Dauphine | PSL
UNIVERSITÉ PARIS

University Paris-Dauphine

Master IASD

Final internship report

Work undertaken at Télécom Paris

Lower bounds and primal-dual methods for affinely constrained convex optimization under metric subregularity

Author:
Oskar Rynkiewicz

Supervisor:
Prof. Olivier Fercoq

September 2020

Contents

1 Introduction

1.1 Context, motivation, objectives, and contributions

In recent years, first-order optimization methods have gathered special attention due to their applicability in a large-scale setting. The sheer amount of data can be exploited only if the used algorithms are capable of efficient processing. One of the ways to certificate a method's performance is by establishing the number of iterations necessary to solve a given optimization problem up to some required accuracy. This leads to lower and upper complexity bounds which indicate the best and the worst performance for a given class of problems. Our work is set in the context of affinely constrained convex optimization problems:

$$\underset{x\in\mathcal{X}}{\text{minimize}}\, f(x) \text{ under constaints } Mx = b.$$

An equivalent formulation including the convex linear composite term is

$$\underset{x\in\mathcal{X}}{\text{minimize}}\, f(x) + \mathrm{I}_b(Mx),$$

where I_b is the indicator function of the singleton $\{b\}$, and f is a smooth, convex function. Another characteristic of the considered problem class is the assumption that the gradient of Lagrangian $\nabla\mathcal{L}(z)$ posses a property called metric subregularity, which describes the behavior of the points z around the solution set.

Problems of this type are a prototype of important problems in the field of Machine Learning, such as regularized empirical risk minimization, Support vector machine, and Machine Learning portfolio problems. The presence of the composite in the formulation indicates a specific, coupled structure which demands the use of the splitting methods, such as the primal-dual splitting algorithms.

First objectives consisted in the study of the primal-dual splitting methods proposed by Vu and Condat, as well as covered the application of the recent convergence analysis technique to establish a global linear convergence in the Chapter 3, in the Theorem 4.

Further objectives consisted of grasping the understating on the property of metric subregularity, and constructing specific problems whose constant of metric subregularity is known by an analytical formula. Then, we aimed to deliberately construct our difficult affine constaints in a way that algorithms take a long time to solve. Having proposed such problem, we developed a lower bound proof determining the lower bound complexity. The research question associated with these tasks is:

> For any first-order algorithm, what is the best possible performance on solving affinely constrained smooth convex optimization problems whose gradient of Lagrangian is metrically subregular?

Related primary contribution: In the Chapter 4, the Theorem 6 established a bound of the linear rate:

$$\left\|x^* - x^{(k)}\right\| \geq O\left(\left(\frac{\kappa}{\kappa + \sqrt{\mu\kappa + 1}}\right)^k\right).$$

Having obtained both the lower and upper bound, we evaluated these bounds together with the practical results of numerical experiments. In the Chapter 5, the Figure 1.1 compared the results and affirmed the presence of the gap.

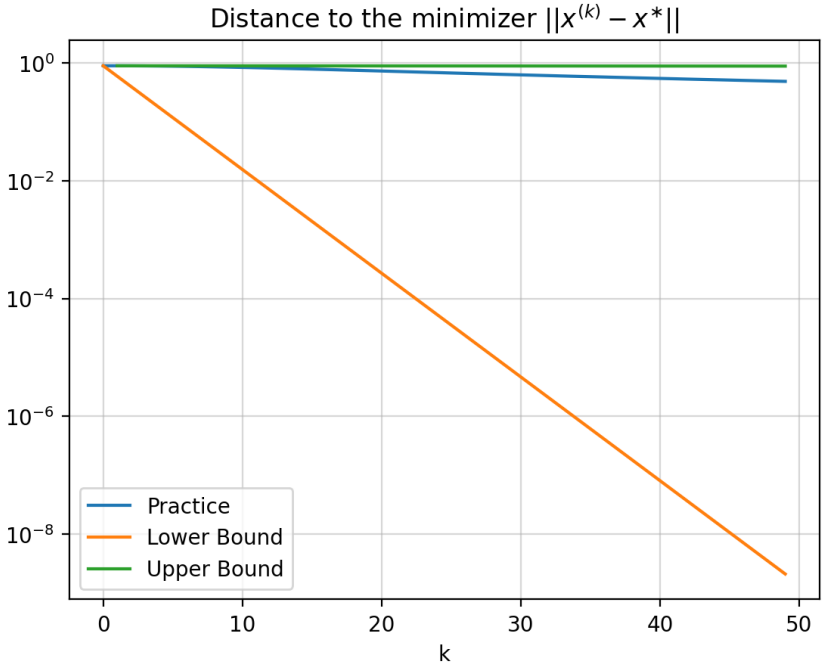

Figure 1.1: The comparison between the practical results and the theoretical bounds.

The organization of the report is as follows. In the remaining Section 1.2 of the introduction we include necessary preliminaries on convergence rates, oracle complexity, convex analysis, and operator theory. In the chapter 2 we introduce the notion of metric subregularity, formulate our optimization problem class in detail, and perform a small literature review. In the Chapter 3, we focus on the upper bound convergence and a primal-dual algorithm. A new complexity lower bound is proposed in Chapter 4. Finally, simple numerical experiments were performed in the Chapter 5.

1.2 Preliminaries

Denote $\mathcal{X}$ the real Euclidean space of dimension n and $\mathcal{Y}$ the real Euclidean space of dimension m. Both $\mathcal{X}$ and $\mathcal{Y}$ have the dot product $\langle \cdot, \cdot \rangle$ and the associated norm $\|\cdot\|$. As $\mathcal{X}$ is a bijection to Euclidean vector space $\mathbb{R}^n$ and $\mathcal{Y}$ is a bijection to $\mathbb{R}^m$ (in both cases taking the zero vector as the origin and the standard basis), we will use $\mathbb{R}^n$ and $\mathcal{X}$ (as well as $\mathbb{R}^m$ and $\mathcal{Y}$) interchangeably.

The identity matrix is denoted as I. A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is a Positive Semi Definite (PSD) matrix if and only if for all $x \in \mathbb{R}^n$, $x^\intercal Ax \geq 0$. For the strict inequality, the matrix is positive definite (PD) for any non-zero vector $x \in \mathbb{R}^n$.

Typically, the iteration complexity bounds for the first-order methods are non-asymptotic because these types of methods often terminate without high precision solution, sometimes before we can see the asymptotic behavior of the algorithm at its tail (after many iterations). Still, asymptotic convergence gives important information on the overall speed of the algorithm (sublinear, linear), and is used to describe methods' convergence rates. The non-asymptotic convergence is described in the Section 1.2.1. Here, we define 2 types of asymptotic convergence of a sequence: Q-convergence and R-convergence.

Definition 1 (Quotient convergence rate)**.** A sequence $\{x_k\}_{k\in\mathbb{N}}$ converges Q-linearly to $\bar{x}$ with rate r if there exists $r \in]0, 1[$ such that the quotient between two successive errors verifies

$$\frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} = r \tag{1.1}$$

For $r = 1$, the Q-convergence is sublinear. Even a single pair of slow successive terms suffices to break the rate of the whole otherwise fast sequence. A generalization of Q-convergence that alleviates this drawback is the R-convergence:

Definition 2 (Root convergence rate)**.** A sequence $\{x_k\}_{k\in\mathbb{N}}$ converges R-linearly to $\bar{x}$ if there exists a sequence $(\sigma_k)_{k\in\mathbb{N}}$ converging to 0 Q-linearly and $\|x_k - \bar{x}\| \leq \sigma_k$ holds for all $k \in \mathbb{N}$ (overall decrease of errors).

1.2.1 Information-based complexity

The black-box model is a basis for the information-based complexity theory that is used extensively in large-scale convex optimization. In such problems, a large number of dimensions n prohibits the use of methods that require costly iterations. A method that converges in a single iteration is useless if the large n renders the iteration insanely expensive. For instance, the standard Interior Point methods are typically unusable in the big data context as their arithmetic complexity is at least $O(n^2)$. In such a large-scale setting, we rather use the first order methods. While less accurate, these methods are dimensions-free, which amounts to a low iteration cost that doesn't explode in high dimensions.

For the first order methods, iterations are relatively cheap and we are primarily concerned with the number of iterations, rather than an arithmetic complexity of an iteration. In other words, we are focused on the amount of information about a problem that an algorithm needs in order to solve it.

An oracle is a proxy between the problem and a method. At each iteration k, an algorithm queries the oracle $\mathcal{O}$ with an iterate $x^{(k)}$ and gets information about the problem, which the method uses to form better $x^{(k+1)}$. With oracle, we can quantify the efficiency of iterative algorithms in terms of the number of iterations it requires (information transfer). As a result, we can bound these quantities, as well as compare the performances of different algorithms on various classes of problems.

Depending on the type of an oracle, it returns different types of information. The oracle of p order returns $f(x^{(k)})$ and the derivatives up to the p-th order (or subderivatives). For the first order methods, we use the first order oracle which outputs $f(x_k)$ and $\nabla f(x_k)$ for a differentiable f. Another type of information commonly included in oracle is the proximity operator, especially for non-smooth problems.

The information-based complexity theory can only work if the algorithm doesn't bypass the oracle. No prior knowledge about the structure of the problem is asserted by the following assumption:

Assumption 1 (Black-box)**.** In the black-box setting, the explicit form of the objective function f and the functional constraints are assumed unknown. The only way an algorithm can gain quantitative information about the problem is through the calls to the oracle.

We consider the algorithms in the context of some problem class consisting of a family of many specific problems that share some characteristics. An algorithm $\mathcal{A}$ is an iterative process that performs sequential calls to the oracle and accumulate information: at the i-th step, a method sends a search point x_i to the oracle which returns information at that point: e.g. $\mathcal{O}(x_i) = (f(x_i), \nabla f(x_i))$ for a first-order oracle and differentiable f. The informational set is updated with the data from the step. Then, a new search point is generated according to the method's search rule that depends on all the previous information, i.e. a call to the oracle is formed by a recurrence:

$$x_i = \mathcal{A}(\mathcal{O}(x_1), \ldots, \mathcal{O}(x_{i-1})) \tag{1.2}$$

The process repeats until some termination criterion $\mathcal{T}$ is satisfied, e.g. the just-formed point x is an ϵ approximate solution to the problem, i.e. $|f(\hat{x}) - f(x^*)| \leq \epsilon$ or $\|\hat{x} - x^*\| \leq \epsilon$ holds for an $\epsilon > 0$, and $\mathcal{A}$ terminates by outputting it.

We want this method to find the solution in a least number of iterations possible. The following definitions summarize all the described aspects in a mathematical way.

Definition 3 (Information based complexity)**.** The complexity $\mathrm{C}_{\mathcal{A}}(\epsilon, f)$ is the number of oracle calls performed by an algorithm $\mathcal{A}$ that are necessary to solve a specific problem with objective f up to accuracy ϵ:

$$\mathrm{C}_{\mathcal{A}}(\epsilon, f) = \min \{k = 1, 2, 3, \ldots \mid f(x_k) - f(x^*) \leq \epsilon\} \tag{1.3}$$

Rather than complexity for a particular problem $\mathrm{C}_{\mathcal{A}}(\epsilon, f)$, we look at the performance of a method on the whole problem class $\mathrm{C}_{\mathcal{A}}(\epsilon)$:

Definition 4 (Oracle complexity for $\mathcal{A}$ on $\mathcal{F}$)**.** The complexity $\mathrm{C}_{\mathcal{A},\mathcal{F}}(\epsilon)$ of $\mathcal{A}$ on the whole problem class with family of target functions $\mathcal{F}$ is the worst case number of oracle calls made by $\mathcal{A}$ on $f \in \mathcal{F}$:

$$\mathrm{C}_{\mathcal{A},\mathcal{F}}(\epsilon) = \max_{f \in \mathcal{F}} \{\mathrm{C}_{\mathcal{A}}(\epsilon, f)\} \tag{1.4}$$

That is, $\mathrm{C}_{\mathcal{A},\mathcal{F}}(\epsilon)$ is the least integer for which $\mathrm{C}_{\mathcal{A}}(\epsilon, f) \leq \mathrm{C}_{\mathcal{A},\mathcal{F}}(\epsilon)$ for all $f \in \mathcal{F}$. Hence, the oracle complexity gives the efficiency of an algorithm on a problem class, which is decided by method's performance on the worst problem in the whole problem class.

We can also have an information-based complexity of a class of problems $\mathcal{F}$, without the context of a given class of algorithms $\mathcal{A}$:

Definition 5 (Oracle complexity of $\mathcal{F}$)**.** The complexity $\mathrm{C}_{\mathcal{F}}(\epsilon)$ of a problem class with target functions $\mathcal{F}$ is determined by the lower bound on the complexities for all black-box algorithms $\mathcal{A}$, i.e. by the minimal complexity obtained while using the best possible (optimal) method:

$$\mathrm{C}_{\mathcal{F}}(\epsilon) = \min_{\mathcal{A}} \{\mathrm{C}_{\mathcal{A},\mathcal{F}}(\epsilon)\} \tag{1.5}$$

$$= \min_{\mathcal{A}} \max_{f\in\mathcal{F}} \{\mathrm{C}_{\mathcal{A}}(\epsilon, f)\} \tag{1.6}$$

$$= \min_{\mathcal{A}} \max_{f\in\mathcal{F}} \min_{k\in\mathbb{Z}^+} \{k = 1, 2, 3, \dots \mid f(\bar{x}) - f(x^*) \leq \epsilon\}\,. \tag{1.7}$$

The complexity of a class amounts to the minimal number of iterations needed by the optimal method on the worst function (particular problem) in a problem class.

Definition 6 (Optimal algorithm $\mathcal{A}$ for a problem class $\mathcal{F}$)**.** An algorithm $\mathcal{A}$ is optimal for some problem class $\mathcal{F}$ if $\mathcal{A}$'s complexity $\mathrm{C}_{\mathcal{A},\mathcal{F}}(\epsilon)$ matches the complexity of the problem class $\mathrm{C}_{\mathcal{F}}(\epsilon)$, i.e. at worst the algorithm has the minimal complexity for which all problems in the class are ϵ-solved. Put differently, the upper complexity bound of an algorithm equals with the lower complexity bound of a problem class.

Definition 7 (Lower complexity bound)**.** A lower complexity bound k of a problem class with objective family $\mathcal{F}$ is obtained when

$$\forall A \in \mathcal{A},\ \exists f \in \mathcal{F} \text{ such that } \mathrm{C}_{\mathcal{A}}(\epsilon, f) \geq k$$

i.e. find the most difficult function f which all algorithms fail to solve with less information than after k oracle calls.

Oracle complexity gives the required number of iterations, which depends on the desired accuracy ϵ (the distance to the optimum). Instead of giving k in terms of ϵ, we can give the accuracy ϵ of an algorithm after k oracle calls. Both formulations are equivalent and we can freely find one from the other. In the later formulation, it follows that instead of the worst case number of iterations needed to attain some accuracy ϵ, we talk about the worst case residual (maximal error) obtained by the algorithm on the problem class after k oracle calls.

Note that the information-based complexity is non-asymptotic concept, but also talk about the obtained results in terms of the asymptotic convergence. For instance, a lower bound with exponent amounts to the linear convergence rate.

1.2.2 Convex Analysis

The convex analysis version of the indicator function will be needed.

Definition 8 (Indicator function)**.** Indicator function $\mathrm{I}_A : \mathcal{X} \to \{0, +\infty\}$ of the set A is given by

$$\mathrm{I}_A(x) = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{else.} \end{cases} \tag{1.8}$$

If the set A in the Definition 8 is a singleton, we simplify the notation and write I_A as $x \in \{a\} \iff x = a$. In particular, we can write the indicator function of the set $\{0\}$ as $\mathrm{I}_0(x) = 0$ if $x = 0$ and $\mathrm{I}_0(x) = +\infty$ when $x \neq 0$ instead of $\mathrm{I}_{\{0\}}(x) = 0$ if $x \in \{0\}$ and $+\infty$ when $x \notin \{0\}$.

Definition 9 (L-Lipschitz continuous gradient)**.** A differentiable function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ has L-Lipschitz continuous gradient ∇f on X with Lipschitz constant $L > 0$ if for all $x, y \in X$ we have

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \tag{1.9}$$

$$\iff \quad |f(y) - f(x) - \langle \nabla f(x), y - x\rangle| \leq \frac{L}{2}\|y - x\|^2 \quad \text{(Descent Lemma).} \tag{1.10}$$

Differentiable functions whose gradient is L-Lipschitz continuous are also called L-smooth.

Definition 10 (Convex set)**.** A set $C \subseteq \mathbb{R}^n$ is convex if for any $x, y \in C$ and $\theta \in [0, 1]$:

$$\theta x + (1 - \theta)y \in C.$$

Example 1. Let $b \in \mathbb{R}^n$. The singleton $C = \{b\}$ is a convex set because for any pair $x, y \in C$, $x = y = b$ and $\theta x + (1 - \theta)y = b \in C$ for $\theta \in [0, 1]$.

Other examples include $\emptyset$, R^n, Euclidean balls, lines, hyperplanes, and halfspaces.

Definition 11 (Convex function)**.** A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if its domain X is a convex set and for all $x, y \in X$ and $\theta \in [0, 1]$,

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y). \tag{1.11}$$

Moreover, if f is differentiable, then we have

$$(1.11) \iff f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle \tag{1.12}$$

$$\iff \nabla^2 f(x) \succeq 0. \tag{1.13}$$

By $\Gamma_0(\mathcal{X})$ we denote the set of all closed proper l.s.c convex functions on $\mathcal{X} \to (-\infty, +\infty]$.

Example 2. The indicator function I_A is a convex function. To see that, apply the Definition 1.11 of convexity with $a, b \in \mathcal{X}$ and consider 4 possible combinations of the memberships of a, b in A.

Definition 12 (μ-strongly convex function)**.** A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex if X is a convex set and there exists a constant $\mu > 0$ such that for any $x, y \in X$ and $\theta \in [0, 1]$ it holds that:

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y) - \frac{\mu}{2}\theta(1 - \theta)\|x - y\|^2. \tag{1.14}$$

Moreover, if f is differentiable, then we have

$$(1.14) \iff f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|x - y\|^2 \tag{1.15}$$

$$\iff \nabla^2 f(x) \succeq \mu I_n. \tag{1.16}$$

Example 3 (Function $\frac{\mu}{2}\|x\|^2$ is strongly convex)**.**

$$\begin{aligned}
&f(\theta x + (1-\theta)y) - \theta(f(x) - (1-\theta)f(y) \\
&= \mu\theta^2||x||^2 + 2\mu\theta(1-\theta)\langle x, y\rangle + \mu(1-\theta)^2||y||^2 - \mu\theta||x||^2 - \mu(1-\theta)||y||^2 \\
&= \mu\theta(\theta-1)||x||^2 + \mu(1-\theta)(1-\theta-1)||y||^2 + 2\mu\theta(1-\theta)\langle x, y\rangle \\
&= -\mu\theta(1-\theta)||x||^2 + 2\mu\theta(1-\theta)\langle x, y\rangle - \mu\theta(1-\theta)||y||^2 \\
&= -\mu\theta(1-\theta)(||x||^2 - 2\langle x, y\rangle + ||y||^2) \\
&= -\mu\theta(1-\theta)||x-y||^2 \\
&\leq -\frac{\mu}{2}\theta(1-\theta)||x-y||^2
\end{aligned} \quad . \tag{1.17}$$

Proposition 1. A function $f : \mathcal{X} \to (-\infty, +\infty]$ is μ-strongly convex with some $\mu > 0$ if and only if $f - \frac{\mu}{2}\|\cdot\|^2$ is convex.

Proposition 2. A sum of an μ-strongly convex function and a convex function is a strongly convex function. To see that, add the inequalities of a strongly convex $g(x)$ and non-strongly convex $h(x)$ and equate them to the strongly convex $f(x)$.

Definition 13 (Subdifferential)**.** Let $f : \mathcal{X} \to [-\infty, +\infty]$. The subdifferential $\partial f : \mathcal{X} \to 2^{\mathcal{X}}$ is the set of the subgradients g of f at some $x \in \mathrm{dom}(f)$:

$$\partial f(x) = \{g \in \mathcal{X} \mid f(y) \geq f(x) + \langle g, (y-x)\rangle, \forall y \in \mathrm{dom}(f) \subseteq \mathcal{X}\}.$$

Example 4. Subdifferential of the indicator function at $\{0\}$ maps to the whole space $\mathcal{G}$ if $x = 0$ or to the empty set otherwise:

$$\partial \mathrm{I}_0(x) = \begin{cases} \mathcal{G} & \text{if } x = 0 \\ \emptyset & \text{else.} \end{cases} \tag{1.18}$$

This can be seen as $\partial \mathrm{I}_0(x) = \{g \mid \mathrm{I}_0(y) \geq \mathrm{I}_0(x) + \langle g, (y-x)\rangle, y \in \{0\}\} = \{g \mid \langle g, x\rangle \geq \mathrm{I}_0(x)\}$, which for $x = 0$ gives $\{g \mid \langle g, 0\rangle \geq 0\} = \mathcal{G}$ and for $x \neq 0$ is $\{g \mid \langle g, x\rangle \geq +\infty\} = \emptyset$.

Definition 14 (Fenchel conjugate)**.** Conjugate f^* of a function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is a convex function defined as:

$$f^*(y) = \sup_{x \in X}\{\langle y, x\rangle - f(x)\} = -\inf_{x \in X}\{f(x) - \langle y, x\rangle\}.$$

Remark. Definition 14 amounts to Legendre transform if f is convex and univariate, neither of which is necessary to take a Fenchel conjugate of f. Since Fenchel conjugate is a generalization of Legendre transform, conjugate f^* is also called Legendre-Fenchel transformation.

Following the work of Moreau [Moreau 1962], an important operator was defined:

Definition 15 (Proximity operator)**.** We define the proximity operator of $f \in \Gamma_0(\mathcal{X})$ for every $x \in \mathcal{X}$ as

$$\mathrm{prox}_f(x) = \underset{y \in \mathcal{X}}{\mathrm{argmin}}\, f(y) + \frac{1}{2}\|y - x\|^2. \tag{1.19}$$

The optimization problem whose solution is defined as proximity operator is always unique provided that f is proper and convex. We say that some function is proximable if its proximal operator can be easily computed, ideally with a closed-form solution.

Example 5 (Prox of the indicator function)**.** The proximal of an indicator function $\mathrm{I}_C : \mathcal{X} \to \{0, +\infty\}$ is:

$$\mathrm{prox}_{I_C}(x) = \operatorname*{argmin}_{y \in \mathcal{X}} \mathrm{I}_b(y) + \frac{1}{2}\|y - x\|^2 = \operatorname*{argmin}_{y \in C} \|y - x\|^2 \tag{1.20}$$

Hence, the proximal operator reduces to Euclidean projection onto C. Indeed, proximal operators can thus be interpreted as generalized projections [Parikh and Boyd 2014]. In the case when C is a singleton, e.g. $C = \{b\}$, then there is no calculation to be made: $\operatorname{argmin}_{y \in \{b\}} \|y - x\|^2$ has single possible vector b to be the proximal of I_b hence $\mathrm{prox}_{I_b} = b$.

1.2.3 Operator theory

Operator theory serves well to generalize convex analysis into set-valued analysis. A set-valued operator A on $\mathcal{X}$ is denoted as $A : \mathcal{X} \to 2^{\mathcal{X}}$. It's essentially a multivalued function with codomain identical to its domain. An operator A maps a point $x \in \mathcal{X}$ to a selection of points in the same space $Ax \subseteq \mathcal{X}$, which we call the image of $x \in \mathcal{X}$ by an operator A, i.e. $Ax = \{y \in \mathcal{X} | y \in Ax\}$. When the cardinality of this subset of $\mathcal{X}$ is at most 1, (Ax is either an empty set or a singleton), we get a special case of a single-valued operator $\mathcal{X} \to \mathcal{X}$.

Graph of an operator is the subset of $\mathcal{X} \times \mathcal{X}$ that identifies (characterizes) the operator: $\mathrm{gra}A = \{(x, y) \in \mathcal{X} \times \mathcal{X} : y \in Ax\}$. The domain and image (range) of an operator are defined as $\mathrm{dom}(A) = \{x \in \mathcal{X} : Ax \neq\}$ and $\mathrm{im}(A) = \bigcup_{x \in \mathcal{X}} Ax$, respectively.

Example 6 (The graph of the operator $\partial \mathrm{I}_0$)**.** The subdifferential $\partial \mathrm{I}_0$ is a set-valued operator which maps to either whole space $\mathcal{G}$ or an empty set $\emptyset$ (Example 4). Its graph can be written:

$$\begin{aligned} \mathrm{gra}(\partial \mathrm{I}_0) &= \{(x, y) : y \in \partial \mathrm{I}_0(x)\} \\ &= \{(x, \emptyset), (0, y) : x \in \mathcal{G} \setminus \{0\} \text{ and } y \in \mathcal{G}\} \end{aligned}$$

The inverse of an operator A is the operator A^{-1} which verifies $A^{-1}y = \{x \in \mathcal{X} \mid y \in Ax\}$, i.e. whose graph is required to be

$$\begin{aligned} \mathrm{gra}(A^{-1}) &= \{(x, y) \in \mathcal{X} \times \mathcal{X} : (y, x) \in \mathrm{gra}(A)\} \\ &= \{(x, y) \in \mathcal{X} \times \mathcal{X} : x \in Ay\}. \end{aligned}$$

Example 7 (The inverse of the operator $\partial \mathrm{I}_0$)**.**

$$\begin{aligned} \mathrm{gra}(\partial \mathrm{I}_0^{-1}) &= \{(x, y) : (y, x) \in \mathrm{gra}(\partial \mathrm{I}_0)\} \\ &= \{(\emptyset, x), (y, 0) : x \in \mathcal{G} \setminus \{0\} \text{ and } y \in \mathcal{G}\} \\ &= \{(y, 0) : y \in \mathcal{G}\} \end{aligned}$$

In other words, the operator $\partial \mathrm{I}_0^{-1}$ is a null (multi)function that collapses any element from $\mathcal{G}$ to 0.

Another characteristic of an operator is its set of zeros: $\operatorname{zer}(A) = \{x \in \operatorname{dom}(A) : 0 \in Ax\} = \{x \in \operatorname{dom}(A) : (x, 0) \in \operatorname{gra}(A)\} = A^{-1}(\{0\})$. The sum $A + B$ of A and B is the operator $(A + B)(x) = \{y + z : y \in Ax, z \in Bx\}$ when $x \in \operatorname{dom}(A) \cup \operatorname{dom}(B)$, and $\emptyset$ elsewhere. Finally, the composition AB of operators A and B is the operator AB whose graph is required to be:

$$\begin{aligned}\operatorname{gra}(AB) &= \{(x, z) \in \mathcal{X} \times \mathcal{X} : \exists y \in \mathcal{X}, y \in Bx, z \in Ay\} \\ &= \{(x, z) \in \mathcal{X} \times \mathcal{X} : \exists (x, y) \in \operatorname{gra}(B), (y, z) \in \operatorname{gra}(A)\}.\end{aligned}$$

Definition 16 (Monotone Operator)**.** An operator A is monotone if for any pair of points x, y from the domain of A, and for all points $u \in Ax$ and $v \in Ay$, it holds that

$$\langle x - y, u - v\rangle \geq 0 \tag{1.21}$$

Definition 17 (Maximal monotone operator)**.** A monotone operator A is maximal there is no operator B such that $\operatorname{gra}(A) \subset \operatorname{gra}(B)$, i.e. if its graph is a maximal element in the graph inclusion ordering.

Definition 18 (Resolvent of an operator A)**.**

$$J_A = (\mathrm{I} + A)^{-1} \tag{1.22}$$

The resolvent can be viewed as a generalization of prox for monotone operators. An operator R is non-expansive if its Lipschitz constant $L = 1$. For $L < 1$, R is a contraction.

Definition 19 (Averaged operator)**.** Operator A is α-averaged if there exists a non-expansive operator R such that $T = \alpha R + (1 - \alpha)\mathrm{I}$

An averaged operator is firmly non-expansive if its $\frac{1}{2}$-averaged. Given $\alpha > 0$, a single-valued operator A is α -cocoercive if αA is firmly-nonexpansive.

Lemma 1 (Composition of averaged operators)**.** *[Ogura and Yamada 2002, Theorem 3] Let $\alpha_1 \in]0, 1[, \alpha_2 \in]0, 1], T_1 \in \mathscr{A}(\alpha_1)$ and $T_2 \in \mathscr{A}(\alpha_2)$. Then $T_1 \circ T_2 \in \mathscr{A}(\alpha')$, where*

$$\alpha' := \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2} \tag{1.23}$$

Definition 20 (Parallel sum)**.** Let A and B be operators $\mathcal{X} \to 2^{\mathcal{X}}$. The parallel sum of A and B is

$$A \square B = \left(A^{-1} + B^{-1}\right)^{-1}. \tag{1.24}$$

There is a connection between the proximity operator and the resolvent of a subdifferential:

Proposition 3. Let $J_{\lambda\partial f}$ be the resolvent of the subdifferential operator ∂f for some $\lambda > 0$, i.e. $J_{\lambda\partial f} = (\mathrm{Id} + \lambda\partial f)^{-1}$. Then,

$$\operatorname{prox}_{\lambda f} = J_{\lambda\partial f}$$

Theorem 1 (Baillon–Haddad)**.** *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a differentiable convex function. Then, ∇f is L-Lipschitz continuous ($L\nabla f$ is nonexpansive) if and only if ∇f is L^{-1}-cocoercive ($L^{-1}\nabla f$ is firmly non-expansive):*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \iff \langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2$$
$$\forall x, y \in \mathrm{dom} f \tag{1.25}$$

Remark. In particular case of $L = 1$, a gradient map is nonexpansive if and only if it's firmly nonxpansive.

2 Problem formulation and related work

In this chapter, we begin with the presentation of metric subregularity. This property has entered the realm of research in optimization only recently, hence we choose to establish this notion first. Then, we fully describe the considered class of optimization problems. Lastly, we will conduct the literature review on first-order optimization algorithms and lower complexity bounds.

2.1 Metric subregularity

The concept of metric regularity has its roots in the work of Hoffman on the estimation of the distance to the set of solutions in a system of linear inequalities [Hoffman 1952]. This idea was generalized into the *regularity at a point* [Ioffe 1979], which is analogous to the metric subregularity [Dontchev and Rockafellar 2009, Section 3H]. The basic idea is that if a map is subregular at some point $\bar{x}$, then we could estimate the distance between points x nearby $\bar{x}$ and the solution set by other distances that can be calculated more easily. We will provide more details, but first let us introduce the definition of the property of metric subregularity:

Definition 21 (Metric subregularity)**.** A set-valued operator $F : \mathcal{X} \to 2^{\mathcal{X}}$ is metrically sub-regular at $\bar{x}$ for $\bar{y}$ with $(\bar{x}, \bar{y}) \in \mathrm{gra}F$ if there exists metric subregularity constant $\kappa \geq 0$ with a neighborhood $\mathcal{N}(\bar{x})$ such that

$$\mathrm{dist}(x, F^{-1}\bar{y}) \leq \kappa \, \mathrm{dist}(\bar{y}, Fx) \qquad \forall x \in \mathcal{N}(\bar{x}). \tag{2.1}$$

The case of $\kappa = +\infty$ translates to the lack of metric subregularity.

The mapping F is set-valued, which makes the metric subregularity a bit complex to visualize. Note that $F^{-1}\bar{y}$ is a set (many points x may map to $\bar{y}$), and for each point x in the given neighborhood $\mathcal{N}(\bar{x})$, we have some set Fx. Hence, quantities $d(x, F^{-1}\bar{y})$ and $d(\bar{y}, Fx)$ are the distances between a point and a set according to

$$d(a, A) = \inf \{\|x - a\| \ : \ a \in A\}.$$

Metric subregularity asserts an existence of a finite constant κ such that the "error distance" $d(x, F^{-1}\bar{y})$ can be bounded (estimated) by the means of a distance $d(\bar{y}, Fx)$. This assertion is equivalent to the ruling out of the critically different behavior of F at points around $\bar{x}$ for $\bar{y}$. It must stay subregular, and the more subregular it is, the smaller the constant κ and better the error bound.

In our work, a concrete case of the Definition 21 of metric subregularity will be uniquely considered: We suppose a single-valued operator $F : \mathcal{X} \to \mathcal{X}$, which means that Fx is at

most a singleton set for any x and we can write $\|\bar{y} - Fx\|$ instead of $d(\bar{y} - Fx)$. Moreover, we consider metric subregularity at x^* for 0:

$$d(x - F^{-1}(0)) \leq \kappa \|Fx\| \qquad \forall x \in \mathcal{N}(x^*).$$

Furthermore, assuming that the solution set $F^{-1}(0)$ contains an unique solution x^*, we get a highly useful estimate of the distance from the solution:

$$\|x - x^*\| \leq \kappa \|Fx\| \qquad \forall x \in \mathcal{N}(x^*) \qquad \forall x \in \mathcal{N}(x^*).$$

As noted in our Problem 1, we associate metric subregularity to the gradient of Lagrangian, hence the operator is $F = \nabla \mathcal{L}$. At last, the form of metric subregularity that will be of our prime interest is

$$\|x - x^*\| \leq \kappa \|\nabla \mathcal{L}(x)\|, \tag{2.2}$$

which essentially amounts to the error bounds property [Kruger 2015], [Drusvyatskiy and Lewis 2018].

In general, the constant of metric subregularity can be called in more descriptive fashion as the modulus of subregularity of F at x for y and denoted $\mathrm{subreg}(F; x \mid y)$. In our case, we will only talk about metric subregularity for $\nabla \mathcal{L}$ at x^* for 0, hence we just use κ.

Checking if a particular problem is metrically subregular is a difficult task in general. Fortunately, for affinely constrained problems we know already that the problem is metrically subregular [Hoffman 1952]. Even more so, in this case the metric subregularity is global, i.e. $\mathcal{N}(x^*) = \mathbb{R}^n$.

Another 2 examples of metrically subregular problems are:

- Global metric subregularity holds when $f + g$ and h^* in Problem 2 are strongly convex [Latafat, N. Freris, and Patrinos 2017, Lemma IV.2]
- Local metric subregularity holds when f, g, h in Problem 2 are piecewise linear quadratic (PLQ) functions [Latafat, N. Freris, and Patrinos 2017, Lemma IV.4]

Some works has been done towards establishing the conditions sufficient for metric subregularity [Bai, Ye, and J. Zhang 2019], but this remains an open problem.

2.2 Problem formulation

2.2.1 Problem class

We are interested in the class of affinely constrained convex minimization problems. That is, minimization of a convex function $f(x)$ under constraints that $Mx = b$. We write these constraints into the objective function and state our optimization problem:

Problem 1 (Affinely constrained convex optimization under metric subregularity)**.** *Let $f : \mathcal{X} \to (-\infty, +\infty]$ be a proper, lower semicontinuous, convex function that is differentiable on $\mathcal{X}$ and whose gradient ∇f is L-Lipschitz continuous for some $L > 0$. Let $b \in \mathcal{Y}$ and let $M : \mathcal{X} \to \mathcal{Y}$ be a linear operator. Let $\mathrm{I}_b \in \Gamma_0(\mathcal{Y})$ be the indicator function*

at $\{b\}$*. Moreover, require that the gradient of the problem's Lagrangian* $\nabla\mathcal{L}$ *is metrically subregular with some finite positive constant* κ*. The problem is given by:*

$$\min_{x\in\mathcal{X}} f(x) + \mathrm{I}_b(Mx) \tag{2.3}$$

We look for a point x which verifies m equality constrains $Mx - b = 0$ and minimizes f. Note that the i-th constraint is given by $M_{i:}x - b_i = 0$, where $M_{i:} \in \mathbb{R}^{1\times n}$ is the i-th row in the matrix $M \in \mathbb{R}^{m\times n}$ and b_i is the i-th component of the vector b. We assume that the solution $x^* \in \mathbb{R}^n$ exists.

We will relate our Problem 1 to a generic template problem of minimizing the sum of three convex terms: a smooth function, a proximable function, and the composite of a proximable function with a linear operator:

Problem 2 (Convex optimization involving Lipschitzian, proximable and linear composite terms)**.** *Let* $f : \mathcal{X} \to (-\infty, +\infty]$ *be convex function that is differentiable on* $\mathcal{X}$ *and whose gradient is* L*-Lipschitz continuous for some* $L > 0$*. Let* $g : \mathcal{X} \to (-\infty, +\infty]$ *and* $h : \mathcal{Y} \to (-\infty, +\infty]$ *be convex, non-differentiable, proximable functions. Let* $M : \mathcal{X} \to \mathcal{Y}$ *be a linear operator. The problem is given by:*

$$\min_{x\in\mathcal{X}} f(x) + g(x) + h(Mx)$$

We make a brief comparison between our Problem 1 and the Problem 2:

- The general smooth term f is present in both formulations. The non-smooth, proximable term g is absent in our problem. In both problems we have a linear operator M.
- Our problem has the non-differentiable, convex, proximable term h, but we set $h = \mathrm{I}_b$ and never consider other forms of h. Our $h = \mathrm{I}_b$ is indeed convex (Example 2) and its proximity operator is indeed simple as $\mathrm{prox}_{\mathrm{I}_b} = b$ (closed-form representation, see Example 5).
- We acknowledge the presence of the convex linear composite term $h \circ M$ in our problem. The composition $\mathrm{I}_b \circ M$ couples the primal and dual problems. This is crucial characteristic which points us towards the use of primal-dual algorithms
- Our requirement of metrically subregular $\nabla\mathcal{L}$ is absent in the general Problem 2. The exclusion of instances without metric subregularity (problems for which the constant of metric subregularity is ∞) narrows down the class of considered problems.

Note that f could be a sum of terms, as long as they are all convex, differentiable and their gradients are L-Lipschitz continuous. For instance, a valid particular problem would be affinely constrained convex minimization of a sum of smooth f with a smooth regularizer $\|x\|_2^2$. Hence, despite omitting the non-smooth term g, we could still have regularization.

The Problem 1 requires only convexity of f, strong convexity isn't required. Strong convexity is a generalization of convexity. In the context of unconstrained minimization, strong convexity has a major impact on the complexity of the problem, enabling linear rates instead of sublinear. The affine constraints complicate the problem and strong

convexity is no longer guarantee of linear rate. Still, we emphasize that strongly convex structure of f isn't required, but permitted in the framework of our model and it will be relevant to this work.

There are weaker conditions than strong convexity, but stronger than convexity. An example of such in-between structure is the error bound property which allows linear convergence in the unconstrained optimization. A generalization of the bound property is metric subregularity. In our Problem 1 we require that the gradient of Lagrangian $\nabla\mathcal{L}$ is metrically subregular. We wish to use the metric subregular structure as prime feature to obtain linear convergence in affinely constrained convex optimization.

2.2.2 Primal, dual, and minimax formulations

Our primal optimization problem corresponds directly to the Problem 1:

$$\min_{x \in \mathcal{X}} f(x) + \mathrm{I}_b(Mx) \tag{2.4}$$

and we have the primal function

$$\mathcal{P}(x) = f(x) + \mathrm{I}_b(Mx). \tag{2.5}$$

The Lagrangian function $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to [-\infty, +\infty]$ associated with the primal problem can be found by noting that the vector of equality constraints is $g(x) = Mx - b = 0$. Then

$$\mathcal{L}(x, y) = f(x) + \langle y, Mx - b \rangle , \tag{2.6}$$

where $x \in \mathcal{X}$ is the primal variable and $y \in \mathcal{Y}$ is the dual variable (Lagrange multiplier vector). The (Lagrange) dual function $\mathcal{D}(y)$ is:

$$\mathcal{D}(y) = \inf_{x \in X} \mathcal{L}(x, y) = \inf_{x \in X} \{f(x) + \langle y, Mx - b \rangle\} = -\langle b, y \rangle - f^*(-M^\intercal y),$$

where $f^*(y)$ denotes the Fenchel conjugate of f (Definition 14). Briefly, the dual function is

$$\mathcal{D}(y) = -f^*(-M^\intercal y) - \langle b, y \rangle . \tag{2.7}$$

The corresponding dual optimization problem is:

$$\max_{y \in \mathcal{Y}} -f^*(-M^\intercal y) - \langle b, y \rangle . \tag{2.8}$$

Remark. The functions (2.5), (2.6), and (2.7) could be alternatively obtained by setting $g = \mathrm{I}_b$ in the classic Fenchel-Rockafellar duality framework [Bauschke and Combettes 2011, Definition 15.19].

The minimax problem corresponding to the Problem 2 is

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x) + g(x) - h^*(y) + \langle Mx, y \rangle .$$

Here, the minimax problem corresponding to our primal and dual is

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x) - \mathrm{I}_b^*(y) + \langle Mx, y \rangle .$$

That is,

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x) - \langle b, y \rangle + \langle Mx, y \rangle \tag{2.9}$$

as the Fenchel conjugate of the indicator function $h = \mathrm{I}_b$ is a linear function $h^*(y) = \mathrm{I}_b^* = \langle b, y \rangle$. We define the minimax function

$$\mathcal{M}(x, y) = f(x) - \langle b, y \rangle + \langle Mx, y \rangle. \tag{2.10}$$

In the primal problem (2.4), there is an $\mathrm{I}_b(Mx)$ term, which is non-differentiable. In the minimax formulation, this linear composition term is no longer present. Advantageously, all the terms in the minimax problem are differentiable.

The primal function $\mathcal{P}(x)$ is convex because f was assumed convex in the Problem 1. The dual function $\mathcal{D}(y)$ is concave. Hence, the minimax function $\mathcal{M}(x, y)$ is convex-concave.

2.2.3 Monotone inclusions formulation

To find the primal solution of (2.4) and the dual solution of (2.8) (or jointly, solution to (2.9)), we can search for saddle points (x^*, y^*) of the Lagrangian because upon strong duality such x^* is the primal solution and y^* is the dual solution.

We search for the saddle points of the Lagrangian. By the optimality condition, the saddle point of $\mathcal{L}(x, y)$ is found by setting the gradient of the Lagrangian to zero. First, with respect to the primal variable:

$$\begin{aligned} \nabla_x \mathcal{L}(x, y) &= 0 \\ \nabla_x f(x) + \nabla_x \langle y, Mx - b \rangle &= 0 \\ \nabla_x f(x) + \nabla_x \langle Mx, y \rangle &= 0 \\ \nabla_x f(x) + M^* y &= 0 \end{aligned}$$

Where M^* is the adjoint of M, i.e. $M^\intercal$. Second, with respect to the dual variable:

$$\begin{aligned} \nabla_y \mathcal{L}(x, y) &= 0 \\ \nabla_y \langle y, Mx - b \rangle &= 0 \\ Mx - b &= 0. \end{aligned}$$

Hence, looking for (x, y) such that $\nabla_x \mathcal{L}(x, y) = 0$ and $\nabla_y \mathcal{L}(x, y) = 0$ amounts to solving:

$$\begin{aligned} \nabla_x f(x) + M^* y &= 0 \\ Mx - b &= 0 \end{aligned}.$$

We will rewrite these formulation as an inclusion problem.

For the first line, we decomposed $\nabla_x \mathcal{L}(x, y)$ into 2 terms ∇f and $M^\intercal y$. Since f is differentiable, the subdifferential $\partial f(x)$ is a singleton, i.e. for any x in the Euclidean space $\mathcal{X}$, the set-mapping given by the subdifferential $\partial f(x)$ has a single element: $\partial f(x) = \{\nabla f(x)\}$. We can claim that

$$\nabla f(x) + M^* y = 0 \text{ is equivalent to the inclusion } 0 \in \partial f(x) + M^* y$$

because $\partial f(x) + M^*y = \{u + M^*y \mid u \in \partial f(x)\} = \{\nabla f(x) + M^*y\}$ as $\partial f(x) = \{\nabla f(x)\}$. Hence, $\nabla f(x) + M^*y = 0 \iff 0 \in \partial f(x) + M^*y$.

For the second line, we state $Mx - b = 0 \iff 0 \in Mx - \{b\}$. Remark that $\partial \mathrm{I}_b^*(y) = \{b\}$.

Therefore, the corresponds inclusion problem becomes

$$\text{find } (x^*, y^*) \in \mathcal{X} \times \mathcal{Y} \text{ such that } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial f(x^*) + M^*y^* \\ Mx^* - \{b\} \end{pmatrix} \tag{2.11}$$

A pair (x^*, y^*) that satisfies the inclusion problem (2.11) is the saddle point of the Lagrangian, x^* is a primal solution, y^* is a dual solution, and strong duality holds. Hence, a method that resolves the inclusion (2.11) amounts to solving the original Problem 1. A major benefit of the inclusion formulation (2.11) is that the composition $h \circ M$ has been split.

2.3 Primal-dual splitting optimization algorithms

For the review on the first-order convex optimization algorithms in unconstrained and set-constrained setting good sources are [Bubeck 2015] and [Nesterov 2018]. Here, we will only review the relevant class of splitting methods, with the focus on the primal-dual splitting methods. An interesting relevant paper is [Necoara, Nesterov, and Glineur 2019], where linear convergence of first order methods in the non-strongly convex case is discussed.

A large-scale convex optimization problems can often be solved efficiently using full splitting methods, i.e. the terms in the problem are splitted into subproblems and the involved operators are used separately. The approach with operators is especially desirable for non-smooth optimization.

Splitting methods are iterative methods that evaluate the operators available in the problem separately. Some examples of these methods are: Forward–Backward splitting method (FBS) [Combettes and Wajs 2005] which generalizes the proximal gradient algorithm, Douglas–Rachford splitting method (DRS) [Eckstein and Bertsekas 1992]: Alternating Direction method of Multipliers (ADMM) [Boyd, Parikh, and Chu 2011].

For our Problem 1 an algorithm has three operators available: gradient operator, proximable operator, and linear operator. As mentioned earlier, f this class has a linear composite term $\mathrm{I}_b \circ M$ which couples primal variable $x \in \mathcal{X}$ with the dual variable $y \in \mathcal{Y}$. This is an obstacle in achieving full splitting because the composite $\mathrm{I}_b \circ M$ couples the operators. This coupling would complicate the workings of the algorithm by introducing implicit inefficient operations like the inverse of a linear operator. Indeed, some optimization tasks can't be immediately solved with the efficient splitting approach.

Oftentimes, splitting can be attained by exploiting the dual structure of the problem. Primal-dual methods achieve full splitting and concurrently solve the primal problem and dual problem [Komodakis and J. Pesquet 2015], thus providing the dual solution in addition to the primal solution. Fortunately, in our case we can use the dual formulation and primal-dual splitting methods to uncouple I_b and M. Therefore, full splitting is achievable and we can potentially leverage all the three operators individually.

Important examples of methods from the class of primal-dual algorithms are: the foundational primal-dual hybrid gradient algorithm (PDHG) [Chambolle and Pock 2011b], Condat [Condat 2013], and Vu [Vu 2013].

Vu's method will be described in detail in the Chapter 3. Condat's method is actually a very similar algorithm to Vu's one. Both researchers proposed the methods independently, while using similar approach and the same proof strategy. Condat's article is slightly more practical: the considered problem belongs to the class like the Problem 2, while Vu took a wider class of problems and took a more abstract approach with more heavy use of operators. We will be mostly working with Vu's method, but both methods are similar enough, that are sometimes being called Condat-Vu (perhaps even Vu-Condat).

Lastly, there exist a body of work on stochastic derivatives of these algorithms. A notable few recent examples are: Stochastic PDHG [Chambolle, Ehrhardt, et al. 2018; Alacaoglu, Fercoq, and Cevher 2019], PURE-CD [Alacaoglu, Fercoq, and Cevher 2020], TriPD [Latafat, N. M. Freris, and Patrinos 2019].

2.4 Lower complexity bounds literature

In the Chapter 4 we will propose a new lower bound. Here, we review the iteration lower complexity bounds established in the literature. All of these results are placed in the framework of the black-box optimization with an oracle, which was described in the Section 1.2.1. The concept of the lower bound from the Definition 7 is especially important.

2.4.1 Unconstrained optimization problems

In the simplest case of convex optimization without constraints, the classical results indicate a sublinear lower complexity bound in the non-strongly convex, smooth case, and linear in the case of strongly convex, smooth optimization problems.

The most common proofs of lower bounds are based on constructing the function which is difficult in terms of the information transfer. For instance, to show the lower bound in the smooth, strongly convex case, one chooses a smooth, strongly convex function which reveals very little information for each oracle call. This way, the obtained number of iterations required to minimize this function constitutes a lower bound of number of iterations that will be needed to solve any other (easier) function in this class. Furthermore, once we determined the lower bound, it can serve as a benchmark for algorithms. If a given method attains the ϵ-solution in the same order of iterations as lower bound indicates, it guarantees that no other method can perform better for any function in the considered class of problems (i.e. this method is optimal).

An example of an *information-difficult* smooth strongly-convex function is:

$$f(x) = \frac{1}{2}(x_1 - 1)^2 + \frac{1}{2}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2 + \frac{\mu}{2}\sum_{i=1}^{n} x_i^2 \tag{2.12}$$

where $\mu > 0$. We can check any algorithm's performance on it. Note that this function is a quadratic which we can solve directly in a straightforward way. The difficulty of

(2.12) doesn't imply that it's difficult to solve analytically, but that it's difficult for an iterative, first-order black-box algorithm that verify the Assumption 2. Typically in the lower bound proofs [Nesterov 2018], we also impose the linear span assumption:

Assumption 2. For all $k = 1, \ldots, n$, the point $x^{(k)}$ generated by an algorithm $\mathcal{A}$ is a linear combination of $x^{(0)}, \nabla f\left(x^{(0)}\right), \nabla f\left(x^{(1)}\right), \ldots, \nabla f\left(x^{(k-1)}\right)$, i.e.

$$x^{(k)} \in x^{(0)} + \operatorname{span}\left\{\nabla f\left(x^{(\ell)}\right)\right\} \quad \forall \ell \in [0, k).$$

This assumption narrows down the class of first-order algorithms in which the bounds hold. However, most algorithms do verify this assumption. The lower bound analysis could be performed without the linear span assumption altogether, by using techniques such as ones in [Nemirovsky 1992], but we won't follow this direction.

The gradient of the particular function (2.12) is $(A + \mu \mathrm{I})x - e_1$, where A (Hessian of f) turns out to be a specific matrix with a tridiagonal structure:

$$A = \begin{bmatrix} 2 & -1 & 0 & \cdots & \cdots & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & -1 & 2 & -1 \\ 0 & \cdots & \cdots & \cdots & 0 & -1 & 1 \end{bmatrix} \tag{2.13}$$

This brings us to the bottom of the informational difficulty of f. Whenever an oracle provides information $\mathcal{O}(x^{(k)})$, the gradient $\nabla f(x^{(k)})$ furnished by an oracle has at most 3 non-null components. Given that most of the coordinates are null, the vector $\nabla f(x^{(k)})$ won't contribute towards finding the corresponding components in the next iterate $x^{(k+1)}$, thus slowing down the convergence to the solution x^*.

We will now show an important Lemma regarding the iterates produced by any method satisfying the Assumption 2 during the optimization of f (2.12).

Lemma 2. *Let $x^{(0)} = 0 \in \mathbb{R}^n$. Any k-th iterate $x^{(k)}$ produced by an algorithm $\mathcal{A}$ satisfying the Assumption 2 has at most k non-null components, i.e. for all $k \in [0, n]$ and for each $i > k$, $x_i^{(k)} = 0$.*

Proof. We proceed with mathematical induction. The base case for $k = 0$ holds automatically by the initialization $x^{(0)} = 0$. All the components are null hence the statement that the 0-th iterate has at most 0 non-null elements is true.

In the induction step we will take some $k < n$ and assume that each ℓ-th iterate with $\ell \in [0, k]$ has only null components after ℓ-th element, i.e. $\forall \ell \leq k, \forall i > \ell, x_i^{(\ell)} = 0$. This assumption is Lemma's statement for the chosen k. Now, we need to show that this implies that the statement holds for $k + 1$. If the chosen $k = 0$:

$$\nabla f(x^{(0)}) = Ax^{(0)} + \mu x^{(0)} - e_1 = -e_1 \implies \forall i > 1, \nabla f(x^{(0)})_i = 0 \tag{2.14}$$

Which implies that $\forall i > 1, x_i^{(1)} = 0$ because $x_i^{(1)}$ are linear combinations of uniquely null elements $\left(\nabla f\left(x^{(0)}\right)\right)_i$. That is, given $k = 0$, $k + 1$ verifies Lemma's statement.

If the chosen $k > 0$, the tridiagonal matrix A results in

$$\nabla f(x^{(k)}) = (A + \mu \mathrm{I})\, x^{(k)} - e_1 \implies \forall i > k+1, \left(\nabla f(x^{(k)})\right)_i = 0 \tag{2.15}$$

Which implies that for all $i > k+1$, $x_i^{(k+1)} = 0$ because $x_i^{(k+1)}$ are linear combinations of null elements $\left(\nabla f\left(x^{(k)}\right)\right)_i$. Overall, (2.14) for $k = 0$ and (2.15) for $k > 0$ show that the iterate $x^{(k+1)}$ has at most $k+1$ non-null elements, which completes the proof by induction. □

Briefly, this Lemma ensures null $x_i^{(k)}$ coordinates of the iterate $x^{(k)}$ for $i > k$. This has major importance in the lower bound proofs based on the construction of the difficult function. A similar technique will be used in the lower bound analysis conducted in the Chapter 4. The theorems that establish state-of-art lower bounds for classical optimization problems rely on the types of techniques describe above. We will now provide some of the crucial lower bound results for the unconstrained optimization.

Theorem 2 (Lower complexity bounds for unconstrained convex smooth optimization)**.** *For any $k \in \left[1, \frac{1}{2}(n-1)\right]$ and any initial point $x_0 \in \mathbb{R}^n$, there exists a problem with a convex, L-smooth objective function f for which any first-order method $\mathcal{A}$ satisfying the Assumption 2 produces points $x^{(k)}$ with the objective lower bounded by $O\left(\frac{1}{k^2}\right)$:*

$$f\left(x^{(k)}\right) - f^* \geq \frac{3L\left\|x^{(0)} - x^*\right\|^2}{32(k+1)^2}$$

and the (disappointing) bound on the minimizing iterates:

$$\left\|x^{(k)} - x^*\right\| \geq \frac{1}{8}\left\|x^{(0)} - x^*\right\|$$

Proof in [Nesterov 2018, Section 2.1.2].

Theorem 3 (Lower complexity bounds for unconstrained smooth strongly convex optimization)**.** *For any $k \in \left[1, \frac{1}{2}(n-1)\right]$, any initial point $x_0 \in \mathbb{R}^n$, and any constants $\mu > 0$, $\frac{L}{\mu} = Q_f > 1$, there exists a problem with a μ-strongly convex, L-smooth objective function f for which any first-order method $\mathcal{A}$ satisfying the Assumption 2 produces points $x^{(k)}$ such that:*

$$\left\|x^{(k)} - x^*\right\| \geq \left(\frac{\sqrt{Q_f} - 1}{\sqrt{Q_f} + 1}\right)^k \left\|x^{(0)} - x^*\right\| \tag{2.16}$$

$$f\left(x^{(k)}\right) - f\left(x^*\right) \geq \frac{\mu}{2}\left(\frac{\sqrt{Q_f} - 1}{\sqrt{Q_f} + 1}\right)^{2k} \left\|x^{(0)} - x^*\right\|^2 \tag{2.17}$$

Proof in [Nesterov 2018, Section 2.1.4].

The corresponding oracle complexities for the smooth problems are $\Omega(1/\sqrt{\varepsilon})$ for the convex case and $\Omega(\sqrt{Q}\log(1/\varepsilon))$ in the strongly convex case.

For the convex, non-smooth cases (only Lipschitzian f, without continuous gradient) the lower bounds are sublinear: $O\left(\frac{1}{\sqrt{k}}\right)$ in the non-strongly convex case and $O\left(\frac{1}{k}\right)$ in the strongly convex case [Bubeck 2015].

Beyond the lower bound results described so far, we could talk about the lower bounds for the constrained optimization problems. There are two types of constrained problems: set constrained problems and problems with functional constraints. The set-constrained problems, such as Euclidean ball constraint, are relatively easy compared to general functional constraints. We focus on the particular type of functional constraints, namely affine constraints, whose "difficulty" lies in-between set-constraints and general functional constraints.

2.4.2 Affinely constrained optimization problems

There aren't many articles that analyze the lower bound for affinely constrained optimization problems. In [Ouyang and Xu 2019], authors considered the deterministic lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems. As a special instance, they cover lower bounds under linear span assumption for affinely constrained problems. They state that the convergence rate $O(1/t)$ is optimal for affinely constrained convex problems. Moreover, they claim that $O(1/t^2)$ is the best possible rate for the strongly convex problems under affine constraints. Accordingly, they found out that the linear convergence of unconstrained strongly convex problems doesn't carry over into the strongly convex case with affine constraints. This confirms that the case with affine constraints is indeed more complex. Still, in this paper no other properties were considered, and by no means it indicates that under metric subregularity the linear rate is unattainable.

In [J. Zhang, Hong, and S. Zhang 2019], the minimax saddle point problem was considered. This class includes cases of affinely constrained smooth convex optimization problems as ours. The deterministic lower iteration complexity bounds for μ_x-strongly convex and μ_y-strongly concave saddle point problems found in [J. Zhang, Hong, and S. Zhang 2019] is

$$\Omega\left(\sqrt{\frac{L_x}{\mu_x}+\frac{L_{xy}^2}{\mu_x\mu_y}\cdot+\frac{L_y}{\mu_y}}\ln\left(\frac{1}{\epsilon}\right)\right), \tag{2.18}$$

which amounts to the linear rate. Interestingly, strong convexity turned out to be insufficient for a linear bound in [Ouyang and Xu 2019] (affinely constrained problem), but strong convexity with strong concavity of the minimax reformulation of the same problem led to a linear bound for Zhang et al.

Xie et al. [Xie et al. 2020] analyzed the stochastic lower bounds for convex-convave minimax optimization problem as ours. They didn't consider metric subregularity, but a linear iteration complexity $\Omega\left((n+Q)\log\left(1/\varepsilon\right)\right)$ was obtained for strongly-convex, strongly-concave case. Hence, the deterministic linear lower bounds from Zhang et al. followed into the stochastic case as long as the minimax optimization problem is strongly-convex, strongly-concave.

In [Woodworth and Srebro 2016], authors used another method of the lower bound analysis that doesn't utilize the tridiagonal matrix like in Nesterov, but leverages the concept of the resisting oracle. They considered both deterministic and stochastic cases, as well as performed the analysis using the prox operator in the oracle calls. Still, the problem considered there is a minimization of the sum of smooth or non-smooth terms. They didn't include the linear composition term and the dual variable isn't incorporated. Hence, the

considered class of composite objectives doesn't include the case of affinely constrained optimization.

3 Linear upper bound of Vu's primal-dual splitting algorithm

In [Liang, Fadili, and Peyré 2016], authors developed a framework to analyze the convergence of monotone operator splitting methods. They obtained results on global and local convergence rates for the inexact Krasnoselskii-Mann iteration, which is the basis to analyze numerous operator splitting methods casted as the Krasnoselskii-Mann fixed-point iteration. For instance, primal-dual splitting methods can be analyzed in that way. A crippled version of the framework will be presented in the Section 3.1. It will be used in the Section 3.2 in the context of a primal-dual splitting method designed by Vu [Vu 2013]. We will show that Vu's method attain an upper complexity bound for affinely constrained convex optimization problems under metric subregularity.

3.1 Linear convergence under metric subregularity

We introduce a framework to analyze the convergence rates through a Krasnosel'skii-Mann fixed point iteration. This section relies on the key paper for this work: *Convergence Rates with Inexact Non-expansive Operators* [Liang, Fadili, and Peyré 2016]. One of the main differences between our scenario and theirs is that we assume exact iterations, rather than inexact ones. That is, there is no error of approximating Tz_k and $\varepsilon_k = 0$. We justify it by the fact that in our case we have exact quantities of all the operators, no estimations are required.

Recall that $\mathcal{X}, \mathcal{Y}$ are the real Euclidean spaces with the dot product $\langle \cdot, \cdot \rangle$ and the norm $\|\cdot\|$. Let $\mathcal{Z}$ be a product space $\mathcal{X} \times \mathcal{Y}$, with the scalar product $\langle (x_1, y_1), (x_2, y_2) \rangle = \langle x_1, x_2 \rangle + \langle y_1, y_2 \rangle$. Denote elements from $\mathcal{Z}$ as $z = (x, y)$, in particular an iterate at k is $z_k = (x_k, y_k) \in \mathcal{Z}$ and its norm is $||z_k|| = \sqrt{\langle z_k, z_k \rangle} = \sqrt{\langle x_k, x_k \rangle + \langle y_k, y_k \rangle}$.

As mentioned, the convergence rates of the splitting methods can be analyzed using the Krasnoselskii-Mann iteration, denoted z_k. We can cast numerous splitting methods as the Krasnoselskii-Mann fixed-point iteration and investigate the rate of convergence at which the Krasnoselskii-Mann iterates approach the fixed point. Then, this results can be linked with the solution x^* of the original interest (since $z_k = (x_k, y_k)$).

The Krasnosnel'skii-Mann iteration [Krasnosel'skiı 1955], [Mann 1953] is a classic result on the convergence of a non-expansive operator via averages to a fixed point, although today we simply describe these operators as averaged operators.

Definition 22 (Krasnosnel'skii-Mann iteration)**.** Let $\lambda_k \in]0, 1]$, and let $T : \mathcal{X} \to \mathcal{X}$ be an λ_k-averaged operator $T_{\lambda_k} = \lambda_k T + (1 - \lambda_k)\mathrm{I}$ such that the set of fixed points $\mathrm{fix}T = \{z \in \mathcal{X} : z = Tz\}$ is non-empty. Then the (exact) Krasnosel'skii-Mann iteration

of T is given by

$$z_{k+1} = z_k + \lambda_k \left(T z_k - z_k\right) = T_{\lambda_k} z_k. \tag{3.1}$$

Moreover, define the error of the iteration

$$e_k = (\mathrm{I} - T)\, z_k = \frac{z_k - z_{k+1}}{\lambda_k} \tag{3.2}$$

and notions

$$T' = \mathrm{I} - T \tag{3.3}$$
$$\tau_k = \lambda_k(1 - \lambda_k). \tag{3.4}$$

Note that $e_k = T' z_k$.

The sequence z_k converges to some fixed point z^* of T. Based on averaged operator T, one can analyze the convergence in terms of the global pointwise iteration-complexity bound and the ergodic iteration-complexity bounds. For instance, global pointwise bound for gradient descent has $T' = \nabla f$ and the optimal rate $O(k^{-2})$ is obtained (ϵ-solution in $O(1/\sqrt{\epsilon})$ iterations). Then, the ergodic iteration bound is optimal $O(1/k)$ ($(1/\epsilon)$ iterations).

Oftentimes, splitting methods like DRS have sub-linear global convergence, but local linear convergence [Demanet and X. Zhang 2016]. With an assumption that T' is metrically subregular (according to the Definition 21), Krasnosel'skii-Mann iteration analysis is useful to show this linear convergence within the neighborhood of subregularity. Our primary concern is the class of affinely constraints problems, for which we know that global metric subregularity holds. Hence, in our case we can automatically benefit the global linear convergence.

Applying the Definition 21 of metric subregularity for the metrically subregular operator T' (i.e. F there is T' here) at some $z^* \in \mathrm{fix}T$ ($\mathrm{zer}T' = \mathrm{fix}T$) for 0 amounts to

$$d(z,\ \mathrm{fix}T) \leq \kappa \left\|T' z\right\|, \quad \forall z \in \mathcal{N}(\bar{z}) \tag{3.5}$$

We will provide a particular version of the [Liang, Fadili, and Peyré 2016, Theorem 3 (Local convergence rate)]. Firstly, notice that they define a ball $\mathbb{B}_a(z^*)$ which is contained in the neighborhood of metric subregularity $\mathcal{N}(z^*)$. Yet, here $\mathcal{N} = \mathcal{Z}$, hence any initial point is metrically sub-regular at z^* and no ball is necessary. As mentioned, we consider the exact case and can drop the ε_k-related terms from the Theorem. Lastly, we assume that $\mathrm{fix}T$ is a singleton, contrary to a general set in the original theorem.

Theorem 4. *[Global linear convergence rate] Let $z^* \in \mathrm{fix}T$ with singleton $\mathrm{fix}T$. Assume T' is globally metrically subregular at z^* for 0 with the constant of metric subregularity $\kappa \in (0, +\infty)$. Then, for any initial point $z_0 \in \mathcal{Z}$, for all $k \in \mathbb{N}$ we have*

$$\left\|z_{k+1} - z^*\right\|^2 \leq \zeta_k \|z_k - z^*\|^2, \quad \textit{where} \quad \zeta_k = \begin{cases} 1 - \frac{\tau_k}{\kappa^2}, & \textit{if } \frac{\tau_k}{\kappa^2} \in]0, 1] \\ \frac{\kappa^2}{\kappa^2 + \tau_k}, & \textit{otherwise} \end{cases} \in [0, 1[. \tag{3.6}$$

1. $\|z_k - z^*\| \to 0$ *if the sequence $(\tau_k)_{k\in\mathbb{N}}$ isn't a summable sequence in $[0, +\infty[$.*
2. $\lim_{k\to+\infty} \|z_k - x^*\|^{-k} < 1$, *which is R-linear convergence.*

3. *If* $0 < \inf_{k\in\mathbb{N}} \lambda_k \leq \sup_{k\in\mathbb{N}} \lambda_k < 1$, *then there exists* $\zeta \in (0,1)$ *such that*

$$\|z_{k+1} - z^*\|^2 \leq \zeta^k \|z_0 - z^*\|^2 \tag{3.7}$$

Proof. Simplified version of the proof of the Theorem 3 in [Liang, Fadili, and Peyré 2016] □

3.2 Applicability of Vu's algorithm

Consider a general monotone inclusion problem:

Problem 3 ([Vu 2013])**.** *Let* $m > 0$, *let each of* $\mathcal{H}, \mathcal{G}_1, \ldots, \mathcal{G}_m$ *be a real Hilbert space, let* $z \in \mathcal{H}$, $b_i \in \mathcal{G}_i$, *and let* $(\omega_i)_{1\leq i\leq m} \in (0,1]$ *such that* $\sum_{i=1}^m \omega_i = 1$. *Moreover, let*

- $A : \mathcal{H} \to 2^{\mathcal{H}}$ *and* $B_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ *be maximally monotone,*
- $C : \mathcal{H} \to \mathcal{H}$ *be* μ*-cocoercive for* $\mu \in (0, +\infty)$,
- $D_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ *be maximally monotone and* v_i*-strongly monotone for* $v_i \in (0, +\infty)$,
- $M_i : \mathcal{H} \to \mathcal{G}_i$ *is a nonzero bounded linear operator*

The primal inclusion problem is:

$$\text{find } \bar{x} \in \mathcal{H} \text{ such that } z \in A\bar{x} + \sum_{i=1}^m \omega_i M_i^* \left((B_i \square D_i)\left(M_i\bar{x} - b_i\right)\right) + C\bar{x} \tag{3.8}$$

While the dual inclusion is:

$$\begin{gathered} \text{find } \bar{y}_1 \in \mathcal{G}_1, \ldots, \bar{y}_m \in \mathcal{G}_m \text{ such that} \\ (\exists x \in \mathcal{H}) \begin{cases} z - \sum_{i=1}^m \omega_i M_i^* \bar{y}_i \in Ax + Cx \\ (\forall i \in \{1, \ldots, m\}) \bar{y}_i \in (B_i \square D_i)\left(M_i x - b_i\right) \end{cases} \end{gathered} \tag{3.9}$$

The primal-dual splitting method by Vu was designed to solve the inclusion Problem 3 involving cocoercive operators. This problem is very general and allows different structures such as the parallel sum □. In our case, the parallel sum won't be used because it's only relevant to block-diagonal constraints.

In Section 5.3 of [Liang, Fadili, and Peyré 2016], the same Problem 3 was considered, only with $z = 0$ and renamed operators: A to C, C to B, and B_i to A_i. The closest setting to the problem 3 is found in [Combettes and J.-C. Pesquet 2012], where instead of the cocoercive operator, C is a single-valued monotone Lipschitzian operator and also instead of the maximally monotone and strongly monotone D_i, monotone operator D_i such that D_i^{-1} is a single-valued Lipschitzian is considered. Both these problems are remarkably general and cover numerous other, more specialized problems. Some of these particular problems are listed in [Vu 2013; Combettes and J.-C. Pesquet 2012].

Originally, Vu's algorithms requires that we let $\tau, (\sigma_i)_{1\leq i\leq m} > 0$ such that

$$\begin{gathered}\eta = \min\left\{\tau^{-1}, \sigma_1^{-1}, \dots, \sigma_m^{-1}\right\}\left(1 - \sqrt{\tau\sum_{i=1}^{m}\sigma_i\omega_i\|M_i\|^2}\right)\\ \beta = \min\left\{\mu, \nu_1, \dots, \nu_m\right\}\\ 2\eta\beta > 1\\ \lambda_k \in \left]0, \frac{4\eta\beta - 1}{2\eta\beta}\right] \quad \forall k \in \mathbb{N}.\end{gathered} \tag{3.10}$$

Moreover, that one lets $(\varepsilon_{1,k})_{k\in\mathbb{N}}$ and $(\varepsilon_{2,k})_{k\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{H}$, and $(\varepsilon_{3,i,k})_{k\in\mathbb{N}}$ and $(\varepsilon_{4,i,k})_{k\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{G}_i$ for $i \in \{1, \dots, m\}$.

Then, Vu's primal-dual splitting routine generates sequences $(x_k)_{k\in\mathbb{N}}$ and $(y_{1,k}, \dots, y_{m,k})_{k\in\mathbb{N}}$ according to below rules for any $k \in \mathbb{N}$:

$$\begin{array}{l} p_{k+1} = J_{\tau A}\left(x_k - \tau\left(\sum_{i=1}^{m}\omega_i M_i^* y_{i,k} + Cx_k + \varepsilon_{1,k} - z\right)\right) + \varepsilon_{2,k}\\ y_{k+1} = 2p_{k+1} - x_k\\ x_{k+1} = x_k + \lambda_k\left(p_{k+1} - x_k\right)\\ \text{for } i = 1, \dots, m\\ \left[\begin{array}{l} q_{i,k+1} = J_{\sigma_i B_i^{-1}}\left(y_{i,k} + \sigma_i\left(M_i y_{k+1} - D_i^{-1} y_{i,k} - \varepsilon_{3,i,k} - r_i\right)\right) + \varepsilon_{4,i,k}\\ y_{i,k+1} = y_{i,k} + \lambda_k\left(q_{i,k+1} - y_{i,k}\right)\end{array}\right.\end{array} \tag{3.11}$$

Recall that J_A is a resolvent of the operator A (Definition 18).

3.2.1 The inclusion problem as a saddle point search

Vu's primal-dual splitting algorithm can be applied to solve convex minimization under affine constraints. To show that, we need to show that the monotone inclusion Problem 3 considered by Vu can be specialized into our inclusion problem (2.11), i.e. the later is a particular case of the former.

Lemma 3. *The general inclusion Problem 3 includes the affinely constrained convex optimization Problem 1.*

Proof. Recall the primal inclusions from the original Vu's Problem 3:

$$\text{find } \bar{x} \in \mathcal{H} \text{ such that } z \in A\bar{x} + \sum_{i=1}^{m}\omega_i M_i^*\left(\left(B_i \square D_i\right)\left(M_i\bar{x} - b_i\right)\right) + C\bar{x}$$

and the dual inclusion

$$\begin{gathered}\text{find } \bar{y}_1 \in \mathcal{G}_1, \dots, \bar{y}_m \in \mathcal{G}_m \text{ such that}\\ (\exists x \in \mathcal{H})\begin{cases} z - \sum_{i=1}^{m}\omega_i M_i^*\bar{y}_i \in Ax + Cx\\ (\forall i \in \{1, \dots, m\})\bar{y}_i \in \left(B_i \square D_i\right)\left(M_i x - b_i\right)\end{cases}\end{gathered} \tag{3.12}$$

We narrow down this inclusion class as follows. We set $z = 0$, $m = 1$ (subscript i takes only 1 hence is dropped), and $\omega = 1$. The operator A is eliminated by setting $A : x \mapsto \{0\}$.

The operator C is chosen to be the subdifferential ∂f. Since f is smooth, $\partial f(x) = \{\nabla f(x)\}$. The gradient ∇f is cocoercive by Baillon–Haddad Theorem 1, which implies that ∂f is cocoercive too and we preserve the cocoercivity required for C.

Furthermore, let $B = D = \partial \mathrm{I}_0$, i.e. both B and D map to the whole space $\mathcal{G}$ at 0 and to an empty set everywhere else (see Example 4). This choice entails that the parallel sum $B \square D$ (Definition 20) becomes $\partial \mathrm{I}_0 \square \partial \mathrm{I}_0 = \left(\partial \mathrm{I}_0^{-1} + \partial \mathrm{I}_0^{-1}\right)^{-1}$, which amounts to $\partial \mathrm{I}_0$ because a sum of null operators (see Example 7 for $\partial \mathrm{I}_0^{-1}$) is still an operator which maps any element to 0, and inverse $\partial \mathrm{I}_0^{-1}$ recovers $\partial \mathrm{I}_0$. Then, the primal and dual inclusions become:

$$\text{find } \bar{x} \in \mathcal{H} \text{ such that } 0 \in \partial f(\bar{x}) + M^* \left(\partial \mathrm{I}_0 \left(M\bar{x} - b\right)\right)$$
$$\text{find } \bar{y} \in \mathcal{G} \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} -M^*\bar{y} \in \partial f(x) \\ \bar{y} \in \mathrm{I}_0 \left(Mx - b\right) \end{cases}$$

The first inclusion requires $-\partial f(\bar{x}) \in M^* \left(\mathrm{I}_0 \left(M\bar{x} - b\right)\right)$, but no element can be contained in an empty set hence $M\bar{x} - b = 0$. There must exists $y \in \partial \mathrm{I}_0 \left(M\bar{x} - b\right)$ such that $-\partial f(\bar{x}) \in M^* y$ has some solution $\bar{x}$. We rewrite the primal inclusion accordingly:

$$\text{find } \bar{x} \in \mathcal{H} \text{ such that } (\exists y \in \mathcal{G}) \begin{cases} M\bar{x} - b = 0 \\ 0 \in \partial f(\bar{x}) + M^* y \end{cases} \tag{3.13}$$

For the dual inclusion, the same reasoning leads to

$$\text{find } \bar{y} \in \mathcal{G} \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} -M^*\bar{y} \in \partial f(x) \\ Mx - b = 0 \end{cases} \tag{3.14}$$

Note the similarity of the primal and dual inclusion. More importantly, the similarity with the inclusion formulation of the saddle search problem (2.11), which stated

$$\text{find } (x^*, y^*) \in \mathcal{X} \times \mathcal{Y} \text{ such that } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial f(x^*) + M^* y^* \\ Mx^* - \{b\} \end{pmatrix}$$

We conclude that if some $(\bar{x}, \bar{y})$ is the solution to the Vu's general inclusion Problem 3 then it's a solution to the inclusion problem (2.11) corresponding to the affine constrained convex minimization Problem 1. □

3.2.2 Iterates and the fixed point operator

The Lemma 3 showed that the iterates produced by Vu's algorithm can solve convex optimization problems under affine constraints. Towards that goal, we modify the original constants (3.10) and iterations (3.11). We follow the same simplifications as before, thus we set $z = 0$, $m = 1$ (subscript i dropped), $\omega = 1$, $A : x \mapsto \{0\}$, $B = D = \partial \mathrm{I}_0$, and $C = \nabla f$. Therefore, Vu's algorithm requires that we let $\tau, \sigma > 0$ such that

$$\begin{gathered} \eta = \min\left\{\tau^{-1}, \sigma^{-1}\right\}\left(1 - \sqrt{\tau\sigma\|M\|^2}\right) \\ \beta = \min\left\{\mu, \nu\right\} \\ 2\eta\beta > 1 \\ \lambda_k \in \left]0, \frac{4\eta\beta - 1}{2\eta\beta}\right] \quad \forall k \in \mathbb{N}. \end{gathered} \tag{3.15}$$

Moreover, we have the error-free case with all the sequences $(\varepsilon_{1,k}, \varepsilon_{2,k}, \varepsilon_{3,k}, \varepsilon_{4,k})_{k\in\mathbb{N}} = 0$ because we need not approximate anything. Then, the simplified Vu's routine generates primal and dual sequences $(x_k)_{k\in\mathbb{N}}$, $(v_k)_{k\in\mathbb{N}}$ according to rules below for any $k \in \mathbb{N}$:

$$\begin{aligned} p_{k+1} &= x_k - \tau\left(M^* y_k + \nabla f(x_k)\right) \\ y_{k+1} &= 2p_{k+1} - x_k \\ x_{k+1} &= x_k + \lambda_k\left(p_{k+1} - x_k\right) \\ q_{k+1} &= y_k + \sigma\left(M y_{k+1} - b\right) \\ y_{k+1} &= y_k + \lambda_k\left(q_{k+1} - y_k\right) \end{aligned} \tag{3.16}$$

Remark. The resolvents disappeared from the algorithm because both $J_{\tau A}$ and $J_{\sigma\partial \mathrm{I}_0^{-1}}$ amount to $J_{\{0\}} = (\mathrm{I} + \{0\})^{-1} = \mathrm{I}^{-1} = \mathrm{I}$.

Following Vu's paper, supplementary operators need to be derived. Firstly, let $\mathcal{K} = \mathcal{H} \oplus \mathcal{G}$ be the Hilbert direct sum endowed with the scalar product $\langle (x_1, y_1), (x_2, y_2)\rangle = \langle x_1, x_2\rangle + \langle y_1, y_2\rangle$ and define operators on $\mathcal{K}$ for the original version of the algorithm:

$$\begin{aligned} &\mathbf{C} : \mathcal{K} \to 2^{\mathcal{K}}, (x, y) \mapsto (Ax) \times \left(b_1 + B_1^{-1} y_1\right) \times \cdots \times \left(b_m + B_m^{-1} y_m\right) \\ &\mathbf{D} : \mathcal{K} \to \mathcal{K}, (x, y) \mapsto \left(\textstyle\sum_i \omega_i M_i^* y_i, -M_1 x, \cdots, -M_m x\right) \\ &\mathbf{E} : \mathcal{K} \to \mathcal{K}, (x, \boldsymbol{y}) \mapsto \left(Cx, D_1^{-1} y_1, \cdots, D_m^{-1} y_m\right) \\ &\mathbf{F} : \mathcal{K} \to \mathcal{K}, (x, \boldsymbol{y}) \mapsto \left(\tfrac{1}{\tau} x - \textstyle\sum_i \omega_i M_i^* y_i, \tfrac{1}{\sigma_1} y_1 - M_1 x, \cdots, \tfrac{1}{\sigma_m} y_m - M_m x\right), \end{aligned}$$

while in our simplified case we obtain

$$\mathbf{C} : \mathcal{K} \to 2^{\mathcal{K}}, (x, y) \mapsto \{0\} \times b \tag{3.17}$$

$$\mathbf{D} : \mathcal{K} \to \mathcal{K}, (x, y) \mapsto (M^* y, -Mx) \tag{3.18}$$

$$\mathbf{E} : \mathcal{K} \to \mathcal{K}, (x, y) \mapsto (\nabla f(x), 0) \tag{3.19}$$

$$\mathbf{F} : \mathcal{K} \to \mathcal{K}, (x, y) \mapsto (\tau^{-1} x - M^* y, \sigma^{-1} y - Mx). \tag{3.20}$$

Using matrix notation, operator $\mathbf{D}$ can be rewritten as Dz and operator $\mathbf{E}$ as Ez with

$$F = \begin{bmatrix} \tau^{-1}\mathrm{I} & -M^* \\ -M & \sigma^{-1}\mathrm{I} \end{bmatrix} \qquad D = \begin{bmatrix} 0 & M^* \\ -M & 0 \end{bmatrix} \qquad E = \begin{bmatrix} \nabla f & 0 \\ 0 & 0 \end{bmatrix} \tag{3.21}$$

Where 0 was used instead of $\{0\}$. Operator $\mathbf{F}$ is self-adjoint, operator $\mathbf{D}$ is maximally monotone and operator $\mathbf{E}$ is L^{-1}-cocoercive (recall that ∇f is L-Lipschitz continuous and the Baillon–Haddad Theorem 1). Then, casting as Krasnoselskii-Mann iteration yields

$$z_{k+1} = \left(\mathrm{I} + \mathbf{F}^{-1} D\right)^{-1} \left(\mathrm{I} - \mathbf{F}^{-1} E\right) z^k \tag{3.22}$$

That is,

$$z_{k+1} = z_k + \lambda_k \left(J_{\mathbf{F}^{-1}(\mathbf{C}+\mathbf{D})}\left(z_k - \mathbf{F}^{-1}\mathbf{E} z_k\right) - z_k\right) \tag{3.23}$$

And the fixed point operator T is averaged with $\frac{2\eta\beta}{4\eta\beta - 1}$ constant by the Lemma 1:

$$\mathbf{T} = J_{\mathbf{F}^{-1}(\mathbf{C}+\mathbf{D})} \circ \left(\mathrm{Id} - \mathbf{F}^{-1}\mathbf{E}\right) \in \frac{\eta\beta}{\eta\beta -} \tag{3.24}$$

Therefore, (3.23) is a special instance of the Krasnosels'kii-Mann iteration from the Definition 22. We successfully casted the iterates into a Krasnoselskii-Mann iteration, thus Vu's algorithm can by applied to solve Problem 1. Moreover, the Theorem 4 guarantees the linear convergence because $T' = \mathrm{I} - T$ is metrically subregular given T from (3.24).

4 Lower bound analysis

In this section we estabilish the lower bound for affinely constrained convex problems under metric subregularity. The primary tasks accomplished in this analysis consist of a formulation of the minimax problem with difficult affine constraints $Mx = b$, bounding the eigenvalues of the matrix $M^\intercal M$ in the Lemma 4, establishing the constant of metric subregularity of $\nabla\mathcal{L}$ in the Lemma 5, finding the explicit solution in Lemma 6, and finding the linear lower bound in the Theorem 6.

4.1 Problem with difficult constraints $Mx = b$

For the lower bound analysis, we will consider an instance of the minimax problem (2.9) with $f(x) = \frac{\mu}{2}\|x\|_2^2$, that is:

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \frac{\mu}{2}\|x\|^2 + \langle Mx - b, y\rangle. \tag{4.1}$$

The corresponding primal formulation is:

$$\min_{x \in \mathcal{X}} \frac{\mu}{2}\|x\|^2 + \mathrm{I}_b(Mx) \tag{4.2}$$

Which corresponds to the minimization of $\frac{\mu}{2}\|x\|^2$ under constraints $Mx = b$, where the affine constraints are expressed by the linear composite term $\mathrm{I}_b(Mx)$.

The dual problem is

$$\min_{y \in \mathcal{Y}} \| - M^\intercal y\|_*^2 + \langle b, y\rangle \,,$$

where $\|\cdot\|_*$ is the dual norm of the euclidean norm $\|\cdot\|$ employed in the primal problem.

The $f(x) = \frac{\mu}{2}\|x\|_2^2$ part of the problem is μ-strongly convex for $\mu > 0$ (see Example 3). Moreover, I_b is convex (Example 2) so (4.2) is strongly convex (sum of a convex term and strongly convex term is a strongly convex function Proposition 2). Thus, for $\mu > 0$, the minmax problem (4.1) is strongly convex, non-strongly concave. Still, we keep the ability to set $\mu = 0$, in which case the f term disappears and the minmax problem is non-strongly convex non-strongly concave.

The affine constraints $Mx = b$ are ought to be made deliberately difficult by imposing a specific matrix M. Crucially, we need M which makes the optimization problem difficult and suitable for the lower bound analysis.

Let $\rho > 0$ be a constant and let $M \in \mathbb{R}^{n\times n}$ be a constraints matrix that verifies

$$\begin{aligned} (Mx)_1 &= (\rho + 1)x_1 \\ (Mx)_i &= (\rho + 1)x_i - x_{i-1} \qquad \text{for } 1 < i \leq n \end{aligned} \tag{4.3}$$

That is,

$$M = \begin{bmatrix} \psi & 0 & & & & \mathbf{0} \\ -1 & \psi & 0 & & & \\ 0 & -1 & \psi & 0 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & 0 & -1 & \psi & 0 \\ \mathbf{0} & & & 0 & -1 & \psi \end{bmatrix} \tag{4.4}$$

Where we defined

$$\psi := \rho + 1 \tag{4.5}$$

Let $b \in \mathbb{R}^n$ be a vector with the constants

$$\begin{aligned} b_1 &= 1 \\ b_i &= 0 \qquad \text{for } 1 < i \leq n \end{aligned} \tag{4.6}$$

From $(Mx)_i = b_i = 0$ and $(Mx)_i = \psi x_i - x_{i-1}$ we notice that $\psi x_i = x_{i+1}$. This indicates that the components of x decrease as i increases, i.e. for large i, the components of x are negligible. We can also look at the constraints (4.3) row-wise: $(Mx)_i = \sum_{j=1}^{n} M_{ij} x_j = m_i^\intercal x$ with $m_i^\intercal \in \mathbb{R}^{1\times n}$ being the i-th row of the matrix M. The difficulty of the constraints M has its source in the fact that for any row $m_i^\intercal$, the constrains allow at most 2 non-null components.

We are going to work with the matrices $M^\intercal M$ and $MM^\intercal$.

$$M^\intercal M = \begin{bmatrix} \psi^2+1 & -\psi & 0 & \cdots & & & & \\ -\psi & \psi^2+1 & -\psi & 0 & \cdots & & & \\ 0 & -\psi & \psi^2+1 & -\psi & 0 & \cdots & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & \cdots & 0 & -\psi & \psi^2+1 & -\psi \\ & & & & \cdots & 0 & -\psi & \psi^2 \end{bmatrix} \tag{4.7}$$

$$MM^\intercal = \begin{bmatrix} \psi^2 & -\psi & 0 & \cdots & & & & \\ -\psi & \psi^2+1 & -\psi & 0 & \cdots & & & \\ 0 & -\psi & \psi^2+1 & -\psi & 0 & \cdots & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & \cdots & 0 & -\psi & \psi^2+1 & -\psi \\ & & & & \cdots & 0 & -\psi & \psi^2+1 \end{bmatrix} \tag{4.8}$$

Matrices with such tridiagonal structure, described briefly as constraints $A_{ij} = 0$ for $|i-j| > 1$, are known as Jacobi matrices (operators). In terms of null and non-null components, both matrices $M^\intercal M$ and $MM^\intercal$ have exactly the same structure as Nesterov's symmetric tridiagonal matrix A_N defined in [Nesterov 2018, Section 2.1], as well as the matrix $A(\eta, \omega)$ in [Xie et al. 2020]. The only difference lies in the values of the non-null elements.

Remark. The matrix $MM^\intercal$ is related to the matrix $M^\intercal M$ via $MM^\intercal = (UV^\intercal)M^\intercal M(VU^\intercal)$, assuming that the singular value decomposition is written as $M = UDV^\intercal$.

4.2 Characterization of the eigenvalues of $M^\intercal M$

The spectrum of $M^\intercal M$ has an important role in this lower bound analysis. In the next Section 4.3, we will see that the smallest eigenvalue of $M^\intercal M$ indirectly determines the value of the constant of metric subregularity in the optimization Problem (2.9).

Proposition 4. Let $\rho, n > 0$. The matrices M, $M^\intercal M$ are invertible. *Proof.* The matrix M has lower triangular structure, hence it's determinant is the product of the elements on its diagonal. By $\det M = (\rho+1)^n > 0$, M is invertible, which implies invertible $M^\intercal M$.

Proposition 5 (If the real matrix M is invertible, then $M^\intercal M$ is a positive definite matrix)**.** For any non-zero vector $x \in \mathbb{R}^n$, $Mx \neq 0$ by the invertibility of M and

$$x^\intercal M^\intercal M x = (Mx)^\intercal Mx = \|Mx\|^2 > 0$$

Proposition 6 (If $\rho > 0$, matrix $M^\intercal M$ is positive definite.)**.** By the application of the Propositions 4 and 5, i.e. by the fact that M is invertible, $M^\intercal M$ is positive definite.

Proposition 7. Positive definite matrix A has positive eigenvalues. *Proof.* Let v be a non-zero eigenvector of A, i.e. $Av = \lambda v$ holds. By the positive definiteness of A, the vector v must verify $v^\intercal Av \geq 0$. Thus, $v^\intercal(\lambda v) \geq 0 \implies v^\intercal v\lambda > 0$. As $v^\intercal v = \|v^\intercal v\|^2 > 0$, the eigenvalues λ are required to be positive by $v^\intercal v\lambda > 0$.

By Proposition 4, we get that $M^\intercal M$ is invertible hence it has only non-null eigenvalues. By the Proposition 6 stating the positive definiteness of $M^\intercal M$, we obtain that its eigenvalues are positive (Proposition 7), which also excludes null eigenvalues. This first characterization allows us to simply say that the *smallest eigenvalue* interests us, rather than the *eigenvalue that is smallest in the absolute value*. We will use Gershgorin circles (discs) to bound λ_{min} and λ_{max}. These discs define the intervals in which the eigenvalues must lie:

Definition 23 (Gershgorin disc)**.** Let A be an $n \times n$ real square matrix with entries a_{ij}. Let R_i be the sum of the absolute values of off-diagonal entries in the i-th row, i.e. $R_i = \sum_{i \neq j} |a_{ij}|$. The i-th Gershgorin disc associated with the i-th row of A is a closed disc centered at a_{ii} with radius R_i, i.e.

$$D_i = D(a_{ii}, R_i) = \{z \in \mathbb{R} \ : \ |z - a_{ii}| \leq R_i\}$$

Theorem 5 (Gershgorin circle theorem)**.** *Let A be an $n \times n$ real square matrix with entries a_{ij}, and let R_i be the absolute sum of off-diagonal elements in the i-th row. Then, each eigenvalue of A lies in at least one of the Gershgorin discs $D(a_{ii}, R_i)$ of A.*

Remark. Gershgorin disc and Gershgorin circle theorem are originally stated for a square complex matrix [Varga 2010]. Here, a simplification for square real matrices was made because we will only deal with matrices whose eigenvalues are real.

Lemma 4. *The eigenvalues $\lambda_1, \ldots, \lambda_n$ of $M^\intercal M$ from (4.7) with some $\rho > 0$ are lower bounded by ρ^2 and upper bounded by $\rho^2+4\rho+4$, i.e. for all the eigenvalues $\lambda \in \{\lambda_1, \ldots, \lambda_n\}$,*

$$\rho^2 \leq \lambda \leq \rho^2 + 4\rho + 4$$

Proof. For the matrix $M^\intercal M$, we can find n Gershgorin discs (Definition 23):

$$\begin{aligned} D_1 &:= D\left((\rho+1)^2+1, |-\rho-1|\right) \\ D_i &:= D((\rho+1)^2+1, |-\rho-1|+|-\rho-1|) \qquad 1<i<n \\ D_n &:= D((\rho+1)^2, |-\rho-1|) \end{aligned}$$

The Gershgorin circle Theorem 5 only asserts that an eigenvalue must be in at least one of the discs, hence we must consider all of them. We can reject the first disc D_1 because it's contained in the D_i discs as $2|-\rho-1| > |-\rho-1|$. This leaves either the last disc D_n or the identical D_i discs to dictate the largest possible range for eigenvalues. The intervals associated with these discs are

$$\begin{aligned} D_i &= \left[\rho^2,\ \rho^2+4\rho+4\right] \\ D_n &= \left[\rho^2+\rho,\ \rho^2+3\rho+2\right] \end{aligned}$$

We identify the bounds on the eigenvalues as

$$\begin{aligned} \lambda_{max} &\leq \max\left\{\rho^2+4\rho+4, \rho^2+3\rho+2\right\} = \rho^2+4\rho+4 \\ \lambda_{min} &\geq \min\left\{\rho^2, \rho^2+\rho\right\} = \rho^2 > 0 \end{aligned}$$

□

Finding the analytical expression for the exact eigenvalues of a general $n \times n$ matrix with $n >> 1$ is not simple and demands lengthy, abstruse formulations. However, this task can sometimes be drastically simplified if the matrix is known to possess some special structure. As mentioned earlier, $M^\intercal M$ is a Jacobi matrix, i.e. it has the tridiagonal structure. The mathematical literature on Jacobi matrices is rich, which puts questions whether there exists an analytical formula for the exact eigenvalues of $M^\intercal M$.

Proposition 8. [Kulkarni, Schmidt, and Tsui 1999, Theorem 2.2.] Let $T_n(a,b,c)$ be an $n \times n$ Toeplitz tridiagonal matrix defined by

$$T_n(a,b,c) = \begin{bmatrix} a & c & & 0 \\ b & \ddots & \ddots & \\ & \ddots & \ddots & c \\ 0 & & b & a \end{bmatrix}$$

The eigenvalues of $T_n(a,b,c)$ are

$$a - 2\sqrt{bc}\cos(k\pi/(n+1)) \quad \text{for } k = 1, 2, \ldots, n \tag{4.9}$$

In the case of Toeplitz tridiagonal matrices $T_n(a,b,c)$, the eigenvalues are determined completely by the Proposition 8. Unfortunately, our matrix $M^\intercal M$ doesn't have the Toeplitz structure due to the last element on the diagonal. Yet, the $n-1 \times n-1$ matrix created by taking the upper left block of $M^\intercal M$ ($M^\intercal M$ without the last row and the last column) has the Toeplitz structure. This shows that $M^\intercal M$ has so-called *pseudo-Toeplitz* structure [Kulkarni, Schmidt, and Tsui 1999]. The location of the eigenvalues of these matrices can be determined via the comparison of the graphs [Fonseca 2006], but it seems that a convenient closed-form solution isn't available.

If we parametrized M's last element and make its diagonal equal to $[\rho{+}1 \quad \cdots \quad \rho{+}1 \quad \omega]$ with $\omega = \sqrt{\rho^2+1}$, we would obtain a matrix $M^\intercal M$ with constant elements on the diagonal. However, this would change the last element on the diagonal above and below the principal diagonal, breaking the constant elements. Hence, we cannot obtain $M^\intercal M$ with a Toeplitz structure by the changes of the values of non-null elements of M. We would still have a pseudo-Toeplitz matrix.

We cannot use the Propostion 8 to find the eigenvalues of the pseudo-Toeplitz $M^\intercal M$ matrix, but we can use it to evaluate the bounds found in Lemma 4. Let $M'^\intercal M'$ be a tridiagonal Toeplitz matrix obtained by adding 1 to the last element on the diagonal of $M^\intercal M$. We view the original matrix $M^\intercal M$ as perturbed $M'^\intercal M'$, where the perturbation is the 1 subtracted from the last element of the diagonal of $M'^\intercal M'$.

Generally, the eigenvalues of tridiagonal Toeplitz matrices can be very sensitive to perturbations of the matrix, which suggests that the spectrum of $M'^\intercal M'$ can be far off the spectrum of $M^\intercal M$. However, the sensitivity varies depending on the structure of the tridiagonal Toeplitz matrix. In particular, the sensitivity of the eigenvalues grows exponentially with the ratio of the absolute values of the sub- and super-diagonal matrix entries [Noschese, Pasquini, and Reichel 2013]. Matrix $M'^\intercal M'$ has the same elements on the sub- and super-diagonal, i.e. unit ratio. This suggests that actually, $M'^\intercal M'$ isn't highly sensitive. Moreover, the perturbation itself is very small because it affects only one element. The relative change of this element depends on ρ. For $\rho < 1$ subtracting 1 has quite a significant effect, with the relative change up to $\approx 50\%$. For $\rho \geq 1$, the relative change quickly gets insignificant as ρ grows, and the perturbation becomes negligible.

By the application of (4.9) with $a = \rho^2 + 2\rho + 2$ and $b = c = -\rho - 1$, we obtain the eigenvalues of $M'^\intercal M'$:

$$\rho^2 + (2\rho+2) - (2\rho-2)\cos\left(\frac{k\pi}{n+1}\right) \qquad \text{for } k = 1, \ldots, n \tag{4.10}$$

From $-1 \leq \cos \leq 1$, we get the bounds for the eigenvalues λ of $M'^\intercal M'$:

$$\rho^2 \leq \lambda_{M'^\intercal M'} \leq \rho^2 + 4\rho + 4 \tag{4.11}$$

Which coincide with the bound obtained in Lemma 4. Notice that the eigenvalues (4.10) depend on n, but this dependence is eliminated when taking the $-1 \leq cos \leq 1$. This is a simplification which hides n from the bounds (4.11). Perhaps we can find bounds which include n, but this isn't crucial because we reasonably expect that the dependence disappears as n grows. After all,

$$\begin{aligned} &\text{for } k = 1 \quad \lim_{n\to\infty} \frac{\pi}{n+1} = 0 \quad \text{and} \quad \lim_{n\to\infty} \cos\left(\frac{k\pi}{n+1}\right) = 1 \\ &\text{for } k = n \quad \lim_{n\to\infty} \frac{n\pi}{n+1} = \pi \quad \text{and} \quad \lim_{n\to\infty} \cos\left(\frac{n\pi}{n+1}\right) = -1 \end{aligned} \tag{4.12}$$

Thus, for large n, the bounds (4.11) are optimal for eigenvalues of $M'^\intercal M'$ and if indeed $M'^\intercal M'$ is robust to the considered perturbation, then these bounds are also optimal for the eigenvalues of $M^\intercal M$. We shall verify this numerically.

Since we have the closed-form solution, we can compute the eigenvalues of $M'^\intercal M'$ exactly. We can also compute the eigenvalues of $M^\intercal M$ numerically. In Figure 4.1, we can see an

example for $\rho = 3$ and $n = 15$. The eigenvalues of the perturbed matrix are very close to the eigenvalues obtained via the closed-form solution for the tridiagonal Toeplitz matrix, which confirms the previous speculations about the sensitivity of the eigenvalues of $M'^{\intercal}M'$. Also, the lower bound ρ^2 and the upper bound $\rho^2 + 4\rho + 4$ already seem tight even for relatively small n.

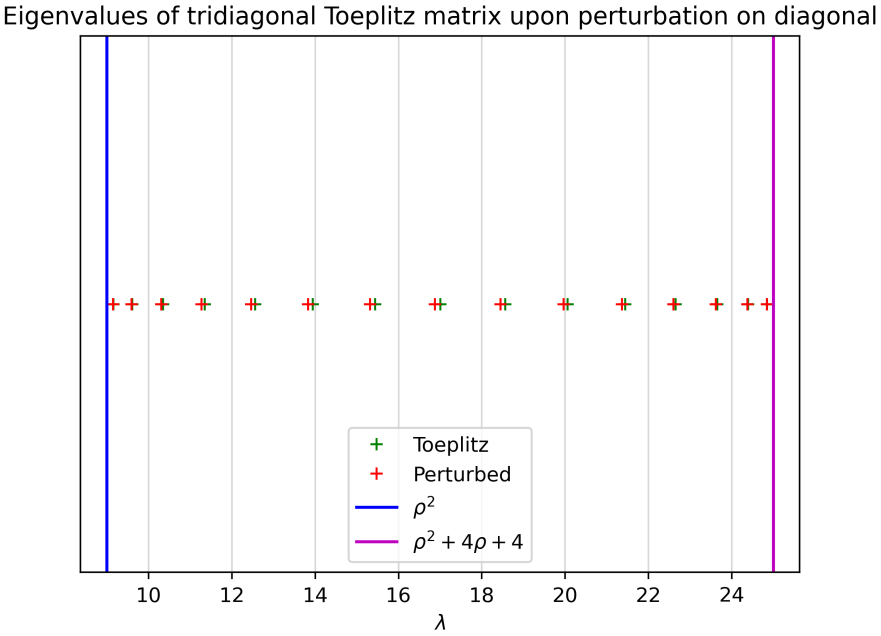

Figure 4.1: The eigenvalues for $\rho = 3$ and $n = 15$. The green *Toeplitz* is $M'^{\intercal}M'$ matrix, while the red *Perturbed* is $M^{\intercal}M$ matrix.

We want to confirm the optimally of the bounds for larger n and more values of ρ. In Figure 4.2, we can observe the largest eigenvalues of $M^{\intercal}M$ and $M'^{\intercal}M'$ for 3 different values of ρ, plotted for the increasing matrix size n.

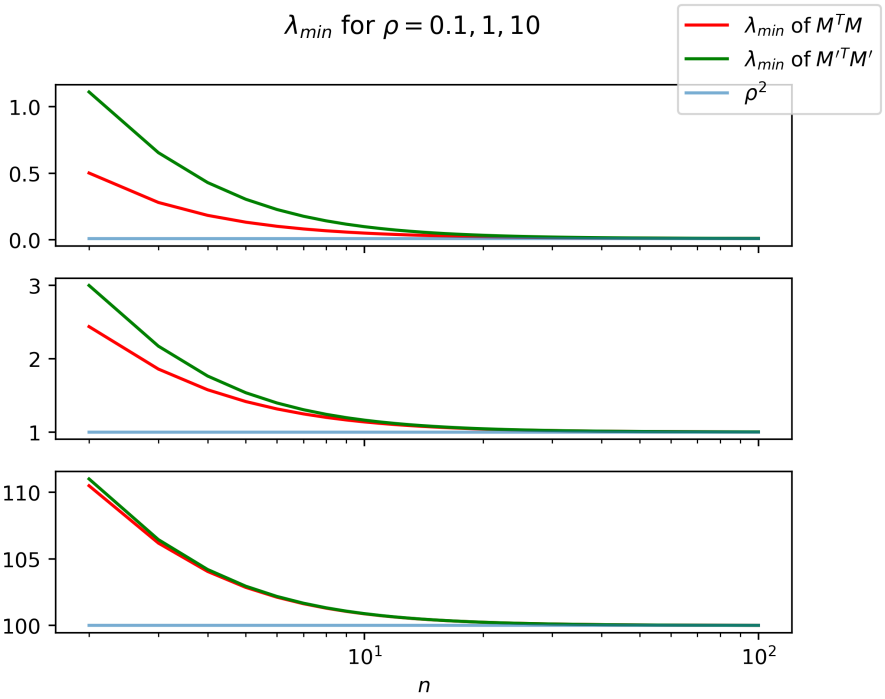

Figure 4.2: The smallest eigenvalues of $M^\intercal M$ and $M'^\intercal M'$ for varying ρ and n.

Similar plots were made for the largest eigenvalues of $M^\intercal M$ and $M'^\intercal M'$ and analogical results were obtained (plot omitted because it's like Figure 4.2, only flipped vertically). For large n, the perturbation turns out to have negligible effect even for small ρ.

Overall, for large matrices the largest eigenvalue of $M^\intercal M$ coincides with the largest eigenvalue of $M'^\intercal M'$, the smallest eigenvalue of $M^\intercal M$ coincide with the smallest eigenvalue of $M'^\intercal M'$. Moreover, the lower bound ρ^2 and the upper bound $\rho^2 + 4\rho + 4$ proposed by Lemma 4 are not only reasonable, but the best ones for sufficiently large n.

4.3 The constant of metric subregularity

The property of metric subregularity described in Section 2.1 will be employed in the analysis. Our main task in this section is determining the constant of metric subregularity. We begin with a necessary result that relates the Euclidean norm of a symmetric matrix and its smallest absolute eigenvalue:

Proposition 9 (Symmetric matrix norm inequality)**.** Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix and let $\lambda_1, ..., \lambda_n \in \mathbb{R}$ be its eigenvalues. We have that for all $x \in \mathbb{R}^n$ and for all $\lambda \in \lambda_1, ..., \lambda_n$,

$$\|Ax\|_2 \geq |\lambda| \, \|x\|_2 \tag{4.13}$$

That is, denoting the smallest eigenvalue as $\lambda_s = \min_{\lambda\in\{\lambda_1,...,\lambda_n\}}|\lambda|$, $\|Ax\|_2 \geq \lambda_s\|x\|_2$.

Proof. Since A is symmetric, spectral theorem applies and there exists a unique orthonormal basis formed by eigenvectors $v_1, \ldots, v_n$ of A. We obtain the spectral decomposition

of A:

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^\intercal \tag{4.14}$$

The outer products $v_i v_i^\intercal$ are the orthogonal projections onto one-dimensional λ_i-eigenspace. Note that

$$\begin{aligned}
\|Ax\|^2 &= \left(\sum_{i=1}^{n} \lambda_i v_i v_i^\intercal x\right)^\intercal \left(\sum_{i=1}^{n} \lambda_i v_i v_i^\intercal x\right) \\
&= \sum_{i=1}^{n} \left(\lambda_i x^\intercal v_i v_i^\intercal \left(\lambda_i v_i v_i^\intercal x + \sum_{\substack{j=1 \\ j\neq i}}^{n} \lambda_j v_j v_j^\intercal x \right) \right) \\
&= \sum_{i=1}^{n} \lambda_i^2 x^\intercal v_i \underbrace{v_i^\intercal v_i}_{1} v_i^\intercal x + \sum_{\substack{j=1 \\ j\neq i}}^{n} \lambda_i \lambda_j x^\intercal v_i \underbrace{v_i^\intercal v_j}_{0} v_j^\intercal x \\
&= \sum_{i=1}^{n} \lambda_i^2 (x^\intercal v_i)(v_i^\intercal x) = \sum_{i=1}^{n} \lambda_i^2 (v_i^\intercal x)^2
\end{aligned}$$

Also, given that any x can by expressed as $\sum_{i=1}^{n} (v_i^\intercal x) v_i$ in the orthonormal base formed by eigenvectors $v_1, \ldots, v_n$ (i.e. as vector projections of x onto v_i),

$$\begin{aligned}
\|x\|^2 = x^\intercal x &= \left(\sum_{i=1}^{n} v_i (v_i^\intercal x)\right)^\intercal \left(\sum_{i=1}^{n} v_i (v_i^\intercal x)\right) \\
&= \sum_{i=1}^{n} \left((x^\intercal v_i) v_i^\intercal \left(v_i (v_i^\intercal x) + \sum_{\substack{j=1 \\ j\neq i}}^{n} v_j (v_j^\intercal x) \right) \right) \\
&= \sum_{i=1}^{n} (x^\intercal v_i) \underbrace{v_i^\intercal v_i}_{1} (v_i^\intercal x) + \sum_{\substack{j=1 \\ j\neq i}}^{n} (x^\intercal v_i) \underbrace{v_i^\intercal v_j}_{0} (v_j^\intercal x) \\
&= \sum_{i=1}^{n} (x^\intercal v_i)(v_i^\intercal x) = \sum_{i=1}^{n} (v_i^\intercal x)^2
\end{aligned}$$

Now, we get an inequality

$$\|Ax\|^2 = \sum_{i=1}^{n} \lambda_i^2 (v_i^\intercal x)^2 \geq \min_{j \in \{1,..,n\}} \lambda_j^2 \sum_{i=1}^{n} (v_i^\intercal x)^2 = \min_{j \in \{1,..,n\}} \lambda_j^2 \|x\|^2 \tag{4.15}$$

Which after taking the square root finishes the proof. □

Remark. In matrix notation, (4.14) is equivalent to $A = Q \Lambda Q^{-1}$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is a similar matrix to A, but in the basis formed by eigenvectors $v_1, \ldots, v_n$ of A, and Q is the orthogonal $n \times n$ matrix of whose i-th column is the eigenvector v_i of A.

Recall that the primal optimization problem is $\min_x \frac{\mu}{2}\|x\|_2^2 + \mathrm{I}_{\{b\}}(Mx)$, which has metrically subregular gradient of Lagrangian, i.e.

$$\|x - x^*\| \leq \kappa \|\nabla \mathcal{L}(x)\|.$$

Our goal now is to establish the constant of metric subregularity κ for our difficult minimax optimization problem. The Lagrangian can be found analytically as:

$$\mathcal{L}(x, y) = \frac{\mu}{2}\|x\|_2^2 + \langle y, Mx - b\rangle \tag{4.16}$$

The gradients of Lagrangian with respect to the primal variable x and dual variable y are respectively:

$$\nabla_x \mathcal{L}(x, y) = \mu x + M^* y \tag{4.17}$$
$$\nabla_y \mathcal{L}(x, y) = Mx - b \tag{4.18}$$

Together, we write

$$\nabla \mathcal{L}(x, y) = (\mu x + M^\intercal y, Mx - b). \tag{4.19}$$

Define a block matrix

$$B := \begin{bmatrix} \mu \mathrm{I} & M^\intercal \\ M & 0 \end{bmatrix}. \tag{4.20}$$

Proposition 10. Matrix B is symmetric. This can be seen by inspecting the block structure of B:

$$B^\intercal = \begin{bmatrix} \mu \mathrm{I} & M^\intercal \\ M & 0 \end{bmatrix}^\intercal = \begin{bmatrix} \mu \mathrm{I} & M^\intercal \\ (M^\intercal)^\intercal & 0 \end{bmatrix} = B.$$

Proposition 11. B is invertible.

Proof. We use an identity for determinants of block matrices

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A) \det\left(D - CA^{-1}B\right), \tag{4.21}$$

which holds when A is invertible. This identity is the counterpart to the case with invertible D (Schur's determinant identity). The block matrix B has invertible upper left matrix $\mu \mathrm{I}$, hence the property (4.21) applies:

$$\det(B) = \det \begin{bmatrix} \mu \mathrm{I} & M^\intercal \\ M & 0 \end{bmatrix} = \det(\mu \mathrm{I}) \det\left(-M \mu \mathrm{I}^{-1} M^\intercal\right) \tag{4.22}$$

By $\det(AB) = \det(A)\det(B)$, we write

$$\det\left(-M \mu \mathrm{I}^{-1} M^\intercal\right) = \det(-M)\ \det\left(\mu \mathrm{I}^{-1}\right)\ \det(M^\intercal) \tag{4.23}$$

The inverse of the identity matrix is invertible. Matrix M is invertible by the Proposition 4, which also asserts that $M^\intercal$ is invertible. An invertible matrix has non-zero determinant, hence $\det(B)$ is equal to the product of 4 non-zero determinants. Thus, $\det(B) \neq 0$, which shows that B is invertible. □

Lemma 5. *Let $M \in \mathbb{R}^{n\times n}$ be a matrix from (4.4) with some $\rho > 0$, let $b \in \mathbb{R}^n$ be a vector from (4.6). Consider the Lagrangian (4.16) of the optimization problem $\min_x \frac{\mu}{2}\|x\|_2^2 + \mathrm{I}_{\{b\}}(Mx)$. Then, the gradient of the Lagrangian $\nabla\mathcal{L}(x,y) = (\mu x + M^\intercal y, Mx - b)$ is metrically subregular at its saddle point (x^*, y^*) for $(0,0)$ with metric subregularity constant given by*

$$\kappa = \frac{2}{\sqrt{\mu^2 + 4\rho^2} - \mu}$$

Proof. Using the fact that $\nabla\mathcal{L}$ at a saddle point (x^*, y^*) is necessarily 0, we obtain

$$\begin{aligned}\nabla\mathcal{L}(x,y) &= \nabla\mathcal{L}(x,y) - \nabla\mathcal{L}(x^*,y^*)\\ &= (\mu x + M^\intercal y - \mu x^* - M^\intercal y^*, Mx - Mx^*)\\ &= \begin{bmatrix}\mu\mathrm{I} & M^\intercal\\ M & 0\end{bmatrix}(x - x^*, y - y^*)\\ &= B(x - x^*, y - y^*)\end{aligned} \tag{4.24}$$

As B is symmetric (Proposition 10), we can use the inequality from the Proposition 9 to bound the Euclidean norm of B. Let $\lambda_1, ..., \lambda_{2n}$ be the eigenvalues of B. For all $\lambda \in \lambda_1, ..., \lambda_{2n}$ the inequality holds:

$$\|Bu\| \geq |\lambda|\,\|u\| \tag{4.25}$$

This inequality requires us to know the smallest eigenvalue in absolute value. If B contained null eigenvalues, the task would be more difficult because we would need to seek the smallest absolute eigenvalue which is non-zero. Fortunately, B is invertible (Proposition 11), which guarantees that it only contains non-null eigenvalues.

Ideally, we would find all the eigenvalues and take the smallest one in absolute value to use the inequality. The eigenvalues are roots of the characteristic polynomial $\det(B - \lambda\mathrm{I}_{2n})$, but solving the characteristic equation analytically is cumbersome given the dimension $2n \times 2n$ of the matrix

$$B - \lambda\mathrm{I} = \begin{bmatrix}(\mu - \lambda)\mathrm{I} & M^\intercal\\ M & -\lambda\mathrm{I}\end{bmatrix}$$

whose determinant would need to be found. It appears that there is no simple, straightforward way to find the eigenvalues of B without dealing with abstruse analytical formulations. Fortunately, we can leverage the block structure of B and the structure of matrices M and $M^\intercal$ that B contains.

Suppose $v \in \mathbb{R}^{2n}$ is a nontrival eigenvector of B with eigenvalue λ, hence $Bv = \lambda v$ holds. Decompose v into 2 blocks such that $v = (v_1 \quad v_2)^\intercal$, where $v_1, v_2 \in \mathbb{R}^n$. We obtain

$$\begin{bmatrix}\mu\mathrm{I} & M^\intercal\\ M & 0\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix} = \lambda\begin{bmatrix}v_1\\ v_2\end{bmatrix}$$

Which amounts to

$$\begin{aligned}\mu v_1 + M^\intercal v_2 &= \lambda v_1\\ Mv_1 &= \lambda v_2.\end{aligned}$$

Thus, with $v_2 = \lambda^{-1}Mv_1$ we have

$$M^\intercal M v_1 = (\lambda^2 - \mu\lambda)v_1.$$

Crucially, the tridiagonal matrix $M^\intercal M$ from (4.7) appears. We obtain that finding the eigenvalues of B amounts to determining the eigenvalues of $M^\intercal M \in \mathbb{R}^{n\times n}$ via relations

$$\lambda_{M^\intercal M} = \lambda_B^2 - \mu\lambda_B \iff \lambda_B = \frac{\mu}{2} \pm \frac{\sqrt{\mu^2 + 4\lambda_{M^\intercal M}}}{2} \tag{4.26}$$

In other words, the eigenvalues of $M^\intercal M$ are a reasonable proxy to the eigenvalues of B. Taking the largest eigenvalue of $M^\intercal M$, (4.26) gives

$$\frac{\mu}{2} - \frac{\sqrt{\mu^2 + 4\max\lambda_{M^\intercal M}}}{2} \le \lambda_B \le \frac{\mu}{2} + \frac{\sqrt{\mu^2 + 4\max\lambda_{M^\intercal M}}}{2} \tag{4.27}$$

Notice that the smallest eigenvalue of B is guaranteed to be negative because eigenvalues $\lambda_{M^\intercal M}$ are bounded away from zero. For the sake of use in the inequality (4.25), we need the smallest eigenvalue of B *in the absolute value*. As B has some negative eigenvalues, we know that $\min\lambda_B$ and $\min|\lambda_B|$ are different quantities.

Fortunately, we can bound $|\lambda_B|$ too. From (4.26) we deduce two intervals in which the eigenvalues λ_B must lie:

$$\frac{\mu}{2} - \frac{\sqrt{\mu^2 + 4\max\lambda_{M^\intercal M}}}{2} \le \lambda_B \le \frac{\mu}{2} - \frac{\sqrt{\mu^2 + 4\min\lambda_{M^\intercal M}}}{2} \tag{4.28}$$

$$\frac{\mu}{2} + \frac{\sqrt{\mu^2 + 4\min\lambda_{M^\intercal M}}}{2} \le \lambda_B \le \frac{\mu}{2} + \frac{\sqrt{\mu^2 + 4\max\lambda_{M^\intercal M}}}{2} \tag{4.29}$$

All we need is a property that $|a| \ge b \iff a \le -b$ or $a \ge b$. From the upper bound in (4.28), we get $|\lambda_B| \ge \frac{\sqrt{\mu^2+4\min\lambda_{M^\intercal M}}}{2} - \frac{\mu}{2}$. From the lower bound in (4.29), we get $|\lambda_B| \ge \frac{\mu}{2} + \frac{\sqrt{\mu^2+4\min\lambda_{M^\intercal M}}}{2}$. Therefore

$$\begin{aligned}\min|\lambda_B| &= \min\left\{\frac{\sqrt{\mu^2 + 4\min\lambda_{M^\intercal M}}}{2} - \frac{\mu}{2}, \frac{\sqrt{\mu^2 + 4\min\lambda_{M^\intercal M}}}{2} + \frac{\mu}{2}\right\} \\ &= \frac{\sqrt{\mu^2 + 4\min\lambda_{M^\intercal M}}}{2} - \frac{\mu}{2}\end{aligned}$$

The Section 4.2 discussed the eigenvalues of $M^\intercal M$. In particular, Lemma 4 provided $\min\lambda_{M^\intercal M} = \rho^2$, hence

$$\lambda_s := \min|\lambda_B| = \frac{\sqrt{\mu^2 + 4\rho^2}}{2} - \frac{\mu}{2} \tag{4.30}$$

As shown in the Section 4.2, for large n (e.g. above 100 which is typical) the bounds are precise hence the actual smallest absolute eigenvalue of B coincide with λ_s. The symmetric matrix norm inequality (4.13) ($\|Bu\| \ge \lambda_s \|u\|$) can be used now. Let $z = (x, v)$, $z^* = (x^*, v^*)$ and substitute $u = z - z^*$ hence

$$\|B(z - z^*)\| \ge \lambda_s\|z - z^*\| \tag{4.31}$$

Using that $B(z - z^*) = \nabla\mathcal{L}(x, v)$ from (4.24),

$$\|\nabla\mathcal{L}(z)\| \ge \lambda_s\|z - z^*\| \tag{4.32}$$

After rearranging we obtain that gradient of Lagrangian $\nabla\mathcal{L}$ is metrically subregular at saddle point z^* for 0:

$$\|z - z^*\| \le \lambda_s^{-1}\|\nabla\mathcal{L}(z)\| \qquad \forall z \in \mathcal{K} \tag{4.33}$$

Which is precisely the Definition 21 of metric subregularity with the operator $F = \nabla\mathcal{L}$ at $\bar{x} = z^*$ for $\bar{y} = 0$, with the constant of metric subregularity $\kappa = \lambda_s^{-1}$ depending on ρ and given by

$$\kappa = \frac{2}{\sqrt{\mu^2 + 4\rho^2} - \mu}$$

The original definition is expressed locally for points x in some neighborhood $\mathcal{N}(\bar{x})$. Here, the neighborhood of regularity is the whole space $\mathcal{K} = \mathcal{H} \oplus \mathcal{G}$ to which z^* belongs, hence $\nabla\mathcal{L}$ has global metric subregularity at z^* for 0. □

Both ρ (in the constraints matrix M) and μ are characteristics of the problem and we don't choose them. However, we have that $\mu > 0$ and assumption $\rho > 0$ is reasonable, in which case we note the range of the metric subregularity constant $\kappa \in (0, \infty)$. In Figure 4.3 we can observe how does κ change in function of ρ. For most of the possible values of ρ, the constant of metric subregularity is either huge (approaching ∞) or tiny (approaching 0). Moreover, the plot indicates that the constant of strong convexity μ doesn't impact κ greatly.

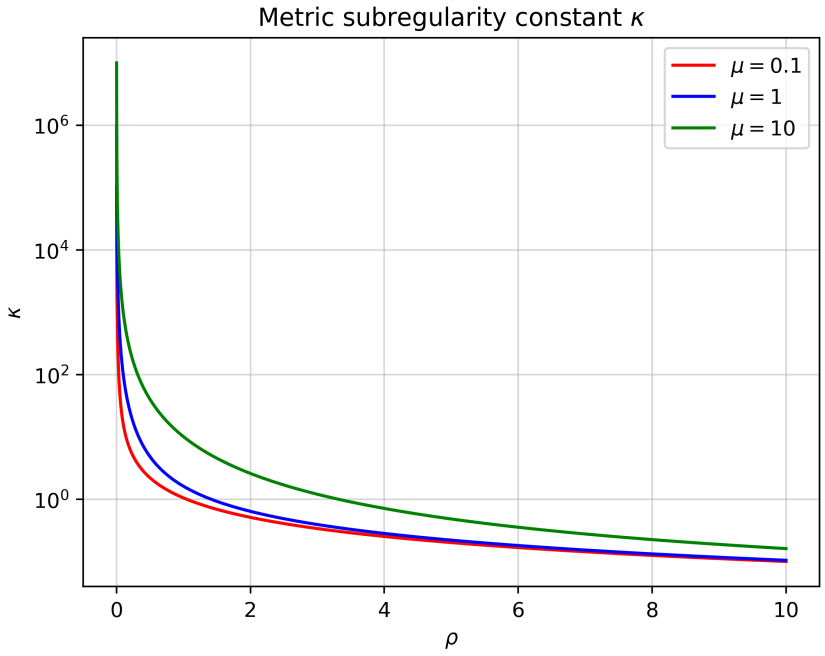

Figure 4.3: The constant of metric subregularity across ρ and various μ.

4.4 Explicit solution

Lemma 6. *The optimal solution of the optimization Problem (2.9) is*

$$x^* = \begin{bmatrix} \psi^{-1} & \psi^{-2} & \cdots & \psi^{-i} & \cdots & \psi^{-n} \end{bmatrix}^\intercal$$
$$y^* = \begin{bmatrix} C_t\psi^{-1} + \psi C_u & C_t\psi^{-2} + \psi^2 C_u & \cdots & C_t\psi^{-i} + \psi^i C_u & \cdots & C_t\psi^{-n} + \psi^n C_u \end{bmatrix}^\intercal$$

Where

$$C_t = \frac{\mu\psi^{n+1}}{-2(\frac{1}{\psi})^{n-2} + \psi^n + \psi^{n+2}}$$
$$C_u = \frac{\mu(\frac{1}{\psi})^{n+1}}{2(\frac{1}{\psi})^{n-2} - \psi^n - \psi^{n+2}}$$
$$\psi = \rho + 1$$

Proof. Define a function r based on the considered Problem (2.9)

$$r(x, y) := \frac{\mu}{2}\|x\|^2 + \langle Mx - b, y\rangle$$

The gradients of r can be found as:

$$\nabla_x r(x, y) = \frac{\mu}{2}\nabla_x\|x\|^2 + \nabla_x \langle Mx - b, y\rangle = \mu x + M^\intercal y$$
$$\nabla_y r(x, y) = \nabla_y \langle Mx - b, y\rangle = Mx - b$$

Letting the gradient to be 0, we obtain that

$$\mu x + M^\intercal y = 0 \tag{4.34}$$
$$Mx - b = 0 \tag{4.35}$$

Hence, the solution x^* verifies $Mx^* = b$ and $\mu x^* + M^\intercal y^* = 0$. Recall M from (4.4), b from (4.6), and $\psi = \rho + 1$. Solving the linear system $Mx^* = b$ amounts to solving equations

$$\begin{aligned} \psi x_1 &= 1 \\ -x_{i-1} + \psi x_i &= 0 \qquad \text{for } 2 \le i \le n \end{aligned}$$

The solution to this system is x^* with components $x_i^* = \psi^{-i}$ for $1 \le i \le n$.

Towards the dual solution y^*, substitute $x = -\mu^{-1}M^\intercal y$ from (4.34) into (4.35) to get

$$MM^\intercal y = -\mu b \tag{4.36}$$

Therefore, the solution y^* we look for must solve the set of linear equations (4.36). Recall the tridiagonal matrix $MM^\intercal$ from (4.8) and b from (4.6). We write $MM^\intercal y = -\mu b$ in the coordinate form as:

$$\begin{aligned} \psi^2 y_1 - \psi y_2 &= -\mu \\ -\psi y_{i-1} + (\psi^2 + 1)y_i - \psi y_{i+1} &= 0 \quad \text{for } 1 < i < n \\ -\psi y_{n-1} + (\psi^2 + 1)y_n &= 0 \end{aligned} \tag{4.37}$$

Which we solve to find y^*. Consider a quadratic equation

$$-\psi q^2 + (\psi^2 + 1)q - \psi = 0$$

Denote the roots of this quadratic as t and u. As the discriminant is $\sqrt{(\psi^2+1)^2 - 4\psi^2} = \sqrt{(\psi^2-1)^2} = \psi^2 - 1$, we find the roots

$$\begin{aligned} t &= \frac{-(\psi^2+1) + (\psi^2-1)}{-2\psi} = \frac{1}{\psi} \\ u &= \frac{-(\psi^2+1) - (\psi^2-1)}{-2\psi} = \psi \end{aligned} \tag{4.38}$$

We chose y such that the i-th coordinate is:

$$y_i = t^i C_t + u^i C_u \quad \text{for } i = 1, \dots, n \tag{4.39}$$

Knowing the coefficients C_t and C_u, the roots t and u can be used to find all the coordinates y_i of solution y^*. Developing the middle equations from (4.37) according to (4.39) gives

$$-\psi(t^{i-1}C_t + u^{i-1}C_u) + (\psi^2+1)(t^i C_t + u^i C_u) - \psi(t^{i+1}C_t + u^{i+1}C_u) = 0$$

Which for $i = 2$ states that

$$tC_t(-\psi t^2 + (\psi^2+1)t - \psi) + uC_u(-\psi u^2 + (\psi^2+1)u - \psi) = 0$$

As $-\psi t^2 + (\psi^2+1)t - \psi = 0$ and $-\psi u^2 + (\psi^2+1)u - \psi = 0$ by the fact that t and u are the roots of the quadratic $-\psi q^2 + (\psi^2+1)q - \psi = 0$, we obtain that for any values of the coefficients C_t and C_u, the equation for $i = 2$ is satisfied. The same reasoning and results apply for all the equations with i such that $1 < i < n$, thus there remains a system of 2 linear equations:

$$\begin{aligned} \psi^2 y_1 - \psi y_2 &= -\mu \\ -\psi y_{n-1} + (\psi^2+1) y_n &= 0 \end{aligned}$$

Again, expanding with $y_i = t^i C_t + u^i C_u$

$$\begin{aligned} -\psi^2(tC_t + uC_u) - \psi(t^2 C_t + u^2 C_u) &= -\mu \\ -\psi(t^{n-1}C_t + u^{n-1}C_u) + (\psi^2+1)(t^n C_t + u^n C_u) &= 0 \end{aligned}$$

We have a system of 2 equations, which we solve for C_t and C_u.

$$\begin{bmatrix} -\psi^2 t - \psi t^2 & -\psi^2 u - \psi u^2 \\ -\psi t^{n-1} + (\psi^2+1)t^n & -\psi u^{n-1} + (\psi^2+1)u^n \end{bmatrix} \begin{bmatrix} C_t \\ C_u \end{bmatrix} = \begin{bmatrix} -\mu \\ 0 \end{bmatrix}$$

Expanding the roots t and u according to (4.38), we obtain the system:

$$\begin{bmatrix} -\psi - \frac{1}{\psi} & -2\psi^3 \\ (\frac{1}{\psi})^n & \psi^{n+2} \end{bmatrix} \begin{bmatrix} C_t \\ C_u \end{bmatrix} = \begin{bmatrix} -\mu \\ 0 \end{bmatrix}$$

Which has the solution

$$C_t = \frac{\mu\psi^{n+1}}{-2(\frac{1}{\psi})^{n-2} + \psi^n + \psi^{n+2}} \tag{4.40}$$

$$C_u = \frac{\mu(\frac{1}{\psi})^{n+1}}{2(\frac{1}{\psi})^{n-2} - \psi^n - \psi^{n+2}} \tag{4.41}$$

□

As n goes to ∞, the constant C_u tends to 0, while C_t tends to a positive constant:

$$C_t \xrightarrow{n\to\infty} \frac{\mu\psi}{1+\psi^2} \tag{4.42}$$

The Lemma 6 gives the explicit solution for a problem where $f(x) = \frac{\mu}{2}\|x\|^2$ is minimized under affine constraints. We can consider the applicability of the Lemma in the case

without the f term, i.e. the minimization of $\mathrm{I}_b(Mx)$. The primal solution x^* doesn't depend on μ and would remain the same. On the other hand, the dual solution y^* would vanish altogether. To see this, consider the equation corresponding to (4.34), but in the case without the f term. It requires that the dual solution y^* verify $M^\intercal y^* = 0$. As $M^\intercal$ is invertible, $M^\intercal y^* = 0$ has only the trivial solution $y^* = 0$. Another way to see it, y^* depends on μ through the coefficients C_t and C_u, but these coefficients tend to 0 when μ goes to 0, a case which corresponds to the disappearing f term. Hence, y^* goes to 0 as well.

4.5 Complexity Lower Bound

We extend the Assumption 2 to take the dual variable into account, as well as the additional matrix multiplication operations.

Assumption 3. For all $k = 1, \ldots, n$, each point $x^{(k)}$ and $y^{(k)}$ generated by an algorithm $\mathcal{A}$ is a linear combination of $x^{(0)}, y^{(0)}, \nabla f\left(x^{(0)}\right), \ldots, \nabla f\left(x^{(k-1)}\right)$, and matrix multiplications $Mx^{(0)}, \ldots, Mx^{(k-1)}$ and $M^\intercal y^{(0)}, \ldots, M^\intercal y^{(k-1)}$, i.e. for all $\ell \in [0, k)$:

$$
\begin{aligned}
x^{(k)} &\in x^{(0)} + \operatorname{span}\left\{\nabla f\left(x^{(\ell)}\right), M^\intercal y^{(\ell)}\right\} \\
y^{(k)} &\in y^{(0)} + \operatorname{span}\left\{Mx^{(\ell)}\right\}
\end{aligned}
$$

The extension of the assumption is a result of the extended oracle. Instead of the vanilla first-order oracle $\mathcal{O}(x) = (\nabla f(x))$, we have

$$
\mathcal{O}(x, y) = (\nabla f(x), Mx, M^\intercal y)\,. \tag{4.43}
$$

Remark. The oracle (4.43) doesn't include the information from proximal operators. As we don't consider the g term, we cannot have prox_g. As f is smooth and oracle already has ∇f, the additional information from prox_f is unlikely to provide faster algorithms. We could consider the additional information from $\operatorname{prox}_{h^*}$, but we choose not to.

Lemma 7. *Let $x^{(0)} = 0 \in \mathbb{R}^n$ and $y^{(0)} = 0 \in \mathbb{R}^m$. Any k-th iterate produced by an algorithm $\mathcal{A}$ satisfying the Assumption 3 has at most k non-null components, i.e. for all $k \in [0, n]$ and for each $i > k$, $x_i^{(k)} = 0$ $y_i^{(k)} = 0$.*

Proof. We proceed with mathematical induction. The base case for $k = 0$ holds automatically by the initialization $x^{(0)} = 0$, $y^{(0)} = 0$. All the components are null hence the statement that the 0-th iterates have at most 0 non-null elements is true.

In the induction step we will take some $k < n$ and assume that each ℓ-th iterate with $\ell \in [0, k]$ has only null components after ℓ-th element, i.e. $\forall \ell \leq k, \forall i > \ell, x_i^{(\ell)} = 0, y_i^{(\ell)} = 0$. This assumption is Lemma's statement for the chosen k. Now, we need to show that this implies that the statement holds for $k + 1$.

If the chosen $k = 0$, $\nabla f(x^{(0)}) = \mathbf{0}$, $Mx^{(0)} = \mathbf{0}$, $M^\intercal y^{(0)} = \mathbf{0}$, which implies that $\forall i > 1, x_i^{(1)} = 0$ and $y_i^{(i)} = 0$ because these points are linear combinations of uniquely null elements. That is, given $k = 0$, $k + 1$ verifies Lemma's statement.

If the chosen $k > 0$,

$$\begin{aligned} \nabla f(x^{(k)}) = \mu x^{(k)} &\implies \forall i > k, \left(\nabla f\left(x^{(k)}\right)\right)_i = 0 \\ \left(Mx^{(k)}\right)_i = -x_{i-1}^{(k)} + (\rho+1)x_i^{(k)} &\implies \forall i > k+1, \left(Mx^{(k)}\right)_i = 0 \\ \left(M^\intercal y^{(k)}\right)_i = -y_{i+1}^{(k)} + (\rho+1)y_i^{(k)} &\implies \forall i > k, \left(M^\intercal y^{(k)}\right)_i = 0 \end{aligned} \tag{4.44}$$

Which implies that for all $i > k+1$, $x_i^{(k+1)}$ and $y_i^{(k+1)}$ are null because they are linear combinations of null elements $\left(\nabla f\left(x^{(k)}\right)\right)_i$, $\left(Mx^{(k)}\right)_i$, $\left(M^\intercal y^{(k)}\right)_i$ (which we know are null for $i > k+1$ by (4.44)). Overall, the iterates $x^{(k+1)}, y^{(k+1)}$ have at most $k+1$ non-null elements, which completes the proof by induction. □

Note that the formula for the geometric series starting at the index m is:

$$\sum_{i=m}^{n} ar^i = \frac{a\left(r^m - r^{n+1}\right)}{1-r}. \tag{4.45}$$

Proposition 12. Let $x \in \mathbb{R}^n$ and $k \in \{0, \dots, n\}$.

$$\sum_{i=1}^{n} (x_i)^2 \geq \sum_{i=k+1}^{n} (x_i)^2 \tag{4.46}$$

Holds because n non-negative terms are greater or equal to the $n-k$ subset of these terms.

Theorem 6 (Lower bound)**.** *Let $\rho > 0$. For any $k \in \left[1, \frac{1}{2}(n-1)\right]$, and any $x^{(0)} \in \mathbb{R}^n$, there exists an affinely constrained convex problem whose gradient of Lagrangian is metrically subregular with constant κ, such that any algorithm satisfying Assumption 3, we have:*

$$\left\|x^* - x^{(k)}\right\| \geq O\left(\left(\frac{\kappa}{\kappa + \sqrt{\mu\kappa + 1}}\right)^k\right). \tag{4.47}$$

Proof. We assume $x^{(0)} = 0$. If it weren't the case, we could shift all the objects in the space of variables and obtain $x^{(0)} = 0$ without loss of generality. Then,

$$\begin{aligned} \left\|x^{(0)} - x^*\right\|^2 = \sum_{i=1}^{n} (x_i^*)^2 &= \sum_{i=1}^{n} \left(\psi^{-2}\right)^i && \text{(Lemma 6)} \\ &= \frac{\psi^{-2}}{1-\psi^{-2}} - \frac{\psi^{-2(n+1)}}{1-\psi^{-2}} && \text{(Geometric progression (4.45))} \\ &\leq \frac{\psi^{-2}}{1-\psi^{-2}} = \frac{1}{\psi^2 - 1}. \end{aligned} \tag{4.48}$$

Also, $\rho > 0 \implies \psi > 1$ and $\psi < \psi^2 < \dots < \psi^k < \dots < \psi^{\frac{n}{2}}$ implies that

$$1 - \left(\frac{\psi}{\psi^n}\right)^2 > 1 - \left(\frac{\psi^2}{\psi^n}\right)^2 > \dots > 1 - \left(\frac{\psi^k}{\psi^n}\right)^2 > \dots > 1 - \left(\frac{\psi^{\frac{n}{2}}}{\psi^n}\right)^2 = 1 - \psi^{-n}.$$

Hence, for all $k \leq \frac{n}{2}$ we have

$$1 - \psi^{-2(n-k)} = 1 - \psi^{2k-2n} = 1 - \left(\frac{\psi^k}{\psi^n}\right)^2 > 1 - \psi^{-n}. \tag{4.49}$$

The main part of the proof is the following inequality:

$$\begin{aligned}
\left\|x^* - x^{(k)}\right\|^2 = \sum_{i=1}^{n} \left(x_i^* - x_i^{(k)}\right)^2 &\geq \sum_{i=k+1}^{n} \left(x_i^* - x_i^{(k)}\right)^2 && \text{(Proposition 12)} \\
&= \sum_{i=k+1}^{n} (x_i^*)^2 && \text{(Lemma 7)} \\
&= \sum_{i=k+1}^{n} \psi^{-2i} && \text{(Lemma 6)} \\
&= \psi^{-2(k+1)} \frac{1 - \psi^{-2(n-k)}}{1 - \psi^{-2}} && \text{(Geometric progression (4.45))} \\
&\geq \psi^{-2(k+1)} \frac{1 - \psi^{-n}}{1 - \psi^{-2}} && \text{(Using (4.49))} \\
&= \psi^{-2k} \left(\frac{\psi^{-2} - \psi^{-2-n}}{1 - \psi^{-2}}\right) \\
&= \psi^{-2k} \left(\frac{1 - \psi^{-n}}{\psi^2 - 1}\right) \\
&= \psi^{-2k} \frac{1}{\psi^2 - 1} \left(1 - \psi^{-n}\right) \\
&\geq \psi^{-2k} \left\|x^{(0)} - x^*\right\|^2 \left(1 - \psi^{-n}\right) && \text{(Using (4.48))}.
\end{aligned}$$

Taking the square root of both sides of the inequality leads to

$$\left\|x^* - x^{(k)}\right\| \geq \psi^{-k} \left\|x^{(0)} - x^*\right\| \sqrt{1 - \psi^{-n}}.$$

Expanding ψ, we get

$$\left\|x^* - x^{(k)}\right\| \geq \left(\frac{1}{1+\rho}\right)^k \left\|x^{(0)} - x^*\right\| \sqrt{1 - (\rho+1)^{-n}}.$$

From Lemma 5, the constant of metric subregularity is

$$\kappa = \frac{2}{\sqrt{\mu^2 + 4\rho^2} - \mu}.$$

The dependence of κ on positive ρ allows us to deduce

$$\rho = \frac{\sqrt{\mu\kappa + 1}}{\kappa}.$$

Hence, the bound amounts to

$$\left\|x^* - x^{(k)}\right\| \geq \left(\frac{\kappa}{\kappa + \sqrt{\mu\kappa + 1}}\right)^k \left\|x^{(0)} - x^*\right\| \sqrt{1 - \left(\frac{\kappa}{\kappa + \sqrt{\mu\kappa + 1}}\right)^n}.$$

Overall, we obtain an exponential lower bound

$$\left\|x^* - x^{(k)}\right\| \geq O\left(\left(\frac{\kappa}{\kappa + \sqrt{\mu\kappa + 1}}\right)^k\right),$$

which proves that for the considered class of problems, the best possible rate of convergence for any algorithm is linear with the constant $\frac{\kappa}{\kappa + \sqrt{\mu\kappa+1}} \in (0, 1)$ depending on the constant of metric subregularity κ and the constant of strong convexity μ (which can be 0). □

4.6 Comparisons and discussion

For non-strongly convex problems with $\mu = 0$, the bound (4.47) is still linear:

$$\left\|x^* - x^k\right\| \geq O\left(\left(\frac{\kappa}{\kappa+1}\right)^k\right). \tag{4.50}$$

Hence $0 < \frac{\kappa}{\kappa+1} < 1$ and the convergence is indeed linear even for non-strongly convex, non-strongly concave problems.

Comparing this with the papers on lower bound analysis for affinely constrained problems (Section 2.4.2), we notice that in literature, the linear lower bound was possible only for strongly-convex, strongly-concave case of the minimax problem, whereas our exponential bound (4.50) hold for non-stronly convex, non-strongly concave problems, at the condition that $\nabla\mathcal{L}$ is metrically subregular.

The complexity bounds given in the Theorem 6 hold only if the number of steps k verifies $k \leq \frac{n}{2}$. This condition is easily satisfied if the problem is large and the number of dimensions n is sufficiently high. If n weren't large, other methods than dimension-free first order methods would likely be better, hence the assumption is reasonable. Moreover, similar condition on k are sometimes made in literature, as for instance in the Nesterov's theorem 2 where $k \in \left[1, \frac{1}{2}(n-1)\right]$.

Our lower bound analysis relies on the metric subregularity in the problem. Besides the iteration count k and the constant of strong convexity of f (which could be 0), the obtained bound (4.47) depends on the metric subregularity constant κ. This is quite different from the other lower bounds in the literature. None of the papers mentioned in the Section 2.4 leverages metric subregularity for the lower bound analysis, nor its bound depends on the constant of metric subregularity. Typically, the obtained bounds depend on the iteration count k and either the condition number $Q = \frac{L}{\mu}$ (for strongly convex, smooth problems as e.g. in the Theorem 3), or only μ, or only L.

The constant of metric subregularity is inherent to a given particular problem, not chosen by us. In general, finding the closed-form representation of κ is unfeasible and even the estimation of it is difficult. In the analysis, we deliberately chosen a specific problem, which not only is difficult in terms of the information transfer, but also allowed us to obtain the κ analytically. Notice that any tridiagonal matrix would satisfy the first of these conditions (the difficulty in terms of the information transfer), but we didn't pick an arbitrary tridiagonal matrix. We defined it as $M^\intercal M$ with consciously structured M (4.4) to preserve the tridiagonality of $M^\intercal M$, while getting the characteristic enabling the second of these conditions (to obtain κ). Namely, this characteristic was a high, tight lower bound on the spectrum of $M^\intercal M$ depending on M's parametrized entries, which then related to the spectrum of the matrix B linked to the gradient of Lagrangian whose constant κ we found. Also, the very representation of $\nabla\mathcal{L}$ via matrix B was possible because $\nabla\mathcal{L}$ happened to be linear for the considered Problem 1. For the more difficult problems (e.g. non-smooth term extension), the gradient of Lagrangian would no longer be linear hence the representation $\nabla\mathcal{L}(x, y) = B(x - x^*, y - y^*)$ fails and the whole technique would need to be rethought. On the whole, the use of metric subregularity in the lower bound analysis for problems whose gradient of Lagrangian is metrically subregular can work well and yield tight lower bounds, but finding κ is challenging and the technique we used to establish it is unlikely to easily extend to other problems.

An interesting extension of the Theorem 6 would be adding the lower bound on $\|Mx - b\|$. Finally, note that the performed lower bound analysis was done only in the deterministic case. A natural extension would be covering the stochastic case, as for instance Xie et al. [Xie et al. 2020] have done for the minimax optimization problems.

5 Numerical experiments

In this chapter, we choose a particular simple optimization problem to verify the theoretical results obtained in Chapters 3 and 4. We will apply Vu's primal-dual splitting algorithm to evaluate its convergence on the chosen problem, and compare it with the theoretical upper and lower bound guarantees. The discussed problem will be a simple one, but there are relevant applications linked to Machine Learning. These aren't considered, but only mentioned for the sake of motivation of the considered simple problem.

The task of basis pursuit is formulated as $\min_x \|x\|_1$ under constrains $Mx = b$ is a special case of LASSO

$$\min_x \frac{1}{2}\|Mx - b\|_2^2 + \lambda\|x\|_1 \tag{5.1}$$

with $\lambda = 0$. Another possible application are SVM. In fact, notice that $\min_\beta f(X\beta)+R(\beta)$ can be written as

$$\min_{\beta,z} f(z) + R(\beta) \quad \text{s.c.} \quad z = X\beta, \tag{5.2}$$

which is an affinely constrained problem which corresponds to the sparse SVM if R is the first norm. Another example is regression with complex regularizers.

In all these cases we have a non-smooth term. An extension with the g term in the Problem 2 would be required to handle these applications. In terms of algorithms and computations, not much would change because the primal-dual splitting methods were designed to work on non-smooth problems. They work on operators and there would be no big issue to have them operate on subdifferential operator rather than gradient. However, estimating the constant of metric subregularity would be challenging.

5.1 Efficient minimum norm solution

Let $M \in \mathbb{R}^{m\times n}$ with $m < n$ and $b \in \mathbb{R}^m$. Consider a system of linear equations $Mx = b$ with infinitely many solutions (underdetermined system). We seek the smallest norm solution which verifies these linear constraints. That is, our norm-minimization problem is:

$$\min_{x\in\mathbb{R}^n} \|x\|_2^2 \quad \text{subject to} \quad Mx = b \tag{5.3}$$

More abstractly, this problem can be formulated as minimization of $f(x) + h \circ M$ with the affine constraints covered in the linear composition term, i.e. $\min \|x\|^2 + \mathrm{I}_{\{b\}}(Mx)$. Thus, (5.3) corresponds to the Problem 1 with $f = \|x\|_2^2$.

On principle, this is an extremely simple problem which has a closed-form solution available. For instance, given

$$M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{5.4}$$

We find that the analytical solution to the set of constraints $Mx = b$ take the form with $x = -y - 1$, $z = 2$, and arbitrary z. Indeed, the system has 2 equations in 3 variables which corresponds to 2 planes whose intersection creates a line. Any of the points on this line satisfies the constraints, but the function $f(x) = \|x\|_2^2$ is minimized only for one such point. Namely, the point x^* which is the smallest solution of $Mx = b$.

From optimality condition we get that the solution to (5.3) is

$$x^* = M^T(MM^T)^{-1}b = M^+b, \tag{5.5}$$

where M^+ is the Moore–Penrose inverse.

The prime instance of 5.3 which we consider is the constraints in an exact way as described in the Section 4.1, i.e. we use the exact difficult problem that was used for the lower bound analysis, with square matrix M from (4.4) and b from (4.6). We typically fix the dimensions $n = m = 200$ for square M. We also test non-square M with $n = 200$, $m = 150$.

We split the experiments into two parts: the algorithmic part related to the upper bound and lower bound part.

5.1.1 Vu's algorithm and upper bounds

The iterates produced by the algorithm are indexed with $k \in [0, \dots, K]$. The problem (5.3) is a strongly convex optimization problem, with f term 2-strongly convex. Indeed, the constant of strong convexity of f is hidden, but it necessarily is $\mu = 2$. For the Lipschitz constant L, we have globally $L \geq 2$ (it follows from the Definition 9 of Lipschitz continuity of ∇f). It implies that ∇f is $\frac{1}{2}$-cocoercive.

Originally, ν-strongly monotone operator D from the general inclusion Problem 3 introduced constant ν, but since we simplified this inclusion problem (see Lemma 3), there is no longer ν. In practice, we fixed $\nu = 10^{500}$ to have something very close to infinity, but not the infinity itself because technically Vu's method requires that $\nu \in (0, +\infty)$. Consequently, $\beta = \frac{1}{L}$. We had to choose the parameters $\tau, \sigma > 0$ such that

$$\begin{gathered} \eta = \min\left\{\tau^{-1}, \sigma^{-1}\right\}\left(1 - \sqrt{\tau\sigma\left\|M\right\|^2}\right) \\ 2\eta\beta > 1 \end{gathered} \tag{5.6}$$

We also needed to pick a sequence of K positive relaxation parameters with θ_k such that for all k

$$0 < \theta_k \leq \frac{4\eta\beta - 1}{2\eta\beta}. \tag{5.7}$$

In total, we faced with the choice of $K + 2$ hyperparameters: K relaxation parameters θ_k, τ, and σ. The choice of these values is active research, and there are no certain values that are the best.

In general, $\theta_k > 1$ indicates overrelaxation and it could be beneficial, but we didn't explore this direction. We took $\theta_k = 1$ for all k, which indicates no relaxation. This is a common choice - the same one has been made e.g. for the classic primal-dual hybrid gradient method [Chambolle and Pock 2011a].

Also, we choose τ, σ in a way such that $2\eta\beta$ wasn't neighter of its extremes, i.e. neither approaching 1, nor very high values. These extreme cases were leading to degenerate behaviour associated with extreme values for the constant of averaging λ. Usually, we found that that very small values of τ, σ in order of 10^{-2} (usully between $0.02 - 0.03$) worked best. In such cases $2\eta\beta$ was above 10.

We assumed a fixed number of iterations and initialized numpy arrays for all the sequences involved in the inner workings of the algorithm: for any $k \in \mathbb{N}$, there are $x^{(k)}, y^{(k)}, p^{(k)} \in \mathbb{R}^n$ (primal space) and $v^{(k)}, q^{(k)} \in \mathbb{R}^m$ (dual space). The $k = 0$-th iterates set to random vectors.

The implemented routine to generate the sequences of primal and dual variables followed the simplified algorithm from (3.16). The two sequences in our prime interest were $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$.

We verified the output of Vu's algorithm using the analytical solution (5.5). In each of the three checked cases of $\rho = 0.1, 1, 5$, the iterates obtained with Vu were correct while checking with 10^{-5} absolute tolerance in difference any of the $n = 200$ components.

In Figures 5.1, 5.2, and 5.3, we can observe the convergence of all the i-th components of $x_i^{(k)}$ across iterations k for $\rho = 0.1$, $\rho = 1$, and $\rho = 5$. As expected, the problem was difficult for the minimization problem, especially the $\rho = 0.1$ case associated with higher metric subregularity (which is bad). We can see that most of the components decrease to 0, i.e. $x_i^{(k)} \approx 0$ for most of i. Recalling the explicit solution $x_i^* = (1 + \rho)^{-i}$ (from Lemma 6) makes it clear why does the plot for $\rho = 1$ has significantly more rapid decrease towards 0 for its components. The case for $\rho = 5$ decreased even more rapidly to zero, having only a few non-null comopnents in the solution. Higher valued $\rho = 5$ slowed down the convergence in comparision to the $\rho = 1$. Going back to Figure 4.3 representing κ in function of ρ, we can see that small and large ρ are linked with degenerate cases of κ.

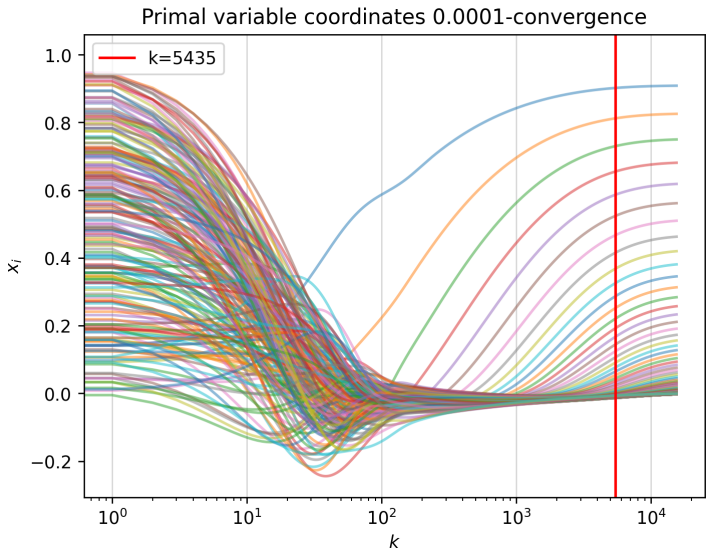

Figure 5.1: The $n = 200$ coordinates across the iterations for $\rho = 0.1$, $\kappa \approx 200$. The 10^{-4}-solution attained after 5435 iterations.

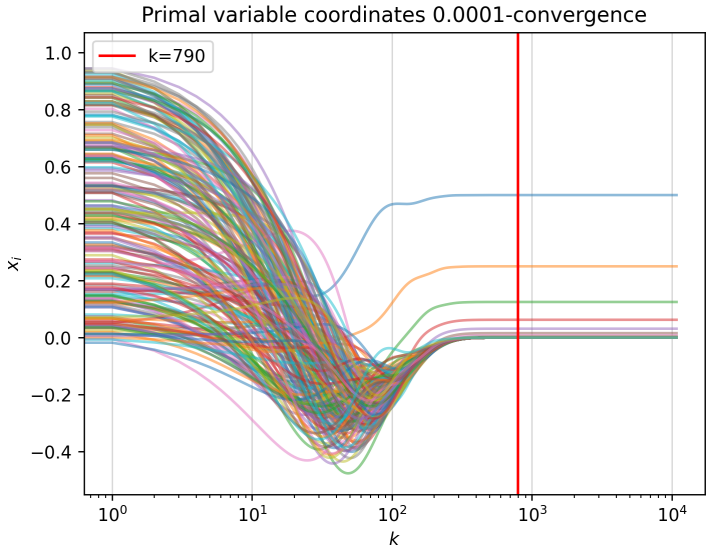

Figure 5.2: The $n = 200$ coordinates across the iterations for $\rho = 1$, $\kappa \approx 2.41$. The 10^{-4}-solution attained after 790 iterations.

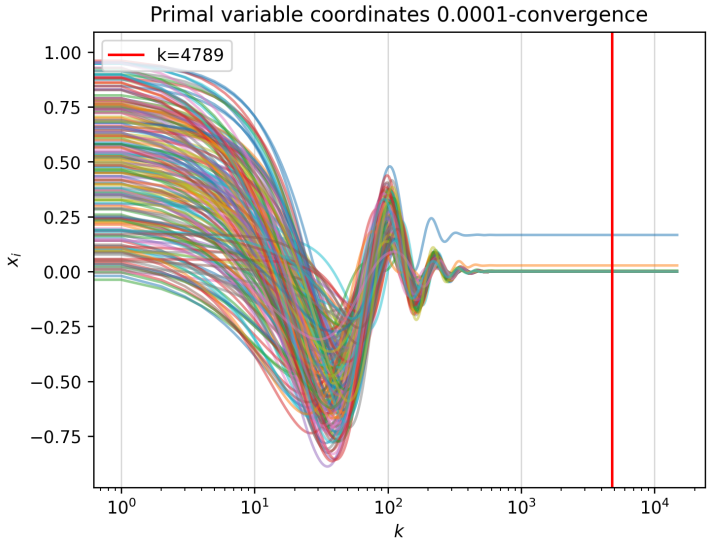

Figure 5.3: The $n = 200$ coordinates across the iterations for $\rho = 5$, $\kappa \approx 0.24$. The 10^{-4}-solution attained after 4789 iterations.

For the theoretical upper bound, we needed to define two quantities related to the Krasnoselskii-Mann iteration:

$$\lambda_k = \frac{1}{\theta_{max}}$$

and

$$\tau_k = \lambda_k(1 - \lambda_k) \tag{5.8}$$

Then, the Theorem 4 applies with the appropriate $\zeta \in [0, 1)$.

Consider the error defined as the norm of the distance from the iterate $x^{(k)}$ to the minimum x^*, i.e. $\left\|x^{(k)} - x^*\right\|$. In Figure 5.4 we see an examples of the theoretical linear

convergence compared with the obtained algorithm's rate. The gap between the lines is a result of the generality of the theoretical bound. The theoretical result holds for a class of problems where e.g. f is 2-strongly convex and ∇f is metrically subregular with constant 2.41. There are plenty of problems which belong to this class, and the theoretic line gives an upper bound on all of them (on the worst case). In practice, particular problem with these constants may be easier, thus it lies below the upper bound.

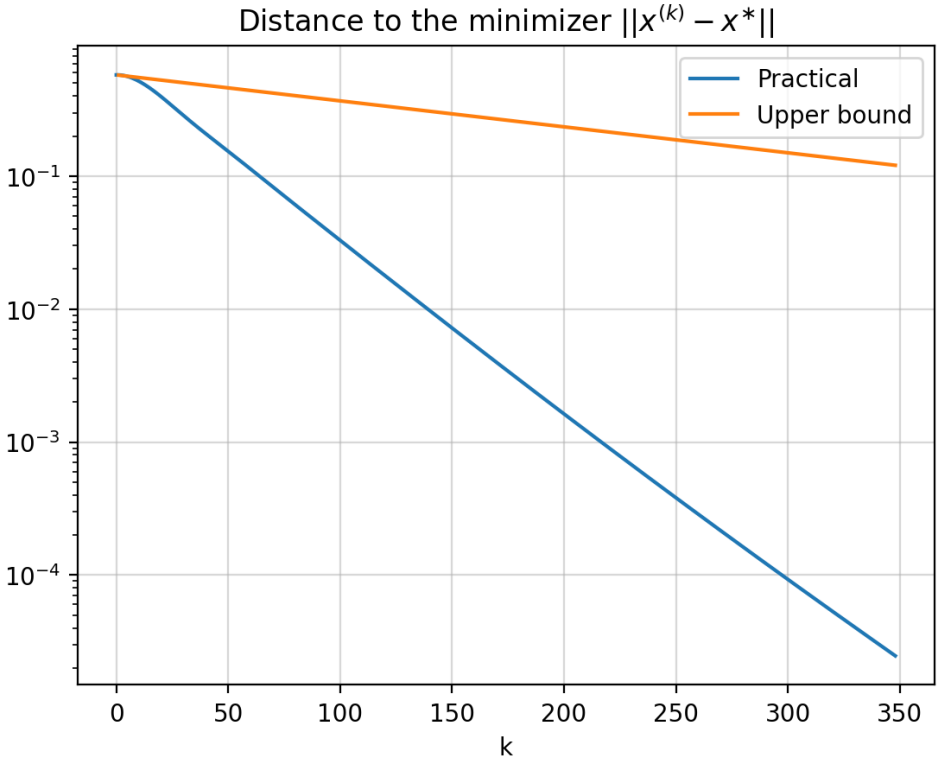

Figure 5.4: Practical Vu's method performance and theoretical upper bound ($\rho = 1$, $\kappa \approx 2.41$).

Also, alternative error using $||e_k||$ was considered, where the error of iteration $e_k = (z^{(k)} - z^{(k+1)})/\theta_k$. We chose $\theta_k = 1$, hence $e_k = z^{(k)} - z^{(k+1)}$. This error has benefit of describing the dual jointly with the primal, not only the x variable. The relevant plot was omitted because it's exactly like in Figure 5.4, only shifted.

5.1.2 Lower bound experiment

In Figure 5.5 we observe a significant gap between the lower bound (yellow) and the upper bound (green), within which lies the line associated with the obtained practical results (blue). We tested with different values of ρ (thus κ), to see if the result holds and the answer is affirmative. Therefore, the lower bound determines the best possible performance that could be attained. Both the theoretical results of the Vu algorithm, as the practical result are in the same order as the lower bound (linear), which is a big positive point. However, in terms of the slope, the method is far off from optimality. Likely new methods need to be developed to close the gap apparent on the plot.

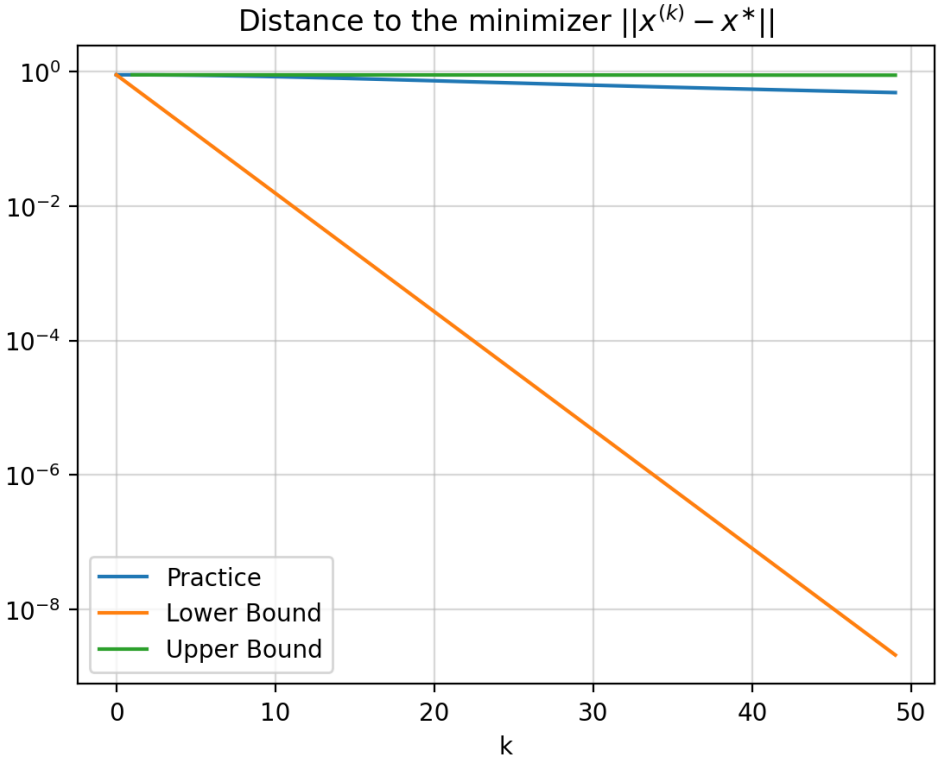

Figure 5.5: The comparison between the bounds ($\rho = 0.5, \kappa \approx 8.47$).

Only 50 iterations were showed without regard to the obtained accuracy ϵ of the solution because more iterations scale the plot down (due to the lower bound continuing to decrease rapidy) and the blue and green lines align.

The bound is decreasing rapidly and the gap appears big, which suggests that the lower bound is tight. However, to judge the quality of the gap, more examples from the same class would need to be constructed. Not only the class must be the same (i.e. Problem 1), but also when comparing problems we need to follow up on the constants that were fixed. In particular, we fix ρ, which fixed κ. To confirm the bound, we would need to construct a different problem from the same class with the same κ (and L and μ). For instance, we could constructed another matrix M with some new dependence on κ. Still, we suspect that the already proposed matrix M is as difficult as it can get, based on the information transfer argument.

Lastly, we conducted a quick experiment regarding efficiency. The analytical solution has a drawback of having a (pseudo) inverse, which is generally a costly operation. On the other hand, first-order methods such as Vu's algorithm are designed to perform well on large scale. Unfortunately, some tests quickly revealed that the implemented algorithm was slower than the inverse. This pitfall can be likely attributed to the use of Python (despite numpy) and insufficient optimization of the code itself.

6 Conclusions

In this work, a new lower bound was proposed, which holds for a class of affinely constrained convex optimization problems whose gradient of Lagrangian is metrically subregular. The novelty of this bound is its dependence on the constant of metric subregularity. The proposed lower bound demonstrated the linear order of convergence. The primal-dual splitting methods are known to match this linear convergence rate and we confirmed this. A new observation was made on the significant gap between the upper and lower bounds.

Four natural future direction would be possible:

- Closing the gap by the study of other splitting methods and proposing the new ones
- Extension of the problem class and incorporating non-smooth terms and establishing the corresponding lower and upper bounds under metric subregularity. This could be complemented with more practical applications such as sparse SVM.
- Improving the obtained exponential lower bound. Potentially there are more difficult problems, achievable by the change of f and M, which would lead to better linear convergence rates.
- Lastly, reconsidering all the results in the stochastic context with huge-scale optimization problems would be worthwhile.

Bibliography

Alacaoglu, A., O. Fercoq, and V. Cevher (2019). "On the convergence of stochastic primal-dual hybrid gradient". In: *arXiv preprint arXiv:1911.00799* (cit. on p. 19).

– (2020). "Random extrapolation for primal-dual coordinate descent". In: *arXiv preprint arXiv:2007.06528* (cit. on p. 19).

Bai, K., J. Ye, and J. Zhang (Jan. 2019). "Directional Quasi-/Pseudo-Normality as Sufficient Conditions for Metric Subregularity". In: *SIAM Journal on Optimization* 29, pp. 2625–2649. DOI: 10.1137/18M1232498 (cit. on p. 14).

Bauschke, H. H. and P. L. Combettes (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 1st. Springer Publishing Company, Incorporated. ISBN: 1441994661 (cit. on p. 16).

Boyd, S., N. Parikh, and E. Chu (2011). *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc (cit. on p. 18).

Bubeck, S. (#nov# 2015). "Convex Optimization: Algorithms and Complexity". In: *Found. Trends Mach. Learn.* 8.3–4, pp. 231–357. ISSN: 1935-8237. DOI: 10.1561/2200000050 (cit. on pp. 18, 21).

Chambolle, A., M. J. Ehrhardt, P. Richtárik, and C.-B. Schonlieb (2018). "Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications". In: *SIAM Journal on Optimization* 28.4, pp. 2783–2808 (cit. on p. 19).

Chambolle, A. and T. Pock (May 2011a). "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging". In: *J. Math. Imaging Vis.* 40.1, pp. 120–145. ISSN: 0924-9907. DOI: 10.1007/s10851-010-0251-1 (cit. on p. 50).

– (2011b). "A first-order primal-dual algorithm for convex problems with applications to imaging". In: *Journal of mathematical imaging and vision* 40.1, pp. 120–145 (cit. on p. 19).

Combettes, P. L. and J.-C. Pesquet (2012). "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators". In: *Set-Valued and variational analysis* 20.2, pp. 307–330 (cit. on p. 26).

Combettes, P. L. and V. R. Wajs (2005). "Signal recovery by proximal forward-backward splitting". In: *Multiscale Modeling & Simulation* 4.4, pp. 1168–1200 (cit. on p. 18).

Condat, L. (2013). "A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms". In: *Journal of Optimization Theory and Applications* 158.2, pp. 460–479 (cit. on p. 19).

Demanet, L. and X. Zhang (2016). "Eventual linear convergence of the Douglas-Rachford iteration for basis pursuit". In: *Math. Comput.* 85, pp. 209–238 (cit. on p. 25).

Dontchev, A. L. and R. T. Rockafellar (2009). *Implicit functions and solution mappings*. Vol. 543. Springer (cit. on p. 13).

Drusvyatskiy, D. and A. Lewis (2018). "Error Bounds, Quadratic Growth, and Linear Convergence of Proximal Methods". In: *Math. Oper. Res.* 43, pp. 919–948 (cit. on p. 14).

Eckstein, J. and D. P. Bertsekas (1992). "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators". In: *Mathematical Programming* 55.1-3, pp. 293–318 (cit. on p. 18).

Fonseca, C. da (Nov. 2006). "On the location of the eigenvalues of Jacobi matrices". In: *Applied Mathematics Letters* 19, pp. 1168–1174. DOI: 10.1016/j.aml.2005.11.029 (cit. on p. 33).

Hoffman, A. (1952). "On approximate solutions of systems of linear inequalities". In: *Journal of research of the National Bureau of Standards* 49, p. 263 (cit. on pp. 13, 14).

Ioffe, A. D. (1979). "Regular Points of Lipschitz Functions". In: *Transactions of the American Mathematical Society* 251, pp. 61–69. ISSN: 00029947 (cit. on p. 13).

Komodakis, N. and J. Pesquet (2015). "Playing with Duality: An overview of recent primal?dual approaches for solving large-scale optimization problems". In: *IEEE Signal Processing Magazine* 32.6, pp. 31–54 (cit. on p. 18).

Krasnosel'skiı, M. A. (1955). "Two remarks on the method of successive approximations". In: *Uspekhi Matematicheskikh Nauk* 10.1, p. 63 (cit. on p. 24).

Kruger, A. (Jan. 2015). "Error Bounds and Metric Subregularity". In: *Optimization* 64, pp. 49–79. DOI: 10.1080/02331934.2014.938074 (cit. on p. 14).

Kulkarni, D., D. Schmidt, and S.-K. Tsui (1999). "Eigenvalues of tridiagonal pseudo-Toeplitz matrices". In: *Linear Algebra and its Applications* 297.1, pp. 63–80. ISSN: 0024-3795. DOI: https://doi.org/10.1016/S0024-3795(99)00114-7 (cit. on p. 33).

Latafat, P., N. Freris, and P. Patrinos (June 2017). "A New Randomized Block-Coordinate Primal-Dual Proximal Algorithm for Distributed Optimization". In: (cit. on p. 14).

Latafat, P., N. M. Freris, and P. Patrinos (2019). "A new randomized block-coordinate primal-dual proximal algorithm for distributed optimization". In: *IEEE Transactions on Automatic Control* 64.10, pp. 4050–4065 (cit. on p. 19).

Liang, J., J. Fadili, and G. Peyré (#sep# 2016). "Convergence Rates with Inexact Non-Expansive Operators". In: *Math. Program.* 159.1–2, pp. 403–434. ISSN: 0025-5610. DOI: 10.1007/s10107-015-0964-4 (cit. on pp. 24–26).

Mann, W. (1953). "Mean value methods in iteration". In: (cit. on p. 24).

Moreau, J. J. (1962). "Fonctions convexes duales et points proximaux dans un espace hilbertien". In: *Comptes rendus hebdomadaires des séances de l'Académie des sciences* 255, pp. 2897–2899 (cit. on p. 9).

Necoara, I., Y. Nesterov, and F. Glineur (May 2019). "Linear Convergence of First Order Methods for Non-Strongly Convex Optimization". In: *Math. Program.* 175.1–2, pp. 69–107. ISSN: 0025-5610. DOI: 10.1007/s10107-018-1232-1 (cit. on p. 18).

Nemirovsky, A. (1992). "Information-based complexity of linear operator equations". In: *Journal of Complexity* 8.2, pp. 153–175. ISSN: 0885-064X. DOI: https://doi.org/10.1016/0885-064X(92)90013-2 (cit. on p. 20).

Nesterov, Y. (2018). *Lectures on Convex Optimization*. 2nd. Springer Publishing Company, Incorporated. ISBN: 3319915770 (cit. on pp. 18, 20, 21, 31).

Noschese, S., L. Pasquini, and L. Reichel (2013). "Tridiagonal Toeplitz matrices: properties and novel applications". In: *Numerical Linear Algebra with Applications* 20.2, pp. 302–326. DOI: 10.1002/nla.1811. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nla.1811 (cit. on p. 34).

Ogura, N. and I. Yamada (Jan. 2002). "Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping". In: *Numerical Functional Analysis and Optimization - NUMER FUNC ANAL OPTIMIZ* 23, pp. 113–137. DOI: 10.1081/NFA-120003674 (cit. on p. 11).

Ouyang, Y. and Y. Xu (2019). "Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems". In: *Mathematical Programming*, pp. 1–35 (cit. on p. 22).

Parikh, N. and S. Boyd (2014). "Proximal algorithms". In: *Foundations and Trends in optimization* 1.3, pp. 127–239 (cit. on p. 10).

Varga, R. S. (2010). *Geršgorin and his circles*. Vol. 36. Springer Science & Business Media (cit. on p. 32).

Vu, B. C. (#apr# 2013). "A Splitting Algorithm for Dual Monotone Inclusions Involving Cocoercive Operators". In: *Adv. Comput. Math.* 38.3, pp. 667–681. ISSN: 1019-7168. DOI: 10.1007/s10444-011-9254-8 (cit. on pp. 19, 24, 26).

Woodworth, B. and N. Srebro (2016). "Tight Complexity Bounds for Optimizing Composite Objectives". In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*. NIPS'16. Barcelona, Spain: Curran Associates Inc., pp. 3646–3654. ISBN: 9781510838819 (cit. on p. 22).

Xie, G., L. Luo, Y. Lian, and Z. Zhang (2020). "Lower Complexity Bounds for Finite-Sum Convex-Concave Minimax Optimization Problems". In: *ICML* (cit. on pp. 22, 31, 48).

Zhang, J., M. Hong, and S. Zhang (2019). *On Lower Iteration Complexity Bounds for the Saddle Point Problems*. arXiv: 1912.07481 [math.OC] (cit. on p. 22).