}}. \tag{10.61}$$

Letting $a = \frac{1}{6}(f'(y) - f'(x) - f''(x)(y - x) - M|y - x|(y - x))^2$, $b = 96M^2$, and $c = M|y - x| - (f''(x) - f''(y))$ yields (10.60). ☐

Inequality (10.60) will be exploited in Section 10.5, in Theorem 10.43.

10.5 Performance guarantees of optimization methods

In the previous sections, we obtained interpolation conditions, i.e., "exact descriptions", of different function classes (see Table 10.1). We now exploit these interpolation conditions, analytically or numerically with the PEP framework presented in Section 10.1, to propose new convergence guarantees for second-order methods on classes of univariate functions. Table 10.2 summarizes results of the literature that we were able to improve (either analytically or numerically) in the univariate case, or for which we proved tightness. We also propose to analyze a few setups that were not considered before.

We selected a subset of existing second-order optimization methods for which we could obtain fair and interesting comparisons with the literature.

All the numerical experiments are performed in Matlab and can be found at

`https://github.com/NizarBousselmi/Second-Order-Univariate-PEP.`

We performed our numerical experiments on a laptop computer (Intel Core i9 CPU, 32 GB RAM). As mentioned, the non-convex PEPs are solved with the non-convex solver of Gurobi [Gur24], which can deal with non-convex quadratic but also non-convex general constraints.

10.5.1 Improved descent lemma of the Cubic Regularized Newton method

The Cubic Regularized Newton method (CNM) [NP06] minimizes at each iteration the quadratic approximation of the function around the current iterate regularized by a cubic penalty. In the univariate case, the next iterate x_{k+1} is computed as the global solution of the following problem:

$$x_{k+1} = \operatorname{Arg\,min}_x f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k)(x - x_k)^2 + \frac{M}{6}|x - x_k|^3. \tag{CNM}$$

In practice, one can obtain x_{k+1} by solving a convex univariate problem [NP06, Section 5]. From the existing multivariate descent lemma and convergence rate from [NP06], we can write the following global rate of convergence on non-convex Hessian Lipschitz functions:

Theorem 10.42 ([NP06], Theorem 1)**.** *The iterates of the Cubic Regularized Newton method* (CNM) *on Hessian M-Lipschitz univariate functions satisfy*

$$f(x_k) - f(x_{k+1}) \geq \frac{M}{12} \max\left\{ \sqrt{\frac{|f'(x_{k+1})|}{M}}, -\frac{2}{3}\frac{f''(x_{k+1})}{M} \right\}^3. \tag{10.62}$$

Moreover, if the function is bounded below by $f_\star$, then

$$\min_{k=1,\dots,N} |f'(x_k)| \leq 4M^{\frac{1}{3}} \left(\frac{3(f(x_0)-f_\star)}{2N}\right)^{\frac{2}{3}}. \tag{10.63}$$

Exploiting the refined description of Hessian Lipschitz functions, and in particular the improved cubic bound of Corollary 10.41 allows to improve this original descent lemma (10.62) for univariate functions by a factor of 5. It results in an improvement of factor $5^{\frac{2}{3}} \approx 2.9$ on the sub-linear global convergence rate (10.63) of CNM on univariate functions.

Theorem 10.43 (Improved descent lemma and gradient convergence rate)**.** *The iterates of the Cubic Regularized Newton method* (CNM) *on Hessian M-Lipschitz univariate functions satisfy*

$$f(x_k) - f(x_{k+1}) \geq \frac{5M}{12}\sqrt{\frac{|f'(x_{k+1})|}{M}}^3. \tag{10.64}$$

Moreover, if the function is bounded below by $f_\star$, then

$$\min_{k=1,\dots,N} |f'(x_k)| \leq \frac{4M^{\frac{1}{3}}}{5^{\frac{2}{3}}} \left(\frac{3(f(x_0)-f_\star)}{2N}\right)^{\frac{2}{3}}. \tag{10.65}$$

Proof. We first establish (10.64). The CNM iterates satisfy the following properties (see [NP06, Equation (2.5), Proposition 1, Lemmas 2 and 3])

$$f'(x_k) + f''(x_k)(x_{k+1}-x_k) + \frac{M}{2}(x_{k+1}-x_k)|x_{k+1}-x_k| = 0, \tag{10.66}$$

$$f''(x_k) + \frac{M}{2}|x_{k+1}-x_k| \geq 0, \tag{10.67}$$

$$f'(x_k)(x_{k+1}-x_k) \leq 0, \tag{10.68}$$

$$M|x_{k+1}-x_k|^2 \geq |f'(x_{k+1})|. \tag{10.69}$$

Exploiting the new term (in blue) from the improved cubic bound (10.60) of Corollary 10.41 yields

$$
\begin{aligned}
f(x_{k+1}) &\overset{(10.60)}{\leq} f(x_k) + f'(x_k)(x_{k+1} - x_k) + \frac{f''(x_k)}{2}(x_{k+1} - x_k)^2 + \frac{M}{6}|x_{k+1} - x_k|^3 \\
&\quad - \frac{M}{3}\left(\frac{\left|f'(x_{k+1}) - f'(x_k) - f''(x_k)(x_{k+1} - x_k) - \frac{M}{2}(x_{k+1} - x_k)|x_{k+1} - x_k|\right|}{M}\right)^{\frac{3}{2}} \\
&\overset{(10.66)}{=} f(x_k) + \frac{f'(x_k)}{2}(x_{k+1} - x_k) - \frac{M}{12}|x_{k+1} - x_k|^3 - \frac{M}{3}\left(\frac{|f'(x_{k+1})|}{M}\right)^{\frac{3}{2}} \\
&\overset{(10.68)}{\leq} f(x_k) - \frac{M}{12}|x_{k+1} - x_k|^3 - \frac{M}{3}\left(\frac{|f'(x_{k+1})|}{M}\right)^{\frac{3}{2}} \\
&\overset{(10.69)}{\leq} f(x_k) - \frac{M}{12}\left(\frac{|f'(x_{k+1})|}{M}\right)^{\frac{3}{2}} - \frac{M}{3}\left(\frac{|f'(x_{k+1})|}{M}\right)^{\frac{3}{2}} \\
&= f(x_k) - \frac{5M}{12}\left(\frac{|f'(x_{k+1})|}{M}\right)^{\frac{3}{2}}
\end{aligned}
$$

establishing the result. Telescoping this new descent lemma (10.64) yields the convergence rate in (10.65)

$$
\begin{aligned}
f(x_0) - f_\star \geq f(x_0) - f(x_N) = \sum_{k=0}^{N-1} f(x_k) - f(x_{k+1}) &\overset{(10.64)}{\geq} \frac{5M}{12}\sum_{k=0}^{N-1}\left(\frac{|f'(x_{k+1})|}{M}\right)^{\frac{3}{2}} \\
&\geq \frac{5MN}{12}\min_{k=1,\ldots,N}\left(\frac{|f'(x_k)|}{M}\right)^{\frac{3}{2}}.
\end{aligned}
$$

□

Finally, the bound on the Hessian follows from the fact that the measure on the gradient of the iterates is always greater than the measure on the Hessian of the iterates for univariate functions.

Theorem 10.44. *The iterates of the Cubic Regularized Newton method* (CNM) *on Hessian M-Lipschitz univariate functions satisfy*

$$
\sqrt{\frac{|f'(x_{k+1})|}{M}} \geq -\frac{2}{3M}f''(x_{k+1}). \tag{10.70}
$$

Proof. Suppose w.l.o.g. that $|x_{k+1} - x_k| = x_{k+1} - x_k \geq 0$ and recall that

$$
|f''(x_{k+1}) - f''(x_k)| \leq M|x_{k+1} - x_k| \tag{10.71}
$$

therefore, we have

$$
\begin{aligned}
|f'(x_{k+1})| &\geq -f'(x_{k+1}) \\
&\overset{(10.45)}{\geq} -f'(x_k) - \frac{M(x_{k+1}-x_k)^2}{4} + \frac{(f''(x_{k+1})-f''(x_k))^2}{4M} \\
&\quad - \frac{1}{2}(f''(x_{k+1})+f''(x_k))(x_{k+1}-x_k) \\
&\overset{(10.66)}{=} \frac{M(x_{k+1}-x_k)^2}{4} + \frac{(f''(x_{k+1})-f''(x_k))^2}{4M} \\
&\quad - \frac{1}{2}(f''(x_{k+1})-f''(x_k))(x_{k+1}-x_k) \\
&= \frac{1}{4M}(f''(x_{k+1})-f''(x_k)-M(x_{k+1}-x_k))^2 \\
\Rightarrow \sqrt{\frac{|f'(x_{k+1})|}{M}} &\geq \frac{1}{2M}|f''(x_{k+1})-f''(x_k)-M|x_{k+1}-x_k|| \\
&\overset{(10.71)}{\geq} \frac{1}{2M}\left(M|x_{k+1}-x_k| - f''(x_{k+1}) + f''(x_k)\right) \\
&= \frac{1}{2M}\left(\frac{2}{3}M|x_{k+1}-x_k| + \frac{1}{3}M|x_{k+1}-x_k| - f''(x_{k+1}) + f''(x_k)\right) \\
&\overset{(10.67),(10.71)}{\geq} -\frac{4}{3}f''(x_k) + \frac{1}{3}(f''(x_k)-f''(x_{k+1})) - f''(x_{k+1}) + f''(x_k)) \\
&= -\frac{2}{3M}f''(x_{k+1}).
\end{aligned}
$$

□

For a single iteration, the improved descent lemma (10.64) is satisfied with equality for the function $f(x) = M\frac{x^3}{6} - \frac{x^2}{2}$ on the starting point $x_0 = 0$. The Cubic Regularized Newton iteration (CNM) on f from x_0 is

$$
\begin{aligned}
x_1 &= \operatorname{Arg\,min}_x f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2 + \frac{M}{6}|x-x_0|^3 \\
&= \operatorname{Arg\,min}_x \frac{1}{2}(-x^2 + \frac{M}{3}|x|^3) = \pm\frac{2}{M}
\end{aligned}
$$

where both signs are global minimum of the cubic bound. Let $x_1 = -\frac{2}{M}$. We have $f(x_0) = 0$, $f(x_1) = -\frac{10}{3M^2}$, $f'(x_1) = \frac{4}{M}$ and therefore $f(x_0) - f(x_1) = \frac{5M}{12}\left(\frac{|f'(x_1)|}{M}\right)^{\frac{3}{2}} = \frac{10}{3M^2}$.

Remark 10.45. Thanks to the improved descent lemma (10.64), we can also improve the results on the convergence rate of CNM on univariate gradient-dominated functions [NP06, Section 4.2]. We can replace quantities $\hat{\omega}$ and

$\tilde{\omega}$ respectively by $\frac{\tilde{\omega}}{5^2}$ and $5^4\tilde{\omega}$ in [NP06, Theorems 6.1 and 6.2].

Remark 10.46. We can also analyze an inexact variant of CNM with arbitrary g and h instead of exact derivatives f' and f''. Then, we obtain an improved descent lemma:

$$\begin{aligned} f(x_{k+1}) \leq & f(x_k) + \frac{2}{3}\sqrt{\frac{12}{M}}|f'(x_k) - g|^{\frac{3}{2}} + \frac{72}{M^2}|f''(x_k) - h|^3 - \frac{M}{36}|x_{k+1} - x_k|^3 \\ & - \frac{1}{3\sqrt{M}}\left(|f'(x_{k+1})| - |f'(x_k) - g| - |(f''(x_k) - h)(x_{k+1} - x_k)|\right)^{\frac{3}{2}} \end{aligned}$$

which can be compared to [CDJ23, Theorem C.1] and has better factors.

Remark 10.47. Since our interpolation conditions only hold for univariate functions, Theorem 10.43 only holds in this setting. However, we believe and conjecture that it also holds in the multivariate case. To support this conjecture, we computed the decrease $f(x_k) - f(x_{k+1})$ of CNM on functions of the form $f(x) : \mathbb{R}^d \to \mathbb{R} : x \mapsto f(x) = \sum_{i=1}^n a_i x_i^3 + \frac{1}{2}x^T Q x + b^T x + c$ for 10^6 values for randomly generated parameters a, Q, b, c and $d = 1, \ldots, 10$, and observed that the descent (10.64) (with the norm of the gradient instead of the absolute value of the derivative) with the new factor 5 always held.

Tightness of multiple steps Even if the improved descent lemma is tight for a single iteration, the worst-case decrease after multiple iterations can be better than predicted by (10.64) only. Indeed, a better bound can result from the use of a more sophisticated combination of the interpolation conditions. The PEP methodology automatically computes numerically this optimal combination. Figure 10.3 compares the worst-case performance for multiple iterations of CNM computed by PEP (blue dots) and as predicted by [NP06, Theorem 1] (red line) and by Theorem 10.43 (blue line). We also plot the gradient residual of CNM for the univariate function $f^{(4)}$ of [CGT10] for which CNM shows a $\mathcal{O}(1/k^{\frac{2}{3}})$ convergence. Figure 10.3 illustrates the two possible kinds of conservatism of a performance guarantee. On one hand, we must use the correct interpolation conditions (blue line instead of red line), and on the other hand, we must combine them optimally (blue dots instead of blue line).

Lower bound on the general worst-case performance Since we solve the PEP in the univariate case, we do not have the general worst-case performance of the method. However, the univariate worst-case is an example of a function for which the method is slow, in other words, it

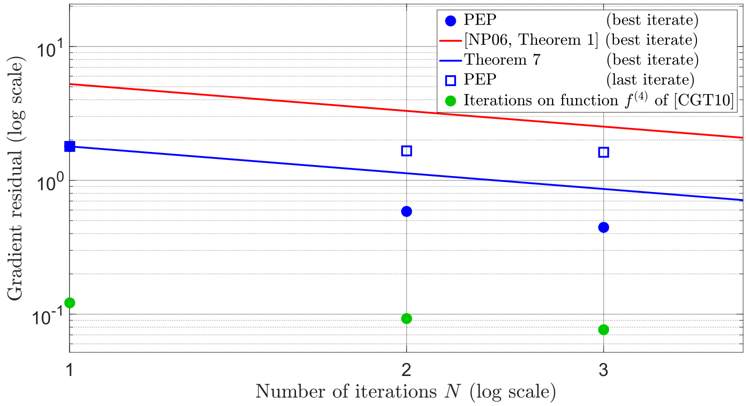


Fig. 10.3 Worst-case performance of the best $\min_{k=1,\ldots,N} |f'(x_k)|$ and last $|f'(x_N)|$ iterates of the Cubic Newton method on Hessian M-Lipschitz functions for varying number of iterations N and $M = f_0 - f_N = 1$. Solving the non-convex PEP of this figure required around 30 minutes.

provides a lower bound on the general worst-case performance. In [CGT10], they exhibited a univariate function $f^{(4)}$ for which CNM has the $\mathcal{O}(1/k^{\frac{2}{3}})$ convergence (green dots in Figure 10.3), thus providing a lower bound on the exact worst-case performance. With PEP, we can provide numerically a worse function (blue dots in Figure 10.3) and therefore close the gap between the best known upper (red line) and lower bounds on the general multivariate worst-case performance of CNM.

Last-iterate convergence rate PEP allows the study of different performance criteria. In particular, in addition to the classical best-iterate analysis of [NP06, Theorem 1], we can also consider the last-iterate convergence of CNM with PEP without additional effort. In Figure 10.3 the last-iterate convergence (blue squares) has a slower order of convergence than the best-iterate convergence. Such last-iterate analysis of CNM has never been done and is straightforward with PEP.

Parameter selection A tight worst-case performance guarantee allows selecting the optimal parameters optimizing the worst-case. We consider the Cubic Regularized Newton method with step size α (CNM-α) defined as the solution of the modified CNM iteration

$$x_{k+1} = \operatorname{Arg\,min}_x f(x) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k)(x - x_k)^2 + \frac{M}{6\alpha}|x - x_k|^3. \tag{10.72}$$

Interpolation conditions together with PEP allow to analyze the worst-case performance of CNM-α for different values of α. In particular, we can select the step size α optimizing the worst-case performance. Figure 10.4 shows the worst-case performance of one iteration of CNM-α for varying step size α and different values of Hessian Lipschitz constant M. Following a similar

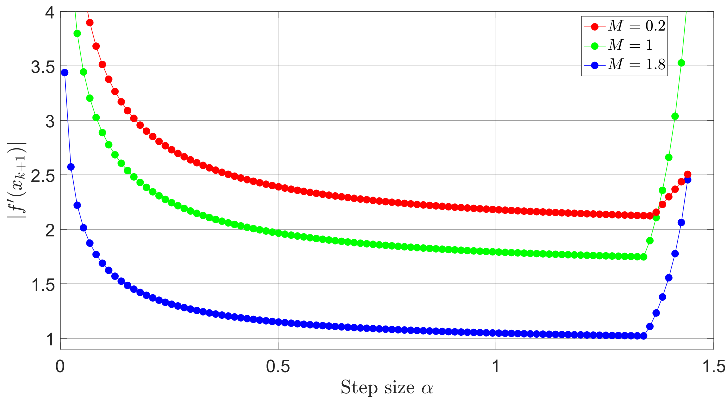


Fig. 10.4 Worst-case performance $|f'(x_{k+1})|$ of the Cubic Regularized Newton method for different step size α on Hessian M-Lipschitz functions with $f_k - f_{k+1} = 1$ and different values of M. Solving the non-convex PEP of this figure required around 15 minutes.

reasoning that [THG17c, Section 3.5], it can be shown that this worst-case scales with $(f_k - f_{k+1})^{\frac{2}{3}}$. Further analysis of such numerical results can lead to the development of new optimized methods.

10.5.2 Gradient Regularized Newton method on convex and strongly convex functions

The Gradient Regularized Newton method 2 (GNM2) [Mis23] is a variant of the Cubic Newton method that achieves a $\mathcal{O}(1/k^2)$ global convergence (faster than the $\mathcal{O}(1/k^{\frac{2}{3}})$ of CNM) and, unlike CNM, has a fully explicit iteration

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k) + \sqrt{\frac{M}{2}|f'(x_k)|}}. \tag{GNM2}$$

Global convergence The proof of its $\mathcal{O}(1/k^2)$ convergence relies on the following descent lemma (proved in the multivariate case):

Lemma 10.48 ([Mis23], Theorem 2.6). *Under the assumption that* $|f'(x_{k+1})| \geq \frac{1}{4}|f'(x_k)|$*, the iterations of* (GNM2) *on convex and M-Lipschitz univariate functions f satisfy*

$$f(x_k) - f(x_{k+1}) \geq \frac{|f'(x_k)|^{\frac{3}{2}}}{48\sqrt{2M}}. \tag{10.73}$$

Again, in the univariate case, we can numerically improve this lemma by a factor of 12 this time. The proof remains another question.

Conjecture 10.49. *Under the assumption that* $|f'(x_{k+1})| \geq \frac{1}{4}|f'(x_k)|$*, the iterations of* (GNM2) *on convex and M-Lipschitz univariate functions f satisfy*

$$f(x_k) - f(x_{k+1}) \geq \frac{|f'(x_k)|^{\frac{3}{2}}}{4\sqrt{2M}}. \tag{10.74}$$

Local convergence GNM2 has a superlinear convergence on strongly convex functions formalized in the following theorem (proved in the multivariate case):

Theorem 10.50 (Theorem 2.7 of [Mis23]). *The iterations of* (GNM2) *on* μ*-strongly convex and Hessian M-Lipschitz univariate functions satisfy*

$$|f'(x_{k+1})| \leq \frac{M}{2\mu^2}|f'(x_k)|^2 + \frac{1}{\mu}\sqrt{\frac{M}{2}}|f'(x_k)|^{\frac{3}{2}}. \tag{10.75}$$

Moreover, if $|f'(x_0)| < \frac{\mu^2}{2M}$*, then*

$$\frac{|f'(x_{k+1})|}{|f'(x_k)|} \leq \sqrt{\frac{2M}{\mu^2}|f'(x_k)|} < 1. \tag{10.76}$$

Figure 10.5 compares the worst-case local convergence of GNM2 on μ-strongly convex Hessian M-Lipschitz univariate functions from (10.75) (red dots) and from PEP (blue dots) for varying number of iterations N, initial gradient $|f'(x_0)| \leq R = \frac{\mu^2}{4M}$, $\mu = 0.1$ and $M = 1$.

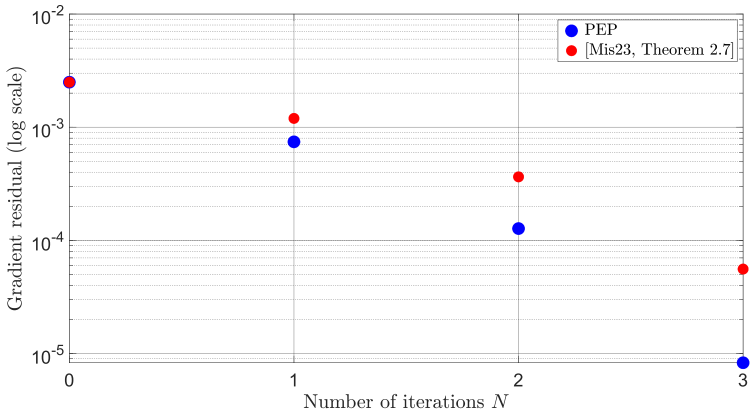


Fig. 10.5 Worst-case performance of GNM2 for varying number of iterations N and $|f'(x_0)| \leq \frac{\mu^2}{4M}$, $M = 1$, and $\mu = 0.1$. Solving the non-convex PEP of this figure required around 1 minute.

10.5.3 Local quadratic convergence of Newton's method on non-convex Hessian Lipschitz functions

Newton's method benefits from local quadratic convergence. The following result from [N$^+$18] formally states this rate of convergence.

Theorem 10.51 (Theorem 1.2.5 of [N$^+$18]). *If*

- *f has a M-Lipschitz continuous Hessian,*
- *$\exists x_\star$ such that $\nabla f(x_\star) = 0$, $\nabla^2 f(x_\star) = \mu I \succ 0$,*
- *$\frac{M}{\mu}||x_0 - x_\star|| \leq \frac{2}{3}$,*

then all Newton iterations $x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$ satisfy

$$||x_{k+1} - x_\star|| \leq \frac{\frac{M}{\mu}||x_k - x_\star||^2}{2\left(1 - \frac{M}{\mu}||x_k - x_\star||\right)}. \tag{10.77}$$

This theorem is well established, but, to the best of our knowledge, its tightness remained an open question. Relying on a PEP analysis, we identified a simple function attaining (10.77), establishing its tightness for any number of iterations.

Theorem 10.52. *Theorem 10.51 (i.e., [N$^+$18, Theorem 1.2.5]) is tight and attained by the following univariate cubic by parts function:*

$$f(x) = -M\frac{|x|^3}{6} + \mu\frac{x^2}{2}. \tag{10.78}$$

Proof. We have $x_\star = 0$ and $x_{k+1} = \frac{-\frac{M}{\mu}x_k|x_k|}{2(1-\frac{M}{\mu}|x_k|)}$ that attains (10.77) when $\frac{M}{\mu}|x_k| < 1$. □

This is another example where the univariate case is sufficiently "rich" to provide the general worst-case behavior. Such surprisingly simple worst-case functions were already observed, for example, for the Gradient method applied to smooth (strongly) convex functions [THG17c].

Gradient method applied to Hessian Lipschitz functions The gradient method has a local linear rate of convergence when applied to Hessian Lipschitz functions [N$^+$18]. Since the tool is not restricted to second-order schemes, we can also examine the tightness of the following theorem on the linear convergence of the gradient method.

Theorem 10.53 ([N$^+$18], Theorem 1.2.4 and Equation (1.2.26))**.** *If*

- *f has M-Lipschitz continuous Hessian,*
- *$\exists x_\star$ such that $\nabla f(x_\star) = 0$, $\mu I \preceq \nabla^2 f(x_\star) \preceq LI$,*
- *$\|x_0 - x_\star\| < \frac{2\mu}{M}$,*

then all Gradient iterations with constant step size $x_{k+1} = x_k - h\nabla f(x_k)$ satisfy

$$\|x_{k+1} - x_\star\| \leq \max\left\{1 - h\left(\mu - \frac{M}{2}\|x_k - x_\star\|\right), h\left(L + \frac{M}{2}\|x_k - x_\star\|\right) - 1\right\}\|x_k - x_\star\|. \tag{10.79}$$

PEP results exactly match the worst-case behavior guaranteed by Theorem 10.53, meaning that it is also unimprovable. Indeed, we can exhibit simple univariate functions attaining this bound for any number of iterations.

Theorem 10.54. *Theorem 10.53 (i.e., [N$^+$18, Equation (1.2.26)] is tight and attained by the two following functions:*

$$f(x) = -M\frac{|x|^3}{6} + \mu\frac{x^2}{2}, \tag{10.80}$$

$$g(x) = M\frac{|x|^3}{6} + L\frac{x^2}{2}. \tag{10.81}$$

Proof. The iterations of the Gradient method on functions f and g are

$$x_{k+1} = x_k - hf'(x_k) = x_k - h\left(-\frac{M}{2}x_k|x_k| + \mu x_k\right) = \left(1 - h\left(\mu - \frac{M}{2}|x_k|\right)\right)x_k$$
$$x_{k+1} = x_k - hg'(x_k) = x_k + h\left(\frac{M}{2}x_k|x_k| + Lx_k\right) = \left(1 + h\left(L + \frac{M}{2}|x_k|\right)\right)x_k.$$

□

Figure 10.6 shows the iterations of the Gradient method on functions f and g. Interestingly, it exhibits the same kind of behavior as for smooth convex functions, i.e., short steps slowly converge whereas long steps overshoot [THG17c, Figure 3].

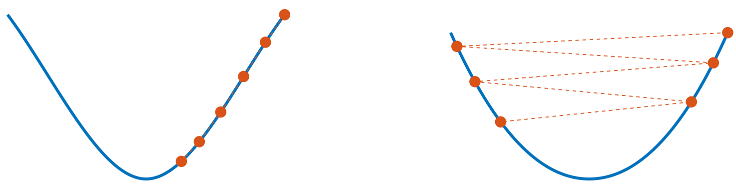


Fig. 10.6 $N = 5$ iterations of gradient method on f (left) and g (right) with step sizes $h = \frac{2}{L+\mu}$ (left), defined in (10.80) and (10.81), and $h = \frac{2.1}{L+\mu}$ (right) and $|x_0 - x_\star| = 0.42$, $M = L = 1$, $\mu = 0.3$.

Fixed damped Newton method A simple variant of Newton's method is the fixed damped Newton method (DNM) adding a damping coefficient $\alpha \geq 0$ to the classical Newton step

$$x_{k+1} = x_k - \alpha\frac{f'(x_k)}{f''(x_k)}. \tag{10.82}$$

This modification expands the region of convergence of the classical Newton's method, although the adaptive damped Newton method is even more efficient [Hil21, IH24, N$^+$18]. We analyze the local behavior of DNM on the same setting as Theorem 10.51 (i.e., Hessian Lipschitz functions) for different initial distances and damping coefficients.

The worst-case performance of DNM on univariate Hessian Lipschitz functions returned by PEP allowed us to identify the analytical expression of functions attaining this worst-case. Depending on the parameters of the settings, we observed that the worst-case performance is exactly attained by one of the functions of the following family of functions parametrized by a

$$f(x) = -M\frac{|x|^3}{6} + \mu\frac{x^2}{2}, \tag{10.83}$$

$$g_a(x) = \begin{cases} -M\frac{x^3}{6} + \mu\frac{x^2}{2} & \text{if } x \leq \frac{\mu+a}{2M}, \\ M\frac{x^3}{6} + a\frac{x^2}{2} + A_1x + B_1 & \text{if } x \geq \frac{\mu+a}{2M}, \end{cases} \tag{10.84}$$

$$h_a(x) = \begin{cases} -M\frac{x^3}{6} + a\frac{x^2}{2} & \text{if } x \leq \frac{a-\mu}{2M}, \\ M\frac{x^3}{6} + \mu\frac{x^2}{2} + A_2x + B_2 & \text{if } \frac{a-\mu}{2M} \leq x \leq 0, \\ -M\frac{x^3}{6} + \mu\frac{x^2}{2} + A_3x + B_3 & \text{if } x \geq 0, \end{cases} \tag{10.85}$$

where A_i and B_i are (tediously but elementary to compute) constants such that the different parts of the function (and its gradient) connect correctly. Given the parameters of the setting, we must consider the worst functions of the families, denoted $g_\star$ and $h_\star$.

Figure 10.7 shows the evolution of the performance of DNM for varying initial distances. We can also observe the impact of the step size on the worst-case performance in Figure 10.8. It allows to select the optimal damping coefficient.

10.5.4 Self-concordant functions

We now analyze the convergence of Newton and damped Newton methods on self-concordant functions using the interpolation conditions of Theorem 10.33. We consider standard-self-concordant functions, i.e., $M = 1$. We consider the evolution of the *Newton decrement* defined as the size of the Newton step in the local norm, namely, in the univariate case

$$\lambda_f(x_k) = \frac{|f'(x_k)|}{\sqrt{f''(x_k)}} \tag{10.86}$$

We only need interpolation without function value since neither the methods nor the performance criterion use function values.

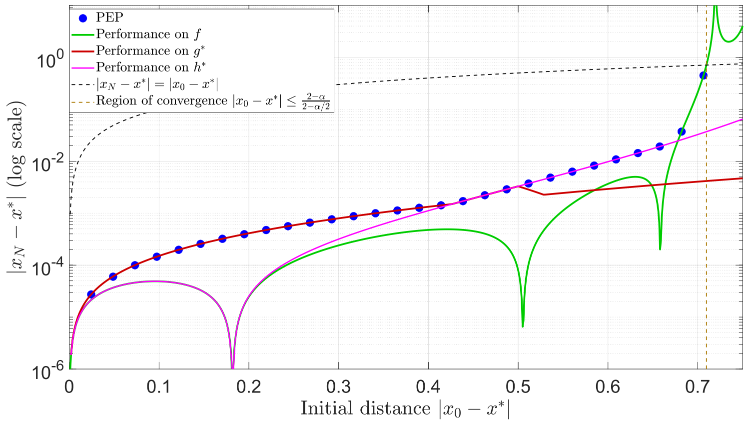


Fig. 10.7 Worst-case performance by PEP (blue dots) of $N = 3$ iterations of damped Newton method for varying initial distance $|x_0 - x_\star|$ and $M = \mu = 1$, $\alpha = 0.9$. We also represented the performance of DNM on functions f (green curve), $g_\star$ (red curve), and $h_\star$ (magenta curve). Solving the non-convex PEP of this figure required around 2 minutes.

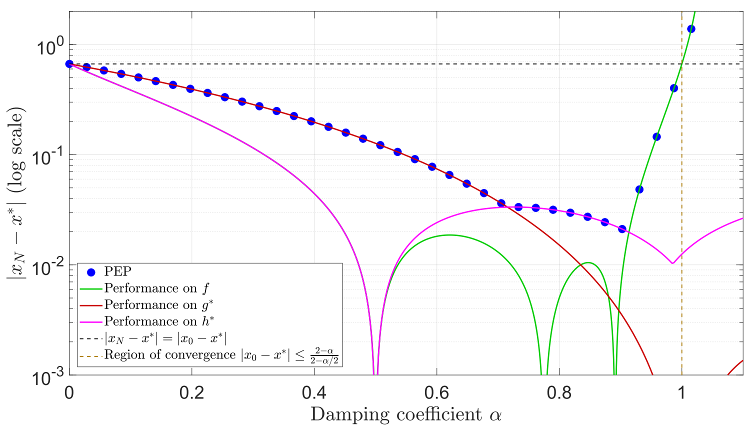


Fig. 10.8 Worst-case performance by PEP (blue dots) of $N = 3$ iterations of damped Newton method with $M = \mu = 1$ and $\frac{M}{\mu}|x_0 - x_\star| = \frac{2}{3}$ on univariate functions for varying damped coefficient. We also represented the performance of DNM on functions f (green curve), $g_\star$ (red curve), and $h_\star$ (magenta curve). Solving the non-convex PEP of this figure required around 1 minute.

Newton's method The following Theorem [Hil21, Equation (11)] provides the tight worst-case performance of a single iteration of Newton's method on univariate functions.

Theorem 10.55 ([Hil21], Equation (11))**.** *Given the initial Newton decrement $\lambda_f(x_k) \leq 1$. The Newton decrement of a single iteration of Newton's method, $\lambda_f(x_{k+1})$, on standard-self-concordant functions satisfies*

$$\lambda_f(x_{k+1}) \leq 4 - \lambda_f(x_k)^2 - 4\sqrt{1 - \lambda_f(x_k)^2}. \tag{10.87}$$

Moreover, there exists a standard-self-concordant function attaining exactly this bound.

This theorem is tight and cannot be improved. However, to illustrate the scope of the tool, we (i) propose an alternative proof to the theorem that consists of explicitly solving the associated PEP, (ii) provide the analytical expression of a standard-self-concordant function attaining the bound of the theorem, (iii) extend numerically the theorem to multiple steps.

Alternative proof of Theorem 10.55 We propose to formulate the PEP associated to Theorem 10.55 and to solve it analytically. The PEP can be written as

$$\begin{aligned} \max_{S=\{(x_0,g_0,h_0),(x_1,g_1,h_1)\}} \; & \frac{|g_1|}{\sqrt{h_1}} \\ \text{s.t. } & x_1 = x_0 - \frac{g_0}{h_0}, \\ & \frac{|g_0|}{\sqrt{h_0}} = R, \\ & S \text{ is standard-self-concordant-interpolable.} \end{aligned} \tag{10.88}$$

For simplicity, we consider variable $\tilde{h}_i = \frac{1}{\sqrt{h_i}}$ instead of h_i. We assume $x_0 = 0$ and $x_1 = 1$ w.l.o.g. since a translation can move x_0 to zero and a linear change of variable $f(x) \to f(cx)$ can move x_1 to one. Therefore, the two equality constraints of (10.88) lead to $g_0 = -R^2$ and $\tilde{h}_0 = \frac{1}{R}$. The interpolation conditions are (see Theorem 10.33)

$$|\tilde{h}_1 - \tilde{h}_2| \leq 1, \tag{10.89}$$

$$g_1 - g_0 \geq \frac{1}{\tilde{h}_0} + \frac{1}{\tilde{h}_1} - \frac{4}{\tilde{h}_0 + \tilde{h}_1 + 1}, \tag{10.90}$$

$$\text{If } \tilde{h}_0 + \tilde{h}_1 > 1 \text{ then } g_0 - g_1 \geq \frac{1}{\tilde{h}_0} + \frac{1}{\tilde{h}_1} - \frac{4}{\tilde{h}_0 + \tilde{h}_1 - 1}. \tag{10.91}$$

We considered that $h_1 \neq 0$, otherwise (10.88) would be unbounded. The problem can now be written as

$$\begin{aligned} \max_{g_1, \tilde{h}_1} \ & |g_1| \tilde{h}_1 \\ \text{s.t. } & |\tilde{h}_1 - \tilde{h}_0| \leq 1, \\ & g_1 - g_0 \geq \frac{1}{\tilde{h}_0} + \frac{1}{\tilde{h}_1} - \frac{4}{\tilde{h}_0 + \tilde{h}_1 + 1}, \\ & \text{If } \tilde{h}_0 + \tilde{h}_1 > 1 \text{ then } g_0 - g_1 \geq \frac{1}{\tilde{h}_0} + \frac{1}{\tilde{h}_1} - \frac{4}{\tilde{h}_0 + \tilde{h}_1 - 1} \end{aligned} \tag{10.92}$$

with $g_0 = -R^2$ and $\tilde{h}_0 = \frac{1}{R}$. One can show that the optimal solution will satisfy

$$g_1 = g_0 - \frac{1}{\tilde{h}_0} - \frac{1}{\tilde{h}_1} + \frac{4}{\tilde{h}_0 + \tilde{h}_1 - 1} \geq 0. \tag{10.93}$$

Therefore, we maximize the performance $p(\tilde{h}_1)$

$$p(\tilde{h}_1) = g_1 \tilde{h}_1 = g_0 \tilde{h}_1 - \frac{\tilde{h}_1}{\tilde{h}_0} - 1 + \frac{4\tilde{h}_1}{\tilde{h}_0 + \tilde{h}_1 - 1} = -R(R+1)\tilde{h}_1 - 1 \frac{4\tilde{h}_1 R}{1 + \tilde{h}_1 R - R}. \tag{10.94}$$

The first-order optimality condition $p'(\tilde{h}_1) = 0$ yields

$$\tilde{h}_1 R = R - 1 + 2\sqrt{\frac{1-R}{1+R}} \tag{10.95}$$

and therefore the optimal value of (10.88) is

$$p(\tilde{h}_1) = g_1 \tilde{h}_1 = 4 - R^2 - 4\sqrt{1 - R^2} \tag{10.96}$$

with R and $p(\tilde{h}_1)$ the Newton decrement before and after the iteration, which is the bound of Theorem 10.55.

Function attaining the bound for a single iteration Relying on the tool, we found a function attaining the bound (10.87) for a single iteration. One can check that Newton iteration from the point $x_k = 0$ on the following functions yields the result

$$f(x) = \begin{cases} Bx - \log(x - A) & \text{if } x \leq \frac{1}{2}(A - R), \\ \left(B + \frac{4}{R+A}\right) x - \log(x + R) & \text{if } x > \frac{1}{2}(A - R), \end{cases} \tag{10.97}$$

where $R = \lambda_f(x_k)$, $A = R\frac{-R^2+(-4+2\sqrt{1-R^2})R+5-\sqrt{1-R^2}}{R^2-1+2\sqrt{1-R^2}}$, and $B = \frac{R^2-1+2\sqrt{1-R^2}}{R(R-1)}$.

Worst-case performance for multiple iterations As in the previous case of Theorem 10.43 for (CNM), even if Theorem 10.55 is tight and unimprovable for a single iteration, the worst-case performance could be better for multiple iterations. We can analyze the worst-case behavior of multiple iterations of Newton's method on self-concordant functions with the tool. Figure 10.9 shows the worst-case Newton decrement $\lambda_f(x_N)$ of $N = 2$ iterations of Newton's method on standard-self-concordant functions for varying initial Newton decrement $\lambda_f(x_0)$. We compare PEP results (blue dots), Theorem 10.55 from [Hil21, Equation (11)] (solid red line), and [N+18, Theorem 5.2.2.1] (broken red line). We observe that the performance is actually better than predicted by previous theoretical bounds and also that the region of convergence is larger.

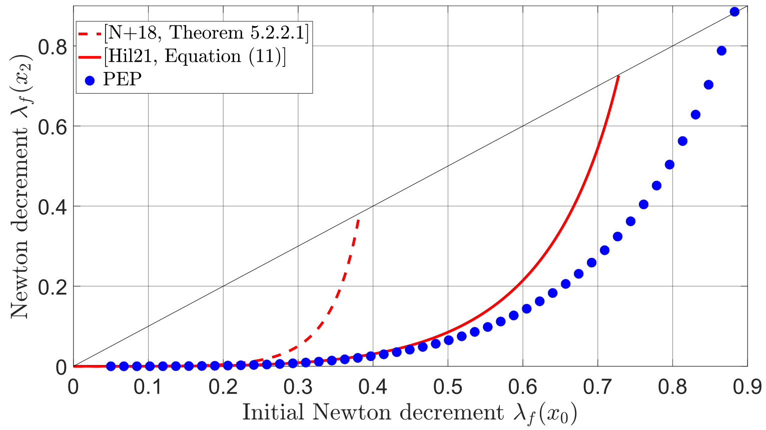


Fig. 10.9 Worst-case performance $\lambda(x_2)$ for varying initial $\lambda(x_0)$ of two iterations of Newton's method (black dots) compared to bounds [N+18, Theorem 5.2.2.1] (dashed red line), [Hil21, Equation (11)] (solid red line), and the PEP results (dots). Solving the non-convex PEP of this figure required around 1 minute.

Damped Newton method We now consider the fixed damped Newton method

$$x_{k+1} = x_k - \gamma_k \frac{f'(x_k)}{f''(x_k)}. \tag{DNM}$$

The following Theorem [Hil21, Equation (11)] provides the tight worst-case performance of a single iteration of damped Newton method on univariate functions for some step sizes γ_k.

Theorem 10.56 ([Hil21], Equation (11)). *A single iteration of damped Newton method* (DNM) *with* $\gamma_k \leq 2\frac{\sqrt{1+\lambda_f(x_k)^3}-1}{\lambda_f(x_k)^3}$ *on standard-self-concordant functions satisfies*

$$\lambda_f(x_{k+1}) \leq \lambda_f(x_k) - \gamma\lambda_f(x_k) + \gamma\lambda_f(x_k)^2. \tag{10.98}$$

Moreover, there exists a standard-self-concordant function attaining exactly this bound.

Note that [Hil21, Equation (11)] covered all values of γ_k and that the case $\gamma_k = \frac{1}{1+\lambda_f(x_k)}$ is always covered by Theorem 10.56. This theorem is tight for a single iteration and, relying on our tool, we can exhibit a function attaining its bound (10.98). Indeed, one can check that a single iteration of (DNM) (with any γ_k) from the point $x_k = 0$ on the following function yields bound (10.98)

$$f(x) = \frac{R-1}{R}x - \log(R-x) \tag{10.99}$$

where $R = \lambda_f(x_k)$.

Moreover, we observed with PEP that bound (10.98) is actually tight even for multiple iterations (unlike Theorem 10.55). Therefore, we propose the following theorem on the rate of convergence of (DNM) on univariate standard-self-concordant functions.

Theorem 10.57. *After N iterations of* (DNM) *with* $\gamma_k \leq 2\frac{\sqrt{1+\lambda_f(x_k)^3}-1}{\lambda_f(x_k)^3}$ *for* $k = 0, \ldots, N-1$ *on standard-self-concordant functions, we have*

$$\lambda_f(x_{k+1}) \leq \lambda_f(x_k) - \gamma\lambda_f(x_k) + \gamma\lambda_f(x_k)^2 \quad \forall k = 0, \ldots, N-1. \tag{10.100}$$

Moreover, there exists a standard-self-concordant function attaining exactly this bound.

Proof. The bound follows directly from Theorem 10.55 and the attaining function is (10.99). □

10.5.5 Quasi-self-concordant functions

The interpolation conditions for the class of quasi-self-concordant functions (Theorem 10.35) involve the exponential function, which makes the associated non-convex PEP challenging to solve. However, Gurobi 11 [Gur24] allows dealing with such non-convex constraints.

Gradient Regularized Newton method The Gradient Regularized Newton method 1 (GNM1) [Doi23] is a variant of the Cubic Newton method that achieves a linear rate of convergence on quasi-self-concordant functions [Doi23, Theorem 3.3] with the following explicit iteration

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k) + M|f'(x_k)|}. \tag{GNM1}$$

GNM1 also exhibits a quadratic local convergence on quasi-self-concordant functions [Doi23, Theorem 4.1]. The proof of this quadratic convergence relies on the following descent lemma (proved in the multivariate case).

Lemma 10.58 ([Doi23], Equation (49))**.** *The iterations of* (GNM1) *on univariate M-quasi-self-concordant functions satisfy*

$$\eta(x_{k+1}) \leq e^{\eta(x_k)}(e^{\eta(x_k)} + \eta(x_k)^2 - \eta(x_k) - 1) \tag{10.101}$$

where $\eta(x) = M\frac{|f'(x)|}{f''(x)}$*.*

Relying on the tool, we observed numerically that this lemma can be improved in the univariate case. And, we can provably improve it by solving analytically the associated PEP.

Lemma 10.59. *The iterations of* (GNM1) *on univariate M-quasi-self-concordant functions satisfy*

$$\eta(x_{k+1}) \leq e^{\frac{\eta(x_k)}{\eta(x_k)+1}}(\eta(x_k) - 1) + 1 \tag{10.102}$$

where $\eta(x) = M\frac{|f'(x)|}{f''(x)}$*.*

Proof. We propose to formulate the associated PEP and to solve it analytically. The PEP can be written as

$$
\begin{aligned}
\max_{S=\{(x_0,g_0,h_0),(x_1,g_1,h_1)\}} \quad & \frac{|g_1|}{h_1} \\
\text{s.t. } & x_1 = x_0 - \frac{g_0}{h_0 + M|g_0|}, \\
& \frac{|g_0|}{h_0} = R, \\
& S \text{ is } M\text{-quasi-self-concordant-interpolable.}
\end{aligned}
\tag{10.103}
$$

We assume $x_0 = 0$ and $g_0 = 1$ w.l.o.g. since a translation can move x_0 to zero and a scaling $f(x) \to cf(x)$ can move g_0 to one. Therefore, the two equality constraints of (10.103) lead to $h_0 = \frac{1}{R}$ and $x_1 = \frac{-R}{MR+1}$. The interpolation conditions are (see Theorem 10.35)

$$g_1 - g_0 \geq \frac{h_0 + h_1}{M} - \frac{2}{M}\sqrt{h_0 h_1 e^{-\frac{M}{2}(x_1 - x_0)}}, \tag{10.104}$$

$$g_0 - g_1 \geq \frac{h_1 + h_0}{M} - \frac{2}{M}\sqrt{h_1 h_0 e^{-\frac{M}{2}(x_0 - x_1)}}. \tag{10.105}$$

Figure 10.1 is an illustration of the feasible domain for (g_1, h_1). One can check that the optimal solution is

$$h_1^* = h_0 e^{-\frac{\eta_0}{\eta_0 + 1}}, \tag{10.106}$$

$$g_1^* = g_0 + \frac{-h_0 + h_1^*}{M}, \tag{10.107}$$

yielding the optimal value

$$M\frac{|g_1^*|}{h_1^*} = e^{\frac{\eta_0}{\eta_0+1}}(\eta_0 - 1) + 1 \tag{10.108}$$

where $\eta_0 = MR = M\frac{|g_0|}{h_0}$. □

Figure 10.10 compares the numerical and analytical solution of PEP (blue dots and curve) with [Doi23, Equation (49)] (red curve) on the worst-case performance of one iteration of (GNM1) on 1-quasi-self-concordant functions.

We can also observe the effect of several iterations of (GNM1) on Figure 10.11.

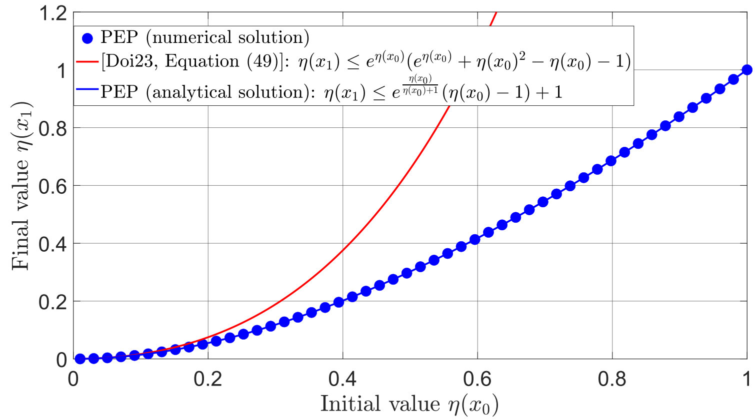


Fig. 10.10 Worst-case performance $\eta(x_1) = M\frac{|f'(x_1)|}{f''(x_k)}$ of one iteration of (GNM1) on 1-quasi-self-concordant functions for different values of $\eta(x_0) = M\frac{|f'(x_0)|}{f''(x_0)}$. We compare the numerical solutions of PEP (blue dots), the known descent lemma of [Doi23, Equation (49)] (red curve), and the analytical solution of PEP (blue curve). Solving the non-convex PEP of this figure required around 1 second.

Newton's method Newton's method also exhibits a local quadratic rate of convergence on quasi-self-concordant and strongly convex functions [Doi23, Theorem 4.1]. The proof of the quadratic convergence relies on the following lemma:

Lemma 10.60 ([Doi23], Equation (53))**.** *A single iteration of Newton's method $x_+ = x - \frac{f'(x)}{f''(x)}$ on M-quasi-self-concordant μ-strongly convex univariate functions satisfy*

$$\frac{M}{\mu}|f'(x_+)| \leq e^{\frac{M}{\mu}|f'(x)|} - \frac{M}{\mu}|f'(x)| - 1. \tag{10.109}$$

Relying on the tool, we can show that this lemma is tight for a single iteration by exhibiting a starting point and a function reaching (10.109).

Lemma 10.61. *Lemma 10.60 (i.e., [Doi23, Equation (53)]) is tight and attained by the starting point $x = 0$ and the function*

$$f(x) = \begin{cases} \frac{\mu}{M^2}e^{-Mx} + \frac{2\mu}{M}x & \text{if } x \leq 0, \\ \frac{\mu}{M^2}e^{Mx} & \text{if } x > 0, \end{cases} \tag{10.110}$$

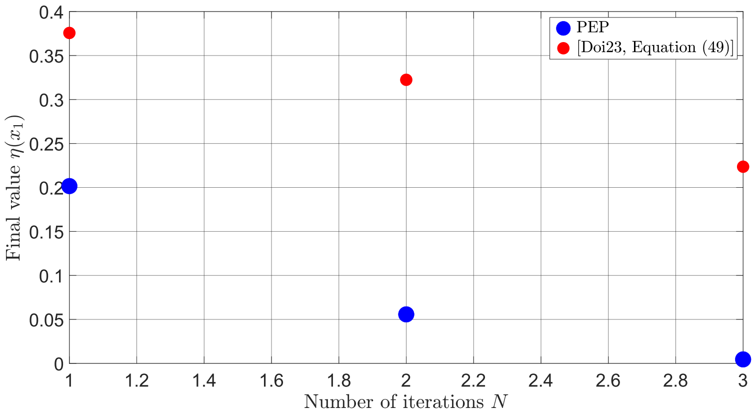


Fig. 10.11 Worst-case performance of GNM1 for varying number of iterations N on 1-quasi-self-concordant functions with $\eta(x_0) = M\frac{|f'(x_0)|}{f''(x_0)} \leq 0.4$. Solving the non-convex PEP of this figure required around 5 seconds.

Proof. We have $f'(0) = \frac{\mu}{M}$, $f''(0) = \mu$, $x_+ = -\frac{1}{M}$, $f'(x_+) = \frac{\mu}{M}(-e+2)$, and

$$\frac{M}{\mu}|f'(x_+)| = e^{\frac{M}{\mu}|f'(0)|} - \frac{M}{\mu}|f'(0)| - 1 = e - 2. \tag{10.111}$$

□

10.5.6 Comparing methods together

As can be seen in Table 10.2 and in the rest of the literature, there is no comparison of the different second-order methods on the same settings, i.e., same function class, performance measure, and initial condition. The tool allows to perform such comparisons on any setting of choice. Figure 10.12 compares the worst-case performance $|f'(x_{k+1})|$ of $N = 1$ iteration of the following methods on μ-strongly convex Hessian M-Lipschitz functions for varying μ and initial gradient $|f'(x_k)| = 1$ and $M = 1$:

- Newton's method: $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$;
- Cubic Regularized Newton method:

$$x_{k+1} = \operatorname{Arg}\min_x f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2}f''(x_k)(x - x_k)^2 + \frac{M}{6}|x - x_k|^3;$$

- Gradient Regularized Newton method 1: $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)+M|f'(x_k)|}$;
- Gradient Regularized Newton method 2: $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)+\sqrt{\frac{M}{2}|f'(x_k)|}}$;
- Adaptive damped Newton method: $x_{k+1} = x_k - \frac{1}{1+M\sqrt{\frac{f'(x_k)^2}{f''(x_k)}}} \frac{f'(x_k)}{f''(x_k)}$.

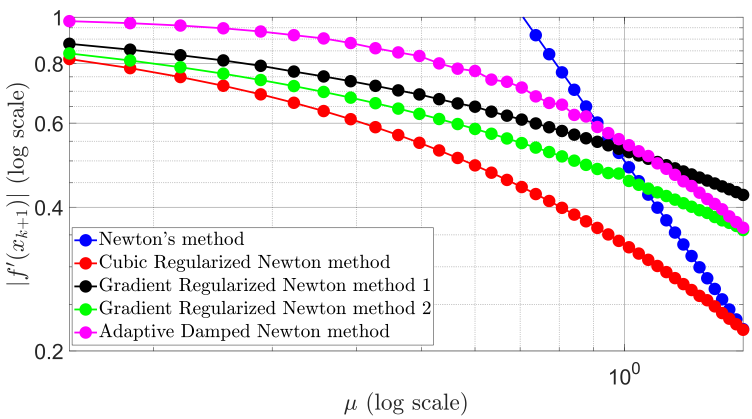


Fig. 10.12 Worst-case performance $|f'(x_{k+1})|$ of different second-order methods on μ-strongly convex Hessian M-Lipschitz functions for varying μ and initial gradient $|f'(x_k)| = 1$ and $M = 1$. Solving the non-convex PEP of this figure required around 15 minutes.

We observe that CNM, GNM1, GNM2, and ADNM share similar behavior with CNM being faster. We also see the local quadratic convergence of NM when μ is sufficiently large. The figure exhibits the contraction factor $\frac{|f'(x_{k+1})|}{|f'(x_k)|}$ after a single iteration from which we can deduce an upper bound on the decrease of the gradient after N iterations.

10.6 Summary

We took a step forward into tightly analyzing the performance of second-order methods, with respect to several performance measures and on several function classes. Indeed, we were able to obtain exact performance guarantees in the *univariate* setting, by (i) providing an exact discrete representation of many univariate function classes of interest, and (ii) proposing a tractable second-order PEP formulation by leveraging advances in solving non-convex PEPs.

This allowed us to (i) get an insight into second-order interpolation conditions, which are already non-trivial when restricted to the univariate case, (ii) prove tightness of existing multivariate bounds by exhibiting worst-case instances, (iii) improve existing convergence rates, either provably or numerically, and (iv) compare methods on a fair basis, since we are not restricted to the analysis of a given criterion adapted to each method.

B

Appendix

B.1 Examples of worst-case instances of smallest dimension

A surprising observation made in [DT14, THG17c] is that a lot of worst-case performance is actually attained by simple univariate functions. This observation is the basis of the development of optimized methods. However, this observation does not hold in general not even for fixed-step first-order methods. We propose a partial list of examples of worst-case performances attained by univariate functions and a list of examples of worst-case performances that cannot be attained by any univariate function. In general, to check whether a PEP admits a univariate worst-case, one could compare the value of the PEP with the value of the worst-case performance among univariate functions computed with non-convex PEP.

Worst-case attained by univariate functions

- Gradient, fast gradient, and optimized gradients methods on smooth convex functions when the performance criterion is the value accuracy of the last iteration $f(x_N) - f^*$ [THG17c, section 4];
- Gradient method on smooth strongly convex functions when the performance criterion is the value accuracy of the last iteration $f(x_N) - f^*$ [THG17c, Section 4.1];

- Fast gradient method on smooth strongly convex functions with a small number of iterations when the performance criterion is the value accuracy of the last iteration $f(x_N) - f^*$ (Section B.2);
- Gradient method on smooth strongly convex functions of the form $F(x) = g(Mx)$ when the performance criterion is the value accuracy of the last iteration $f(x_N) - f^*$ [BHG24];
- Gradient and Newton Method on Hessian Lipschitz functions when the performance criterion is the distance to optimality of the last iteration $\|x_N - x^*\|$ [RBHG25].

Worst-case not attained by univariate functions

- Fast gradient method on smooth strongly convex functions with a large number of iterations when the performance criterion is the value accuracy of the last iteration $f(x_N) - f^*$ (Section B.2);
- Gradient method on smooth hypoconvex functions when the performance criterion is the gradient residual of the last iteration $\|\nabla f(x_N)\|$ [RGP24];
- Gradient method with exact line search on smooth strongly convex functions when the performance criterion is the value accuracy of the last iteration $f(x_N) - f^*$ [dKGT17];
- A single iteration of Newton's method on self-concordant functions when the performance criterion is the Newton decrement [Hil21].

B.2 Worst-case performance of fast gradient method on smooth strongly convex functions

In this section, we consider the fast gradient method (FGM) [Nes83]

$$\begin{cases} x_{k+1} &= y_k - \frac{1}{L}\nabla f(y_k), \\ \theta_{k+1} &= \frac{1+\sqrt{4\theta_k^2+1}}{2}, \\ y_{k+1} &= x_{k+1} + \frac{\theta_k - 1}{\theta_{k+1}}(x_{k+1} - x_k), \end{cases} \tag{FGM}$$

and observe that after a certain number of iterations, there is no univariate function that attains the worst-case performance of FGM on smooth strongly convex functions. We consider the performance criterion $f(x_N) - f(x^*)$ and the initial condition $\|x_0 - x^*\| \leq R$.

Convex case The convex case was covered in [THG17c, Section 4.2]. Let us define the following family of (possibly strongly convex) Huber loss

$$f_{\mu,\tau}(x) = \begin{cases} \frac{\mu}{2}x^2 + a_\tau|x| + b_\tau, & \text{if } |x| \geq \tau, \\ \frac{L}{2}x^2, & \text{otherwise,} \end{cases} \tag{B.1}$$

where $a_\tau = (L-\mu)\tau$, $b_\tau = -\frac{L-\mu}{2}\tau^2$. It has been numerically observed [THG17c, Conjecture 5] that the worst-case performance of FGM on L-smooth convex functions is attained by f_{0,τ_0} with

$$\tau_0 = \frac{R}{2\sum_{k=0}^{N-1} h_{N,k} + 1}, \tag{B.2}$$

and the following convergence rate

$$f(x_N) - f^* = \frac{LR^2}{2} \frac{1}{2\sum_{k=0}^{N-1} h_{N,k} + 1} = \frac{LR}{2}\tau_0. \tag{B.3}$$

Remark B.1. Once the form of function $f_{0,\tau}$ is fixed, we can compute τ_0 with

$$\tau_0 = \arg\max_\tau f_{0,\tau}(x_N(\tau)) - f_{0,\tau}(x^*), \tag{B.4}$$

where $f_{0,\tau}(x^*) = 0$.

Strongly convex case Similarly, we can compute the worst τ in the strongly convex case $\mu > 0$, and we obtain

$$\tau_\mu = \frac{-R\big(1 + \mu K_N(\mu)\big)^2}{(L-\mu)K_N(\mu)\big(\mu K_N(\mu) + 2\big) - 1}, \tag{B.5}$$

where $K_N(\rho)$ is a polynomial of degree N that captures the whole information about the behavior of a given method on all quadratic $\rho\frac{x^2}{2}$. It has the following expression

$$K_N(\rho) = \frac{1}{\rho}\sum_{l=1}^{N}(-\rho)^l \sum_{\substack{p\in\mathbb{N}^{l+1} \\ N=p_0>\ldots>p_l\geq 0}} \prod_{k=0}^{l-1} h_{p_k,p_{k+1}} \tag{B.6}$$

or

$$K_N(\rho) = -\sum_{k=0}^{N-1}\left(h_{N,k} + \rho h_{N,k}K_k(\rho)\right) \text{ with } K_0(\rho) = 0 \tag{B.7}$$

where the vectors p in (B.6) can be seen as all the strictly decreasing natural sequences of length $l+1$ from N to a strictly lower positive number.

The performance of FGM on the function f_{μ,τ_μ} is

$$f(x_N) - f(x^*) = \frac{LR}{2}\tau_\mu. \tag{B.8}$$

However, unlike the convex case, it does not always match the worst-case performance of FGM on L-smooth μ-strongly convex functions computed numerically with PEP. Figure B.1 compares the worst-case performance of FGM on L-smooth μ-strongly convex functions computed with PEP with the performance of FGM on f_{μ,τ_μ} for different number of iterations N.

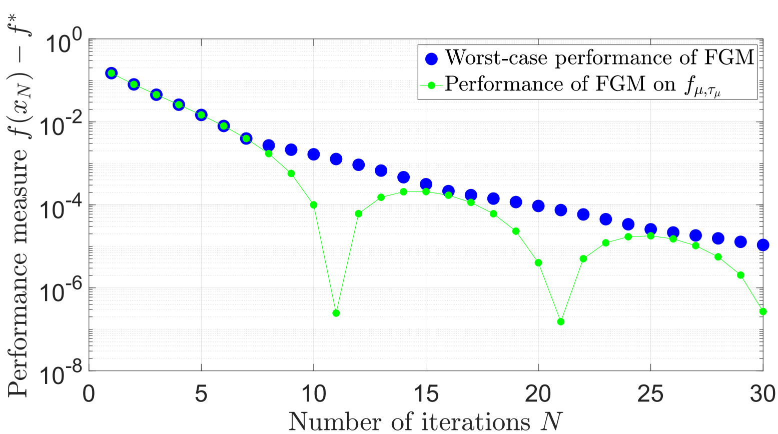


Fig. B.1 Worst-case performance of FGM on L-smooth μ-strongly convex functions computed with PEP (blue dots) and performance of FGM on f_{μ,τ_μ} (green dots) for different number of iterations N, $L = R = 1$, and $\mu = 0.1$.

We observe that the performance on f_{μ,τ_μ} matches the worst-case performance computed by PEP for $N = 1, \dots, 7$. For $N \geq 8$, there is a gap between the two values, meaning that f_{μ,τ_μ} is no longer a worst-case function.

Given Figure B.1, there could still exist a univariate function attaining the worst-case performance. However, using the non-convex PEP methodology allows to conclude that there are no such univariate functions and that the worst-case functions are multivariate. We were not able to identify them.

B.3 Proof of Theorems 10.26 and 10.27

We exploit Theorem 10.18 to obtain interpolation conditions with function values of the classes $\int \mathcal{F}_{M,\alpha,(+)}$, which leads to interpolation conditions without function values of the classes $\int^{(2)} \mathcal{F}_{M,\alpha,(+)}$ by Lemma 10.4. First, we show that these classes satisfy the assumptions required by Theorem 10.18 in Section B.3.1, and secondly, we prove Theorems 10.26 and 10.27 in Section B.3.2.

B.3.1 Assumptions of Theorem 10.18

Proposition B.2. *Let $\alpha, M \geq 0$. Then, $\mathcal{F}_{M,\alpha,(+)}$, $\int \mathcal{F}_{M,\alpha,(+)}$ and $\int^{(2)} \mathcal{F}_{M,\alpha,(+)}$ are respectively order* 0, 1 *or* 2 *connectable (Assumption 10.14). In addition, when $\alpha \leq 1$, $\int^{(2)} \mathcal{F}_{M,\alpha,(+)}$ is also convex (Assumption 10.6).*

Proof of Proposition B.2. (Connectability) Definition 10.20, involving derivatives, is pointwise and holds everywhere except for a finite set of points. Hence, by definition, $\mathcal{F}_{M,\alpha,(+)}$ is order 0-connectable, and the function f, juxtaposition of several functions in $\mathcal{F}_{M,\alpha,(+)}$ might be non-differentiable only at the junctions between intervals. The same argument holds for $\int \mathcal{F}_{M,\alpha,(+)}$ and $\int^{(2)} \mathcal{F}_{M,\alpha,(+)}$ (and $\int^{(k)} \mathcal{F}_{M,\alpha,(+)}$), except that it involves higher-order derivatives.

(Convexity) When $\alpha \leq 1$, $\int \mathcal{F}_{M,\alpha,(+)}$ is convex since, given $f_1, f_2 \in \int \mathcal{F}_{M,\alpha,(+)}$, $\lambda \in [0,1]$, and $f = \lambda f_1 + (1-\lambda) f_2$ it holds:

$$\begin{aligned} |f''(x)| &= |\lambda f_1''(x) + (1-\lambda) f_2''(x)| \\ &\leq \lambda |f_1''(x)| + (1-\lambda)|f_2''(x)| \\ &\leq M_1(\lambda f_1'(x)^\alpha + (1-\lambda) f_2'(x)^\alpha) \\ &\leq M_1(\lambda f_1'(x) + (1-\lambda) f_2'(x))^\alpha \end{aligned} \tag{B.9}$$

by concavity of x^α, $x \geq 0$, $\alpha \leq 1$. Moreover, $\forall x \in \mathbb{R}$, if $f_1(x), f_2(x) \geq 0$

then $f(x) \geq 0$. The same arguments hold for $\int^{(2)} \mathcal{F}_{M,\alpha,(+)}$ (with f''' and f'' instead of f'' and f' in (B.9)). $\square$

The other properties involved in Theorem 10.18, i.e., extremal interpolability and, when $\alpha > 1$, extremal connectability of $\mathcal{F}_{M,\alpha,(+)}$ and $\int \mathcal{F}_{M,\alpha,(+)}$ (Assumptions 10.9 and 10.11), require having access to interpolation conditions for both classes.

We now prove $\mathcal{F}_{M,\alpha,(+)}$ is extremally interpolable and provide its extremal interpolants:

Proposition B.3. *Let $\alpha, M \geq 0$.*

It holds that $\mathcal{F}_{M,\alpha,(+)}$ is extremally interpolable, and the extremal interpolants (Definition 10.8) of an $\mathcal{F}_{M,\alpha,(+)}$-interpolable set $S = \{(x_i, f_i)\}_{i=1,2}$, where $x_1 < x_2$, are given by:

- $[f_{\min}$, *Case 1*$]$ *If $\alpha \geq 1$, if we consider $\mathcal{F}_{M,0}$, or if $\alpha < 1$ and $\tilde{f}_2 + \tilde{f}_1 \geq M(x_2 - x_1)$:*

$$f_{\min}(x) = \begin{cases} \nu\left(\tilde{f}_1 - \frac{\beta(\alpha)}{|\beta(\alpha)|} M(x - x_1)\right) & x \in [x_1, z] \\ \nu\left(\tilde{f}_2 + \frac{\beta(\alpha)}{|\beta(\alpha)|} M(x - x_2)\right) & x \in [z, x_2] \end{cases}, \text{ where } z = \frac{x_1 + x_2}{2} + \frac{\beta(\alpha)}{|\beta(\alpha)|} \frac{\tilde{f}_1 - \tilde{f}_2}{2M}.$$

- $[f_{\min}$, *Case 2*$]$ *Else:*

$$f_{\min}(x) = \begin{cases} \nu\left(\tilde{f}_1 - \frac{\beta(\alpha)}{|\beta(\alpha)|} M(x - x_1)\right) & x \in [x_1, x_1 + \frac{\tilde{f}_1}{M}] \\ 0 & x \in [x_1 + \frac{\tilde{f}_1}{M}, x_2 - \frac{\tilde{f}_2}{M}] \\ \nu\left(\tilde{f}_2 + \frac{\beta(\alpha)}{|\beta(\alpha)|} M(x - x_2)\right) & x \in [x_2 - \frac{\tilde{f}_2}{M}, x_2] \end{cases}.$$

- $[f_{\max}$, *Case 1*$]$ *If $\alpha \leq 1$, or if $\alpha > 1$ and $\tilde{f}_2 + \tilde{f}_1 \geq M(x_2 - x_1)$:*

$$f_{\max}(x) = \begin{cases} \nu\left(\tilde{f}_1 + \frac{\beta(\alpha)}{|\beta(\alpha)|} M(x - x_1)\right) & x \in [x_1, y] \\ \nu\left(\tilde{f}_2 - \frac{\beta(\alpha)}{|\beta(\alpha)|} M(x - x_2)\right) & x \in [y, x_2] \end{cases}, \text{ where } y = \frac{x_1 + x_2}{2} - \frac{\beta(\alpha)}{|\beta(\alpha)|} \frac{\tilde{f}_1 - \tilde{f}_2}{2M}.$$

- $[f_{\max}$, *Case 2*$]$ *Else:*

$$f_{\max}(x) = \begin{cases} \nu\left(\tilde{f}_1 + \frac{\beta(\alpha)}{|\beta(\alpha)|} M(x - x_1)\right) & x \in [x_1, x_1 + \frac{\tilde{f}_1}{M}] \\ \infty & x \in [x_1 + \frac{\tilde{f}_1}{M}, x_2 - \frac{\tilde{f}_2}{M}] \\ \nu\left(\tilde{f}_2 - \frac{\beta(\alpha)}{|\beta(\alpha)|} M(x - x_2)\right) & x \in [x_2 - \frac{\tilde{f}_2}{M}, x_2] \end{cases},$$

where

$$\tilde{f}_i = \begin{cases} f_i^{1-\alpha} & \text{if } \alpha \neq 1, \\ \log(f_i) & \text{if } \alpha = 1, \end{cases} \tag{B.10}$$

$$\nu(x) = \begin{cases} x^{\beta(\alpha)} & \text{if } \alpha \neq 1, \\ e^x & \text{if } \alpha = 1. \end{cases} \tag{B.11}$$

Proof. We first show that $f_{\min}, f_{\max}$ are the extremal interpolating envelopes of S as defined in (10.16). It holds that $f_{\min}, f_{\max}$ interpolate S, since $\nu(\tilde{f}_i) = f_i$, $i = 1, 2$. For any $x \in [x_1, x_2]$, they are also the extremal f satisfying (10.35) with respect to S, i.e.,

$$f_{\min}(x) = \min_{f \geq 0} f \text{ s.t. } |\tilde{f} - \tilde{f}_i| \leq M|x - x_i|,\ i = 1, 2, \tag{B.12}$$

$$f_{\max}(x) = \max_{f \geq 0} f \text{ s.t. } |\tilde{f} - \tilde{f}_i| \leq M|x - x_i|,\ i = 1, 2, \tag{B.13}$$

where we remove the constraint $f \geq 0$ for $\mathcal{F}_{M,0}$. For instance, at all $x \in [x_1, x_2]$, $f_{\min}(x)$ takes the maximal value between 0 and the lower boundaries $\tilde{f} \geq \tilde{f}_i - M|x - x_i|$. When $\alpha \geq 1$ (hence $\frac{\beta(\alpha)}{|\beta(\alpha)|} < 0$), or if $\alpha < 1$ and $\tilde{f}_1 + \tilde{f}_2 \geq M(x_2 - x_1)$, these boundaries are always larger than 0, and $f_{\min}$ consists of two pieces. Else, the constraint $f \geq 0$ becomes tight, and $f_{\min}$ consists of three parts.

On the contrary, at all $x \in [x_1, x_2]$, $f_{\min}(x)$ takes the minimal value of the upper boundaries $\tilde{f} \leq \tilde{f}_i + M|x - x_i|$. When $\alpha \leq 1$, or if $\alpha > 1$ and $\tilde{f}_1 + \tilde{f}_2 \geq M(x_2 - x_1)$, these boundaries cross each other before meeting their respective asymptote, and $f_{\max}$ consists of two pieces. Else, $f_{\max}$ reaches ∞ on a third interval.

To prove extremal interpolability of $\mathcal{F}_{M,\alpha,(+)}$, it remains to show $f_{\min}, f_{\max} \in \mathcal{F}_{M,\alpha,(+)}$. This holds since $\mathcal{F}_{M,\alpha,(+)}$ is order 0-connectable and $f_{\min}, f_{\max}$ are functions by parts satisfying on each interval:

$$|\nu\left(\tilde{f}_i \pm M(x - x_i)\right)'| = M|\beta(\alpha)|\nu\left(\tilde{f}_i \pm M(x - x_i)\right)^{\alpha} \quad \forall i = 1, 2, \tag{B.14}$$

since $f_{\min}$ and $f_{\max}$ are larger than 0 except for the extremal interpolants of $\mathcal{F}_{M,0}$ and since both belong to $\bar{\mathcal{C}}^0$ (or $\mathcal{C}^0$ if $\alpha \leq 1$), and are piecewise $\bar{\mathcal{C}}^1$ (or $\mathcal{C}^1$ when $\alpha \leq 1$). □

Remark B.4. The cases $\mathcal{F}_{M,0}$ and $\mathcal{F}_{M,1,\alpha}$ are the only ones for which the extremal interpolants take a single expression.

Remark B.5. Since S is $\mathcal{F}_{M,\alpha,(+)}$-interpolable and satisfies (10.35), all quantities in Proposition B.3 are well-defined, i.e., $x_1 \leq z \leq x_2$ and $x_1 \leq y \leq x_2$.

We build on these extremal interpolants to prove $\mathcal{F}_{M,\alpha,(+)}$ is extremally connectable when $\alpha > 1$ (the case $\alpha \leq 1$ follows from Proposition B.2).

Proposition B.6. *Let $M \geq 0$ and $\alpha > 1$. It holds that $\int \mathcal{F}_{M,\alpha,(+)}$ is extremally connectable.*

Proof. Let $S = \{(x_i, f_i^0, f_i^1)\}_{i=1,2}$ and $g_{\min / \max}$ be defined as in (10.16), interpolating $\tilde{S} = \{(x_i, f_i^1\}_{i=1,2}$. Let $a \geq 0$ and define the following function

$$g(x,a) = \min\{g_{\max}(x), \max\{g_{\min}(x), a\}\}. \tag{B.15}$$

By construction, g is defined as the juxtaposition of functions in $\mathcal{F}_{M,\alpha,+}$ on at most three intervals: either a constant function, either $g_{\min}$ or $g_{\max}$. Hence, since $\mathcal{F}_{M,\alpha,+}$ is order 0-connectable, $g \in \mathcal{F}_{M,\alpha,+}$. In addition, g interpolates $\tilde{S}$ since $g_{\max}(x_i) = g_{\min}(x_i) = f_i^1$, $\forall i = 1,2$. Let

$$f(x,a) := f_1^0 - \int_{-\infty}^{x_1} g(z,a)\mathrm{d}z + \int_{-\infty}^{x} g(z,a)\mathrm{d}z. \tag{B.16}$$

By Lemma 10.5, $f \in \int \mathcal{F}_{M,\alpha,+}$, $f'(x_i,a) = f_i^1$, $\forall i = 1,2$, and $f(x_1,a) = f_1^0$. In addition,

$$f(x_2,a) = f_1^0 + \int_{x_1}^{x_2} g(z,a)\mathrm{d}z. \tag{B.17}$$

Since $f_1^0 + \int_{x_1}^{x_2} g_{\min}(z)\mathrm{d}z \leq f_2^0 \leq f_1^0 + \int_{x_1}^{x_2} g_{\max}(z)\mathrm{d}z$ and $\int_{x_1}^{x_2} g(z,a)\mathrm{d}z$ depends continuously on a, with a minimum in $\int_{x_1}^{x_2} g_{\min}(z)\mathrm{d}z$ and a maximum in $\int_{x_1}^{x_2} g_{\max}(z)\mathrm{d}z$, there exists some $a \geq 0$ for which $f(x_2,a) = f_2^0$, hence $\int \mathcal{F}_{M,\alpha,+}$ is extremally completable. □

B.3.2 Proof of Theorems 10.26 and 10.27

Proof of Theorem 10.26. $\mathcal{F}_{M,1,(+)}$ and $\int \mathcal{F}_{M,1,(+)}$ satisfy all assumptions in Theorem 10.18. Hence, S is interpolable if and only if it satisfies

$$|\log(g_i) - \log(g_j)| \leq M|x_i - x_j|, \tag{B.18}$$

and $\forall x_i < x_j$:

$$\begin{cases} f_j - f_i \geq \frac{g_i+g_j}{M} - \frac{2}{M}\sqrt{g_i g_j} e^{-\frac{M}{2}(x_j-x_i)} \\ f_j - f_i \leq -\frac{g_i+g_j}{M} + \frac{2}{M}\sqrt{g_i g_j} e^{\frac{M}{2}(x_j-x_i)} \end{cases} \tag{B.19}$$

$$\Leftrightarrow f_i - f_j \leq \frac{g_i + g_j}{M} - \frac{2}{M}\sqrt{g_i g_j} e^{-\frac{M}{2}(x_i-x_j)}. \tag{B.20}$$

Imposing (10.40) on all pairs is strictly equivalent to this condition.

In addition, (10.40) implies (B.18) when applied to all pairs. Indeed, summing (10.40) imposed on (i,j) and (j,i) yields

$$\frac{g_i + g_j}{\sqrt{g_i g_j}} \leq e^{-\frac{M}{2}(x_j-x_i)} + e^{\frac{M}{2}(x_j-x_i)} \tag{B.21}$$

and using $\frac{1}{t} + t \leq \frac{1}{\sqrt{s}} + \sqrt{s} \Leftrightarrow \frac{1}{s} \leq t^2 \leq s$ when $s \geq 1$, $t \geq 0$ with $t = \sqrt{\frac{g_j}{g_i}}$ and $s = e^{M(x_j-x_i)}$ yields

$$e^{-M(x_j-x_i)} \leq \frac{g_j}{g_i} \leq e^{M(x_j-x_i)}, \tag{B.22}$$

which is equivalent to (B.18). □

Proof of Theorem 10.27. $\mathcal{F}_{M,\alpha,+}$ and $\int \mathcal{F}_{M,\alpha,+}$ satisfy all assumptions in Theorem 10.18. Hence, S is interpolable if and only if it satisfies (10.41), and $\forall x_i < x_j$:

- Integration of $f_{\min}(x)$, Case 1:

If $\alpha > 1$, or if $\alpha < 1$ and $\tilde{g}_i + \tilde{g}_j \geq M(x_j - x_i) = \frac{\beta(\alpha)}{|\beta(\alpha)|}M(x_j - x_i)$, or considering $\mathcal{F}_{M,0}$:

$$f_j - f_i \geq \frac{\beta(\alpha)}{|\beta(\alpha)|M(\beta(\alpha)+1)}\left(\tilde{g}_i^{\beta(\alpha)+1} + \tilde{g}_j^{\beta(\alpha)+1} - \frac{1}{2^{\beta(\alpha)}}(\tilde{g}_i + \tilde{g}_j - \frac{\beta(\alpha)}{|\beta(\alpha)|}M(x_j - x_i))^{\beta(\alpha)+1}\right)$$

- Integration of $f_{\min}(x)$, Case 2: If $\alpha < 1$ and $\tilde{g}_i + \tilde{g}_j \leq M(x_j - x_i)$:

$$f_j - f_i \geq \frac{1}{M(\beta(\alpha)+1)}\left(\tilde{g}_i^{\beta(\alpha)+1} + \tilde{g}_j^{\beta(\alpha)+1}\right)$$

- Integration of $f_{\max}(x)$, Case 1:

If $\alpha < 1$, or if $\alpha > 1$ and $\tilde{g}_i + \tilde{g}_j \geq M(x_j - x_i) = \frac{\beta(\alpha)}{|\beta(\alpha)|}M(x_i - x_j)$:

$$f_j - f_i \leq \frac{\beta(\alpha)}{|\beta(\alpha)|M(\beta(\alpha)+1)}\left(-\tilde{g}_i^{\beta(\alpha)+1} - \tilde{g}_j^{\beta(\alpha)+1} + \frac{1}{2^{\beta(\alpha)}}(\tilde{g}_i + \tilde{g}_j + \frac{\beta(\alpha)}{|\beta(\alpha)|}M(x_j - x_i))^{\beta(\alpha)+1}\right)$$

$$\Leftrightarrow f_i - f_j \geq \frac{\beta(\alpha)}{|\beta(\alpha)|M(\beta(\alpha)+1)}\left(\tilde{g}_i^{\beta(\alpha)+1} + \tilde{g}_j^{\beta(\alpha)+1} - \frac{1}{2^{\beta(\alpha)}}(\tilde{g}_i + \tilde{g}_j - \frac{\beta(\alpha)}{|\beta(\alpha)|}M(x_i - x_j))^{\beta(\alpha)+1}\right).$$

The second case for $f_{\max}$ leads to no further constraint. Considering $\mathcal{F}_{M,\alpha,+}$, imposing (10.43) and (10.44) on all pairs is strictly equivalent to these conditions, since $\frac{\beta(\alpha)}{|\beta(\alpha)|} = -1$ if $\alpha > 1$, $\frac{\beta(\alpha)}{|\beta(\alpha)|} = 1$ otherwise.

Considering $\mathcal{F}_{M,0}$, there are no subcases to consider, and imposing (10.43) on all pairs is strictly equivalent to these conditions. Furthermore, satisfaction of (10.43) on both pairs (i,j) and (j,i) implies satisfaction of $|g_i - g_j| \leq M|x_i - x_j|$, since adding both inequalities gives:

$$0 \geq g_i^2 + g_j^2 - \frac{1}{4}(g_i + g_j - M(x_j - x_i))^2 - \frac{1}{4}(g_i + g_j + M(x_j - x_i))^2$$

$$\Leftrightarrow 0 \geq \frac{(g_i - g_j)^2}{2} - \frac{M^2}{2}(x_j - x_i)^2 \Leftrightarrow |g_i - g_j| \leq M|x_i - x_j|.$$

□

B.4 Proof of Theorem 10.38

By Proposition B.2, $\mathcal{H}_M$ is convex and order 2 connectable. We will also see in Lemma B.7 that $\mathcal{G}_M = \int \mathcal{F}_M$ is extremally interpolable, hence assumptions in Theorem 10.18 are satisfied.

For the sake of readability, we shorten the notation of the pair $\big((x_i, g_i, h_i, f_i), (x_j, g_j, h_j, f_j)\big)$ to (i,j) and introduce:

$$\Delta x_{ij} = x_j - x_i, \tag{B.23}$$

$$\Delta x_{xi} = x - x_i, \tag{B.24}$$

$$\Delta h_{ij} = h_j - h_i, \tag{B.25}$$

$$T_{ij}^g = g_j - g_i - h_i(x_j - x_i), \tag{B.26}$$

$$T_{ij}^f = f_j - f_i - g_i(x_j - x_i) - \frac{h_i}{2}(x_j - x_i)^2. \tag{B.27}$$

We will use the following identities:

$$\Delta x_{ij} = -\Delta x_{ji}, \tag{B.28}$$

$$\Delta h_{ij} = -\Delta h_{ji}, \tag{B.29}$$

$$T_{ij}^g = -T_{ji}^g + \Delta h_{ij} \Delta x_{ij}, \tag{B.30}$$

$$T_{ij}^f = -T_{ji}^f + T_{ij}^g \Delta x_{ij} - \frac{1}{2}\Delta h_{ij} \Delta x_{ij}^2. \tag{B.31}$$

Interpolation conditions for $\mathcal{F}_M$ are given in Proposition 10.30. Building on these conditions, we obtain their extremal interpolants:

Lemma B.7 (Minimal and maximal gradient functions of $\mathcal{H}_M$)**.** *Consider an interval $[x_1, x_2]$ and a $\mathcal{H}_M$-interpolable pair of points $S = \{(x_i, f_i, g_i, h_i)\}_{i=1,2}$. Then, in $[x_1, x_2]$, the expressions of $g_{\min}$ and $g_{\max}$ as defined in* (10.16) *are given by:*
When $\Delta h_{12} \neq \pm M\Delta x_{12}$,

$$g_{\min}(x) = \begin{cases} g_1 + h_1\Delta x_{x1} - \frac{M}{2}\Delta x_{x1}^2 & \text{if } x \in [x_1, y_1] \\ g_1 + h_1\Delta x_{x1} - \frac{M}{2}(y_1 - x_1)^2 & \\ \quad + \frac{M}{2}(x^2 - y_1^2) + M(x_1 - 2y_1)(x - y_1) & \text{if } x \in [y_1, y_2] \\ g_2 + h_2\Delta x_{x2} - \frac{M}{2}\Delta x_{x2}^2 & \text{if } x \in [y_2, x_2], \end{cases} \tag{B.32}$$

where

$$y_1 = x_2 - \frac{T_{12}^g + \frac{M}{2}\Delta x_{12}^2}{\Delta h_{12} + M\Delta x_{12}} - \frac{\Delta h_{12} + M\Delta x_{12}}{4M}, \tag{B.33}$$

$$y_2 = x_2 - \frac{T_{12}^g + \frac{M}{2}\Delta x_{12}^2}{\Delta h_{12} + M\Delta x_{12}} + \frac{\Delta h_{12} + M\Delta x_{12}}{4M}, \tag{B.34}$$

and

$$g_{\max}(x) = \begin{cases} g_1 + h_1\Delta x_{x1} + \frac{M}{2}\Delta x_{x1}^2 & \text{if } x \in [x_1, z_1] \\ g_1 + h_1\Delta x_{x1} + \frac{M}{2}(z_1 - x_1)^2 & \\ \quad - \frac{M}{2}(x^2 - z_1^2) - M(x_1 - 2z_1)(x - z_1) & \text{if } x \in [z_1, z_2] \\ g_2 + h_2\Delta x_{x2} + \frac{M}{2}\Delta x_{x2}^2 & \text{if } x \in [z_2, x_2], \end{cases} \tag{B.35}$$

where

$$z_1 = x_2 + \frac{T_{12}^g - \frac{M}{2}\Delta x_{12}^2}{-\Delta h_{12} + M\Delta x_{12}} - \frac{(-\Delta h_{12} + M\Delta x_{12})}{4M}, \tag{B.36}$$

$$z_2 = x_2 + \frac{T_{12}^g - \frac{M}{2}\Delta x_{12}^2}{-\Delta h_{12} + M\Delta x_{12}} + \frac{(-\Delta h_{12} + M\Delta x_{12})}{4M}. \tag{B.37}$$

When $\Delta h_{12} = \pm M \Delta x_{12}$:

$$g_{\max}(x) = g_{\min}(x) = g_1 + h_1 \Delta x_{x1} \pm \frac{M}{2} \Delta x_{x1}^2 = g_2 + h_2 \Delta x_{x2} \pm \frac{M}{2} \Delta x_{x2}^2. \tag{B.38}$$

Proof. Suppose $\Delta h_{12} \neq \pm \Delta x_{12}$. For any $x \in [x_1, x_2]$, let

$$g'_{\min}(x) = \begin{cases} h_1 - M\Delta x_{x1} & \text{if } x \in [x_1, y_1] \\ h_1 + M(x + x_1 - 2y_1) & \text{if } x \in [y_1, y_2] \\ h_2 - M\Delta x_{x2} & \text{if } x \in [y_2, x_2], \end{cases} \tag{B.39}$$

It holds that $g'_{\min}(x)$ is the first derivative of $g_{\min}(x)$ since $\forall x \in [x_1, x_2]$, $g_{\min}(x) = \int_{x_1}^{x} g'_{\min}(t)\mathrm{d}t$. If $x \in [x_1, y_2]$, this is straightforward, and if $x \in [y_2, x_2]$ it follows from:

$$\begin{aligned}
\int_{x_1}^{x} g'_{\min}(t)\mathrm{d}t &= g_1 + h_1(y_2 - x_1) - \frac{M}{2}(y_1 - x_1)^2 + \frac{M}{2}(y_2^2 - y_1^2) \\
&\quad + M(x_1 - 2y_1)(y_2 - y_1) + h_2(x - y_2) - \frac{M}{2}(x^2 - y_2^2) + Mx_2(x - y_2) \\
&= g_2 + h_2\Delta x_{x2} - \frac{M}{2}\Delta x_{x2}^2 - T_{12}^g - \Delta h_{12}(y_2 - x_2) \\
&\quad - M\left((y_2 - x_2)\Delta x_{12} + \frac{\Delta x_{12}^2}{2} - (y_2 - y_1)^2\right) \\
&= g_2 + h_2\Delta x_{x2} - \frac{M}{2}\Delta x_{x2}^2 - (T_{12}^g + M\frac{\Delta x_{12}^2}{2}) \\
&\quad - (\Delta h_{12} + M\Delta x_{12})(y_2 - x_2) + \frac{\Delta h_{12} + M\Delta x_{12}}{4M} \\
&= g_2 + h_2\Delta x_{x2} - \frac{M}{2}\Delta x_{x2}^2.
\end{aligned}$$

We first observe that $g_{\min}(x)$ interpolates S since

$$g_{\min}(x_1) = g_1,\ g_{\min}(x_2) = g_2,\ g'_{\min}(x_1) = h_1 \text{ and } g'_{\min}(x_2) = g_2.$$

We now show that for any $x \in [x_1, x_2]$, $g_{\min}(x)$ and $g'_{\min}(x)$ are the minimal g and associated h satisfying (10.45) with respect to S, which completes the proof. To this end, consider the following optimization problem: $g_{\min}(x) =$

$$\begin{cases} \min_{g,h} & g \\ \text{s.t.} & g_1 - g \geq \frac{1}{4M}(h-h_1)^2 - \frac{M}{4}\Delta x_{x1}^2 - \frac{1}{2}(h+h_1)\Delta x_{x1} \qquad (\mu_0) \\ & g_2 - g \geq \frac{1}{4M}(h-h_2)^2 - \frac{M}{4}\Delta x_{x2}^2 - \frac{1}{2}(h+h_2)\Delta x_{x2} \qquad (\mu_1) \\ & g - g_2 \geq \frac{1}{4M}(h_2-h)^2 - \frac{M}{4}\Delta x_{x2}^2 + \frac{1}{2}(h_2+h)\Delta x_{x2} \qquad (\lambda_0) \\ & g - g_1 \geq \frac{1}{4M}(h_1-h)^2 - \frac{M}{4}\Delta x_{x1}^2 + \frac{1}{2}(h_1+h)\Delta x_{x1} \qquad (\lambda_1), \end{cases} \tag{B.40}$$

where μ_0, μ_1, λ_0 and λ_1 are the associated dual variables. We consider its KKT conditions:

Stat. cond. in f:
$$1 + \mu_0 + \mu_1 - \lambda_0 - \lambda_1 = 0 \tag{B.41}$$

Stat. cond. in g:
$$\begin{aligned} &\mu_0\left(\frac{1}{2M}(h-h_1) - \frac{1}{2}\Delta x_{x1}\right) + \mu_1\left(\frac{1}{2M}(h-h_2) - \frac{1}{2}\Delta x_{x2}\right) \\ &+ \lambda_0\left(\frac{1}{2M}(h-h_2) + \frac{1}{2}\Delta x_{x2}\right) + \lambda_1\left(\frac{1}{2M}(h-h_1) + \frac{1}{2}\Delta x_{x1}\right) = 0 \end{aligned} \tag{B.42}$$

Dual feasibility:
$$\mu_0, \mu_1, \lambda_0, \lambda_1 \geq 0 \tag{B.43}$$

Primal feasibility:
$$g_1 - g \geq \frac{1}{4M}(h-h_1)^2 - \frac{M}{4}\Delta x_{x1}^2 - \frac{1}{2}(h+h_1)\Delta x_{x1} \tag{B.44}$$
$$g_2 - g \geq \frac{1}{4M}(h-h_2)^2 - \frac{M}{4}\Delta x_{x2}^2 - \frac{1}{2}(h+h_2)\Delta x_{x2} \tag{B.45}$$
$$g - g_2 \geq \frac{1}{4M}(h_2-h)^2 - \frac{M}{4}\Delta x_{x2}^2 + \frac{1}{2}(h_2+h)\Delta x_{x2} \tag{B.46}$$
$$g - g_1 \geq \frac{1}{4M}(h_1-h)^2 - \frac{M}{4}\Delta x_{x1}^2 + \frac{1}{2}(h_1+h)\Delta x_{x1} \tag{B.47}$$

Compl. slackness:
$$\mu_0\left(g_1 - g - \left(\frac{1}{4M}(h-h_1)^2 - \frac{M}{4}\Delta x_{x1}^2 - \frac{1}{2}(h+h_1)\Delta x_{x1}\right)\right) = 0 \tag{B.48}$$
$$\mu_1\left(g_2 - g - \frac{1}{4M}(h-h_2)^2 - \left(\frac{M}{4}\Delta x_{x2}^2 - \frac{1}{2}(h+h_2)\Delta x_{x2}\right)\right) = 0 \tag{B.49}$$
$$\lambda_0\left(g - g_2 - \left(\frac{1}{4M}(h_2-h)^2 - \frac{M}{4}\Delta x_{x2}^2 + \frac{1}{2}(h_2+h)\Delta x_{x2}\right)\right) = 0 \tag{B.50}$$
$$\lambda_1\left(g - g_1 - \left(\frac{1}{4M}(h_1-h)^2 - \frac{M}{4}\Delta x_{x1}^2 + \frac{1}{2}(h_1+h)\Delta x_{x1}\right)\right) = 0. \tag{B.51}$$

Let $x \in [x_1, y_1]$. Then, $\mu_0 = \mu_1 = \lambda_0 = 0$, $\lambda_1 = 1$, $h = h_1 - M\Delta x_{x1}$ and $g = g_1 + h_1 \Delta x_{x1} - \frac{M}{2}\Delta x_{x1}^2$ is a valid solution. Indeed, all conditions follow directly from the definition of g and h, except for (B.45), which becomes equivalent to (10.45) and is thus satisfied by assumption on S, and (B.46), which becomes:

$$\begin{aligned}
&T_{12}^g - \frac{3M}{4}\Delta x_{x1}^2 - \frac{M}{4}\Delta x_{x2} - \frac{M}{2}\Delta x_{x1}\Delta x_{x2} + \frac{\Delta h_{12}^2}{4M} + \frac{\Delta h_{12}}{2}(2x - x_1 - x_2) \leq 0 \\
&\Leftrightarrow \Delta x_{x2}(\Delta h_{12} + M\Delta x_{12}) \leq -T_{12}^g - \frac{\Delta h_{12}^2}{4M} - \frac{3M\Delta x_{12}^2}{4} - \frac{\Delta h_{12}\Delta x_{12}}{2} \\
&\Leftrightarrow x \leq y_1.
\end{aligned}$$

Let $x \in [y_2, x_2]$. Then, $\mu_0 = \mu_1 = \lambda_1 = 0$, $\lambda_0 = 1$, $h = h_2 - M\Delta x_{x2}$ and $g = g_2 + h_2 \Delta x_{x2} - \frac{M}{2}\Delta x_{x2}^2$ is a valid solution. Again, all conditions follow directly from the definition of g and h, except for (B.44), which becomes equivalent to (10.45) as applied to the pair (j, i) and is thus satisfied by assumption on S, and (B.47), which becomes:

$$\begin{aligned}
&T_{12}^g - \frac{3M}{4}\Delta x_{x2}^2 + \frac{M}{4}\Delta x_{x1} + \frac{M}{2}\Delta x_{x1}\Delta x_{x2} - \frac{\Delta h_{12}^2}{4M} + \frac{\Delta h_{12}}{2}(2x + \frac{x_1}{2} - \frac{3x_2}{2}) \geq 0 \\
&\Leftrightarrow \Delta x_{x2}(\Delta h_{12} + M\Delta x_{12}) \geq -T_{12}^g + \frac{\Delta h_{12}^2}{4M} - \frac{M\Delta x_{12}^2}{4} + \frac{\Delta h_{12}\Delta x_{12}}{2} \\
&\Leftrightarrow x \geq y_2.
\end{aligned}$$

Else, $\mu_0 = \mu_1 = 0$,

$$\lambda_0 = \frac{2M(x - y_1)}{\Delta h_{12} + M\Delta x_{12}} \underset{x \geq y_1, (10.54)}{\geq} 0,\ \lambda_1 = 1 - \lambda_0 \underset{x \leq y_2}{\geq} 1 - \frac{2M(y_2 - y_1)}{\Delta h_{12} + M\Delta x_{12}} \underset{\text{Def. of } y_2}{=} 0$$

$$h = h_1 + M(x + x_1 - 2y_1)$$

and

$$\begin{aligned}
g &= g_1 + \frac{M}{4}(x + x_1 - 2y_1)^2 - \frac{M}{4}\Delta x_{x1}^2 + \frac{1}{2}(2h_1 + M(x + x_1 - 2y_1))\Delta x_{x1} \\
&= g_1 + h_1\Delta x_{x1} - \frac{M}{2}(y_1 - x_1)^2 + \frac{M}{2}(x^2 - y_1^2) + M(x_1 - 2y_1)(x - y_1),
\end{aligned}$$

is a valid solution. Indeed, all conditions follow directly from the definition of g, h, λ_0 and λ_1 except for (B.44) and (B.45), which become equivalent to

(10.54) and are thus satisfied by assumption on S.

Suppose now $\Delta h_{12} = \pm M \Delta x_{12}$. Note that this implies $T^g_{12} = \pm \frac{M}{2} \Delta x^2_{12}$. Indeed, Condition (10.45) applied to both pairs becomes:

$$T^g_{12} \geq \pm \frac{M}{2} \Delta x^2_{12},$$
$$T^g_{2,1} \geq \pm \frac{M}{2} \Delta x^2_{12} \underset{\text{(B.30)}}{\Leftrightarrow} T^g_{12} \leq \Delta h_{12} \Delta x_{12} \mp \frac{M}{2} \Delta x^2_{12} = \pm \frac{M}{2} \Delta x^2_{12}.$$

In this case, $\mu_1 = \lambda_0 = \lambda_1 = 0$, $\mu_0 = 1$,

$$g = g_1 + h_1 \Delta x_{x1} + \frac{M}{2} \Delta x^2_{x1} \underset{T^g_{12} = \frac{M}{2} \Delta x^2_{12}}{=} g_2 + h_2 \Delta x_{x2} + \frac{M}{2} \Delta x^2_{x2}$$
$$h = h_1 + M \Delta x_{x1} = h_2 + M \Delta x_{x2}.$$

are always a solution satisfying the KKT conditions associated to (B.40). Indeed, all conditions follow directly from these two definitions of g and h.

Finally, suppose $\Delta h_{12} = -M \Delta x_{12}$, which implies $T^g_{12} = -\frac{M}{2} \Delta x^2_{12}$ since (10.45) applied to both pairs becomes:

$$T^g_{12} \geq -\frac{M}{2} \Delta x^2_{12},$$
$$T^g_{21} \geq -\frac{M}{2} \Delta x^2_{12} \underset{\text{(B.30)}}{\Leftrightarrow} T^g_{12} \leq \Delta h_{12} \Delta x_{12} + \frac{M}{2} \Delta x^2_{12} = -\frac{M}{2} \Delta x^2_{12}.$$

In this case, let $g'_{\min}(x) = h_1 - M \Delta x_{x1} = h_2 - M \Delta x_{x2}$. Clearly, $g'_{\min}(x)$ is the first derivative of $g_{\min}(x)$, and this function interpolates S. Moreover, in this case, $\mu_0 = \mu_1 = \lambda_1 = 0$, $\lambda_0 = 1$,

$$g = g_1 + h_1 \Delta x_{x1} - \frac{M}{2} \Delta x^2_{x1} \underset{T^g_{12} = -\frac{M}{2} \Delta x^2_{12}}{=} g_2 + h_2 \Delta x_{x2} - \frac{M}{2} \Delta x^2_{x2}$$
$$h = h_1 - M \Delta x_{x1} = h_2 - M \Delta x_{x2}.$$

are always a solution satisfying the KKT conditions associated with Problem (B.40). Indeed, all conditions follow directly from these two definitions of g and h.

The computation of $g_{\max}(x)$ follows the same argument, except with the associated derivative being:

$$g'_{\max}(x) = \begin{cases} h_1 + M\Delta x_{x1} & \text{if } x \in [x_1, z_1] \\ h_1 - M(x + x_1 - 2z_1) & \text{if } x \in [z_1, z_2] \\ h_2 + M\Delta x_{x2} & \text{if } x \in [z_2, x_2]. \end{cases} \tag{B.52}$$

□

Remark B.8. Observe that y_1, y_2, z_1 and z_2 are well-defined in the sense that

$$x_1 \leq y_1 \leq y_2 \leq x_2 \text{ and } x_1 \leq z_1 \leq z_2 \leq y_2. \tag{B.53}$$

Indeed, we have

$$\begin{aligned}
y_1 \geq x_1 &\Leftrightarrow -T^g_{12} + \Delta x_{12}\Delta h_{12} \geq -\frac{M}{2}\Delta x_{12}^2 + \frac{1}{4M}(\Delta h_{12} + M\Delta x_{12})^2, \\
y_2 \leq x_2 &\Leftrightarrow T^g_{12} \geq -\frac{M}{2}\Delta x_{12}^2 + \frac{1}{4M}(\Delta h_{12} + M\Delta x_{12})^2, \\
z_1 \geq x_1 &\Leftrightarrow T^g_{12} \geq -\frac{M}{2}\Delta x_{12}^2 + \frac{1}{4M}(\Delta h_{12} + M\Delta x_{12})^2, \\
z_2 \leq x_2 &\Leftrightarrow -T^g_{12} + \Delta x_{12}\Delta h_{12} \geq -\frac{M}{2}\Delta x_{12}^2 + \frac{1}{4M}(\Delta h_{12} + M\Delta x_{12})^2,
\end{aligned}$$

i.e., (10.45) applied to the pairs $(2,1)$ and $(1,2)$.

Exploiting Lemmas 10.18 and B.7, we are now ready to proceed to the proof of Theorem 10.38. Indeed, in Proposition B.9, we prove necessity and sufficiency of (10.54), (10.55), (10.59) and (10.45) to $\mathcal{H}_M$-interpolability of a set $S = \{(x_i, g_i, h_i, f_i)\}_{i=1,2}$, relying on the integration of the extremal gradients derived in Lemma B.7. By Lemma 10.40, these conditions are equivalent to the conditions of Theorem 10.38 (see Remark B.10). And, by Lemma 10.16, these interpolation conditions hold for any set $S = \{(x_i, f_i, g_i, h_i)\}_{i\in[N]}$. Therefore, Proposition B.9 concludes the proof of Theorem 10.38.

Proposition B.9. *A set $S = \{(x_i, g_i, h_i, f_i)\}_{i=1,2}$ is $\mathcal{H}_M$-interpolable if and only if, $\forall i, j = 1, 2$,*
1) If $\Delta h_{ij} + M|\Delta x_{ij}| \neq 0$, then

$$|\Delta h_{ij}| \leq M|\Delta x_{ij}|, \tag{B.54}$$

$$T^f_{ij} \geq -\frac{M}{6}|\Delta x_{ij}|^3 + \frac{\left(T^g_{ij} + \frac{M}{2}\Delta x_{ij}|\Delta x_{ij}|\right)^2}{2\left(\Delta h_{ij} + M|\Delta x_{ij}|\right)} + \frac{\left(\Delta h_{ij} + M|\Delta x_{ij}|\right)^3}{96M^2}, \tag{B.55}$$

$$T^f_{ij} \leq \frac{M}{6}|\Delta x_{ij}|^3 - \frac{\left(T^g_{ij} - \frac{M}{2}\Delta x_{ij}|\Delta x_{ij}|\right)^2}{2\left(M|\Delta x_{ij}| - \Delta h_{ij}\right)} - \frac{\left(M|\Delta x_{ij}| - \Delta h_{ij}\right)^3}{96M^2}, \tag{B.56}$$

$$T^g_{ij} \geq -\frac{M}{2}\Delta x^2_{ij} + \frac{1}{4M}(\Delta h_{ij} + M\Delta x_{ij})^2. \tag{B.57}$$

2) If $\Delta h_{ij} \pm M|\Delta x_{ij}| = 0$*, then*

$$T^g_{ij} = \mp\frac{M}{2}\Delta x_{ij}|\Delta x_{ij}|, \tag{B.58}$$

$$T^f_{ij} = \mp\frac{M}{6}|\Delta x_{ij}|^3. \tag{B.59}$$

Proof. Suppose $x_1 < x_2$. By Lemma 10.18, it holds that S is $\mathcal{F}_M$-interpolable if and only if it satisfies $|h_j - h_i| \leq M|x_j - x_i|$ (Proposition 10.21), (B.57) and

$$\int_{x_1}^{x_2} g_{\min}(x)\mathrm{d}x \leq f_2 - f_1 \leq \int_{x_1}^{x_2} g_{\max}(x)\mathrm{d}x, \tag{B.60}$$

where $g_{\min}$ and $g_{\max}$ are given in Lemma B.7.

Consider first the case $\Delta h_{12} \neq \pm M|\Delta x_{12}|$. Integration of extremal interpolants gives:

$$T^f_{12} \geq -\frac{M}{6}\Delta x^3_{12} + \frac{(T^g_{12} + \frac{M}{2}\Delta x^2_{12})^2}{2(\Delta h_{12} + M\Delta x_{12})} + \frac{(\Delta h_{12} + M\Delta x_{12})^3}{96M^2}, \tag{B.61}$$

$$T^f_{12} \leq \frac{M}{6}\Delta x^3_{12} - \frac{(T^g_{12} - \frac{M}{2}\Delta x^2_{12})^2}{2(M\Delta x_{12} - \Delta h_{12})} - \frac{(M\Delta x_{12} - \Delta h_{12})^3}{96M^2}. \tag{B.62}$$

Suppose now $x_2 < x_1$. Then, taking Lemma B.7 with x_2 and x_1 permuted and following the same argument as in the case $x_1 < x_2$, we obtain the following inequalities to be satisfied:

$$T^f_{21} \geq -\frac{M}{6}\Delta x^3_{21} + \frac{(T^g_{21} + \frac{M}{2}\Delta x^2_{21})^2}{2(\Delta h_{21} + M\Delta x_{21})} + \frac{(\Delta h_{21} + M\Delta x_{21})^3}{96M^2}, \tag{B.63}$$

$$T^f_{21} \leq \frac{M}{6}\Delta x^3_{21} + \frac{(T^g_{21} - \frac{M}{2}\Delta x^2_{21})^2}{2(\Delta h_{21} - M\Delta x_{21})} + \frac{(\Delta h_{21} - M\Delta x_{21})^3}{96M^2}. \tag{B.64}$$

By (B.28), (B.29), (B.30) and (B.31), Condition (B.63) becomes:

$$
\begin{aligned}
&-T_{12}^{f}+T_{12}^{g}\Delta x_{12}-\frac{1}{2}\Delta h_{12}\Delta x_{12}^{2}\geq \frac{M}{6}\Delta x_{12}^{3}-\frac{(\Delta h_{12}+M\Delta x_{12})^{3}}{96M^{2}}\\
&\qquad\qquad\qquad\qquad+\frac{(-T_{12}^{g}+\Delta h_{12}\Delta x_{12}+\frac{M}{2}\Delta x_{12}^{2})^{2}}{-2(\Delta h_{12}+M\Delta x_{12})}\\
\Leftrightarrow &-T_{12}^{f}+T_{12}^{g}\Delta x_{12}-\frac{1}{2}\Delta h_{12}\Delta x_{12}^{2}\geq \frac{M}{6}\Delta x_{12}^{3}-\frac{(T_{12}^{g}+\frac{M}{2}\Delta x_{12}^{2})^{2}}{2(\Delta h_{12}+M\Delta x_{12})}\\
&\qquad\qquad\qquad\qquad+T_{12}^{g}\Delta x_{12}-\frac{1}{2}\Delta h_{12}\Delta x_{12}^{2}-\frac{(\Delta h_{12}+M\Delta x_{12})^{3}}{96M^{2}}\\
\Leftrightarrow &\, T_{12}^{f}\leq -\frac{M}{6}\Delta x_{12}^{3}+\frac{(T_{12}^{g}+\frac{M}{2}\Delta x_{12}^{2})^{2}}{2(\Delta h_{12}+M\Delta x_{12})}+\frac{(\Delta h_{12}+M\Delta x_{12})^{3}}{96M^{2}}.
\end{aligned}
\tag{B.65}
$$

Similarly, (B.64) becomes:

$$
\begin{aligned}
&-T_{12}^{f}+T_{12}^{g}\Delta x_{12}-\frac{1}{2}\Delta h_{12}\Delta x_{12}^{2}\leq -\frac{M}{6}\Delta x_{12}^{3}-\frac{(\Delta h_{12}-M\Delta x_{12})^{3}}{96M^{2}}\\
&\quad+\frac{(-T_{12}^{g}+\Delta h_{12}\Delta x_{12}-\frac{M}{2}\Delta x_{12}^{2})^{2}}{-2(\Delta h_{12}-M\Delta x_{12})}\\
&\Leftrightarrow T_{12}^{f}\geq \frac{M}{6}\Delta x_{12}^{3}+\frac{(T_{12}^{g}-\frac{M}{2}\Delta x_{12}^{2})^{2}}{2(\Delta h_{12}-M\Delta x_{12})}+\frac{(\Delta h_{12}-M\Delta x_{12})^{3}}{96M^{2}}.
\end{aligned}
\tag{B.66}
$$

Conditions (B.61) and (B.66) are equivalent to a single constraint, independently of the sign of Δx_{12}:

$$
T_{12}^{f}\geq -\frac{M}{6}|\Delta x_{12}|^{3}+\frac{\left(T_{12}^{g}+\frac{M}{2}\Delta x_{12}|\Delta x_{12}|\right)^{2}}{2\left(\Delta h_{12}+M|\Delta x_{12}|\right)}+\frac{(\Delta h_{12}+M|\Delta x_{12}|)^{3}}{96M^{2}}.
$$

Similarly, Conditions (B.62) and (B.65) can be expressed as:

$$
T_{12}^{f}\leq \frac{M}{6}|\Delta x_{12}|^{3}-\frac{\left(T_{12}^{g}-\frac{M}{2}\Delta x_{12}|\Delta x_{12}|\right)^{2}}{2\left(M|\Delta x_{12}|-\Delta h_{12}\right)}-\frac{(M|\Delta x_{12}|-\Delta h_{12})^{3}}{96M^{2}}.
$$

We have obtained interpolation conditions to be satisfied by the pair $(1,2)$ for S to be $\mathcal{H}_M$-interpolable. Extending the argument to the pair $(2,1)$ allows concluding that a $\mathcal{H}_M$-interpolable set $S=\{(x_i,g_i,h_i,f_i)\}_{i=1,2}$ necessarily satisfies, for all $i,j=1,2$, Conditions (B.55) and (B.56).

Suppose now $x_2 > x_1$ and $\Delta h_{12} = \pm M\Delta x_{12}$, $T^g_{12} = \pm\frac{M}{2}\Delta x^2_{12}$. Then,

$$f_2 - f_1 = \int_{x_1}^{x_2} g_{\min}(x)\mathrm{d}x \Leftrightarrow T^f_{12} = \pm\frac{M}{6}\Delta x^3_{12}. \tag{B.67}$$

Finally, suppose $x_1 < x_2$ and $\Delta h_{12} = \pm M\Delta x_{12} \Leftrightarrow \Delta h_{21} = \pm M\Delta x_{21}$, implying $T^g_{12} = \mp\frac{M}{2}\Delta x^2_{12}$. By Lemma B.7 applied to the pair $(2,1)$, it holds that

$$T^f_{21} = \pm\frac{M}{6}\Delta x^3_{21} \underset{\text{(B.30),(B.29)}}{\Leftrightarrow} T^f_{12} = \mp\Delta x^3_{12}. \tag{B.68}$$

Combining (B.67) and (B.68), it holds that $\forall i,j = 1,2$ if $\Delta h_{ij} = \pm M|\Delta x_{ij}|$, then necessarily $T^f_{ij} = \pm\frac{M}{6}|\Delta x_{ij}|^3$ (and $T^g_{ij} = \pm\frac{M}{2}\Delta x_{ij}|\Delta x_{ij}|$) for S to be $\mathcal{H}_M$-interpolable. □

B.5 Proof of Lemma 10.40

Proof of Lemma 10.40. Condition (10.55) evaluated at (j,i) is equivalent to Condition (10.59) evaluated at (i,j):

$$\begin{aligned}
&T^f_{ji} \geq -\frac{M}{6}|\Delta x_{ji}|^3 + \frac{\left(T^g_{ji} + \frac{M}{2}\Delta x_{ji}|\Delta x_{ji}|\right)^2}{2\left(\Delta h_{ji} + M|\Delta x_{ji}|\right)} + \frac{\left(\Delta h_{ji} + M|\Delta x_{ji}|\right)^3}{96M^2} \\
&\Leftrightarrow -T^f_{ij} + T^g_{ij}\Delta x_{ij} - \frac{1}{2}\Delta h_{ij}\Delta x^2_{ij} \geq -\frac{M}{6}|\Delta x_{ij}|^3 - \frac{\left(-T^g_{ij} + \Delta h_{ij}\Delta x_{ij} - \frac{M}{2}\Delta x_{ij}|\Delta x_{ij}|\right)^2}{2\left(\Delta h_{ij} - M|\Delta x_{ij}|\right)} \\
&\qquad\qquad - \frac{\left(\Delta h_{ij} - M|\Delta x_{ij}|\right)^3}{96M^2} \\
&\Leftrightarrow T^f_{ij} \leq \frac{M}{6}|\Delta x_{ij}|^3 + \frac{\left(T^g_{ij} - \frac{M}{2}\Delta x_{ij}|\Delta x_{ij}|\right)^2}{2\left(\Delta h_{ij} - M|\Delta x_{ij}|\right)} + T^g_{ij}\Delta x_{ij} - \frac{1}{2}\Delta h_{ij}\Delta x^2_{ij} - T^g_{ij}\Delta x_{ij} \\
&\qquad - \frac{1}{2}\Delta h_{ij}\Delta x^2_{ij} + \frac{\left(\Delta h_{ij} - M|\Delta x_{ij}|\right)^3}{96M^2} \\
&\Leftrightarrow T^f_{ij} \leq \frac{M}{6}|\Delta x_{ij}|^3 - \frac{\left(T^g_{ij} - \frac{M}{2}\Delta x_{ij}|\Delta x_{ij}|\right)^2}{2\left(M|\Delta x_{ij}| - \Delta h_{ij}\right)} - \frac{\left(M|\Delta x_{ij}| - \Delta h_{ij}\right)^3}{96M^2}.
\end{aligned}$$

Hence, if S satisfies (10.55) $\forall i,j = 1,2$, then it satisfies (10.59) $\forall i,j = 1,2$.

We now show that satisfaction $\forall i,j = 1,2$ of (10.54) and (10.55) (hence satisfaction of (10.59)) imply satisfaction of (10.45) $\forall i, j = 1, 2$. Observe that (10.45) is satisfied by both pairs (i, j) and (j, i) if and only if

$$T^g_{ij} - K \geq 0 \text{ and } T^g_{ji} - K \geq 0 \underset{(B.30)}{\Leftrightarrow} -T^g_{ij} + \Delta h_{ij}\Delta x_{ij} - K \geq 0, \tag{B.69}$$

where $K = -\frac{M}{2}\Delta x^2_{ij} + \frac{1}{4M}(\Delta h_{ij} + M\Delta x_{ij})^2$. Condition (B.69) is equivalent to

$$(T^g_{ij} - K)(-T^g_{ij} + \Delta h_{ij}\Delta x_{ij} - K) \geq 0, \tag{B.70}$$

where the equivalence follows from the fact that $T^g_{ij} < K$ and $-T^g_{ij} + \Delta h_{ij}\Delta x_{ij} < K$ is in contradiction with Condition (10.54), since it implies:

$$2K > \Delta h_{ij}\Delta x_{ij} \Leftrightarrow -\frac{M}{2}\Delta x^2_{ij} + \frac{1}{2M}\Delta h^2_{ij} > 0.$$

Substracting (10.55) and (10.59) yields:

$$\begin{aligned}
\frac{M}{3}|\Delta x_{ij}|^3 \geq{}& \frac{\left(T^g_{ij} - \frac{M}{2}\Delta x_{ij}|\Delta x_{ij}|\right)^2}{2\left(M|\Delta x_{ij}| - \Delta h_{ij}\right)} + \frac{\left(T^g_{ij} + \frac{M}{2}\Delta x_{ij}|\Delta x_{ij}|\right)^2}{2\left(\Delta h_{ij} + M|\Delta x_{ij}|\right)} + \\
&\frac{\left(M|\Delta x_{ij}| - \Delta h_{ij}\right)^3 + \left(\Delta h_{ij} + M|\Delta x_{ij}|\right)^3}{96M^2} \\
\Leftrightarrow{}& \frac{M|\Delta x_{ij}|}{(M|\Delta x_{ij}|)^2 - \Delta h^2_{ij}}\left(-(T^g_{ij})^2 + T^g_{ij}\Delta h_{ij}\Delta x_{ij} - \frac{M^2\Delta x^2_{ij}}{16} - \frac{3\Delta h^2_{ij}\Delta x^2_{ij}}{8} + \frac{\Delta h^4_{ij}}{16M^2}\right) \geq 0 \\
\underset{(10.54)}{\Leftrightarrow}{}& (T^g_{ij} - \frac{\Delta h^2_{ij}}{4M} + \frac{M}{4}\Delta x^2_{12} - \frac{\Delta h_{ij}\Delta x_{ij}}{2})(-T^g_{ij} - \frac{\Delta h^2_{ij}}{4M} + \frac{M}{4}\Delta x^2_{ij} + \frac{\Delta h_{ij}\Delta x_{ij}}{2}) \geq 0 \\
\Leftrightarrow{}& (T^g_{ij} - K)(-T^g_{ij} + \Delta h_{ij}\Delta x_{ij} - K) \geq 0,
\end{aligned}$$

and (10.45) is satisfied.

Finally, since $\Delta h_{ij} \geq -M|\Delta x_{ij}|$, (10.55) implies $T^f_{ij} \geq -\frac{M}{6}|\Delta x_{ij}|^3$, and since $\Delta h_{ij} \leq M|\Delta x_{ij}|$, (10.59) implies $T^f_{ij} \leq \frac{M}{6}|\Delta x_{ij}|^3$. □

Remark B.10. By Lemma 10.40, proving that (10.54) and (10.55) are interpolation conditions for $\mathcal{H}_M$ is equivalent to proving that (10.54), (10.55), (10.59) and (10.45) are interpolation conditions for $\mathcal{H}_M$.

Remark B.11. Conditions (10.45) and (10.59) can be removed from the final interpolation conditions since they are implied by the two remaining conditions. On the contrary, (10.54) cannot be removed from the interpolation conditions since it is not implied by (10.55). Indeed, consider for example the set $S = \{(0,0,0,0), (\frac{1}{4}, \frac{1}{3}, 0, 0)\}$ satisfying (10.55) but not (10.54) for $M = 1$.

However, since (10.45) implies (10.54) (it suffices to sum the condition applied to the pairs (i, j) and (j, i) to recover (10.54)), another set of interpolation conditions equivalent to the ones of Theorem 10.38 consists of (10.54) and (10.45).

PART III
Conclusion

11

Research outcomes and perspective

We extended and exploited the Performance Estimation Problem framework to analyze and obtain tight worst-case guarantees on new function classes and optimization methods. We summarize the research outcomes and discuss future perspectives.

11.1 Research outcomes

In both Part I and II, we extended the PEP framework to the analysis of new settings and then exploited this extension to analyze some of them.

In Part I, 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 values spectrum. We also obtained interpolation conditions for quadratic functions. Then, we showed a limitation on the classes of linear operators that could be represented via Gram matrices in a convex way.

With these interpolation conditions, the Performance Estimation Problem framework can be extended to analyze first-order optimization methods applied to problems involving linear operators. Therefore, we analyzed and obtained new numerical and analytical worst-case guarantees on the gradient, Chambolle-Pock, Condat-Vũ, and Barzilai-Borwein methods.

In Part II, we proposed a principled technique to obtain interpolation conditions for classes of univariate functions. We exploited this technique to obtain interpolation conditions for univariate (convex) Hessian Lipschitz functions and convex generalized self-concordant functions. To incorporate these non-convex conditions into the PEP framework, we used the recent non-convex PEP extension, allowing to use any non-convex constraints.

Using these new interpolation conditions and the non-convex PEP framework, we analyzed Newton's method and several of its regularized variants to obtain, again, numerical and analytical worst-case guarantees on different classes of univariate functions.

11.2 Perspective

11.2.1 Interpolation conditions

Obtaining exact worst-case guarantees (with or without PEP) requires interpolation conditions for all function classes involved in the analysis. In this work, we developed interpolation conditions for several classes of linear operators and second-order classes of univariate functions. For linear operators, the following questions remain open:

- Can we develop interpolation conditions for general square matrices, and which properties could we require?
- Is there another convex representation of classes of linear operators that does not rely on Gram representations?

The technique proposed in Chapter 10 could be used to develop interpolation conditions for the following classes of univariate functions:

- (L_0, L_1)-smooth functions [ZHSJ19];
- Higher-order classes of functions, e.g., functions with Lipschitz third derivatives.

However, the technique only provides interpolation conditions for univariate classes of functions. The interpolation conditions of classes considered in Chapter 10 and mentioned above in the general multivariate case are still open.

11.2.2 Analysis and design of methods

Using these new interpolation conditions, we analyzed different optimization methods and settings. However, we could further exploit them to improve our understanding of the large variety of methods, including more instances of primal-dual and second-order methods than the ones considered here. This understanding can further lead to the design of optimized methods. Given the current knowledge of interpolation conditions and the non-convex PEP, many interesting settings can also be almost directly analyzed (see Chapter 9).

11.2.3 Why are there so many univariate worst-cases?

We have observed that, surprisingly, there often exists a univariate worst-case instance of a setting (see, e.g., Section B.1). In other words, a univariate function is often sufficiently "rich" to trigger the worst-case behavior of an optimization method on a function class. However, it is not always the case, sometimes we need multivariate functions. Understanding this phenomenon (via the Performance Estimation or another theory) could lead to a better view of the worst-case behavior of optimization methods and thus to their design.

11.2.4 Further extension and usage of PEP

Recently, several extensions and exploitations of the PEP framework have been developed (see, e.g., the `PEPit` documentation website). As mentioned above, the framework can still be extended and exploited to analyze new types of settings.

However, it could also now be sufficiently developed and user-friendly (thanks to the implementations `PESTO` [THG17b] and `PEPit` [GMG+24] in *Matlab* and *Python*) to be widely used as an assistant for optimization theoretists. Ideally, people could develop their new optimization methods on a function class already available in `PEPit`, perform their usual theoretical analysis and numerical experiments, and at the end, put their method in the PEP machinery to observe numerically the exact worst-case guarantee.

Bibliography

[AB98] Charalambos D. Aliprantis and Owen Burkinshaw. Principles of real analysis. *Gulf Professional Publishing*, 1998.

[AB22] Aviad Aberdam and Amir Beck. An accelerated coordinate gradient descent algorithm for non-separable composite optimization. *Journal of Optimization Theory and Applications*, 193(1):219–246, 2022.

[Abb24] Hadi Abbaszadehpeivasti. *Performance analysis of optimization methods for machine learning*. PhD thesis, Tilburg University, 2024.

[AdKZ22] Hadi Abbaszadehpeivasti, Etienne de Klerk, and Moslem Zamani. The exact worst-case convergence rate of the gradient method with fixed step lengths for L-smooth functions. *Optimization Letters*, 16(6):1649–1661, 2022.

[AdKZ23a] Hadi Abbaszadehpeivasti, Etienne de Klerk, and Moslem Zamani. Conditions for linear convergence of the gradient method for non-convex optimization. *Optimization Letters*, 17(5):1105–1125, 2023.

[AdKZ23b] Hadi Abbaszadehpeivasti, Etienne de Klerk, and Moslem Zamani. Convergence rate analysis of randomized and cyclic coordinate descent for convex optimization through semidefinite programming. *Applied Set-Valued Analysis and Optimization*, 5(2):141–153, 2023.

[AdKZ24] Hadi Abbaszadehpeivasti, Etienne de Klerk, and Moslem Zamani. On the rate of convergence of the difference-of-convex algorithm (DCA). *Journal of Optimization Theory and Applications*, 202(1):475–496, 2024.

[Alt18] Jason Altschuler. *Greed, hedging, and acceleration in convex optimization*. PhD thesis, Massachusetts Institute of Technology, 2018.

[AP25] Jason M. Altschuler and Pablo A. Parrilo. Acceleration by stepsize hedging: Multi-step descent and the silver stepsize schedule. *Journal of the ACM*, 72(2):1–38, 2025.

[ApS24] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 10.1.*, 2024.

[BAP23] Luis M. Briceño-Arias and Nelly Pustelnik. Theoretical and numerical comparison of first order algorithms for cocoercive equations and smooth convex optimization. *Signal Processing*, 206:108900, 2023.

[Bar21] Mathieu Barré. *Worst-case analysis of efficient first-order methods*. PhD thesis, Université Paris sciences et lettres, 2021.

[BB88] Jonathan Barzilai and Jonathan M Borwein. Two-point step size gradient methods. *IMA journal of numerical analysis*, 8(1):141–148, 1988.

[BCN18] Léon Bottou, Frank E. Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM review*, 60(2):223–311, 2018.

[BDH19] Oleg Burdakov, Yuhong Dai, and Na Huang. Stabilized Barzilai-Borwein method. *Journal of Computational Mathematics*, pages 916–936, 2019.

[BDL+25] Nizar Bousselmi, Zhicheng Deng, Jie Lu, Julien M. Hendrickx, and François Glineur. On the convex interpolation for linear operators. *preprint, to be submitted*, 2025.

[Bec17] Amir Beck. *First-order methods in optimization*. SIAM, 2017.

[Bec22] Amir Beck. Proximal methods materials. 2022. URL: `https://sites.google.com/site/amirbeck314/home?authuser=0`.

[Ber12] Dimitri Bertsekas. *Dynamic programming and optimal control: Volume I*, volume 4. Athena scientific, 2012.

[BHG24] Nizar Bousselmi, Julien M. Hendrickx, and François Glineur. Interpolation conditions for linear operators and applications to performance estimation problems. *SIAM Journal on Optimization*, 34(3):3033–3063, 2024.

[BJM$^+$12] Francis Bach, Rodolphe Jenatton, Julien Mairal, Guillaume Obozinski, et al. Optimization with sparsity-inducing penalties. *Foundations and Trends® in Machine Learning*, 4(1):1–106, 2012.

[BLDPM24] Andrea Bisoffi, Lidong Li, Claudio De Persis, and Nima Monshizadeh. Controller synthesis for input-state data with measurement errors. *IEEE Control Systems Letters*, 2024.

[BPHG24] Nizar Bousselmi, Nelly Pustelnik, Julien M. Hendrickx, and François Glineur. Comparison of proximal first-order primal and primal-dual algorithms via performance estimation. In *2024 32nd European Signal Processing Conference (EUSIPCO)*, pages 2647–2651. IEEE, 2024.

[BT09] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[BTB23] Mathieu Barré, Adrien B. Taylor, and Francis Bach. Principled analyses and design of first-order methods with inexact proximal operators. *Mathematical Programming*, 201(1):185–230, 2023.

[BTd20] Mathieu Barré, Adrien Taylor, and Alexandre d'Aspremont. Complexity guarantees for Polyak steps with momentum. In *Conference on learning theory*, pages 452–478. PMLR, 2020.

[BUG23] Sebastian Banert, Manu Upadhyaya, and Pontus Giselsson. The chambolle–pock method converges weakly with $\theta > 1/2$ and $\tau\sigma\|L\|^2 < 4/(1+2\theta)$. *arXiv preprint arXiv:2309.03998*, 2023.

[BV04] Stephen P. Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[CD94] Shaobing Chen and David Donoho. Basis pursuit. In *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 41–44. IEEE, 1994.

[CDJ23] El Mahdi Chayti, Nikita Doikov, and Martin Jaggi. Unified convergence theory of stochastic and variance-reduced cubic Newton methods. *arXiv preprint arXiv:2302.11962*, 2023.

[CGT00] Andrew R. Conn, Nicholas I.M. Gould, and Philippe L. Toint. *Trust region methods*. SIAM, 2000.

[CGT10] Coralia Cartis, Nicholas I.M. Gould, and Philippe L. Toint. On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization problems. *Siam journal on optimization*, 20(6):2833–2852, 2010.

[CGT11] Coralia Cartis, Nicholas I.M. Gould, and Philippe L. Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Mathematical Programming*, 127(2):245–295, 2011.

[CH22] Sebastien Colla and Julien M. Hendrickx. Automated performance estimation for decentralized optimization via network size independent problems. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 5192–5199. IEEE, 2022.

[CH23] Sebastien Colla and Julien M. Hendrickx. Automatic performance estimation for decentralized optimization. *IEEE Transactions on Automatic Control*, 68(12):7136–7150, 2023.

[CHZ16] Peijun Chen, Jianguo Huang, and Xiaoqun Zhang. A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions. *Fixed Point Theory and Applications*, 2016(54), 2016.

[CKCH23] Laurent Condat, Daichi Kitahara, Andrés Contreras, and Akira Hirabayashi. Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists. *SIAM Review*, 65(2):375–435, 2023.

[CLL17] Chong-Yung Chi, Wei-Chiang Li, and Chia-Hsiang Lin. *Convex optimization for signal processing and communications: from fundamentals to applications*. CRC press, 2017.

[CM07] Sunil Chopra and Peter Meindl. *Supply chain management. Strategy, planning & operation*. Springer, 2007.

[Col24] Sébastien Colla. *Computer-Aided Analysis of Decentralized Optimization Methods*. PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2024.

[Con13] Laurent Condat. A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *Journal of optimization theory and applications*, 158(2):460–479, 2013.

[CP11a] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

[CP11b] Patrick L. Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[CP16a] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[CP16b] Antonin Chambolle and Thomas Pock. On the ergodic convergence rates of a first-order primal–dual algorithm. *Mathematical Programming*, 159(1):253–287, 2016.

[CSV09] Andrew R. Conn, Katya Scheinberg, and Luis N. Vicente. *Introduction to derivative-free optimization*. SIAM, 2009.

[Dan16] George B. Dantzig. Linear programming and extensions. In *Linear programming and extensions*. Princeton university press, 2016.

[DG24] Shuvomoy Das Gupta. *Advances in Computer-Assisted Design and Analysis of First-Order Optimization Methods and Related*

Problems. PhD thesis, Massachusetts Institute of Technology, 2024.

[DGFST24] Shuvomoy Das Gupta, Robert M. Freund, Xu Andy Sun, and Adrien Taylor. Nonlinear conjugate gradient methods: worst-case convergence rates via computer-assisted analyses. *Mathematical Programming*, 2024.

[DGVPR24a] Shuvomoy Das Gupta, Bart P.G. Van Parys, and Ernest K. Ryu. Branch-and-bound performance estimation programming: A unified methodology for constructing optimal optimization methods. *Mathematical Programming*, 204(1):567–639, 2024.

[DGVPR24b] Shuvomoy Das Gupta, Bart PG Van Parys, and Ernest K. Ryu. Branch-and-bound performance estimation programming: A unified methodology for constructing optimal optimization methods. *Mathematical Programming*, 204(1):567–639, 2024.

[dKGT17] Etienne de Klerk, François Glineur, and Adrien B. Taylor. On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions. *Optimization Letters*, 11(7):1185–1199, 2017.

[dKGT20] Etienne de Klerk, Francois Glineur, and Adrien B. Taylor. Worst-case convergence analysis of inexact gradient and Newton methods through semidefinite programming performance estimation. *SIAM Journal on Optimization*, 30(3):2053–2082, 2020.

[DKW82] Chandler Davis, William M. Kahan, and Hans F. Weinberger. Norm-preserving dilations and their applications to optimal error bounds. *SIAM Journal on Numerical Analysis*, 19(3):445–469, 1982.

[DL02] Yu-Hong Dai and Li-Zhi Liao. R-linear convergence of the Barzilai and Borwein gradient method. *IMA Journal of numerical analysis*, 22(1):1–10, 2002.

[DMN24] Nikita Doikov, Konstantin Mishchenko, and Yurii Nesterov. Super-universal regularized Newton method. *SIAM Journal on Optimization*, 34(1):27–56, 2024.

[Doi21] Nikita Doikov. *New second-order and tensor methods in convex optimization*. PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2021.

[Doi23] Nikita Doikov. Minimizing quasi-self-concordant functions by gradient regularization of Newton method. *arXiv preprint arXiv:2308.14742*, 2023.

[Dou66] Ronald G. Douglas. On majorization, factorization, and range inclusion of operators on hilbert space. *Proceedings of the American Mathematical Society*, 17(2):413–415, 1966.

[DS19] Yoel Drori and Ohad Shamir. Presentation on "Performance estimation of stochastic first-order methods". *ICCOPT*, 2019.

[DST15] Yoel Drori, Shoham Sabach, and Marc Teboulle. A simple algorithm for a class of nonsmooth convex–concave saddle-point problems. *Operations Research Letters*, 43(2):209–214, 2015.

[dST$^+$21] Alexandre d'Aspremont, Damien Scieur, Adrien Taylor, et al. Acceleration methods. *Foundations and Trends in Optimization*, 5(1-2):1–245, 2021.

[DT14] Yoel Drori and Marc Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145(1):451–482, 2014.

[DT16] Yoel Drori and Marc Teboulle. An optimal variant of Kelley's cutting-plane method. *Mathematical Programming*, 160(1):321–351, 2016.

[DT20] Yoel Drori and Adrien B. Taylor. Efficient first-order methods for convex minimization: a constructive approach. *Mathematical Programming*, 184(1):183–220, 2020.

[DTdB22] Radu-Alexandru Dragomir, Adrien B. Taylor, Alexandre d'Aspremont, and Jérôme Bolte. Optimal complexity and certification of Bregman first-order methods. *Mathematical Programming*, pages 41–83, 2022.

[Eld80] Lars Eldén. Perturbation theory for the least squares problem with linear equality constraints. *SIAM Journal on Numerical Analysis*, 17(3):338–350, 1980.

[Fen49] Werner Fenchel. On conjugate convex functions. *Canadian Journal of Mathematics*, 1(1):73–77, 1949.

[Gal10] Jean H. Gallier. Notes on the Schur complement. 2010. URL: `https://repository.upenn.edu/cgi/viewcontent.cgi?article=1637&context=cis_papers`.

[GDT23a] Baptiste Goujaud, Aymeric Dieuleveut, and Adrien Taylor. Counter-examples in first-order optimization: a constructive approach. *IEEE Control Systems Letters*, 7:2485–2490, 2023.

[GDT23b] Baptiste Goujaud, Aymeric Dieuleveut, and Adrien Taylor. On fundamental proof structures in first-order optimization. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 3023–3030. IEEE, 2023.

[GEIGM22] Charles Guille-Escuret, Adam Ibrahim, Baptiste Goujaud, and Ioannis Mitliagkas. Gradient descent is optimal under lower restricted secant inequality and upper error bound. *Advances in Neural Information Processing Systems*, 35:24893–24904, 2022.

[GLG22] Eduard Gorbunov, Nicolas Loizou, and Gauthier Gidel. Extragradient method: $\mathcal{O}(1/k)$ last-iterate convergence for monotone variational inequalities and connections with cocoercivity. In *International Conference on Artificial Intelligence and Statistics*, pages 366–402. PMLR, 2022.

[GM76] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & mathematics with applications*, 2(1):17–40, 1976.

[GMG$^+$24] Baptiste Goujaud, Céline Moucer, François Glineur, Julien M. Hendrickx, Adrien B. Taylor, and Aymeric Dieuleveut. PEPit: computer-assisted worst-case analyses of first-order optimization methods in Python. *Mathematical Programming Computation*, 16(3):337–367, 2024.

[Gou24] Baptiste Goujaud. *Constructive approaches to worst-case complexity analyses of gradient methods for convex optimization: contributions, new insights, and novel results*. PhD thesis, Institut Polytechnique de Paris, 2024.

[Gri24] Benjamin Grimmer. Provably faster gradient descent via long steps. *SIAM Journal on Optimization*, 34(3):2588–2608, 2024.

[GSW23] Benjamin Grimmer, Kevin Shu, and Alex L. Wang. Accelerated gradient descent via long steps. *arXiv preprint arXiv:2309.09961*, 2023.

[GTD22] Baptiste Goujaud, Adrien Taylor, and Aymeric Dieuleveut. Optimal first-order methods for convex functions with a quadratic upper bound. *arXiv preprint arXiv:2205.15033*, 2022.

[GTD23] Baptiste Goujaud, Adrien Taylor, and Aymeric Dieuleveut. Provable non-accelerations of the heavy-ball method. *arXiv preprint arXiv:2307.11291*, 2023.

[GTG22] Eduard Gorbunov, Adrien Taylor, and Gauthier Gidel. Last-iterate convergence of optimistic gradient method for monotone variational inequalities. *Advances in neural information processing systems*, 35:21858–21870, 2022.

[GTHG23] Eduard Gorbunov, Adrien Taylor, Samuel Horváth, and Gauthier Gidel. Convergence of proximal point and extragradient-based methods beyond monotonicity: the case of negative comonotonicity. In *International Conference on Machine Learning*, pages 11614–11641. PMLR, 2023.

[Gur24] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2024.

[GY20] Guoyong Gu and Junfeng Yang. Tight sublinear convergence rate of the proximal point algorithm for maximal monotone inclusion problems. *SIAM Journal on Optimization*, 30(3):1905–1921, 2020.

[GY24] Guoyong Gu and Junfeng Yang. Tight ergodic sublinear convergence rate of the relaxed proximal point algorithm for monotone variational inequalities. *Journal of Optimization Theory and Applications*, 202(1):373–387, 2024.

[Hil21] Roland Hildebrand. Optimal step length for the Newton method: case of self-concordant functions. *Mathematical Methods of Operations Research*, 94(2):253–279, 2021.

[HJ12] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[HKP+22] Slavomír Hanzely, Dmitry Kamzolov, Dmitry Pasechnyuk, Alexander Gasnikov, Peter Richtárik, and Martin Takác. A damped Newton method achieves global $\mathcal{O}(1/k^2)$ and local quadratic convergence rate. *Advances in Neural Information Processing Systems*, 35:25320–25334, 2022.

[HM12] Uwe Helmke and John B. Moore. *Optimization and dynamical systems*. Springer Science & Business Media, 2012.

[IH24] Anastasia Ivanova and Roland Hildebrand. Optimal step length for the maximal decrease of a self-concordant function by the Newton method. *Optimization Letters*, 18(3):847–854, 2024.

[JDGR23] Uijeong Jang, Shuvomoy Das Gupta, and Ernest K. Ryu. Computer-assisted design of accelerated composite optimization methods: Optista. *arXiv preprint arXiv:2305.15704*, 2023.

[Kam25] Yassine Kamri. *Contributions to Performance Estimation Problems for the Analysis and Design of First-Order Optimization Methods*. PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2025.

[KF16] Donghwan Kim and Jeffrey A. Fessler. Optimized first-order methods for smooth convex minimization. *Mathematical programming*, 159(1):81–107, 2016.

[KF17] Donghwan Kim and Jeffrey A. Fessler. On the convergence analysis of the optimized gradient method. *Journal of optimization theory and applications*, 172(1):187–205, 2017.

[KF18a] Donghwan Kim and Jeffrey A. Fessler. Another look at the fast iterative shrinkage/thresholding algorithm (FISTA). *SIAM Journal on Optimization*, 28(1):223–250, 2018.

[KF18b] Donghwan Kim and Jeffrey A. Fessler. Generalizing the optimized gradient method for smooth convex minimization. *SIAM Journal on Optimization*, 28(2):1920–1950, 2018.

[KF21] Donghwan Kim and Jeffrey A. Fessler. Optimizing the efficiency of first-order methods for decreasing the gradient of smooth convex functions. *Journal of optimization theory and applications*, 188(1):192–219, 2021.

[KGK24] Valery O. Krivchenko, Alexander Vladimirovich Gasnikov, and Dmitry A. Kovalev. Convex-concave interpolation and application of PEP to the bilinear-coupled saddle point problem. *Russian Journal of Nonlinear Dynamics*, 20(5):875–893, 2024.

[KHG23] Yassine Kamri, Julien M. Hendrickx, and François Glineur. On the worst-case analysis of cyclic coordinate-wise algorithms on smooth convex functions. In *2023 European Control Conference (ECC)*. IEEE, 2023.

[KHG25] Yassine Kamri, Julien M. Hendrickx, and François Glineur. Numerical design of optimized first-order algorithms. *arXiv preprint arXiv:2507.20773*, 2025.

[Kim21] Donghwan Kim. Accelerated proximal point method for maximally monotone operators. *Mathematical Programming*, 190(1):57–87, 2021.

[Kim24] Jungbin Kim. A proof of exact convergence rate of gradient descent. part II. performance criterion $(f(x_N) - f_*)/\|x_0 - x_*\|^2$. *arXiv preprint arXiv:2412.04427*, 2024.

[LG24] Alan Luner and Benjamin Grimmer. Performance estimation for smooth and strongly convex sets. *arXiv preprint arXiv:2410.14811*, 2024.

[Lie21] Felix Lieder. On the convergence rate of the Halpern-iteration. *Optimization letters*, 15(2):405–418, 2021.

[LQT+24] Haochuan Li, Jian Qian, Yi Tian, Alexander Rakhlin, and Ali Jadbabaie. Convex and non-convex optimization under generalized smoothness. *Advances in Neural Information Processing Systems*, 36, 2024.

[LRP16] Laurent Lessard, Benjamin Recht, and Andrew Packard. Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM Journal on Optimization*, 26(1):57–95, 2016.

[Mis23] Konstantin Mishchenko. Regularized Newton method with global convergence. *SIAM Journal on Optimization*, 33(3):1440–1462, 2023.

[MM20] Yura Malitsky and Konstantin Mishchenko. Adaptive gradient descent without descent. In *International Conference on Machine Learning*, pages 6702–6712. PMLR, 2020.

[MTB23] Céline Moucer, Adrien Taylor, and Francis Bach. A systematic approach to Lyapunov analyses of continuous-time models in convex optimization. *SIAM Journal on Optimization*, 33(3):1558–1586, 2023.

[N⁺18] Yurii Nesterov et al. *Lectures on convex optimization*, volume 137. Springer, 2018.

[Nem95] Arkadi Nemirovski. Information-based complexity of convex programming. *Lecture notes*, 834, 1995.

[Nem04] Arkadi Nemirovski. Prox-method with rate of convergence $\mathcal{O}(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

[Nes83] Yurii Nesterov. A method for solving the convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. In *Dokl. Akad. Nauk SSSR*, volume 269, pages 543–547, 1983.

[NN94] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[NNG19] Ion Necoara, Yurii Nesterov, and Francois Glineur. Linear convergence of first order methods for non-strongly convex optimization. *Mathematical Programming*, 175(1):69–107, 2019.

[NP06] Yurii Nesterov and Boris T. Polyak. Cubic regularization of Newton method and its global performance. *Mathematical programming*, 108(1):177–205, 2006.

[NW06] Jorge Nocedal and Stephen J. Wright. Numerical optimization. *Springer Series in Operations Research and Financial Engineering*, 2006.

[PBBZP16] Nelly Pustelnik, Amel Benazza-Benhayia, Yuling Zheng, and Jean-Christophe Pesquet. Wavelet-based image deconvolution and reconstruction. *Wiley encyclopedia of electrical and electronics engineering*, 2016.

[PE10] Daniel P. Palomar and Yonina C. Eldar. *Convex optimization in signal processing and communications*. Cambridge university press, 2010.

[Ped20] Fabian Pedregosa. Residual polynomials and the Chebyshev method. 2020. URL: `http://fa.bianp.net/blog/2020/polyopt/`.

[Pol87] Boris T. Polyak. Introduction to optimization. 1987.

[Pol09] Roman A. Polyak. Regularized Newton method for unconstrained convex optimization. *Mathematical programming*, 120:125–145, 2009.

[PR23] Jisun Park and Ernest K. Ryu. Accelerated infeasibility detection of constrained optimization and fixed-point iterations. In *International Conference on Machine Learning*, pages 27294–27345. PMLR, 2023.

[RBCH23] Anne Rubbens, Nizar Bousselmi, Sébastien Colla, and Julien M. Hendrickx. Interpolation constraints for computing worst-case bounds in performance estimation problems. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 3015–3022. IEEE, 2023.

[RBHG25] Anne Rubbens, Nizar Bousselmi, Julien M. Hendrickx, and François Glineur. Performance estimation of second-order optimization methods on classes of univariate functions. *arXiv preprint arXiv:2506.22764*, 2025.

[RCH25] Anne Rubbens, Sébastien Colla, and Julien M Hendrickx. Computer-aided analyses of stochastic first-order methods, via interpolation conditions for stochastic optimization. *arXiv preprint arXiv:2507.05466*, 2025.

[RGP24] Teodor Rotaru, François Glineur, and Panagiotis Patrinos. Exact worst-case convergence rates of gradient descent: a complete analysis for all constant stepsizes over nonconvex and convex functions. *arXiv preprint arXiv:2406.17506*, 2024.

[RH24] Anne Rubbens and Julien M. Hendrickx. A constraint-based approach to function interpolation, with application to performance estimation for weakly convex optimisation. *arXiv preprint arXiv:2405.08405*, 2024.

[RHT25] Anne Rubbens, Julien M. Hendrickx, and Adrien Taylor. A constructive approach to strengthen algebraic descriptions of function and operator classes. *preprint arXiv:2504.14377*, 2025.

[RL05] Peter J. Rousseeuw and Annick M. Leroy. *Robust regression and outlier detection*. John Wiley & Sons, 2005.

[Rod22] Anton Rodomanov. *Quasi-Newton methods with provable efficiency guarantees*. PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2022.

[ROF92] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[RPG25a] Teodor Rotaru, Panagiotis Patrinos, and François Glineur. Tight analysis of difference-of-convex algorithm (DCA) improves convergence rates for proximal gradient descent. *arXiv preprint arXiv:2503.04486*, 2025.

[RPG25b] Teodor Rotaru, Panagiotis Patrinos, and François Glineur. Tight convergence rates in gradient mapping for the difference-of-convex algorithm. *arXiv preprint arXiv:2506.01791*, 2025.

[RS25] Vinit Ranjan and Bartolomeo Stellato. Verification of first-order methods for parametric quadratic optimization. *Mathematical Programming*, 2025.

[RTBG20] Ernest K. Ryu, Adrien B. Taylor, Carolina Bergeling, and Pontus Giselsson. Operator splitting performance estimation:

Tight contraction factors and optimal parameter selection. *SIAM Journal on Optimization*, 30(3):2251–2271, 2020.

[Sho12] Naum Zuselevich Shor. *Minimization methods for non-differentiable functions*, volume 3. Springer Science & Business Media, 2012.

[SL16] Ziqiang Shi and Rujie Liu. A better convergence analysis of the block coordinate descent method for large scale machine learning. *arXiv preprint arXiv:1608.04826*, 2016.

[SNW11] Suvrit Sra, Sebastian Nowozin, and Stephen J. Wright. *Optimization for machine learning*. MIT press, 2011.

[STD19] Tianxiao Sun and Quoc Tran-Dinh. Generalized self-concordant functions: a recipe for newton-type methods. *Mathematical Programming*, 178(1):145–213, 2019.

[Tay17] Adrien B. Taylor. *Convex interpolation and performance estimation of first-order methods for convex optimization.* PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2017.

[Tay24] Adrien Taylor. Towards principled and systematic approaches to the analysis and design of optimization algorithms. *PSL Research University*, 2024.

[TB19] Adrien Taylor and Francis Bach. Stochastic first-order methods: non-asymptotic and computer-aided analyses via potential functions. In *Conference on Learning Theory*, pages 2934–2992. PMLR, 2019.

[TD23] Adrien Taylor and Yoel Drori. An optimal gradient method for smooth strongly convex minimization. *Mathematical Programming*, 199(1):557–594, 2023.

[THG17a] Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Exact worst-case performance of first-order methods for composite convex optimization. *SIAM Journal on Optimization*, 27(3):1283–1313, 2017.

[THG17b] Adrien B. Taylor, Julien M. Hendrickx, and François 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), pages 1278–1283. IEEE, 2017.

[THG17c] Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161:307–345, 2017.

[THG18] Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Exact worst-case convergence rates of the proximal gradient method for composite convex minimization. *Journal of Optimization Theory and Applications*, 178(2):455–476, 2018.

[Toi24] Philippe L. Toint. Examples of slow convergence for adaptive regularization optimization methods are not isolated. *arXiv preprint arXiv:2409.16047*, 2024.

[TVSL18] Adrien Taylor, Bryan Van Scoy, and Laurent Lessard. Lyapunov functions for first-order methods: Tight automated convergence guarantees. In *International Conference on Machine Learning*, pages 4897–4906. PMLR, 2018.

[UBTG25] Manu Upadhyaya, Sebastian Banert, Adrien B. Taylor, and Pontus Giselsson. Automated tight Lyapunov analysis for first-order methods. *Mathematical Programming*, 209(1):133–170, 2025.

[Val43] Frederick A. Valentine. On the extension of a vector function so as to preserve a Lipschitz condition. *Bulletin of the American Mathematical Society*, 49:100–108, 1943.

[VG24] Pierre Vernimmen and François Glineur. Convergence analysis of an inexact gradient method on smooth convex functions. In *ESANN 2024, European Symposium on Artificial Neural Networks, Computational Intellignece and Machine Learning*, 2024.

[VG25] Pierre Vernimmen and François Glineur. Empirical and computer-aided robustness analysis of long-step and accelerated methods in smooth convex optimization. *arXiv preprint arXiv:2506.09730*, 2025.

[VKK+25] Artem Vasin, Valery Krivchenko, Dmitry Kovalev, Fedyor Stonyakin, Nazari Tupitsa, Pavel Dvurechensky, Mohammad Alkousa, Nikita Kornilov, and Alexander Gasnikov. On solving minimization and min-max problems by first-order methods with relative error in gradients. *arXiv preprint arXiv:2503.06628*, 2025.

[VSFL17] Bryan Van Scoy, Randy A. Freeman, and Kevin M. Lynch. The fastest known globally convergent first-order method for minimizing strongly convex functions. *IEEE Control Systems Letters*, 2(1):49–54, 2017.

[VSL25] Bryan Van Scoy and Laurent Lessard. The fastest known first-order method for minimizing twice continuously differentiable smooth strongly convex functions. *IEEE Control Systems Letters*, 2025.

[VSSPL23] Bryan Van Scoy, John W. Simpson-Porco, and Laurent Lessard. Automated Lyapunov analysis of primal-dual optimization algorithms: An interpolation approach. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 1306–1311. IEEE, 2023.

[Vũ13] Bang Công Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 38(3):667–681, 2013.

[Wel73] John C. Wells. Differentiable functions on banach spaces with lipschitz derivatives. *Journal of Differential Geometry*, 8(1):135–152, 1973.

[Wri23] Stephen J. Wright. Optimization in machine learning and data science. *Collections*, 56(03), 2023.

[Yan18] Ming Yan. A new primal–dual algorithm for minimizing the sum of three functions with a linear operator. *Journal of Scientific Computing*, 76:1698–1717, 2018.

[YHL16] Peng Yi, Yiguang Hong, and Feng Liu. Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems. *Automatica*, 74:259–269, 2016.

[ZAdK24] Moslem Zamani, Hadi Abbaszadehpeivasti, and Etienne de Klerk. The exact worst-case convergence rate of the alternating direction method of multipliers. *Mathematical Programming*, 208(1):243–276, 2024.

[ZG23] Moslem Zamani and François Glineur. Exact convergence rate of the last iterate in subgradient methods. *arXiv preprint arXiv:2307.11134*, 2023.

[ZGH24] Moslem Zamani, François Glineur, and Julien M. Hendrickx. On the set of possible minimizers of a sum of convex functions. *IEEE Control Systems Letters*, 2024.

[ZHSJ19] Jingzhao Zhang, Tianxing He, Suvrit Sra, and Ali Jadbabaie. Why gradient clipping accelerates training: A theoretical justification for adaptivity. In *International Conference on Learning Representations*, 2019.