# Exact computation of an error bound for a generalized linear complementarity problem with unique solution

Jean-Pierre Dussault, Jean Charles Gilbert

# Exact computation of an error bound for a generalized linear complementarity problem with unique solution

Jean-Pierre Dussault [†] and Jean Charles Gilbert [‡]

Wednesday 20th October, 2021 (17:57)

This paper considers a generalized form of the standard linear complementarity problem with unique solution and provides a more precise expression of an upper error bound discovered by Chen and Xiang in 2006. This expression has at least two advantages. It makes possible the exact computation of the error bound factor and it provides a satisfactory upper estimate of that factor in terms of the data bitlength when the data is formed of rational numbers. Along the way, we show that, when any rowwise convex combination of two square matrices is nonsingular, the $\ell_\infty$ norm of the inverse of these rowwise convex combinations is maximized by an extreme diagonal matrix.

**Keywords:** Complexity, data bitlength, error bound, extreme diagonal matrix, linear complementarity problem, matrix inverse norm, P-matrix, rowwise convex combination of matrices, separable function, strong duality.

**AMS MSC 2020:** 15A09, 15A60, 49N15, 65F35, 65Y20, 90C33, 90C46, 90C60.

**Table of contents**



[†]Département d'Informatique, Faculté des Sciences, Université de Sherbrooke, Canada. E-mail: Jean-Pierre.Dussault@Usherbrooke.ca.

[‡]INRIA Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France. E-mail: Jean-Charles.Gilbert@inria.fr.

# 1 Introduction

Error bounds play a prominent role in the analysis of mathematical problems and the algorithms to solve them, in particular in numerical optimization [28]. This paper focuses on error bounds discovered by Chen and Xiang [10; 2006] for the linear complementarity problem with a **P**-matrix and simplifies the expression of its upper factor. The paper also deduces some consequences of this new expression.

In its standard form [11], the *linear complementarity problem* (LCP) reads

$$0 \leqslant x \perp (Mx + q) \geqslant 0, \tag{1.1}$$

where the unknown is $x \in \mathbb{R}^n$ (the set of real vectors with $n$ components), while $M \in \mathbb{R}^{n\times n}$ (the set of real matrices of order $n$) and $q \in \mathbb{R}^n$ are data. Inequalities on vectors must be understood componentwise (for example $x \geqslant 0$ in (1.1) means $x_i \geqslant 0$ for all $i \in [1\,{:}\,n]$, the set of the first $n$ integers). The compact writing of the problem in (1.1) means that one has to find a vector $x \in \mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x \geqslant 0\}$ such that $Mx + q \geqslant 0$ and $x^{\mathsf{T}}(Mx + q) = 0$ ("$^{\mathsf{T}}$" is used to denote vector or matrix transposition).

A matrix $M \in \mathbb{R}^{n\times n}$ is said to be a **P**-*matrix* if all its principal minors are positive (i.e., the determinant $\det M_{II} > 0$, for all $I \subseteq [1\,{:}\,n]$; by convention $\det M_{\varnothing\varnothing} = 1$). One denotes by **P** the class of **P**-matrices. It is known that problem (1.1) has a unique solution, whatever $q$ is, if and only if $M \in \mathbf{P}$ [32; 1958]. There are many other characterizations of the **P**-matricity [11], including algorithmic ones [2, 3].

For the sake of generality and for taking advantage of its symmetric formulation, which allows us to shorten some proofs, this paper considers an LCP in a slightly more general form than (1.1), namely

$$0 \leqslant (Ax + a) \perp (Bx + b) \geqslant 0, \tag{1.2}$$

where $A$, $B \in \mathbb{R}^{n\times n}$ and $a$, $b \in \mathbb{R}^n$ are the data (see for example [11, 27, 33]). Throughout this work, we assume that problem (1.2) has a unique solution $\bar{x}$. Conditions on $A$ and $B$ ensuring this property, whatever the vectors $a$ and $b$ are, will be recalled in proposition 3.1.

An error bound associated with a set $S$ is an estimate of the distance to $S$ by quantities that are easier to evaluate than this distance, usually those that are used to define the set. The set considered in this paper is the solution set of the LCP (1.2), which has been said to be reduced to the singleton $\{\bar{x}\}$, while the quantity used to estimate the distance to $\bar{x}$ is defined as follows.

Let $\|\cdot\|$ denote an arbitrary norm on $\mathbb{R}^n$. The *natural residual* [22, 23] associated with the linear complementarity problem (1.2) is the function $r : \mathbb{R}^n \to \mathbb{R}^n$ whose value at $x \in \mathbb{R}^n$ is given by

$$r(x) := \min(Ax + a, Bx + b), \tag{1.3}$$

where the minimum operator "min" acts componentwise (for two vectors $u$, $v \in \mathbb{R}^n$ and $i \in [1\,{:}\,n]$: $[\min(u, v)]_i = \min(u_i, v_i)$). It is clear that $x$ solves (1.2) if and only if $r(x) = 0$, since $\min(Ax + a, Bx + b) = 0$ if and only if, for all $i \in [1\,{:}\,n]$, $(Ax + a)_i \geqslant 0$, $(Bx + b)_i \geqslant 0$ and either $(Ax + a)_i$ or $(Bx + b)_i$ vanishes. Therefore $\|r(x)\|$ is a possible measure of the proximity of $x$ to $\bar{x}$. In this paper, we consider error bounds of the form

$$\forall\, x \in \mathbb{R}^n : \quad \check{\beta}\, \|r(x)\| \leqslant \|x - \bar{x}\| \leqslant \beta\, \|r(x)\|,$$

where $\check{\beta}$ and $\beta$ are positive constants (independent of $x$), that we call the *lower* and *upper error bound factors*, respectively. Error bounds for the LCP have been the subject of many contributions, see [30, 26, 24, 22, 23, 21, 10, 9, 12, 20, 8], the references therein, as well as the subsequent papers citing them.

For $P$ and $Q \in \mathbb{R}^{n\times n}$, we define

$$[P, Q] := \{X \in \mathbb{R}^{n\times n} : P \leqslant X \leqslant Q\},$$

where the inequalities act again componentwise (i.e., $P \leqslant X \leqslant Q$ means $P_{ij} \leqslant X_{ij} \leqslant Q_{ij}$ for all $i, j \in [1:n]$). Hence, for the identity matrix $I$, $[0, I]$ is a compact notation for the set of diagonal matrices with diagonal elements in the interval $[0, 1]$. Note also that the set of *extreme points* of $[0, I]$, denoted by $\mathrm{ext}[0, I]$, is the set of diagonal matrices with diagonal elements in $\{0, 1\}$ (see [31; p. 162] for the definition of an extreme point of a convex set; one can use [15; proposition 2.12] for a rigorous proof of this claim).

For $D \in [0, I]$, we denote by

$$C_D := (I - D)A + DB \tag{1.4}$$

the *rowwise convex combination* of the matrices $A$ and $B \in \mathbb{R}^{n\times n}$. We show in proposition 3.1 that the LCP (1.2) has a unique solution whatever the vectors $a$ and $b$ are, if and only if

$$\forall\, D \in [0, I] : \quad C_D \text{ is nonsingular.} \tag{1.5}$$

Therefore, this assumption is made throughout this paper. When this assumption holds, we show in proposition 3.2, as a straightforward extension of [10; 2006, (2.3)], that the following lower and upper error bounds hold:

$$\forall\, x \in \mathbb{R}^n : \quad \left(\max_{D\in[0,I]} \|C_D\|\right)^{-1} \|r(x)\| \leqslant \|x - \bar{x}\| \leqslant \left(\max_{D\in[0,I]} \|C_D^{-1}\|\right) \|r(x)\|, \tag{1.6}$$

where $\|\cdot\|$ denotes a norm on $\mathbb{R}^n$ and the induced matrix norm.

In this paper, we are interested in giving more precision on the way the lower and upper error bound factors appearing in (1.6) can be computed, when the $\ell_\infty$ norm is used. If the lower bound factor is easy to evaluate (see section 3.2), the upper bound factor

$$\beta := \max_{D\in[0,I]} \|C_D^{-1}\|_\infty \tag{1.7}$$

raises more difficulty. This concern makes perfect sense because, as far as we know, this upper error bound factor is the best one obtained so far for the LCP (1.1) with $M \in \mathbf{P}$; in particular, it is smaller, hence better, than the one of Mathias and Pang [24] (see [10; theorem 2.3]). We shall show that the evaluation of $\beta$ can be simplified since one has

$$\beta = \max_{D\in\mathrm{ext}[0,I]} \|C_D^{-1}\|_\infty. \tag{1.8}$$

This extends to higher dimension the simple observation that, when $n = 1$, the map $D \in [0, 1] \mapsto \|C_D^{-1}\|_\infty$ is monotone, so that it attains its maximum on $[0, 1]$ at a point in $\{0, 1\}$. For $n > 1$, however, $D_{kk} \in [0, 1] \mapsto \|C_D^{-1}\|_\infty$ may be nonmonotone (see the example in section 4.2), so that an analysis along this line is not straightforward. Furthermore, this

map can be neither convex nor concave (see example 3.3). For these reasons, we shall present a specific, rather long and indirect, proof of (1.8). The simplification (1.8) of (1.7) may look minor at first glance, but it may be interesting for reasons that are discussed in section 4.4: it simplifies the computation of $\beta$ for small $n$ and it may be crucial for giving an upper estimate of $\beta$ in terms of the data bitlength in some complexity analysis.

The paper is organized as follows. The next background section presents two results that will play an important role in getting the expression (1.8) of $\beta$: the first one deals with the norm of a matrix inverse and the second deals with min-max duality in optimization. Section 3 is dedicated to the proof of (1.8). Section 4 illustrates the result and its proof by several examples or counter-examples. We conclude by some thoughts on complexity issues.

### Notation

The unit closed ball associated with a norm $\|\cdot\|$ is denoted by $\bar{B} := \{x : \|x\| \leqslant 1\}$ and the unit sphere by $\partial B := \{x : \|x\| = 1\}$.

## 2 Background

*Once a result is discovered, one may then look for a more direct proof. When first hunting for certainty it is reasonable to use whatever tools one possess.*

Jonathan M. Borwein [6; 2016].

This section presents two results that will play a major part in our strategy to get the desired result in section 3. The first one (lemma 2.1) gives an expression of $\|\mathcal{A}^{-1}\|$, for a nonsingular matrix $\mathcal{A} \in \mathbb{R}^{n\times n}$, in terms of an optimization problem. Consequences of this expression for the $\ell_\infty$ norm are given in corollary 2.2 and in the technical lemma 2.3. The second result (lemma 2.4) highlights conditions to have strong duality on a product space $X \times [1:p]$ for a pairing function that has a separable property. These results could be found elsewhere, but presenting them here with the level of details that is needed below and with their proof is probably helpful for the reader.

### 2.1 Norm of a matrix inverse

For a given nonsingular matrix function $z \in \mathbb{R}^p \mapsto \mathcal{A}(z) \in \mathbb{R}^{n\times n}$, analyzing the map $\|[\mathcal{A}(\cdot)]^{-1}\|$ is often more difficult than analyzing $\|[\mathcal{A}(\cdot)\|$. It is possible, however, to toggle from one map to the other thanks to the following lemma.

**Lemma 2.1 (norm of a matrix inverse)** *If $\mathcal{A} \in \mathbb{R}^{n\times n}$ is a nonsingular square matrix and if $\|\cdot\|$ denotes a vector norm and its induced matrix norm, then*

$$\min_{\|v\|=1} \|\mathcal{A}v\| = \|\mathcal{A}^{-1}\|^{-1}. \tag{2.1a}$$

*In addition, $\bar{v}$ solves the problem in the left-hand side of (2.1a) if and only if $\bar{w} :=$*

$\|\mathcal{A}^{-1}\|\,\mathcal{A}\bar{v}$ *solves the problem in the left-hand side of*

$$\max_{\|w\|=1} \|\mathcal{A}^{-1}w\| = \|\mathcal{A}^{-1}\|. \tag{2.1b}$$

PROOF. 1) Let us first prove (2.1a). By the nonsingularity of $\mathcal{A}$, the following identity holds

$$\alpha := \min_{\|v\|=1} \|\mathcal{A}v\| = \min_{\|\mathcal{A}^{-1}w\|=1} \|w\|. \tag{2.2}$$

Note that $\alpha > 0$ by the nonsingularity of $\mathcal{A}$ and the compacity of $\partial B$. Then $\|\mathcal{A}^{-1}w\| \leqslant 1$ for any $w$ verifying $\|w\| = \alpha$ (since otherwise $\tilde{w} := w/\|\mathcal{A}^{-1}w\|$ would verify $\|\mathcal{A}^{-1}\tilde{w}\| = 1$ and $\|\tilde{w}\| = \|w\|/\|\mathcal{A}^{-1}w\| < \alpha$, contradicting the fact that $\alpha$ is the optimal value of the problem in right-hand side of (2.2)). This implies that

$$\max_{\|w\|=\alpha} \|\mathcal{A}^{-1}w\| = 1,$$

since any solution $w$ to the problem in the right-hand side of (2.2) is such that $\|w\| = \alpha$ and $\|\mathcal{A}^{-1}w\| = 1$. The identity (2.1a) now follows by

$$1 = \max_{\|w\|=\alpha} \|\mathcal{A}^{-1}w\| = \alpha \max_{\|w\|=1} \|\mathcal{A}^{-1}w\| = \alpha\,\|\mathcal{A}^{-1}\|.$$

2) Let $\bar{v}$ solve the problem in the left-hand side of (2.1a) and define $\bar{w} := \mathcal{A}\bar{v}/\alpha$. Then $\|\bar{w}\| = 1$ and $\|\mathcal{A}^{-1}\bar{w}\| = \|\bar{v}\|/\alpha = \|\mathcal{A}^{-1}\|$, which shows that $\bar{w}$ solves the problem in the left-hand side of (2.1b).

Reciprocally, suppose that $\bar{w}$ solves the problem in the left-hand side of (2.1b) and set $\bar{v} := \alpha\mathcal{A}^{-1}\bar{w}$. Then, $\|\bar{v}\| = \alpha\|\mathcal{A}^{-1}\bar{w}\| = 1$ and $\|\mathcal{A}\bar{v}\| = \alpha$, so that $\bar{v}$ solves the problem in the left-hand side of (2.1a). □

For the $\ell_2$-norm, the identity (2.1a) can also be obtained by using the relation between the smallest $\lambda_{\min}$ and the largest $\lambda_{\max}$ eigenvalues of inverse symmetric matrices:

$$\min_{\|v\|_2=1} \|\mathcal{A}v\|_2^2 = \lambda_{\min}(\mathcal{A}^{\mathsf{T}}\mathcal{A}) = \frac{1}{\lambda_{\max}(\mathcal{A}^{-1}\mathcal{A}^{-\mathsf{T}})} = \frac{1}{\|\mathcal{A}^{-\mathsf{T}}\|_2^2} = \|\mathcal{A}^{-1}\|_2^{-2}.$$

Nevertheless, the infinity norm is used below, as well as the link highlighted in lemma 2.1 between the vectors $\bar{v}$ giving the minimum in (2.1a) and the vectors $\bar{w}$ giving the maximum in (2.1b).

As we just said, in the sequel, the infinity vector and its induced matrix norms, both denoted by $\|\cdot\|_\infty$, are used. For this reason, we consider this case in corollary 2.2 below and bring some precision. We denote by $e^i$ the $i$th basis vector of $\mathbb{R}^n$ and set $e := \sum_{i\in[1:n]} e^i$, which is the vector of all ones. By definition and computation [18; § 5.6.5] (see also (2.8a)-(2.8d) in the proof of corollary 2.2 below), for a matrix $\mathcal{A} \in \mathbb{R}^{n\times n}$, one has

$$\|\mathcal{A}\|_\infty := \max_{\|w\|_\infty=1} \|\mathcal{A}w\|_\infty = \max_{i\in[1:n]} \|\mathcal{A}_{i:}\|_1, \tag{2.3}$$

where $\mathcal{A}_{i:} := (e^i)^\mathsf{T}\mathcal{A}$ denotes the $i$th row of $\mathcal{A}$ and $\|v\|_1 := \sum_{i\in[1:n]} |v_i|$ denotes the $\ell_1$-norm of $v \in \mathbb{R}^n$. We also denote by "sign" the maximal monotone multifunction $\mathbb{R} \multimap \mathbb{R}$ that is the subdifferential of the absolute value function: it associates with $t \in \mathbb{R}$ the following set of $\mathbb{R}$:

$$\operatorname{sign} t := \begin{cases} \{-1\} & \text{if } t < 0 \\ [-1,1] & \text{if } t = 0 \\ \{1\} & \text{if } t > 0. \end{cases} \tag{2.4}$$

One finds other definitions of sign(0), in particular to make it a single-valued function, but our choice of definition is important for the sequel, like in the formulas (2.5b) below. Recall that $\|\cdot\|_1$ is the dual norm of $\|\cdot\|_\infty$ with respect to the Euclidean scalar product, which means that

$$\|v\|_1 = \max_{\|w\|_\infty=1} v^\mathsf{T} w = \max_{\|w\|_\infty=1} |v^\mathsf{T} w|. \tag{2.5a}$$

The solution sets of these maximum problems are

$$\operatorname*{Arg\,max}_{\|w\|_\infty=1} v^\mathsf{T} w = (\operatorname{sign} v) \cap \partial B_\infty \quad \text{and} \quad \operatorname*{Arg\,max}_{\|w\|_\infty=1} |v^\mathsf{T} w| = (\pm \operatorname{sign} v) \cap \partial B_\infty, \tag{2.5b}$$

where, for a vector $v \in \mathbb{R}^n$, $\operatorname{sign} v := (\operatorname{sign} v_1) \times \cdots \times (\operatorname{sign} v_n) \subseteq \mathbb{R}^n$ (hence $\operatorname{sign} 0_{\mathbb{R}^n} = B_\infty$), $\pm \operatorname{sign} v := (\operatorname{sign} v) \cup (-\operatorname{sign} v)$ and the boundary $\partial B_\infty$ of $B_\infty$ is present only to deal with the case where $v = 0$. For a nonsingular square matrix $\mathcal{A}$, we adopt the following notation

$$\mathcal{W}_\infty(\mathcal{A}) := \operatorname*{Arg\,max}_{\|w\|_\infty=1} \|\mathcal{A}^{-1}w\|_\infty = \{w \in \partial B_\infty : \|\mathcal{A}^{-1}w\|_\infty = \|\mathcal{A}^{-1}\|_\infty\}, \tag{2.6a}$$

$$\mathcal{V}_\infty(\mathcal{A}) := \operatorname*{Arg\,min}_{\|v\|_\infty=1} \|\mathcal{A}v\|_\infty = \{v \in \partial B_\infty : \|\mathcal{A}v\|_\infty = \|\mathcal{A}^{-1}\|_\infty^{-1}\}. \tag{2.6b}$$

The second equality in (2.6a) comes from the definition of the induced matrix norm $\|\cdot\|_\infty$ in (2.3), while the second equality in (2.6b) is deduced from the identity (2.1a). The next corollary gives other expressions of these sets.

**Corollary 2.2 ($\ell_\infty$-norm of a matrix inverse)** *Suppose that $\mathcal{A} \in \mathbb{R}^{n\times n}$ is a nonsingular matrix. Set $\beta := \|\mathcal{A}^{-1}\|_\infty$ and $\alpha := 1/\beta$. Then,*

$$\mathcal{W}_\infty(\mathcal{A}) = \bigcup \left\{ \pm \operatorname{sign}(\mathcal{A}^{-\mathsf{T}} e^i) : \|(\mathcal{A}^{-1})_{i:}\|_1 = \beta \right\}, \tag{2.7a}$$

$$\mathcal{V}_\infty(\mathcal{A}) = \alpha\, \mathcal{A}^{-1}\Big(\mathcal{W}_\infty(\mathcal{A})\Big). \tag{2.7b}$$

Proof. [(2.7a)] Observe first that

$$\beta = \max_{\|w\|_\infty=1} \|\mathcal{A}^{-1}w\|_\infty \qquad \text{[definition of the matrix norm } \|\cdot\|_\infty\text{]} \tag{2.8a}$$

$$= \max_{\|w\|_\infty=1} \max_{i\in[1:n]} |(e^i)^\mathsf{T}\mathcal{A}^{-1}w| \qquad \text{[definition of the vector norm } \|\cdot\|_\infty\text{]} \tag{2.8b}$$

$$= \max_{i\in[1:n]} \max_{\|w\|_\infty=1} |(e^i)^\mathsf{T}\mathcal{A}^{-1}w| \qquad \text{[the two max's commute]} \tag{2.8c}$$

$$= \max_{i\in[1:n]} \|\mathcal{A}^{-\mathsf{T}} e^i\|_1 \qquad \text{[(2.5a)]}. \tag{2.8d}$$

We can now establish the identity (2.7a).

[$\subseteq$] If $\bar{w} \in \mathcal{W}_\infty(\mathcal{A})$, $\bar{w}$ solves the problem in (2.8a)-(2.8b), by definition. Let $\bar{\imath} \in [1:n]$ be a solution to the inner problem $\max\{|(e^i)^\mathsf{T}\mathcal{A}^{-1}\bar{w}| : i \in [1:n]\}$ appearing in (2.8b). Then, the pair $(\bar{w}, \bar{\imath})$ maximizes the map $(w, i) \in \partial B_\infty \times [1:n] \mapsto |(e^i)^\mathsf{T}\mathcal{A}^{-1}w|$. It follows that $\bar{\imath}$ solves to the problems in (2.8c)-(2.8d) and $\bar{w}$ is a solution to the inner problem $\max\{|(e^{\bar{\imath}})^\mathsf{T}\mathcal{A}^{-1}w| : \|w\|_\infty = 1\}$ appearing in (2.8c). Hence, by (2.5b), $\bar{w} \in \pm\operatorname{sign}(\mathcal{A}^{-\mathsf{T}}e^{\bar{\imath}})$ and, by (2.8d), $\beta = \|\mathcal{A}^{-\mathsf{T}}e^{\bar{\imath}}\|_1 = \|(\mathcal{A}^{-1})_{\bar{\imath}:}\|_1$.

[$\supseteq$] Suppose now that $\bar{w} \in \pm\operatorname{sign}(\mathcal{A}^{-\mathsf{T}}e^{\bar{\imath}})$ for some $\bar{\imath} \in [1:n]$ satisfying $\|\mathcal{A}^{-\mathsf{T}}e^{\bar{\imath}}\|_1 = \beta$.

- By this last identity, $\bar{\imath}$ solves the problem in (2.8d), hence the problem in (2.8c).
- By the nonsingularity of $\mathcal{A}^{-\mathsf{T}}$, one component of $\mathcal{A}^{-\mathsf{T}}e^{\bar{\imath}}$ does not vanish, so that $\bar{w} \in \pm\operatorname{sign}(\mathcal{A}^{-\mathsf{T}}e^{\bar{\imath}}) \cap \partial B_\infty$. By (2.5b), this implies that $\bar{w}$ solves the problem $\max\{|(e^{\bar{\imath}})^\mathsf{T}\mathcal{A}^{-1}w| : \|w\|_\infty = 1\}$.

It results from these last two observations and (2.8a)-(2.8c), that $\bar{w}$ solves the problem in (2.8a)-(2.8b). We have shown that $\bar{w} \in \mathcal{W}_\infty(\mathcal{A})$.

[(2.7b)] This is a consequence of the last claim in lemma 2.1, according to which $\bar{v} \in \mathcal{V}_\infty(\mathcal{A})$ if and only if $\bar{v} = \alpha\,\mathcal{A}^{-1}\bar{w}$ with $\bar{w} \in \mathcal{W}_\infty(\mathcal{A})$. □

We conclude this section by synthesizing in the following lemma a mechanism that, despite its innocuous appearance, plays a major part in the proof of proposition 3.4 below. As shown in the lemma's proof, this mechanism is only operational when some element of $\mathcal{A}^{-1}$ vanishes, but this fact is revealed indirectly, through a property of a vector $v \in \mathcal{V}_\infty(\mathcal{A})$.

**Lemma 2.3 (technical)** *Suppose that $\mathcal{A} \in \mathbb{R}^{n\times n}$ is nonsingular, that $\alpha := \|\mathcal{A}^{-1}\|_\infty^{-1}$ and that $v \in \mathcal{V}_\infty(\mathcal{A})$ has the property that $|(\mathcal{A}v)_k| < \alpha$ for some $k \in [1:n]$. Then, there exists a $v' \in \mathcal{V}_\infty(\mathcal{A})$ such that $(\mathcal{A}v')_k = 0$.*

Proof. Let $\beta := 1/\alpha$. Since $v \in \mathcal{V}_\infty(\mathcal{A})$, the vector defined by $w := \beta\mathcal{A}v$ is in $\mathcal{W}_\infty(\mathcal{A})$, by (2.7b). By assumption, $(\mathcal{A}v)_k \in (-\alpha, \alpha)$, so that $w_k \in (-1, 1)$. These two facts on $w$ and (2.7a) imply that there must be some index $i$ such that

$$w \in \pm\operatorname{sign}(\mathcal{A}^{-\mathsf{T}}e^i), \qquad \|(\mathcal{A}^{-1})_{i:}\|_1 = \beta \qquad \text{and} \qquad (\mathcal{A}^{-1})_{ik} = 0.$$

Define the vector $w' \in \mathbb{R}^n$ by vanishing the $k$th component of $w$:

$$w'_i := \begin{cases} w_i & \text{if } i \neq k \\ 0 & \text{otherwise.} \end{cases}$$

Then, we also have $w' \in \pm\operatorname{sign}(\mathcal{A}^{-\mathsf{T}}e^i)$, implying that $w' \in \mathcal{W}_\infty(\mathcal{A})$. The sought vector is $v' := \alpha\mathcal{A}^{-1}w'$. Indeed, on the one hand, $v' \in \mathcal{V}_\infty(\mathcal{A})$ by (2.7b). On the other hand, $\mathcal{A}v' = \alpha w'$ implying that $(\mathcal{A}v')_k = 0$, as desired. □

## 2.2 Strong duality for separable functions

Let be given a set $X$ and $p$ functions $\varphi_i : X \to \bar{\mathbb{R}}$. Usually, equality does not hold in the weak duality inequality [17, 16, 5, 15]

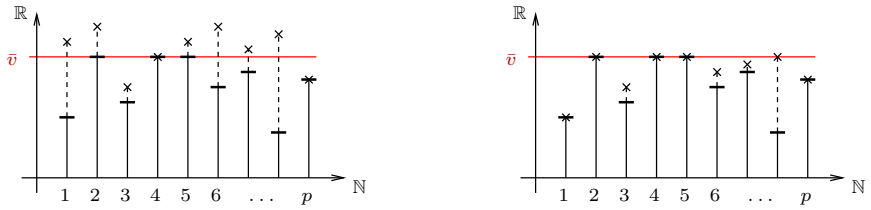
$$\inf_{x\in X} \max_{i\in[1:p]} \varphi_i(x) \geqslant \max_{i\in[1:p]} \inf_{x\in X} \varphi_i(x). \tag{2.9}$$

Take for example, $X = \mathbb{R}$, $p = 2$, $\varphi_1(x) = (x+1)^2$ and $\varphi_2(x) = (x-1)^2$, in which case the left-hand side value is 1, while the right-hand side value is 0 (see [16; lemma 4.5] for a way of modifying (2.9) that ensures equality). The situation is very different, more elementary and more favorable, when $X$ is a Cartesian product $X = X_1 \times \cdots \times X_p$ of sets $X_i$ and each function $\varphi_i$ only depends on the $i$th component $x_i \in X_i$ of $x = (x_1, \ldots, x_p) \in X$; then equality holds above with some other interesting properties. This particular situation, which occurs below, is analyzed in the next lemma. In this one, the problems

$$\inf_{x\in X} \max_{i\in[1:p]} \varphi_i(x_i) \qquad \text{and} \qquad \max_{i\in[1:p]} \inf_{x_i\in X_i} \varphi_i(x_i).$$

are called the *primal* and *dual* problems, respectively. A *primal* (resp. *dual*) *solution* is a solution to this primal (resp. dual) problem.

The following pictures illustrate this particular situation and help to have a good understanding of the next lemma, which is useful for the sequel.




The first $p$ integers $i$ are given in abscissa and the values of the $\varphi_i(x_i)$'s are given in ordinate. We shall see that there is no duality gap for this problem: the common primal and dual optimal value, denoted by $\bar{v}$, is represented by the horizontal (red) line in the two pictures. For each integer $i \in [1:p]$, the horizontal short bar gives the value $\inf\{\varphi_i(x_i) : x_i \in X_i\}$, which may or may not be equal to $\bar{v}$, but is always less than this optimal value, by the right-hand side of (2.10a). In the left-hand side picture, the crosses are the values $\varphi_i(x_i)$ for some $x \in X$ (hence, for a particular abscissa, a cross is always above the horizontal short bar), while in the right-hand side picture, the crosses are the values $\varphi_i(\bar{x}_i)$ for some primal solution $\bar{x} \in X$, which must not exceed $\bar{v}$ by left-hand side of (2.10a). The dual problem consists in determining the indices $i \in [1:p]$ for which the short bar values $\inf\{\varphi_i(x_i) : x_i \in X_i\}$ are maximal; in the pictures, the dual solution set is $\{2, 4, 5\}$. The primal problem consists in determining $x \in X$ such that the *maximum of the* $\varphi_i(x_i)$'s (the ordinate of the highest crosses) is as small as possible.

The next lemma not only shows the lack of duality gap for the separable case, but also describes the sets of primal and dual solutions. It also shows how to construct a primal solution from the dual solutions, as well as a dual solution from the primal solutions.

**Lemma 2.4 (strong duality for separable functions)** *Let $X := X_1 \times \cdots \times X_p$ be the Cartesian product of nonempty sets $X_i$ and let $\varphi_i : X_i \to \mathbb{R}$, $i \in [1:p]$, be arbitrary functions. An $x \in X$ is written $x = (x_1, \ldots, x_p)$, with $x_i \in X_i$ for $i \in [1:p]$.*

1) (No duality gap) *The following identity holds*

$$\inf_{x\in X} \max_{i\in[1:p]} \varphi_i(x_i) = \max_{i\in[1:p]} \inf_{x_i\in X_i} \varphi_i(x_i). \tag{2.10a}$$

*Denote by $\bar{v}$ the common value of the two sides of this identity.*

2) (Set of primal solutions) *The set of primal solutions is the possibly empty set $\bar{X} := \bar{X}_1 \times \cdots \times \bar{X}_p$, where*

$$\bar{X}_i := \{x_i \in X_i : \varphi_i(x_i) \leqslant \bar{v}\}. \tag{2.10b}$$

3) (Set of dual solutions) *The set of dual solutions is the nonempty set*

$$\bar{I} := \{i \in [1:p] : \varphi_i(x_i) \geqslant \bar{v} \text{ for all } x_i \in X_i\}. \tag{2.10c}$$

4) (Saddle-point property) *The following properties are equivalent:*

   *(i)* $(\bar{x}, \bar{\imath}) \in \bar{X} \times \bar{I}$,

   *(ii)* $(\bar{x}, \bar{\imath})$ *is a saddle-point of the map* $(x, i) \in X \times [1:p] \mapsto \varphi_i(x_i)$, *meaning that*

$$\forall\, (x, i) \in X \times [1:p] : \qquad \varphi_i(\bar{x}_i) \leqslant \varphi_{\bar{\imath}}(\bar{x}_{\bar{\imath}}) \leqslant \varphi_{\bar{\imath}}(x_{\bar{\imath}}), \tag{2.10d}$$

   *(iii)* $\bar{x}_{\bar{\imath}}$ *minimizes* $\varphi_{\bar{\imath}}$ *on* $X_{\bar{\imath}}$ *and* $\bar{\imath}$ *maximizes* $\varphi_i(\bar{x}_i)$ *on* $[1:p]$.

5) (Deducing a primal solution from the dual solutions) *Suppose that, for any dual solution $\bar{\imath} \in \bar{I}$, the problem $\inf\{\varphi_{\bar{\imath}}(x_{\bar{\imath}}) : x_{\bar{\imath}} \in X_{\bar{\imath}}\}$ has a solution $\hat{x}_{\bar{\imath}}$, then the primal problem has a solution $\bar{x} \in X$ satisfying*

$$\bar{x}_{\bar{I}} = \hat{x}_{\bar{I}} \qquad \text{and} \qquad \bar{I} = \operatorname*{Arg\,max}_{i \in [1:p]} \varphi_i(\bar{x}). \tag{2.10e}$$

6) (Deducing a dual solution from the primal solutions) *Suppose that $\bar{X} \neq \varnothing$. Then, $\bar{\imath} \in \bar{I}$ if and only if, for all $\bar{x} \in \bar{X}$, $\bar{\imath}$ maximizes $i \in [1:p] \mapsto \varphi_i(\bar{x}_i)$,*

PROOF. 1) By the weak duality property (2.9) and the fact that $\varphi_i$ only depends on the $i$th component of $x$, the inequality "$\geqslant$" holds in (2.10a). Let us prove the reverse inequality. Let $\varepsilon > 0$. For any $i \in [1:p]$, there is an $x_i^\varepsilon \in X_i$ such that

$$\varphi_i(x_i^\varepsilon) \leqslant \inf_{x_i \in X_i} \varphi_i(x_i) + \varepsilon.$$

Therefore,

$$\max_{i \in [1:p]} \varphi_i(x_i^\varepsilon) \leqslant \max_{i \in [1:p]} \inf_{x_i \in X_i} \varphi_i(x_i) + \varepsilon. \tag{2.11}$$

It is here that the separability assumption intervenes. Since the left-hand side of (2.11) is the value at $x^\varepsilon := (x_1^\varepsilon, \ldots, x_p^\varepsilon)$ of the function $x = (x_1, \ldots, x_p) \in X \mapsto \max_{i \in [1:p]} \varphi_i(x_i)$, the following inequality certainly holds

$$\inf_{x \in X} \max_{i \in [1:p]} \varphi_i(x_i) \leqslant \max_{i \in [1:p]} \varphi_i(x_i^\varepsilon).$$

Combining with (2.11), we get

$$\inf_{x \in X} \max_{i \in [1:p]} \varphi_i(x_i) \leqslant \max_{i \in [1:p]} \inf_{x_i \in X_i} \varphi_i(x_i) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the inequality "$\leqslant$" holds in (2.10a).

2) The point $\bar{x} = (\bar{x}_1, \dots, \bar{x}_p)$ is a primal solution if and only if

$$\max_{i\in[1:p]} \varphi_i(\bar{x}_i) \leqslant \inf_{x\in X} \max_{i\in[1:p]} \varphi_i(x_i) = \bar{v} \qquad \text{or} \qquad \forall\, i \in [1:p]:\ \varphi_i(\bar{x}_i) \leqslant \bar{v}.$$

This fact reads $\bar{x} \in \bar{X}$, for the given $\bar{X}$.

3) It is clear that $\bar{I}$ is nonempty, since the dual problem consists in taking the maximum of $p$ values in $\mathbb{R} \cup \{-\infty\}$. An index $\bar{\imath} \in [1:p]$ is a solution to the dual problem if and only if

$$\inf_{x_{\bar{\imath}}\in X_{\bar{\imath}}} \varphi_{\bar{\imath}}(x_{\bar{\imath}}) \geqslant \max_{i\in[1:p]} \inf_{x_i\in X_i} \varphi_i(x_i) = \bar{v} \qquad \text{or} \qquad \forall\, x_{\bar{\imath}} \in X_{\bar{\imath}}:\ \varphi_{\bar{\imath}}(x_{\bar{\imath}}) \geqslant \bar{v}.$$

This fact reads $\bar{\imath} \in \bar{I}$, for the given $\bar{I}$.

4) By (2.10a), there is no duality gap. Then, by the standard minmax duality theory [14; proposition VI.1.2] (see also [15; theorem 14.3]), the equivalence $(i) \Leftrightarrow (ii)$ follows. Now, condition $(iii)$ is just another way of expressing (2.10d).

5) According to (2.10c), for $i \notin \bar{I}$, there is an $\tilde{x}_i \in X_i$ such that $\varphi_i(\tilde{x}_i) < \bar{v}$. Let $\bar{x} \in X$ be defined by

$$\bar{x}_i = \begin{cases} \hat{x}_i & i \in \bar{I} \\ \tilde{x}_i & \text{otherwise.} \end{cases}$$

For $i \in \bar{I} \neq \varnothing$, one has

$$\begin{aligned} \varphi_i(\bar{x}_i) &= \varphi_i(\hat{x}_i) \qquad [\text{definition of } \bar{x}] \\ &= \inf\{\varphi_i(x_i) : x_i \in X_i\} \qquad [\text{assumption on } \hat{x}_i] \\ &= \bar{v} \qquad [i \in \bar{I} \text{ and definition of } \bar{v}]. \end{aligned}$$

And, for $i \notin \bar{I}$, one has

$$\begin{aligned} \varphi_i(\bar{x}_i) &= \varphi_i(\tilde{x}_i) \qquad [\text{definition of } \bar{x}] \\ &< \bar{v} \qquad [i \notin \bar{I} \text{ and definition of } \tilde{x}_i]. \end{aligned}$$

Since $\varphi_i(\bar{x}_i) \leqslant \bar{v}$ for all $i \in [1:p]$, (2.10b) shows that $\bar{x}$ is a primal solution. Furthermore, $\bar{x}_{\bar{\imath}} = \hat{x}_{\bar{\imath}}$ for all $\bar{\imath} \in \bar{I}$, so that the first identity in (2.10e) holds. We also have that $\operatorname{Arg\,max}\{\varphi_i(\bar{x}_i) : i \in [1:p]\} = \bar{I}$, which is the second identity in (2.10e).

6) The implication "$\Rightarrow$" follows from the implication $(i) \Rightarrow (iii)$ in point 4.

We prove the implication "$\Leftarrow$" by contraposition. Suppose that $\bar{\imath}$ is not a dual solution. Then, by (2.10c), one can find $\tilde{x}_{\bar{\imath}} \in X_{\bar{\imath}}$ such that $\varphi_{\bar{\imath}}(\tilde{x}_{\bar{\imath}}) < \bar{v}$. For some primal solution $\hat{x}$, which exists by assumption, we construct the point $\bar{x} \in X$ whose $i$th component is given by

$$\bar{x}_i = \begin{cases} \hat{x}_i & \text{if } i \neq \bar{\imath} \\ \tilde{x}_{\bar{\imath}} & \text{if } i = \bar{\imath}. \end{cases}$$

Since $\varphi_i(\bar{x}_i) \leqslant \bar{v}$ for all $i \in [1:n]$ (this is because $\varphi_i(\hat{x}_i) \leqslant \bar{v}$ for all $i \in [1:n]$ and $\varphi_{\bar{\imath}}(\tilde{x}_i) < \bar{v}$), the vector $\bar{x}$ is also a primal solution, by (2.10b). However, $\varphi_{\bar{\imath}}(\bar{x}_{\bar{\imath}}) = \varphi_{\bar{\imath}}(\tilde{x}_{\bar{\imath}}) < \bar{v} = \max_i \varphi_i(\bar{x}_i)$ shows that $\bar{\imath}$ does not maximize $i \in [1:p] \mapsto \varphi_i(\bar{x}_i)$. □

# 3 Finitely computable error bounds for the LCP

The goal of this section is to give some localization of the solution set of problem (1.7). More specifically, its main result, proposition 3.4, shows that a solution can always be found in $\mathrm{ext}[0, I]$, the set of extreme points of $[0, I]$ (one may find solutions with a diagonal element in $(0, 1)$ however; examples are given in sections 4.1 and 4.3).

## 3.1 On the generalized LCP

For the sake of precision and for the reader's convenience, we adapt to the generalized LCP (1.2) some results that are well known for the standard LCP (1.1). The next proposition gives conditions ensuring the uniqueness of the solution to (1.2) whatever the vectors $a$ and $b$ are; one of them (condition $(iv)$) is the general assumption (1.5). The equivalence $(i) \Leftrightarrow (ii)$ is related to [27; 1990, proposition 2], which considers another form of generalized LCP, but our proof is different. The equivalence $(ii) \Leftrightarrow (iv)$ extends [1; lemma 2.1], which assumes that $A$ is a positive diagonal matrix.

**Proposition 3.1 (well-posedness of the generalized LCP)** *The following properties are equivalent*

- $(i)$ *the LCP (1.2) has a unique solution whatever the vectors $a$ and $b$ are,*
- $(ii)$ *$A$ is nonsingular and $BA^{-1} \in \mathbf{P}$,*
- $(iii)$ *$B$ is nonsingular and $AB^{-1} \in \mathbf{P}$,*
- $(iv)$ *$(I - D)A + DB$ is nonsingular for all $D \in [0, I]$.*

Proof. $[(i) \Leftrightarrow (ii)]$ Let us show that $(i)$ implies that $A$ is nonsingular. Since $A$ is a square matrix, it suffices to prove its injectivity. Let $u$ be such that $Au = 0$. Consider the LCP in $x$:

$$0 \leqslant Ax \perp (Bx + |Bu|) \geqslant 0,$$

where $|Bu|$ is the vector made of the absolute values of the components of $Bu$. This LCP admits the solutions $x = 0$ and $x = u$. Hence $u = 0$ by the assumed uniqueness property of (1.2).

Assume now that $A$ is nonsingular. Then, the solutions $y$ to the following standard LCP

$$0 \leqslant y \perp (BA^{-1}y + b - BA^{-1}a) \geqslant 0. \tag{3.1}$$

are in bijection with the solutions $x$ to (1.2) through the relation $y = Ax + a$. Therefore, existence and uniqueness of the solution to (1.2) is equivalent to the existence and uniqueness of the solution to (3.1), a property that is known to be equivalent to $BA^{-1} \in \mathbf{P}$ (recalled after the definition of the $\mathbf{P}$-matricity in the introduction).

$[(i) \Leftrightarrow (iii)]$ This follows from the equivalence $(i) \Leftrightarrow (ii)$, by symmetry of the generalized LCP (1.2).

$[(ii) \Leftrightarrow (iv)]$ When $(iv)$ holds, $A$ is nonsingular (take $D = 0$ in $(iv)$). Now, when $A$ is nonsingular, $(iv)$ is equivalent to the property

$$\forall\, D \in [0, I] : \quad (I - D) + D(BA^{-1}) \text{ is nonsingular.}$$

By [1; lemma 2.1], this last property is equivalent to $BA^{-1} \in \mathbf{P}$. □

We now derive the lower and upper error bounds (1.6) for the LCP (1.2). This is a straighforward adaptation to problem (1.2) of the error bound of Chen and Xiang [10], who consider problem (1.1).

**Proposition 3.2 (error bounds for the generalized LCP)** *Consider the LCP (1.2). Denote by $\|\cdot\|$ an arbitrary norm on $\mathbb{R}^n$ and its induced matrix norm, denote by $r$ the natural residual (1.3) and define $C_D$ by (1.4). If (1.5) holds, then the lower and upper error bounds (1.6) hold.*

Proof. Chen and Xiang [10; 2006, § 2] start by observing that, for $u$, $v$, $\bar{u}$ and $\bar{v} \in \mathbb{R}^n$, one has

$$\min(u, v) - \min(\bar{u}, \bar{v}) = (I - D)(u - \bar{u}) + D(v - \bar{v}), \tag{3.2a}$$

where $D$ is the diagonal matrix, dependent on $(u, v, \bar{u}, \bar{v})$, that has its diagonal element $D_{ii}$, for $i \in [1:n]$, defined by

$$D_{ii} := \begin{cases} \text{arbitrary in } [0,1] & \text{if } u_i = v_i \text{ and } \bar{u}_i = \bar{v}_i, \\ 0 & \text{if } u_i \leqslant v_i \text{ and } \bar{u}_i \leqslant \bar{v}_i, \text{ with one strict inequality,} \\ 1 & \text{if } u_i \geqslant v_i \text{ and } \bar{u}_i \geqslant \bar{v}_i, \text{ with one strict inequality,} \\ \frac{u_i - \bar{u}_i - \min(u_i, v_i) + \min(\bar{u}_i, \bar{v}_i)}{(u_i - \bar{u}_i) - (v_i - \bar{v}_i)} & \text{otherwise.} \end{cases}$$

The four cases are clearly disjoint. The fourth and last case above corresponds to the cases where ($u_i \leqslant v_i$ and $\bar{u}_i > \bar{v}_i$) or ($u_i \geqslant v_i$ and $\bar{u}_i < \bar{v}_i$) ; in these cases, the denominator of the fraction defining $D_{ii}$ is nonzero. They observe that $D \in [0, I]$. Indeed, this is imposed in the first three cases above and for the fourth case, one has

- if $u_i \leqslant v_i$ and $\bar{u}_i > \bar{v}_i$, one has $D_{ii} = (\bar{u}_i - \bar{v}_i)/((\bar{u}_i - \bar{v}_i) + (v_i - u_i)) \in [0, 1]$ ;
- si $u_i \geqslant v_i$ and $\bar{u}_i < \bar{v}_i$, one has $D_{ii} = (u_i - v_i)/((u_i - v_i) + (\bar{v}_i - \bar{u}_i)) \in [0, 1]$.

Now, for a solution $\bar{x}$ to the LCP (1.2), define $u := Ax + a$, $v := Bx + b$, $\bar{u} := A\bar{x} + a$ and $\bar{v} := B\bar{x} + b$. Clearly, $\min(u, v) = r(x)$, $\min(\bar{u}, \bar{v}) = 0$ and, therefore, (3.2a) yields

$$r(x) = [(I - D)A + DB](x - \bar{x}) = C_D(x - \bar{x}). \tag{3.2b}$$

One deduces from this identity that

$$\|r(x)\| \leqslant \|C_D\| \, \|x - \bar{x}\|, \tag{3.2c}$$

Furthermore, since $D \in [0, I]$, $C_D$ is nonsingular by the assumption (1.5), so that one also deduces from (3.2b):

$$\|x - \bar{x}\| \leqslant \left\|C_D^{-1}\right\| \, \|r(x)\|, \tag{3.2d}$$

Of course $D$ in (3.2c) and (3.2d) depends on $x$ and $\bar{x}$, but, since $D \in [0, I]$, one certainly has (1.6) by taking the maximum in $D \in [0, I]$ in these two estimates. □

### 3.2 Computation of the lower error bound factor

Before focusing in section 3.3 on the main objective of this paper, which is the simplification of the upper error bound factor in (1.6), let us mention that the lower error bound factor in (1.6), with the $\ell_\infty$ norm, namely

$$\left(\max_{D\in[0,I]} \|C_D\|_\infty\right)^{-1}, \tag{3.3}$$

can be easily computed.

Observe first that, for two vectors $u$ and $v$ and a vector norm $\|\cdot\|$, one has

$$\max_{t\in[0,1]} \|(1-t)u+tv\|) = \max(\|u\|,\|v\|) = \max_{t\in\{0,1\}} \|(1-t)u+tv\|). \tag{3.4}$$

Next,

$$\begin{aligned}
\max_{D\in[0,I]} \|C_D\|_\infty &= \max_{D\in[0,I]} \max_{i\in[1:n]} \|(1-D_{ii})A_{i:}+D_{ii}B_{i:}\|_1 \qquad [(1.4) \text{ and } (2.3)]\\
&= \max_{i\in[1:n]} \max_{D_{ii}\in[0,1]} \|(1-D_{ii})A_{i:}+D_{ii}B_{i:}\|_1 \qquad [\text{the two max's commute}]\\
&= \max_{i\in[1:n]} \max(\|A_{i:}\|_1,\|B_{i:}\|_1) \qquad [(3.4)]\\
&= \max(\|A\|_\infty,\|B\|_\infty). \qquad (3.5)
\end{aligned}$$

This shows that (3.3) can be easily computed. Observe that, by (3.4), the maximum in (3.3) is obtained for $D\in \text{ext}[0,I]$. In view of (3.5), it is also obtained for $D=0$ or $D=I$.

### 3.3 Computation of the upper error bound factor

Before stating the precise result, let us consider the next example, which shows that the function that is maximized in (1.7) can be rather nonlinear; in particular, it may neither be convex nor concave. Therefore, the optimal value of the optimization problem in (1.7) and its solutions may not be easy to compute.

**Example 3.3 (nonlinearity of $D\mapsto \|C_D^{-1}\|_\infty$)** *For $A=I$ and $B\in\mathbf{P}$, the map $D\in[0,I]\mapsto \|C_D^{-1}\|_\infty$ has no guaranteed convexity or concavity property.*

Consider the LCP (1.2) with

$$A=I \qquad \text{and} \qquad B=\begin{pmatrix}1+\varepsilon & 1\\ 1 & 1\end{pmatrix}\in\mathbf{P},$$

where $\varepsilon>0$ is a small number. One computes easily

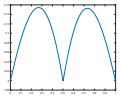
$$C_D=\begin{pmatrix}1+\varepsilon D_{11} & D_{11}\\ D_{22} & 1\end{pmatrix} \qquad \text{and} \qquad C_D^{-1}=\frac{1}{\Delta}\begin{pmatrix}1 & -D_{11}\\ -D_{22} & 1+\varepsilon D_{11}\end{pmatrix},$$

where $\Delta := 1+\varepsilon D_{11}-D_{11}D_{22}$ is the determinant of $C_D$, which is positive on $[0,I]$. Using (2.3), one gets

$$\|C_D^{-1}\|_\infty = \frac{\max(1+D_{11},1+\varepsilon D_{11}+D_{22})}{\Delta}.$$

By looking at $\|C_D^{-1}\|_\infty$ on the segment $D_{11} \in [0,1] \mapsto (D_{11}, 1-D_{11})$, the diagonal of the square, we get the maximum of two functions, which has the following form

$$D_{11} \in [0,1] \mapsto \frac{\max(1+D_{11}, 2-(1-\varepsilon)D_{11})}{1-(1-\varepsilon)D_{11}+D_{11}^2}.$$



For a small $\varepsilon > 0$, the first part of the max is a concave function of $D_{11} \in [0,1]$ around its maximum at $D_{11} = \sqrt{3} - 1 \simeq 0.7321$ (for $\varepsilon = 0$). The graph of this function is given in the figure above (for $\varepsilon = 10^{-3}$); note that the first part of the max gives the second arch in the figure. Therefore, for the given matrix $M$, the map $D \in [0, I] \mapsto \|C_D^{-1}\|_\infty$ is neither convex nor concave.

Actually, the maximum of $\|C_D^{-1}\|_\infty$ is obtained for $D = \mathrm{Diag}(1,1)$, which is not on the segment $D_{11} \in [0,1] \mapsto (D_{11}, 1-D_{11})$, on which the map presents two hills, as shown by the figure above. □

By the compactness of $[0, I]$ and the continuity of $D \in [0, I] \mapsto \|C_D^{-1}\|_\infty$, the maximization problem (1.7), recalled below

$$\beta := \max_{D \in [0,I]} \|C_D^{-1}\|_\infty, \tag{3.6}$$

has a solution, say $\bar{D}$. Since $C_{\bar{D}}$ is nonsingular, $\beta$ given by (3.6) is finite and positive. Then, one can also define the positive number

$$\alpha := \beta^{-1}. \tag{3.7}$$

The goal of this section is to show that the value $\beta$ can be obtained by restricting the feasible set of problem (3.6) to $\mathrm{ext}[0, I]$, the set of extreme diagonal matrices of $[0, I]$:

$$\beta = \max_{D \in \mathrm{ext}[0,I]} \|C_D^{-1}\|_\infty. \tag{3.8}$$

The key mechanism of the proof of the main of this paper, proposition 3.4 below, is based on lemma 2.3 and is illustrated in section 4.3 in the particular case where $A = I$ and $B = M$.

**Proposition 3.4 (validity of (3.8))** *Suppose that $A$ and $B \in \mathbb{R}^{n\times n}$ satisfy (1.5) and that $\bar{D}$ solves the optimization problem in (3.6). Then, if $\bar{D}_{kk} \in (0,1)$ for some $k \in [1\!:\!n]$, $\bar{D}$ remains optimal if $\bar{D}_{kk}$ is changed to any value in $[0,1]$. In particular, the value of $\beta$ defined by (3.6) is also given by (3.8).*

Before starting the analysis, let us observe that the objective of problem (3.6) is made of the composition of the nonlinear smooth function $D \mapsto C_D^{-1}$ and the convex function $\|\cdot\|_\infty$, but this objective is maximized, not minimized, so that the theory developed for the class of composite problems [7, 29, 4] does not apply. For this reason, we provide a specific proof of proposition 3.4. This one is postponed to page 16.

Part of the analysis is based on the following rewriting of $\beta$, defined by (3.6) (some more justifications are given after (3.9e), $C_D$ is defined by (1.5)):

$$\max_{D \in [0,I]} \|C_D^{-1}\|_\infty = \max_{D \in [0,I]} \left( \min_{\|v\|_\infty = 1} \|C_D v\|_\infty \right)^{-1} \qquad [(2.1a)] \tag{3.9a}$$

$$= \left( \min_{D\in[0,I]} \min_{\|v\|_\infty=1} \|C_D v\|_\infty \right)^{-1} \tag{3.9b}$$

$$= \left( \min_{\|v\|_\infty=1} \min_{D\in[0,I]} \|C_D v\|_\infty \right)^{-1} \qquad \text{[the two min's commute]} \tag{3.9c}$$

$$= \left( \min_{\|v\|_\infty=1} \min_{D\in[0,I]} \max_{i\in[1:n]} |(C_D v)_i| \right)^{-1} \qquad \text{[definition of } \|\cdot\|_\infty\text{]} \tag{3.9d}$$

$$= \left( \min_{\|v\|_\infty=1} \max_{i\in[1:n]} \min_{D_{ii}\in[0,1]} |(1-D_{ii})(Av)_i + D_{ii}(Bv)_i| \right)^{-1}, \tag{3.9e}$$

where we have been able to switch $\min_D$ and $\max_i$ from (3.9d) to (3.9e), without duality gap, thanks to point 1 of lemma 2.4 and the fact that $[0,I] = [0,1] \times \cdots \times [0,1]$ ($n$ times) is a Cartesian product and that $|(C_D v)_i| = |(1-D_{ii})(Av)_i + D_{ii}(Bv)_i|$ only depends on $D_{ii}$. In (3.9c), we have a minimum in $v$ (i.e., the infimum is attained), since by (3.9b) the function $(D,v) \mapsto \|C_D v\|_\infty$ has a minimizer $(\bar{D},\bar{v})$ on $[0,I] \times \partial B_\infty$, which implies that $\bar{v}$ solves the problem in (3.9c) (this property of nested optimization problems is discussed around [15; corollary 1.10]). Let us deduce some consequences of the identities in (3.9).

According to (3.6), the value of the left-hand side in (3.9a) is $\beta > 0$ and, according to (3.7), the optimal values of the optimization problems inside the parentheses in (3.9b)-(3.9c) is $\alpha > 0$, so that

$$\alpha = \min_{D\in[0,I]} \min_{\|v\|_\infty=1} \|C_D v\|_\infty, \tag{3.10a}$$

$$= \min_{\|v\|_\infty=1} \min_{D\in[0,I]} \|C_D v\|_\infty. \tag{3.10b}$$

Therefore, one can write

$$\boxed{\bar{D} \text{ solves (3.6)} \iff \exists\, \bar{v} \text{ such that } (\bar{D},\bar{v}) \text{ solves problems (3.10).}} \tag{3.11}$$

We also have

$$\left.\begin{array}{r} \bar{D} \text{ solves (3.6)} \\ \bar{v} \in \mathcal{V}_\infty(C_{\bar{D}}) \end{array}\right\} \iff (\bar{D},\bar{v}) \text{ solves problems (3.10).} \tag{3.12}$$

This is because, when $\bar{D}$ solves (3.6) and $\bar{v} \in \mathcal{V}_\infty(C_{\bar{D}})$ (i.e., $\bar{v}$ minimizes $\|C_{\bar{D}} v\|_\infty$ on $\partial B_\infty$ by (2.6b)), $(\bar{D},\bar{v})$ solves the problems in (3.10). Reciprocally, when $(\bar{D},\bar{v})$ solves the problems in (3.10), then $\bar{D}$ solves (3.6) by (3.11) and $\bar{v}$ minimizes $\|C_{\bar{D}} v\|_\infty$ on $\partial B_\infty$, which also reads $\bar{v} \in \mathcal{V}_\infty(C_{\bar{D}})$ by (2.6b).

Pursuing along the vein that exploits (3.9), we see that the optimal value of the optimization problems inside the parentheses in (3.9d)-(3.9e) is also $\alpha > 0$, so that, for a $\bar{v}$ such that $(\bar{D},\bar{v})$ solves the problems in (3.10) for some $\bar{D} \in [0,I]$, one has

$$\alpha = \min_{D\in[0,I]} \max_{i\in[1:n]} |(C_D \bar{v})_i|, \tag{3.13a}$$

$$= \max_{i\in[1:n]} \min_{D_{ii}\in[0,1]} |(1-D_{ii})\bar{v}_i + D_{ii}(M\bar{v})_i|. \tag{3.13b}$$

We shall also use the following implication:

$$\left.\begin{array}{r}(\bar{D},\bar{v}) \text{ solves problems (3.10)}\\ \bar{D}' \text{ solves (3.13a)}\end{array}\right\} \implies \left\{\begin{array}{l}(\bar{D}',\bar{v}) \text{ solves problems (3.10)}\\ \bar{D}' \text{ solve (3.6).}\end{array}\right. \tag{3.14}$$

Indeed, by the left-hand side of the implication, $(\bar{D},\bar{v})$ minimizes $\|C_D v\|_\infty$ on $[0,I]\times\partial B_\infty$ and $\bar{D}'$ minimizes $\|C_D\bar{v}\|_\infty$ on $[0,I]$. Then, $(\bar{D}',\bar{v})$ minimizes $\|C_D v\|_\infty$ on $[0,I]\times\partial B_\infty$ or, equivalently, $(\bar{D}',\bar{v})$ solves problems (3.10). Next, $\bar{D}'$ solves (3.6), by (3.11).

We conclude this preliminary discussion with an elementary lemma.

**Lemma 3.5 (elementary)** *Suppose that $\nu$ and $\mu\in\mathbb{R}$, that $\alpha>0$ and that*

$$\min_{\delta\in[0,1]} |(1-\delta)\nu+\delta\mu| = \alpha. \tag{3.15}$$

*Then, the solution set of the optimization problem in (3.15) is $\{0\}$, $\{1\}$ or $[0,1]$.*

Proof. By (3.15):

$$\forall\,\delta\in[0,1]: \quad |(1-\delta)\nu+\delta\mu| \geqslant \alpha.$$

Note first that $\nu$ and $\mu$ must not vanish and must have the same sign, since, otherwise, the minimum value in (3.15) would be zero, which would contradict $\alpha>0$. Then, three cases can occur.

- If $|\nu|=|\mu|$, then $\nu=\mu$ since $\nu$ and $\mu$ have the same sign. In that case, the objective of the optimization problem in (3.15) is the constant $|\nu|$, so that its solution set is $[0,1]$.
- If $|\nu|<|\mu|$, then the solution set of the optimization problem in (3.15) is $\{0\}$.
- If $|\nu|>|\mu|$, then the solution set of the optimization problem in (3.15) is $\{1\}$. □

Proof of proposition 3.4. Suppose that $A$ and $B\in\mathbb{R}^{n\times n}$ satisfy (1.5) and that $\bar{D}$ solves the optimization problem in (3.6). Since the last claim of the proposition is clear, we only focus on the first part of it, assuming that $\bar{D}_{kk}\in(0,1)$ for some $k\in[1\!:\!n]$. By (3.11)-(3.12), there is a $\bar{v}\in\mathcal{V}_\infty(C_{\bar{D}})$ such that $(\bar{D},\bar{v})$ solves the problems in (3.10). The goal of the proof is now to show that one can replace $\bar{D}_{kk}$ by any value in $[0,1]$, to form a diagonal matrix $\bar{D}'$ that is still a solution to (3.6). Sometimes (case 1 below), this goal will be reached with the chosen initial $\bar{v}$; other times (case 2 below), it will be necessary to change the optimal $\bar{D}$ and $\bar{v}\in\mathcal{V}_\infty(C_{\bar{D}})$ several times (infinitely often is not excluded, with a limit argument) to reach the goal. Before introducing these cases, we highlight the principal argument that is used in the proof.

*Principal argument.* Recall that the optimal value of (3.6) is denoted by $\beta:=\|C_{\bar{D}}^{-1}\|_\infty$, which is positive, and that $\alpha:=1/\beta=\|C_{\bar{D}}\bar{v}\|_\infty$ for the chosen $\bar{v}\in\mathcal{V}_\infty(C_{\bar{D}})$. Now, we want to determine the other values that the $k$th diagonal element $\bar{D}_{kk}$ of the optimal $\bar{D}$ can take, if any, while keeping the optimality of the resulting diagonal matrix. Here is the mechanism that allows us to change $\bar{D}_{kk}$. By point 2 of lemma 2.4, for the current $\bar{v}$ and for any value $\bar{D}'_{kk}$ taken in the interval

$$[a_k,b_k]:=\{D_{kk}\in[0,1]: |(1-D_{kk})(A\bar{v})_k+D_{kk}(B\bar{v})_k|\leqslant\alpha\}, \tag{3.16}$$

the diagonal matrix $\bar{D}'$ defined by

$$\bar{D}'_{ii} := \begin{cases} \bar{D}'_{kk} & \text{if } i = k \\ \bar{D}_{ii} & \text{otherwise,} \end{cases}$$

is a solution to problem (3.13a). By (3.14), we get that $\bar{D}'$ is a solution to problem (3.6). In conclusion, for any $\bar{v} \in \mathcal{V}_\infty(C_{\tilde{D}})$ with $\tilde{D}$ solving (3.6), the interval $[a_k, b_k]$ defined by (3.16) is a set of optimal values for $\bar{D}_{kk}$. These intervals depend on $\bar{v}$. Our objective is to show that the union of these intervals $[a_k, b_k]$ for some well chosen $\bar{v} \in \mathcal{V}_\infty(C_{\tilde{D}})$ and solutions $\tilde{D}$ to (3.6) is $[0, 1]$ (the reasoning only holds when $\bar{D}_{kk} \in (0, 1)$). In case 1 below, one has $[a_k, b_k] = [0, 1]$, immediately. In case 2 below, the objective is realized by changing $\bar{v}$ and $\tilde{D}$, alternatively, possibly infinitely often.

By optimality of $\bar{D}$ for (3.13a), we have $|(C_{\bar{D}}\bar{v})_k| \leqslant \alpha$. Therefore, either $\min\{|(C_D\bar{v})_k| : D_{kk} \in [0, 1]\} = \alpha$ or $\min\{|(C_D\bar{v})_k| : D_{kk} \in [0, 1]\} < \alpha$. We now examine these two complementary cases.

1) *Case where*

$$\min_{D_{kk}\in[0,1]} |(1 - D_{kk})(A\bar{v})_k + D_{kk}(B\bar{v})_k| = \alpha. \tag{3.17}$$

By lemma 3.5, with $\nu = (A\bar{v})_k$ and $\mu = (B\bar{v})_k$, the solution set of problem (3.17) is either $\{0\}$, $\{1\}$ or $[0, 1]$. From (3.16) and (3.17), this solution set is also the interval $[a_k, b_k]$. Therefore, by the *principal argument* described above, for the considered vector $\bar{v}$ solving (3.10b), the $k$th element of the optimal $\bar{D}$ can be $\{0\}$, $\{1\}$ or any value in $[0, 1]$. Since $\bar{D}_{kk} \in (0, 1)$, by asumption, one has $[a_k, b_k] = [0, 1]$, which concludes the proof in this case.

2) *Case where*

$$\min_{D_{kk}\in[0,1]} |(1 - D_{kk})(A\bar{v})_k + D_{kk}(B\bar{v})_k| < \alpha. \tag{3.18}$$

In that case, the interval $[a_k, b_k]$ defined by (3.16) is not guaranteed to contain 0 or 1. By modifying the vector $\bar{v}$, however, we show that one can find intervals of substitutes for $\bar{D}_{kk}$, maintaining the optimality of the diagonal matrix, that cover all the interval $[0, 1]$; this is the desired result.

It suffices to extend the interval $[a_k, b_k]$ of optimal values for $\bar{D}_{kk}$ to the left so that it contains 0, because, by symmetry, the interval $[a_k, b_k]$ can then also be extended to the right so that it contains 1 (switch $A$ and $B$ and replace $D$ by $I - D$).

One can assume that $a_k > 0$, since otherwise there is nothing to prove. This implies that $\alpha < \|A\|_\infty$ (because, by optimality of $\bar{D}$, one has $\alpha = \|C_{\bar{D}}\bar{v}\|_\infty \leqslant \|C_0\bar{v}\|_\infty = \|A\bar{v}\|_\infty \leqslant \|A\|_\infty$ and $\alpha \neq \|A\|_\infty$ since otherwise $|(A\bar{v})_k| \leqslant \|A\bar{v}\|_\infty = \alpha$ and $a_k = 0$ by (3.16)).

We do this extension by an iterative procedure whose iterates, indexed by $j \in \mathbb{N}$, are pairs $(\bar{D}^j, \bar{v}^j)$ verifying

$$(\bar{D}^j, \bar{v}^j) \text{ solves the problems in (3.10)}, \tag{3.19a}$$

$$\bar{D}^j_{ii} = \bar{D}_{ii} \text{ for } i \neq k, \tag{3.19b}$$

$$(1 - \bar{D}^j_{kk})(A\bar{v}^j)_k + \bar{D}^j_{kk}(B\bar{v}^j)_k = 0, \tag{3.19c}$$

$$0 < \bar{D}^{j+1}_{kk} \leqslant (1 - \alpha/(2\|A\|_\infty))\bar{D}^j_{kk}. \tag{3.19d}$$

The iterative process is interrupted as soon as 0 is in the interval

$$[a_k^j, b_k^j] := \{D_{kk} \in [0,1] : |(1-D_{kk})(A\bar{v}^j)_k + D_{kk}(B\bar{v}^j)_k| \leqslant \alpha\}, \tag{3.20}$$

that is, as soon as $a_k^j = 0$. It will be clear from the construction of these intervals that their union will be formed of solutions for $\bar{D}_{kk}$. Actually, the reasoning below does not control directly $a_k^j$ but it controls $\bar{D}_{kk}^j \in [a_k^j, b_k^j]$, which tends to zero by (3.19d).

- Let us determine $(\bar{D}^0, \bar{v}^0)$ and verify (3.19a)-(3.19c) for $j = 0$ ((3.19d) for $j = 0$ will be verified when $\bar{D}_{kk}^1$ will be determined, in the next point).

  When (3.18) holds, point 2 of lemma 2.4 ensures that changing $\bar{D}_{kk}$ in order to have $|(C_{\bar{D}}\bar{v})_k| < \alpha$ will not change the optimality of $\bar{D}$, so that we can actually assume that $|(C_{\bar{D}}\bar{v})_k| < \alpha$. Then, lemma 2.3 with $\mathcal{A} = C_{\bar{D}}$ and $v = \bar{v} \in \mathcal{V}_\infty(C_{\bar{D}})$ tells us that one can find a $\bar{v}^0 \in \mathcal{V}_\infty(C_{\bar{D}})$ such that $(C_{\bar{D}}\bar{v}^0)_k = 0$, which reads $(1-\bar{D}_{kk})(A\bar{v}^0)_k + \bar{D}_{kk}(B\bar{v}^0)_k = 0$. Therefore, setting $\bar{D}^0 := \bar{D}$, we see that (3.19b) and (3.19c) hold. Furthermore, (3.19a) also holds since, by (3.12), $\bar{D}$ solves (3.6) and $\bar{v}^0 \in \mathcal{V}_\infty(C_{\bar{D}})$ imply that $(\bar{D}^0, \bar{v}^0)$ solves the problems in (3.10).

- Let us now show how to construct $(\bar{D}^{j+1}, \bar{v}^{j+1})$ from $(\bar{D}^j, \bar{v}^j)$, if this is necessary.

  Assume that $a_k^j > 0$ (otherwise, there is no reason to pursue the iterative process). Then, $0 \neq (A\bar{v}^j)_k \neq (B\bar{v}^j)_k$ (otherwise $a_k^j = 0$) and by definition of $a_k^j$ in (3.20):

$$(1-a_k^j)(A\bar{v}^j)_k + a_k^j(B\bar{v}^j)_k = \alpha\,\mathrm{sign}((A\bar{v}^j)_k). \tag{3.21}$$

  Indeed, if $(1-a_k^j)(A\bar{v}^j)_k + a_k^j(B\bar{v}^j)_k = \alpha$, one has, by definition of $a_k^j$, $(1-t)(A\bar{v}^j)_k + t(B\bar{v}^j)_k > \alpha$ for all $t \in [0, a_k^j)$, in particular $(1-0)(A\bar{v}^j)_k + 0(B\bar{v}^j)_k > \alpha$, so that $(A\bar{v}^j)_k > \alpha > 0$. Similarly, $(1-a_k^j)(A\bar{v}^j)_k + a_k^j(B\bar{v}^j)_k = -\alpha$ implies that $(A\bar{v}^j)_k < 0$.

  Now, define the diagonal matrix $\bar{D}^{j+1} \in [0, I]$ by

$$\bar{D}_{ii}^{j+1} \in \begin{cases} (a_k^j + \bar{D}_{kk}^j)/2 & \text{if } i = k \\ \bar{D}_{ii} & \text{otherwise,} \end{cases} \tag{3.22}$$

  so that (3.19b) is verified with $j+1$ instead of $j$. Adding side by side (3.19c) and (3.21), and using the definition (3.22) of $\bar{D}^{j+1}$, we get

$$(1-\bar{D}_{kk}^{j+1})(A\bar{v}^j)_k + \bar{D}_{kk}^{j+1}(B\bar{v}^j)_k = \frac{1}{2}\,\alpha\,\mathrm{sign}((A\bar{v}^j)_k). \tag{3.23}$$

  Subtracting side by side (3.19c) from (3.23), using $(A\bar{v}^j)_k \neq (B\bar{v}^j)_k$, $(A\bar{v}^j)_k - (B\bar{v}^j)_k = (A\bar{v}^j)_k/\bar{D}_{kk}^j$ by (3.19c) again and finally $|(A\bar{v}^j)_k| \leqslant \|A\|_\infty$ yields

$$\bar{D}_{kk}^j - \bar{D}_{kk}^{j+1} = \frac{(\alpha/2)\,\mathrm{sign}(A\bar{v}^j)_k}{(A\bar{v}^j)_k - (B\bar{v}^j)_k} = \frac{\alpha/2}{|(A\bar{v}^j)_k|}\,\bar{D}_{kk}^j \geqslant \frac{\alpha}{2\|A\|_\infty}\bar{D}_{kk}^j,$$

  which is (3.19d).

  We still have to determine $\bar{v}^{j+1}$ and to verify (3.19a) and (3.19c) with $j+1$ instead of $j$. By (3.22) and (3.19c), $\bar{D}_{kk}^{j+1} \in [a_k^j, \bar{D}_{kk}^j] \subseteq [a_k^j, b_k^j]$. This implies that, like $\bar{D}^j$,

$\bar{D}^{j+1}$ solves problem (3.13a) with $\bar{v} = \bar{v}^j$ (point 2 of lemma 2.4) and, by (3.19a) and (3.14), $(\bar{D}^{j+1}, \bar{v}^j)$ solves the problems in (3.10). Now, by (3.23),

$$|(1 - \bar{D}_{kk}^{j+1})(A\bar{v}^j)_k + \bar{D}_{kk}^{j+1}(B\bar{v}^j)_k| < \alpha \quad \text{or} \quad |(C_{\bar{D}^{j+1}}\bar{v}^j)_k| < \alpha.$$

Then, lemma 2.3 with

$$A = C_{\bar{D}^{j+1}} \qquad \text{and} \qquad v = \bar{v}^j \in \mathcal{V}_\infty(C_{\bar{D}^{j+1}})$$

(the last membership comes from the fact that $(\bar{D}^{j+1}, \bar{v}^j)$ solves the problems in (3.10) and the implication "$\Leftarrow$" in (3.12)) tells us that one can find a

$$\bar{v}^{j+1} \in \mathcal{V}_\infty(C_{\bar{D}^{j+1}}) \text{ such that } (C_{\bar{D}^{j+1}}\bar{v}^{j+1})_k = 0.$$

The first membership implies (3.19a) with $j+1$ replacing $j$, by the implication "$\Rightarrow$" of (3.12) (note that $\bar{D}^{j+1}$ solves (3.6) by the implication "$\Leftarrow$" of (3.12)). The second identity reads (3.19c) with $j+1$ replacing $j$.

By the two previous points, the iterative procedure defining $(\bar{D}^j, \bar{v}^j)$, for $j \in \mathbb{N}$, is well defined, unless it is interrupted by the fact that $a_k^j = 0$ for some $j \in \mathbb{N}$, which is a desirable property since then $\bar{D}$ is solution to (3.6) with any $\bar{D}_{kk} \in [0, b_k]$.

If the procedure does not terminate, one has $\bar{D}_{kk}^j \to 0$ by (3.19d) and $\bar{D}$ is optimal for any $\bar{D}_{kk} \in [\bar{D}_{kk}^j, b_k]$. Since the set of solutions to problem (3.6) is closed, we get that $\bar{D}$ is optimal for any $\bar{D}_{kk} \in [0, b_k]$.

# 4 Discussion

## 4.1 Optimal diagonal elements in $[0, 1]$

It may occur that the $k$th diagonal element of a solution $\bar{D}$ to (3.6) can be any number in $[0, 1]$. This is clearly the case when $A = B$, since then $C_D$ given by (1.4) is the matrix $A$ and, therefore, is independent of $D \in [0, I]$. Similarly, and more generally, if the $k$th row of $A - B$ vanishes, $C_D$ is independent of $D_{kk} \in [0, 1]$, so that any number in $[0, 1]$ can be taken for the $k$th element of an optimal diagonal matrix. See also section 4.3 for a less trivial example.

## 4.2 No Cartesian product structure

The goal of this section is to show that there is no guaranteed Cartesian product structure for the elements of the diagonal matrices solving problem (3.6) and that this lack of guaranteed Cartesian product structure also holds for the solution pairs $(\bar{D}, \bar{v})$ to the problems in (3.10). The example also illustrates case 1 of the proof of proposition 3.4, which occurs when (3.17) holds.

Consider the **P**-matrix

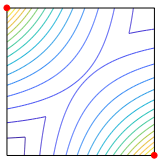
$$M = \begin{pmatrix} 1/2 & 1 \\ -1 & 1/2 \end{pmatrix}. \tag{4.1}$$

We have

$$M_D = \begin{pmatrix} 1 - D_{11}/2 & D_{11} \\ -D_{22} & 1 - D_{22}/2 \end{pmatrix} \qquad \text{and} \qquad M_D^{-1} = \frac{1}{\Delta} \begin{pmatrix} 1 - D_{22}/2 & -D_{11} \\ D_{22} & 1 - D_{11}/2 \end{pmatrix},$$

where $\Delta := \det M_D = 1 - \frac{1}{2}(D_{11} + D_{22}) + \frac{5}{4} D_{11} D_{22}$. Therefore, by (2.3),

$$\begin{aligned} \|M_D^{-1}\|_\infty &= \tfrac{1}{\Delta} \max\left(1 + D_{11} - \tfrac{1}{2} D_{22}, 1 - \tfrac{1}{2} D_{11} + D_{22}\right) \\ &= \begin{cases} (1 + D_{11} - \frac{1}{2} D_{22})/\Delta & \text{if } D_{11} \geqslant D_{22} \\ (1 - \frac{1}{2} D_{11} + D_{22})/\Delta & \text{otherwise.} \end{cases} \end{aligned}$$



The right-hand side plot shows the level curves of $D \in [0,1]^2 \mapsto \|M_D^{-1}\|_\infty$ and its two (red) dots are the points at which this function is maximized. Observe that $D_{11} \mapsto \|M_D^{-1}\|_\infty$ is monotone increasing for $D_{22} = 0$ and monotone decreasing for $D_{22} = 1$.

It is pure calculation to see that $(1 + D_{11} - \frac{1}{2} D_{22})/\Delta$ is increasing with $D_{11}$ and decreasing with $D_{22}$ on $[0,1]^2$, so that the first part of the max above is maximized for $D = \mathrm{Diag}(1,0)$, in which case $\|M_D^{-1}\|_\infty = 4$. Symmetrically, the second part of the max above, namely $(1 - \frac{1}{2} D_{11} + D_{22})/\Delta$, is obtained by switching $D_{11}$ and $D_{22}$ in the first part of the max. It is therefore maximized for $D = \mathrm{Diag}(0,1)$ for which $\|M_D^{-1}\|_\infty = 4$. Therefore, for the matrix $M$ in (4.1), problem (3.6) has two solutions $\bar{D} = \mathrm{Diag}(1,0)$ and $\bar{D} = \mathrm{Diag}(0,1)$ and the optimal value of (3.6) is $\beta = 4$. Now $\|M_0^{-1}\|_\infty = \|I\|_\infty = 1$ and $\|M_I^{-1}\|_\infty = \|M^{-1}\|_\infty = 6/5$, so that neither $\mathrm{Diag}(0,0)$ nor $\mathrm{Diag}(1,1)$ is a solution. This shows that *the solution set of* (3.6) *is not the image by* Diag *of a Cartesian product.*

The matrix (4.1) allows us to illustrate case 1 of the proof of proposition 3.4, which occurs when (3.17) holds.

- For $\bar{D} = \mathrm{Diag}(1,0)$, one gets $\mathcal{V}_\infty(M_{\bar{D}}) = \{(-1, \frac{1}{4}), (1, -\frac{1}{4})\}$. For $\bar{v} = (-1, \frac{1}{4})$, $M\bar{v} = (-\frac{1}{4}, \frac{9}{8})$ and the intervals defined by (3.16) are

  $$\begin{aligned} [a_1, b_1] &= \{D_{11} \in [0,1] : |(1 - D_{11})(-1) + D_{11}(-1/4)| \leqslant 1/4\} = \{1\}, \\ [a_2, b_2] &= \{D_{22} \in [0,1] : |(1 - D_{22})(1/4) + D_{22}(9/8)| \leqslant 1/4\} = \{0\}, \end{aligned}$$

  which give indeed the diagonal elements of the considered $\bar{D}$. For $\bar{v} = (1, -\frac{1}{4})$, $M\bar{v} = (\frac{1}{4}, -\frac{9}{8})$ and the intervals defined by (3.16) are

  $$\begin{aligned} [a_1, b_1] &= \{D_{11} \in [0,1] : |(1 - D_{11})(1) + D_{11}(1/4)| \leqslant 1/4\} = \{1\}, \\ [a_2, b_2] &= \{D_{22} \in [0,1] : |(1 - D_{22})(-1/4) + D_{22}(-9/8)| \leqslant 1/4\} = \{0\}, \end{aligned}$$

  which also give the diagonal elements of the considered $\bar{D}$.

- For $\bar{D} = \mathrm{Diag}(0,1)$, one gets $\mathcal{V}_\infty(M_{\bar{D}}) = \{(-\frac{1}{4}, -1), (\frac{1}{4}, 1)\}$. For $\bar{v} = (-\frac{1}{4}, -1)$, $M\bar{v} = (-\frac{9}{8}, -\frac{1}{4})$ and the intervals defined by (3.16) are

  $$\begin{aligned} [a_1, b_1] &= \{D_{11} \in [0,1] : |(1 - D_{11})(-1/4) + D_{11}(-9/8)| \leqslant 1/4\} = \{0\}, \\ [a_2, b_2] &= \{D_{22} \in [0,1] : |(1 - D_{22})(-1) + D_{22}(-1/4)| \leqslant 1/4\} = \{1\}, \end{aligned}$$

  which give indeed the diagonal elements of the considered $\bar{D}$. For $\bar{v} = (\frac{1}{4}, 1)$, $M\bar{v} = (\frac{9}{8}, \frac{1}{4})$ and the intervals defined by (3.16) are

  $$\begin{aligned} [a_1, b_1] &= \{D_{11} \in [0,1] : |(1 - D_{11})(1/4) + D_{11}(9/8)| \leqslant 1/4\} = \{0\}, \\ [a_2, b_2] &= \{D_{22} \in [0,1] : |(1 - D_{22})(1) + D_{22}(1/4)| \leqslant 1/4\} = \{1\}, \end{aligned}$$

which also give the diagonal elements of the considered $\bar{D}$.

This example also shows that there is no Cartesian product structure in the solutions $(\bar{D}, \bar{v}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^n$ to the problems in (3.10).

### 4.3 Illustration of the proof of proposition 3.4

This section provides a nontrivial example of matrix $M$ for which any number in $[0, 1]$ can be chosen for a diagonal element of an optimal $\bar{D}$ (compare with section 4.1, where this possibility is trivially verified). For this, it is necessary to have an element of $M_{\bar{D}}^{-1}$ that vanishes (see also the proof of lemma 2.3). Along the way, the example illustrates case 2 of the proof of proposition 3.4, which occurs when (3.18) holds.

Consider the **P**-matrix

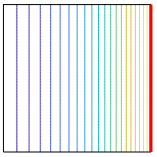
$$M = \begin{pmatrix} 1/2 & 0 \\ 3 & 5 \end{pmatrix}.$$

We have

$$M_D = \begin{pmatrix} 1 - D_{11}/2 & 0 \\ 3D_{22} & 1 + 4D_{22} \end{pmatrix} \qquad \text{and} \qquad M_D^{-1} = \begin{pmatrix} \frac{1}{1-D_{11}/2} & 0 \\ \frac{-3D_{22}}{(1-D_{11}/2)(1+4D_{22})} & \frac{1}{1+4D_{22}} \end{pmatrix}.$$

Clearly, $\bar{D}_{11} = 1$, since this makes $\|(M_D^{-1})_{1:}\|_1$ and $\|(M_D^{-1})_{2:}\|_1$ maximal for $D_{11} \in [0, 1]$, whatever $\bar{D}_{22}$ is in $[0, 1]$ (the right-hand side plot shows the level curves of $D \in [0, 1]^2 \mapsto \|M_D^{-1}\|_\infty$ and the rightmost vertical line, made of the maximizers of this function). Therefore,

$$M_{\bar{D}}^{-1} = \begin{pmatrix} 2 & 0 \\ \frac{-6\bar{D}_{22}}{1+4D_{22}} & \frac{1}{1+4D_{22}} \end{pmatrix}.$$



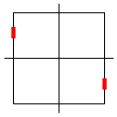
Then, $\|(M_{\bar{D}}^{-1})_{1:}\|_1 = 2$, while $\|(M_{\bar{D}}^{-1})_{2:}\|_1 = (1 + 6\bar{D}_{22})/(1 + 4\bar{D}_{22})$ ranges in $[1, 7/5]$, so that $\beta = 2$ by (2.3), $\alpha := 1/\beta = 1/2$ and the optimal set of $\bar{D}$'s is $\mathrm{Diag}(\{1\}, [0, 1])$.

To illustrates case 2 of the proof of proposition 3.4, which occurs when (3.18) holds, consider the following solution pair $(\bar{D}, \bar{v})$ to the problems in (3.10):

$$\bar{D} = \begin{pmatrix} 1 & 0 \\ 0 & 3/4 \end{pmatrix} \qquad \text{and} \qquad \bar{v} = \begin{pmatrix} 1 \\ -17/32 \end{pmatrix}.$$

For the given $\bar{D}$, one computes easily the set

$$\mathcal{V}_\infty(M_{\bar{D}}) = (\{-1\} \times [7/16, 11/16]]) \cup (\{1\} \times [-11/16, -7/16]]),$$



to which the given $\bar{v}$ must belong. Observe that (3.18) holds for $k = 2$:

$$\min_{D_{22}\in[0,1]} |(1 - D_{22})\bar{v}_2 + D_{22}(M\bar{v})_2| = \min_{D_{22}\in[0,1]} |-17/32 + (7/8)D_{22}| = 0 < \alpha = 1/2.$$

In particular, one has $|(1 - \bar{D}_{22})\bar{v}_2 + \bar{D}_{22}(M\bar{v})_2| = 1/8 < \alpha$ (the right-hand side dot in figure 4.1 has the coordinates $(\bar{D}, |M_{\bar{D}}\bar{v}|) = (3/4, 1/8)$). Here are two iterations of the proof of proposition 3.4 (case 2); see figure 4.1.

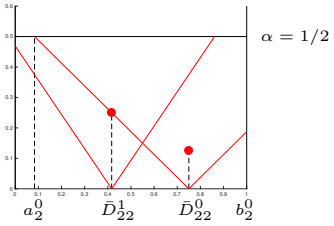



Figure 4.1: Illustration of the proof of proposition 3.4. See the description in the text.

- The first iterate $(\bar{D}^0, \bar{v}^0)$ satisfying (3.19a)-(3.19d) is defined with $\bar{D}^0 = \bar{D}$ and a $\bar{v}^0 = (1, -9/16) \in \mathcal{V}_\infty(M_{\bar{D}^0})$ given by lemma 2.3. Il follows that the interval defined by (3.20) is $[a_2^0, b_2^0] = [1/12, 1]$.
- The second iterate $(\bar{D}^1, \bar{v}^1)$ satisfying (3.19a)-(3.19d) is defined as follows. Only the second ($k = 2$) component of $\bar{D}^0$ is modified to get $\bar{D}^1$: one sets $\bar{D}_{11}^1 = \bar{D}_{11} = 1$ and $\bar{D}_{22}^1 = \frac{1}{2}(a_2^0 + \bar{D}_{22}^0) = 5/12$ (see (3.22)). Then $|(M_{\bar{D}^1}\bar{v}^0)_2| = 1/4 < \alpha = 1/2$ (the left-hand side point in figure 4.1 is $(\bar{D}_{22}^1, |(M_{\bar{D}^1}\bar{v}^0)_2|)$) and a $\bar{v}^1 = (1, -15/32) \in \mathcal{V}_\infty(M_{\bar{D}^1})$ is given by lemma 2.3. Il follows that the interval defined by (3.20) is $[a_2^1, b_2^1] = [0, 31/36]$.

Gathering the two intervals computed during these two iterations, we get $[0, 31/36] \cup [1/12, 1] = [0, 1]$. Hence, for this special case, the procedure terminates after two stages.

### 4.4 Complexity issues

The simplification (1.8) of the error bound factor $\beta$ given by (1.7) allows us to compute it by evaluating the map $D \in [0, I] \mapsto \|C_D^{-1}\|_\infty$ at the $2^n$ extreme points of $[0, I]$, which are the diagonal matrices $D$ with diagonal entries in $\{0, 1\}$. We believe that this is an improvement for small values of $n$. Nevertheless, for large $n$, this exponential number of evaluations can make this exact extensive computation approach very time consuming. Now, it is not unlikely that, for special classes of matrices, the simplification (1.8) can yield an efficient way of computing the error bound factor. Finally, we are also exploring the possibility to simplify this extensive evaluation by a specific algorithm based on the developments made in this paper.

Another interest of the simplified formula (1.8) of $\beta$ deals with the complexity analysis of some algorithms for solving the generalized linear complementarity problem (1.2) with matrices $A$ and $B$ verifying (1.5) (equivalent to the **P**-matricity of $M$ if $(A, B) = (I, M)$) and integer (or rational) data. When the complexity is expressed in terms of the data bitlength and when the error bound (1.6) intervenes, the question may arise to know whether the error bound factor can be bounded above by a formula using the data bitlength or the bitlength of the matrix $A$ and $B$, denoted $\mathfrak{L}(A, B)$ say, since the data bitlength is certainly larger than $\mathfrak{L}(A, B)$. It is known from [25; paragraph straddling pages 209-210] (probably also implicit in [19]), that, for an arbitrary nonsingular matrix $M \in \mathbb{R}^{n\times n}$,

$$\|M^{-1}\|_\infty \leqslant n\, 2^{\mathfrak{L}(M)+1}.$$

Thanks to the formula (1.8) of $\beta$, the error bound factor is equal to $\|C_{\bar{D}}^{-1}\|_\infty$, for some $\bar{D} \in \mathrm{ext}[0, I]$. Therefore, the rows of $C_{\bar{D}}$ defined by (1.4) are those of $A$ or $B$. As a result, one certainly has

$$\mathfrak{L}(C_{\bar{D}}) \leqslant \mathfrak{L}(A, B). \tag{4.2}$$

As a result, with $\bar{D}$ solving the optimization problem in (1.8), one has

$$\max_{D\in[0,I]} \|C_D^{-1}\|_\infty = \|C_{\bar{D}}^{-1}\|_\infty \leqslant n\, 2^{\mathfrak{L}(C_{\bar{D}})+1} \leqslant n\, 2^{\mathfrak{L}(A,B)+1}.$$

Without (4.2), the upper bound would have been in terms of $\mathfrak{L}(C_{\bar{D}})$, which could be infinite since the optimal diagonal matrix $\bar{D}$ could have irrational numbers in some entries. Therefore, thanks to (4.2), for the generalized linear complementarity problem (1.2), with matrices $A$ and $B$ verifying (1.5), one has the error bound

$$\forall\, x \in \mathbb{R}^n : \quad \|x - \bar{x}\|_\infty \leqslant n\, 2^{\mathfrak{L}(A,B)+1}\, \|\min(Ax+a, Bx+b)\|_\infty, \tag{4.3}$$

where $\bar{x}$ is the unique solution to the generalized LCP.

This subject is further explored in [13].

## References


[1] M. Aganagić (1984). Newton's method for linear complementarity problems. *Mathematical Programming*, 28, 349–362. [doi]. 11, 12

[2] I. Ben Gharbia, J.Ch. Gilbert (2013). An algorithmic characterization of P-matricity. *SIAM Journal on Matrix Analysis and Applications*, 34(3), 904–916. [doi]. 2

[3] I. Ben Gharbia, J.Ch. Gilbert (2019). An algorithmic characterization of P-matricity II: adjustments, refinements, and validation. *SIAM Journal on Matrix Analysis and Applications*, 40(2), 800–813. [doi]. 2

[4] J.F. Bonnans, R. Cominetti, A. Shapiro (1999). Second order optimality conditions based on parabolic second order tangent sets. *SIAM Journal on Optimization*, 9(2), 466–492. [doi]. 14

[5] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, C. Sagastizábal (2006). *Numerical Optimization – Theoretical and Practical Aspects* (second edition). Universitext. Springer Verlag, Berlin. [authors] [editor] [doi]. 7

[6] J.M. Borwein (2016). A very complicated proof of the minimax theorem. *Minimax Theory and its Applications*, 1(1), 21–27. 4

[7] J.V. Burke, R.A. Poliquin (1992). Optimality conditions for non-finite valued convex composite functions. *Mathematical Programming*, 57(1), 103–120. [doi]. 14

[8] T. Chen, W. Li, X. Wu, S. Vong (2015). Error bounds for linear complementarity problems of MB-matrices. *Numerical Algorithms*, 70, 341–356. [doi]. 3

[9] X. Chen, S. Xiang (2006). Perturbation bounds of P-matrix linear complementarity problems. *SIAM Journal on Optimization*, 18(4), 1250–1265. [doi]. 3

[10] X. Chen, S. Xiang (2006). Computation of error bounds for P-matrix linear complementarity problems. *Mathematical Programming*, 106, 513–525. [doi]. 2, 3, 12

[11] R.W. Cottle, J.-S. Pang, R.E. Stone (2009). *The Linear Complementarity Problem*. Classics in Applied Mathematics 60. SIAM, Philadelphia, PA, USA. [doi]. 2

[12] P.-F. Dai (2011). Error bounds for linear complementarity problems of DB-matrices. *Linear Algebra and its Applications*, 434(3), 830–840. [doi]. 3

[13] J.-P. Dussault, J.Ch. Gilbert (2021). A complexity result for the linear complementarity problem with a P-matrix (to appear). Technical report. 23

[14] I. Ekeland, R. Temam (1999). *Convex Analysis and Variational Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Translated from the French. [doi]. 10

[15] J.Ch. Gilbert (2021). *Fragments d'Optimisation Différentiable – Théorie et Algorithmes*. Lecture Notes (in French) of courses given at ENSTA and at Paris-Saclay University, Saclay, France. [hal-03347060, pdf]. 3, 7, 10, 15

[16] E.G. Golshtein, N.V. Tretyakov (1996). *Modified Lagrangians and Monotone Maps in Optimization*. Discrete Mathematics and Optimization. John Wiley & Sons, New York. 7, 8

[17] J.-B. Hiriart-Urruty, C. Lemaréchal (1996). *Convex Analysis and Minimization Algorithms* (second edition). Grundlehren der mathematischen Wissenschaften 305-306. Springer. 7

[18] R.A. Horn, C. Jonhson (1985). *Matrix Analysis*. Cambridge University Press, Cambridge, U.K. 5

[19] L.G. Khachiyan (1979). A polynomial algorithm in linear programming. *Soviet Mathematics Doklady*, 20, 191–194. 22

[20] W. Li, H. Zheng (2014). Some new error bounds for linear complementarity problems of H-matrices. *Numerical Algorithms*, 67, 257–269. [doi]. 3

[21] Z.-Q. Luo, O.L. Mangasarian, J. Ren, M.V. Solodov (1994). New error bounds for the linear complementarity problem. *Mathematics of Operations Research*, 19(4), 880–892. [doi]. 3

[22] Z.-Q. Luo, P. Tseng (1992). Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem. *SIAM Journal on Optimization*, 2(1), 43–54. [doi]. 2, 3

[23] O.L. Mangasarian, J. Ren (1994). New improved error bounds for the linear complementarity problem. *Mathematical Programming*, 66(2), 241–255. [doi]. 2, 3

[24] R. Mathias, J.-S. Pang (1990). Error bounds for the linear complementarity problem with a P-matrix. *Linear Algebra and its Applications*, 132, 123–136. [doi]. 3

[25] A. Nemirovski (2012). *Introduction to linear optimization*. ISYE 6661. Georgia Institute of Technology, H. Milton Stewart School of Industrial and Systems Engineering [internet]. 22

[26] J.-S. Pang (1986). Inexact Newton methods for the nonlinear complementarity problem. *Mathematical Programming*, 36, 54–71. 3

[27] J.-S. Pang (1990). Newton's method for B-differentiable equations. *Mathematics of Operations Research*, 15, 311–341. [doi]. 2, 11

[28] J.-S. Pang (1997). Error bounds in mathematical programming. *Mathematical Programming*, 79, 299–332. 2

[29] J.-P. Penot (1994). Optimality conditions in mathematical programming and composite optimization. *Mathematical Programming*, 67, 225–245. [doi]. 14

[30] S.M. Robinson (1981). Some continuity properties of polyhedral multifunctions. *Mathematical Programming Study*, 14, 206–214. [doi]. 3

[31] R.T. Rockafellar (1970). *Convex Analysis*. Princeton Mathematics Ser. 28. Princeton University Press, Princeton, New Jersey. 3

[32] H. Samelson, R.M. Thrall, O. Wesler (1958). A partition theorem for the Euclidean $n$-space. *Proceedings of the American Mathematical Society*, 9, 805–807. [editor]. 2

[33] Xinzhen Zhang, Fengming Ma, Yiju Wang (2005). A Newton-type algorithm for generalized linear complementarity problem over a polyhedral cone. *Applied Mathematics and Computation*, 169(1), 388–401. [doi]. 2