

Optimization

A Journal of Mathematical Programming and Operations Research

ISSN: 0233-1934 (Print) 1029-4945 (Online) Journal homepage: www.tandfonline.com/journals/gopt20

Feasibility problems via paramonotone operators in a convex setting

J. Camacho, M.J. Cánovas, J.E. Martínez-Legaz & J. Parra

To cite this article: J. Camacho, M.J. Cánovas, J.E. Martínez-Legaz & J. Parra (2024) Feasibility problems via paramonotone operators in a convex setting, Optimization, 73:10, 3055-3086, DOI: [10.1080/02331934.2024.2313690](https://doi.org/10.1080/02331934.2024.2313690)

To link to this article: <https://doi.org/10.1080/02331934.2024.2313690>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 12 Feb 2024.



Submit your article to this journal [↗](#)



Article views: 810



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Feasibility problems via paramonotone operators in a convex setting

J. Camacho ^a, M.J. Cánovas ^a, J.E. Martínez-Legaz ^b and J. Parra ^a

^aCenter of Operations Research, Miguel Hernández University of Elche, Elche (Alicante), Spain;

^bDepartament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, and BGSMath, Barcelona, Spain

ABSTRACT

This paper is focussed on some properties of paramonotone operators on Banach spaces and their application to certain feasibility problems for convex sets in a Hilbert space and convex systems in the Euclidean space. In particular, it shows that operators that are simultaneously paramonotone and bimonotone are constant on their domains, and this fact is applied to tackle two particular situations. The first one, closely related to simultaneous projections, deals with a finite amount of convex sets with an empty intersection and tackles the problem of finding the smallest perturbations (in the sense of translations) of these sets to reach a nonempty intersection. The second is focussed on the distance to feasibility; specifically, given an inconsistent convex inequality system, our goal is to compute/estimate the smallest right-hand side perturbations that reach feasibility. We advance that this work derives lower and upper estimates of such a distance, which become the exact value when confined to linear systems.

ARTICLE HISTORY

Received 20 April 2023

Accepted 21 January 2024

KEYWORDS

Distance function; convex inequalities; distance to feasibility; paramonotone operators; displacement mapping

MATHEMATICS SUBJECT CLASSIFICATIONS

47N10; 47H05; 52A20; 90C31; 49K40

1. Introduction

The present paper is focussed on paramonotone operators with applications to certain feasibility problems for convex sets in a Hilbert space and convex inequality systems in $\mathbb{R}^n$. As usual, throughout the paper $\subset$ is understood as $\subseteq$. To start with, we recall some basic properties of operators in Banach spaces. Let X be a real Banach space, with topological dual X^* , and denote by $\langle \cdot, \cdot \rangle$ the corresponding canonical pairing. A set-valued operator $T : X \rightrightarrows X^*$ is said to be *monotone* if

$$\langle x - y, x^* - y^* \rangle \geq 0 \quad \text{whenever } (x, x^*), (y, y^*) \in \text{gph } T,$$

where $\text{gph } T := \{(x, x^*) \in X \times X^* : x^* \in T(x)\}$ is the graph of T . In the case when both T and $-T$ are monotone, then T is called *bimonotone*. If T is monotone and, in addition, $\text{gph } T$ is maximal in the sense of inclusion order, it is said to be *maximally monotone*. A well-known example of a maximally monotone operator is

CONTACT J. Parra  parra@umh.es

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

the subdifferential operator of a proper, lower semicontinuous (lsc, for short), convex function $f : X \rightarrow]-\infty, +\infty]$, denoted by ∂f (see Section 2 for details). Monotone operators are fundamental tools of nonlinear analysis and optimization; see, e.g. the books [1–7]. A monotone operator T is called *paramonotone* if the following implication holds:

$$\left. \begin{array}{l} (x, x^*), (y, y^*) \in \text{gph} T \\ \langle x - y, x^* - y^* \rangle = 0 \end{array} \right\} \Rightarrow (x, y^*), (y, x^*) \in \text{gph} T.$$

The term paramonotonicity was introduced in [8] (although the condition was previously presented in [9] without a name). The initial motivation for the introduction of paramonotone operators comes from its crucial role regarding interior point methods for variational inequalities (see again [8,9], and also [10]). Some important examples of paramonotone operators are gathered in Section 2. At this moment, let us mention that subdifferentials of proper lsc convex functions enjoy this property (see [10, Proposition 2.2] in the Euclidean space and [11, Fact 3.1] for its extension to Banach spaces).

Looking at the applications of Sections 4 and 5, we are interested in operators of the form $T_1 \cap (-T_2)$, where $T_1, T_2 : X \rightrightarrows X^*$ are paramonotone, which are also paramonotone and, additionally, bimonotone; this fact entails that $T_1 \cap (-T_2)$ is constant on its domain (as shown in Proposition 3.5); recall that the domain of an operator T is given by $\text{dom} T := \{x \in X \mid T(x) \neq \emptyset\}$. Observe that

$$\text{dom} (T_1 \cap (-T_2)) = \{x \in X \mid 0 \in (T_1 + T_2)(x)\},$$

which, in the particular case $T_i = \partial f_i$, $i = 1, 2$, where the f_i 's are proper, lsc and convex, is known to coincide with

$$\arg \min (f_1 + f_2),$$

i.e. with the set of (global) minimizers of $f_1 + f_2$, provided that a regularity condition ensuring $\partial f_1 + \partial f_2 = \partial(f_1 + f_2)$ is satisfied. These comments easily generalize to the sum of a finite number of functions (see Section 3 for details) and are applied to particular problems of the form

$$\text{minimize}_{x \in X} \sum_{i=1}^m f_i(x), \tag{1}$$

where all f_i 's are proper lsc convex functions on X .

Now we present two applications discussed in the paper. The first one is developed in a Hilbert space Y whose norm, associated with the corresponding inner product, is denoted by $\|\cdot\|$. It deals with a finite number of nonempty closed convex sets $S_1, S_2, \dots, S_m \subset Y$ such that $\bigcap_{i=1}^m S_i = \emptyset$ and is focussed on the

optimization problem given by

$$\text{minimize}_{x \in Y} \sum_{i=1}^m \alpha_i d(x, S_i)^p, \quad (2)$$

where $\alpha_i > 0$, $i = 1, \dots, m$; without loss of generality we assume $\sum_{i=1}^m \alpha_i = 1$, $p \geq 1$ and $d(x, S_i)$ denotes the distance from point x to the set S_i , $i = 1, \dots, m$, that is, $d(x, S_i) := \inf_{y \in S_i} \|x - y\|$; this infimum is actually a minimum, as, according to the Hilbert projection theorem, it is attained (at a unique point). The following proposition establishes that (2) is equivalent to the problem:

$$\begin{aligned} &\text{minimize} \quad \|u\|_{\alpha,p} \\ &\text{subject to} \quad \bigcap_{i=1}^m (S_i + u_i) \neq \emptyset, \\ &\quad \quad \quad u = (u_1, \dots, u_m) \in Y^m, \end{aligned} \quad (3)$$

where $\|u\|_{\alpha,p}$ denotes the weighted p -norm in space Y^m defined as

$$\|u\|_{\alpha,p} = \left(\sum_{i=1}^m \alpha_i \|u_i\|^p \right)^{1/p}.$$

This equivalence was already observed in [12, Section 4] for Euclidean spaces; we include a proof for the sake of completeness.

Proposition 1.1: *A point $\bar{u} = (\bar{u}_1, \dots, \bar{u}_m) \in Y^m$ is an optimal solution to (3) if and only if there exists an optimal solution $\bar{x}$ to (2) such that $\bar{u}_i = \bar{x} - P_i(\bar{x})$, $i = 1, \dots, m$, with $P_i(\bar{x}) := \arg \min_{y \in S_i} \|\bar{x} - y\|$ being the best approximation of $\bar{x}$ in S_i .*

Proof: Let $\bar{u} = (\bar{u}_1, \dots, \bar{u}_m) \in Y^m$ be an optimal solution to (3), and take $\bar{x} \in \bigcap_{i=1}^m (S_i + \bar{u}_i)$. There exist $s_i \in S_i$, $i = 1, \dots, m$, such that $\bar{x} = s_i + \bar{u}_i$. For every $x \in Y$, we have

$$\begin{aligned} \sum_{i=1}^m \alpha_i d(\bar{x}, S_i)^p &\leq \sum_{i=1}^m \alpha_i \|\bar{x} - s_i\|^p = \sum_{i=1}^m \alpha_i \|\bar{u}_i\|^p \leq \sum_{i=1}^m \alpha_i \|x - P_i(x)\|^p \\ &= \sum_{i=1}^m \alpha_i d(x, S_i)^p. \end{aligned}$$

To justify the latter inequality, observe that $\bigcap_{i=1}^m (S_i + x - P_i(x)) \neq \emptyset$, because from the equalities $x = P_i(x) + x - P_i(x)$, $i = 1, \dots, m$, it immediately follows that $x \in \bigcap_{i=1}^m (S_i + x - P_i(x))$. Thus, $\bar{x}$ is an optimal solution to (2). Furthermore, setting $x = \bar{x}$, we also deduce that $d(\bar{x}, S_i) = \|\bar{x} - s_i\|$, $i = 1, \dots, m$, that is, $s_i = P_i(\bar{x})$, so that $\bar{u}_i = \bar{x} - s_i = \bar{x} - P_i(\bar{x})$.

Conversely, let $\bar{x}$ be an optimal solution to (2), u be a feasible solution to (3), and take $x \in \bigcap_{i=1}^m (S_i + u_i)$. Then, there exist $s_i \in S_i$, $i = 1, \dots, m$, such that $x = s_i + u_i$, and we have

$$\begin{aligned} \|u\|_{\alpha,p}^p &= \sum_{i=1}^m \alpha_i \|u_i\|^p = \sum_{i=1}^m \alpha_i \|x - s_i\|^p \geq \sum_{i=1}^m \alpha_i d(x, S_i)^p \geq \sum_{i=1}^m \alpha_i d(\bar{x}, S_i)^p \\ &= \sum_{i=1}^m \alpha_i \|\bar{x} - P_i(\bar{x})\|^p, \end{aligned}$$

which shows that the point $\bar{u} = (\bar{u}_1, \dots, \bar{u}_m)$, with $\bar{u}_i := \bar{x} - P_i(\bar{x})$, $i = 1, \dots, m$, is an optimal solution to (3). ■

According to Proposition 1.1, problem (2) is equivalent to that of finding the smallest translations of the sets S_i that achieve a nonempty intersection.

All the results in Section 4, with only two exceptions, correspond to the case $m = 2$, and extending them to arbitrary m appears to be problematic. The two exceptions are the last two results in the section. The first of them (Theorem 4.5) shows that, for $p \geq 2$, the displacement mappings are constant on the optimal solution set. As a consequence, Corollary 4.6 ensures that the smallest translations that, applied to mutually disjoint convex sets, get translated sets with a nonempty intersection are unique.

The second application, developed in Section 5, deals with convex inequality systems in $\mathbb{R}^n$ parameterized with respect to the right-hand side

$$\sigma(b) := [g_i(x) \leq b_i, i = 1, \dots, m],$$

where $x \in \mathbb{R}^n$ is the vector of decision variables and, for each $i \in \{1, \dots, m\}$, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is a (finite-valued) convex function on $\mathbb{R}^n$, and $(b_i)_{i=1,\dots,m} \equiv b \in \mathbb{R}^m$. Taking a nominal $\bar{b} \in \mathbb{R}^m$ such that $\sigma(\bar{b})$ is inconsistent (i.e. there is no $x \in \mathbb{R}^n$ satisfying all inequalities of $\sigma(\bar{b})$), our aim is to estimate the distance in $\mathbb{R}^m$ endowed with any p -norm, with $p \geq 2$, from $\bar{b}$ to the set of parameters b such that $\sigma(b)$ is consistent. This *distance to feasibility* can be computed by solving the following problem:

$$\text{minimize}_{x \in \mathbb{R}^n} \sum_{i=1}^m [g_i(x) - b_i]_+^p,$$

which also adapts to the format of (1). Some new results on these issues are shown later in Section 5. Sharper results are presented for linear systems when $p = 2$. In order to integrate these new contributions into the existing literature, let us comment that a counterpart result on the distance from a consistent system to infeasibility is established in [13, Proposition 3.10] in the context of convex systems under affine perturbations and by appealing to the Fenchel-Legendre conjugate of the supremum function $\sup_{i=1,\dots,m} (g_i(x) - b_i)$; see also

[14, Theorem 12] from a linearization approach. Generically, the distance to feasibility/infeasibility, sometimes referred to as ‘distance to ill-posedness’, constitutes an important stability measure of systems and, in particular, it appears in the computation of condition numbers; see, e.g. [15] and the pioneer work [16], developed in the framework of conic linear systems (see also [17,18] for repercussions in the complexity of certain algorithms).

At this point, we summarize the structure of the paper. Section 2 gathers some background on convex sets, convex functions and monotone operators, which is appealed to in the remaining sections. Section 3 explores some new properties of paramonotone operators and, in particular, analyzes the simultaneous fulfilment of paramonotonicity and bimonotonicity. The problem of simultaneous projections – see (2) and (3) – is tackled in Section 4, while the distance to feasibility of convex systems under right-hand side perturbations is dealt with in Section 5.

2. Preliminaries

Let X be a real Banach space and $f : X \rightarrow]-\infty, +\infty]$ be a proper lsc convex function. We denote by $\text{dom}f := \{x \in X \mid f(x) < +\infty\}$ the domain of f . Recall that the subdifferential operator of f , $\partial f : X \rightrightarrows X^*$, assigns to each $x \in \text{dom}f$ the (possibly empty) set $\partial f(x)$ formed by all $x^* \in X^*$ (called subgradients) such that

$$f(y) - f(x) \geq \langle y - x, x^* \rangle, \quad \text{for all } y \in X;$$

here and in the sequel, $\langle \cdot, \cdot \rangle : X \times X^* \rightarrow \mathbb{R}$ stands for the duality pairing, so that $\langle x, x^* \rangle$ denotes the value of the continuous linear functional $x^* \in X^*$ at $x \in X$. When $x \notin \text{dom}f$ we define $\partial f(x) := \emptyset$; in this way the domain of the set-valued mapping, $\text{dom}\partial f$, is always contained in $\text{dom}f$. Associated with f , its Fenchel conjugative function $f^* : X^* \rightarrow]-\infty, +\infty]$ is given by

$$f^*(x^*) = \sup \{ \langle x, x^* \rangle - f(x) \mid x \in X \}.$$

Recall that the Young-Fenchel inequality writes as $f^*(x^*) + f(x) \geq \langle x, x^* \rangle$ for all $x \in X$.

For completeness, we gather in the following theorem some well-known results about ∂f and f^* in Banach spaces used in the paper. They can be traced out from different references dealing with convex analysis in infinite dimensional spaces. Here, we mainly cite the books [2,7,19,20]. From now on, $\text{int}A$ denotes the interior of $A \subset X$ and the zero vector of X^* is denoted by just 0.

Theorem 2.1: *Let $f : X \rightarrow]-\infty, +\infty]$ be a proper lsc convex function. Then we have:*

- (i) [2, Proposition 4.1.5] *f is continuous at x if and only if $x \in \text{int dom}f$;*
- (ii) [19, Theorem 3.2.15] $\text{int dom}f \subset \text{dom}\partial f$;

- (iii) [19, Proposition 3.2.17] $x \in \arg \min f$ if and only if $0 \in \partial f(x)$;
- (iv) [19, Exercise 4.2.15] ∂f is maximally monotone;
- (v) [11, Fact 3.1] (see [10, Proposition 2.2] for finite dimensions) ∂f is paramonotone;
- (vi) [19, Proposition 5.31] For any $x \in X$, we have the equivalence

$$x^* \in \partial f(x) \Leftrightarrow f^*(x^*) + f(x) = \langle x, x^* \rangle;$$

(indeed, this statement does not require convexity);

- (vii) [19, Theorem 3.4.2] (see also [21, Theorem 3]) Let $g : X \rightarrow]-\infty, +\infty]$ be any proper convex function. If $(\text{int dom } f) \cap \text{dom } g \neq \emptyset$, then we have the subdifferential sum rule

$$\partial f(x) + \partial g(x) = \partial(f + g)(x), \quad \text{whenever } x \in \text{dom } \partial f \cap \text{dom } \partial g,$$

(indeed, ‘ $\supset$ ’ is the nontrivial inclusion, as $\partial f(x) + \partial g(x) \subset \partial(f + g)(x)$ comes directly from the definition of subdifferential; moreover, the lower semicontinuity of f is not needed).

Recall that, for an arbitrary monotone operator $T : X \rightrightarrows X^*$, a lsc convex function $h : X \times X^* \rightarrow]-\infty, +\infty]$ is said to be a *representative function* of T if

$$h(x, x^*) \begin{cases} = \langle x, x^* \rangle, & \text{if } (x, x^*) \in \text{gph } T, \\ > \langle x, x^* \rangle, & \text{elsewhere.} \end{cases}.$$

Operators for which a representative function exists are called *representable monotone*. For a detailed study of representable monotone operators we refer to [22], where this notion was introduced.

Remark 2.2: From Theorem 2.1(vi), observe that, if $f : X \rightarrow]-\infty, +\infty]$ is a proper lsc convex function, the function $h_f : X \times X^* \rightarrow]-\infty, +\infty]$ defined by

$$h_f(x, x^*) = f(x) + f^*(x^*), \quad \text{for } (x, x^*) \in X \times X^* \quad (4)$$

is a representative function of ∂f . More generally, every maximally monotone operator is representable, as far as its well-known Fitzpatrick function is a representative function (see, e.g. [2, Section 9.1.2] for details). An easy consequence of this fact is that intersections of arbitrary collections of maximally monotone operators are representable, too. According to [22, Corollary 32], in finite-dimensional spaces only such intersections are representable. This is no longer true in infinite dimensional spaces, as proved in [23, Theorem 11.2].

The rest of this section is devoted to recall some results about metric projections and, in order to ensure existence and uniqueness of the best approximation to closed convex sets, we assume that Y is a Hilbert space. Here, $\|\cdot\|$ denotes the norm associated with the corresponding inner product $\langle \cdot, \cdot \rangle$. Given any

nonempty closed convex set $S \subset Y$, we denote by $P_S : Y \rightarrow Y$ the *metric projection* on S , which assigns to each $x \in Y$ its (unique) best approximation in S , denoted by $P_S(x)$, i.e. $P_S(x)$ is the unique point of S such that

$$\|x - P_S(x)\| = d(x, S) = \min \{\|x - s\| : s \in S\}.$$

(Observe that we write $P_S : Y \rightarrow Y$ instead of $P_S : Y \rightrightarrows Y$ due to its single-valuedness.) It is well-known that the function $x \mapsto d(x, S)$, denoted for convenience by $d_S : Y \rightarrow [0, +\infty[$, is a continuous convex function. Recall that, for a continuous convex function, $f : Y \rightarrow \mathbb{R}$, applying [2, Corollary 4.2.5], we deduce that f is Gâteaux differentiable at a point x if and only if $\partial f(x)$ reduces to a singleton, i.e. $\partial f(x) = \{\nabla f(x)\}$; see [2, Section 2] for details. In our applications, the facts that the subdifferentials $\partial d_S(x)$ or $\partial d_S^2(x)$ reduce to a singleton are crucial. Accordingly, condition (i) in the following proposition is stated directly in these terms (instead of Gâteaux differentiability). From now on, $N_S(x)$ denotes the normal cone to S at $x \in S$, which is given by

$$N_S(x) := \{x^* \in Y \mid \langle s - x, x^* \rangle \leq 0, \text{ for all } s \in S\}, \quad (5)$$

B denotes the closed unit ball in Y and $\text{bd}S$ the boundary of S .

Proposition 2.3: *Let Y be a Hilbert space and $\emptyset \neq S \subset Y$ a closed set. Then, we have*

- (i) [2, Corollary 4.2.5 and Theorem 4.5.7] *S is convex if and only if $\partial d_S^2(x)$ is singleton for all $x \in Y$; in such a case,*

$$\nabla d_S^2(x) = 2(x - P_S(x)).$$

- (ii) [19, Proposition 4.1.5] (see also [24, Section 1]) *If S is convex, then*

$$\partial d_S(x) = \begin{cases} \{0\}, & \text{if } x \in \text{int}S, \\ N_S(x) \cap B, & \text{if } x \in \text{bd}S, \\ \{\|x - P_S(x)\|^{-1}(x - P_S(x))\}, & \text{if } x \notin S. \end{cases}$$

3. On paramonotone and bimonotone operators

This section provides some results, appealed to in Sections 4 and 5, about operators which are simultaneously paramonotone and bimonotone on a real Banach space X . To start with, we provide some basic results on these two properties separately.

Proposition 3.1: *Let $T : X \rightrightarrows X^*$ be a representable monotone operator. The following statements are equivalent:*

- (i) *T is paramonotone;*

(ii) For any representative function h of T , the following implication holds:

$$\left. \begin{array}{l} (x, x^*), (y, y^*) \in \text{gph} T \\ \langle x - y, x^* - y^* \rangle = 0 \end{array} \right\} \Rightarrow h(x, y^*) + h(y, x^*) = h(x, x^*) + h(y, y^*). \quad (6)$$

(iii) There exists a representative function h of T such that (6) holds.

Proof: (i) $\Rightarrow$ (ii). Consider any representative function of T , h , and take $(x, x^*), (y, y^*) \in \text{gph} T$, with $\langle x - y, x^* - y^* \rangle = 0$. Then, the paramonotonicity entails $y^* \in T(x)$ and $x^* \in T(y)$, yielding

$$h(x, y^*) + h(y, x^*) = \langle x, y^* \rangle + \langle y, x^* \rangle = \langle x, x^* \rangle + \langle y, y^* \rangle = h(x, x^*) + h(y, y^*);$$

(ii) $\Rightarrow$ (iii). Straightforward.

(iii) $\Rightarrow$ (i). Let h be a representative function of T satisfying (6). Let $x, y \in X$, $x^* \in T(x)$, $y^* \in T(y)$ and suppose $\langle x - y, x^* - y^* \rangle = 0$. Hence,

$$h(x, y^*) + h(y, x^*) = h(x, x^*) + h(y, y^*) = \langle x, x^* \rangle + \langle y, y^* \rangle = \langle x, y^* \rangle + \langle y, x^* \rangle.$$

Since $h(x, y^*) \geq \langle x, y^* \rangle$ and $h(y, x^*) \geq \langle y, x^* \rangle$, these inequalities actually hold as equalities, yielding $y^* \in T(x)$ and $x^* \in T(y)$. ■

Remark 3.2: Observe that the paramonotonicity of the subdifferential operator ∂f of a proper lsc convex function f can be alternatively deduced from Proposition 3.1. Just consider the representative function h_f introduced in (4), which is separable and, hence, one always has

$$h_f(x, y^*) + h_f(y, x^*) = h_f(x, x^*) + h_f(y, y^*).$$

Other examples of paramonotone operators are mappings of the form $I - A$ where I is the identity mapping and A is nonexpansive (see [25, Theorem 6.1]); see also [10, Section 3] for the analysis of paramonotonicity of affine functions in $\mathbb{R}^n$.

We next present two easy characterizations of bimonotone operators.

Proposition 3.3: For an operator $T : X \rightrightarrows X^*$, the following conditions are equivalent:

- (i) T is bimonotone;
- (ii) $\langle x - y, x^* - y^* \rangle = 0$, whenever $(x, x^*), (y, y^*) \in \text{gph} T$;
- (iii) There exist monotone operators T_1 and T_2 such that $T = T_1 \cap (-T_2)$.

Proof: (i) $\Leftrightarrow$ (ii) is trivial.

(i) $\Rightarrow$ (iii). Write $T = T \cap (-(-T))$.

(iii) $\Rightarrow$ (i). If T_1 is monotone, so is T , since $T \subset T_1$. Analogously, since

$$-T = T_2 \cap (-T_1),$$

$-T$ is monotone, too. ■

In the same way as Proposition 3.1 characterizes paramonotone operators in terms of representative functions, the following proposition provides an analogous characterization of bimonotone operators.

Proposition 3.4: *For a representable monotone operator $T : X \rightrightarrows X^*$, the following conditions are equivalent:*

- (i) T is bimonotone;
- (ii) For any representative function h of T , the following implication holds

$$h(x, x^*) + h(y, y^*) = \langle x, x^* \rangle + \langle y, y^* \rangle \Rightarrow h(x, x^*) + h(y, y^*) = \langle x, y^* \rangle + \langle y, x^* \rangle. \quad (7)$$

- (iii) There exists a representative function h of T such that (7) holds.

Proof: (i) $\Rightarrow$ (ii). Consider any representative function h of T , and assume that

$$h(x, x^*) + h(y, y^*) = \langle x, x^* \rangle + \langle y, y^* \rangle.$$

Hence, $h(x, x^*) = \langle x, x^* \rangle$ and $h(y, y^*) = \langle y, y^* \rangle$, that is, $(x, x^*), (y, y^*) \in \text{gph} T$, yielding $\langle x - y, x^* - y^* \rangle = 0$. Consequently,

$$\langle x, y^* \rangle + \langle y, x^* \rangle = \langle x, x^* \rangle + \langle y, y^* \rangle = h(x, x^*) + h(y, y^*).$$

(ii) $\Rightarrow$ (iii). Straightforward.

(iii) $\Rightarrow$ (i). Let $x, y \in X$, $(x, x^*), (y, y^*) \in \text{gph} T$. We then have

$$h(x, x^*) + h(y, y^*) = \langle x, x^* \rangle + \langle y, y^* \rangle,$$

and hence $\langle x, y^* \rangle + \langle y, x^* \rangle = h(x, x^*) + h(y, y^*) = \langle x, x^* \rangle + \langle y, y^* \rangle$, from which the equality $\langle x - y, x^* - y^* \rangle = 0$ immediately follows. ■

From now on, the symbol ' $\perp$ ' represents orthogonality; specifically, given any subsets $A \subset X$ and $B \subset X^*$, $A \perp B$ means that $\langle x, x^* \rangle = 0$ for any $(x, x^*) \in A \times B$, whereas $A^\perp := \{x^* \in X^* \mid \langle x, x^* \rangle = 0, \text{ for all } x \in A\}$ and $B^\perp := \{x \in X \mid \langle x, x^* \rangle = 0, \text{ for all } x^* \in B\}$.

In view of the applications to be developed in the next sections, we are mainly interested in operators which are both paramonotone and bimonotone. Our next result characterizes such operators.

Proposition 3.5: For $T : X \rightrightarrows X^*$, the following conditions are equivalent:

- (i) T is paramonotone and bimonotone;
- (ii) T is monotone and constant on its domain;
- (iii) $(\text{dom} T - \text{dom} T)^\perp (\text{range } T - \text{range } T)$ and $\text{gph } T = \text{dom} T \times \text{range } T$.

Proof: (i) $\Rightarrow$ (ii). Let $x, y \in \text{dom} T$, and take $x^* \in T(x)$, $y^* \in T(y)$. By bimonotonicity, we have $\langle x - y, x^* - y^* \rangle = 0$. Hence, by paramonotonicity, $y^* \in T(x)$ and $x^* \in T(y)$. This proves that $T(x) = T(y)$.

(ii) $\Rightarrow$ (i). The paramonotonicity of T is an obvious consequence of its being constant on its domain. To prove bimonotonicity, let $(x, x^*), (y, y^*) \in \text{gph } T$. Monotonicity implies $\langle x - y, x^* - y^* \rangle \geq 0$, and we can interchange x^* and y^* , since $T(x) = T(y)$; therefore $\langle x - y, x^* - y^* \rangle = 0$.

(ii) $\Leftrightarrow$ (iii) Comes from the fact that T is constant on $\text{dom} T$ if and only if $\text{gph } T = \text{dom} T \times \text{range } T$. ■

The following proposition deals with the case when a paramonotone and bimonotone operator has a dense domain. It turns out that such operators are single-valued. Similarly, if a paramonotone and bimonotone operator has a dense range, then its inverse is single-valued. We also provide a characterization of those paramonotone and bimonotone operators that are maximally monotone.

Proposition 3.6: Let $T : X \rightrightarrows X^*$ be paramonotone and bimonotone. Then, we have

- (i) If $\text{dom} T$ is dense in X , then $\text{range } T$ is a singleton;
- (ii) If $\text{range } T$ is dense in X^* , then $\text{dom} T$ is a singleton;
- (iii) T is maximally monotone if and only if $\text{dom} T$ and $\text{range } T$ are closed affine varieties and

$$\text{dom} T - \text{dom} T = (\text{range } T - \text{range } T)^\perp. \quad (8)$$

Proof: (i) Assume, reasoning by contradiction, that there exist (x, x^*) and $(x, \tilde{x}^*)$ in $\text{gph } T$ with $x^* \neq \tilde{x}^*$ and take $u \in X$ with $\langle u, x^* - \tilde{x}^* \rangle \neq 0$. Under the current assumption, we can take a sequence $\{x^r\}_{r \in \mathbb{N}} \subset \text{dom} T$ converging to $x + u$. For r large enough we have $\langle x^r - x, x^* - \tilde{x}^* \rangle \neq 0$ and $x^* \in T(x) = T(x^r)$ because of Proposition 3.5. This contradicts (i) $\Rightarrow$ (ii) in Proposition 3.3. We have thus proved that T is single-valued on its domain, which, by Proposition 3.5, means that $\text{range } T$ is a singleton.

(ii) follows analogously to (i) by considering (x, x^*) and $(\tilde{x}, x^*)$ in $\text{gph } T$ with $x \neq \tilde{x}$, taking again Proposition 3.5 into account.

(iii) Assume that T is maximally monotone. Take $x_0 \in \text{dom} T$, $x_0^* \in \text{range } T$, and let S and S_* be the linear subspaces generated by $\text{dom} T - \text{dom} T$ and

$\text{range } T - \text{range } T$, respectively. Define $\widehat{T} : X \rightrightarrows X^*$ by

$$\widehat{T}(x) := \begin{cases} x_0^* + S^\perp & \text{if } x \in x_0 + \text{cl } S, \\ \emptyset & \text{otherwise.} \end{cases}$$

We have

$$\begin{aligned} \text{dom } \widehat{T} - \text{dom } \widehat{T} &= (x_0 + \text{cl } S) - (x_0 + \text{cl } S) = \text{cl } S - \text{cl } S = \text{cl } S = (S^\perp)^\perp \\ &= (S^\perp - S^\perp)^\perp = \left((x_0^* + S^\perp) - (x_0^* + S^\perp) \right)^\perp \\ &= (\text{range } \widehat{T} - \text{range } \widehat{T})^\perp, \end{aligned}$$

which proves (8) for operator $\widehat{T}$. Moreover,

$$\text{gph } \widehat{T} = (x_0 + \text{cl } S) \times (x_0^* + S^\perp) = \text{dom } \widehat{T} \times \text{range } \widehat{T}.$$

Therefore, by equivalence (i) $\Leftrightarrow$ (iii) in Proposition 3.5, the operator $\widehat{T}$ is paramonotone and bimonotone; in particular, $\widehat{T}$ is monotone. On the other hand, by the same equivalence, we have

$$\text{range } T - \text{range } T \subset (\text{dom } T - \text{dom } T)^\perp = S^\perp;$$

hence $S_* \subset S^\perp$ and

$$\begin{aligned} \text{gph } T &= \text{dom } T \times \text{range } T \subset (x_0 + \text{cl } S) \times (x_0^* + \text{cl } S_*) \subset \text{dom } \widehat{T} \times (x_0^* + S^\perp) \\ &= \text{dom } \widehat{T} \times \text{range } \widehat{T} = \text{gph } \widehat{T}. \end{aligned}$$

Thus, by the maximal monotonicity of T , we have $T = \widehat{T}$, from which we deduce that $\text{dom } T = \text{dom } \widehat{T} = x_0 + \text{cl } S$ and $\text{range } T = \text{range } \widehat{T} = x_0^* + S^\perp$, thus proving that $\text{dom } T$ and $\text{range } T$ are closed affine varieties.

Let us see the converse implication. Let $(x, x^*) \in X \times X^*$ be such that

$$\langle x - y, x^* - y^* \rangle \geq 0 \quad \text{for all } (y, y^*) \in \text{gph } T = \text{dom } T \times \text{range } T \quad (9)$$

(the latter equality following again from Proposition 3.5). Since $\text{range } T$ is an affine variety, we can easily prove that $x - y \in (\text{range } T - \text{range } T)^\perp = \text{dom } T - \text{dom } T$. More in detail, replace y^* in (9) with $x_0^* \pm \lambda v$ for any given $x_0^* \in \text{range } T$ and any $v \in \text{range } T - \text{range } T$, with $\lambda > 0$, then divide both sides of the resulting specification of (9) by λ and let $\lambda \rightarrow +\infty$ to obtain $\langle x - y, \pm v \rangle \geq 0$. Therefore, given that $\text{dom } T$ is an affine variety, we deduce that $x \in \text{dom } T$. Similarly, using that $\text{range } T - \text{range } T = (\text{dom } T - \text{dom } T)^\perp$, we obtain that $x^* \in \text{range } T$. Thus, $(x, x^*) \in \text{dom } T \times \text{range } T = \text{gph } T$, which proves that T is maximally monotone. ■

The following propositions involve a finite number of paramonotone operators and are intended to provide a unified framework to deal with the applications

of Sections 4 and 5. First, we introduce the following lemma, which has an easy proof.

Lemma 3.7: *If $T_1 : X \rightrightarrows X^*$ and $T_2 : X \rightrightarrows X^*$ are paramonotone, then so are $T_1 + T_2$ and $T_1 \cap (-T_2)$.*

Proposition 3.8: *Let $T_i : X \rightrightarrows X^*$, $i = 1, \dots, m$, be paramonotone operators. Then the intersection mappings*

$$\tilde{T}_i := T_i \cap \left(- \sum_{j \neq i} T_j \right), \quad i = 1, \dots, m.$$

are monotone and constant in their common domain

$$\mathcal{A} := \left\{ x \in X \mid 0 \in \sum_{j=1}^m T_j(x) \right\}.$$

Proof: Fix $i \in \{1, \dots, m\}$. From Lemma 3.7, $\sum_{j \neq i} T_j$ is paramonotone and, hence, the same lemma establishes that $\tilde{T}_i$ is paramonotone. Then, from equivalence (i) $\Leftrightarrow$ (iii) in Proposition 3.3, $\tilde{T}_i$ is bimonotone. Hence Proposition 3.5 (ii) yields that $\tilde{T}_i$ is constant in $\text{dom} \tilde{T}_i$. Finally, one easily sees that $\text{dom} \tilde{T}_i$ coincides with $\mathcal{A}$. ■

Now, we particularize Proposition 3.8 by considering finitely many proper lsc convex functions, $f_i : X \rightarrow]-\infty, +\infty]$, $i = 1, \dots, m$, and the corresponding subdifferential operators $T_i := \partial f_i$, $i = 1, \dots, m$. We assume the following regularity condition in order to apply the subdifferential sum rule (see Theorem 2.1 (vii)): there exists some index $i_0 \in \{1, \dots, m\}$, such that

$$\text{dom} f_{i_0} \cap \left(\bigcap_{i \neq i_0} \text{intdom} f_i \right) \neq \emptyset, \quad (10)$$

which is equivalent to the existence of some $\bar{x} \in \bigcap_{i=1, \dots, m} \text{dom} f_i$ such that the $m-1$ of the functions f_i , $i \in \{1, \dots, m\} \setminus \{i_0\}$ are continuous at $\bar{x}$ (see Theorem 2.1 (i)).

In this particular case, we are considering the operators

$$\tilde{T}_i := \partial f_i \cap \left(- \sum_{j \neq i} \partial f_j \right), \quad i = 1, \dots, m, \quad (11)$$

whose common domain, appealing to statements (iii) and (vii) in Theorem 2.1, writes as

$$\mathcal{A} = \left\{ x \in X \mid 0 \in \sum_{i=1}^m \partial f_i(x) \right\} = \arg \min \sum_{i=1}^m f_i.$$

We summarize the previous comments in the following proposition.

Proposition 3.9: *Let $f_i : X \rightarrow]-\infty, +\infty]$, $i = 1, \dots, m$, be proper lsc convex functions and assume that for some $i_0 \in \{1, \dots, m\}$ condition (10) holds. Then, operators $\tilde{T}_i$, $i \in \{1, \dots, m\}$, defined in (11) are constant on their common domain*

$$\mathcal{A} = \arg \min \sum_{i=1}^m f_i.$$

Remark 3.10: Proposition 3.9 can be applied to specific operators in order to derive some classical statements which can be found in the literature, as the one of [26, Lemma 2] involving $\partial f \cap (-N_C)$, and regarding the optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C, \end{array}$$

in the case when $f : \mathbb{R}^n \rightarrow]-\infty, +\infty]$ is a proper lsc convex function and C is a closed convex subset of $\mathbb{R}^n$. Specifically, if S denotes the set of optimal solutions of such a problem, [26, Lemma 2] states that $\partial f(x) \cap (-N_C(x))$ is independent of $x \in S$. To derive this statement from Proposition 3.9, just observe that the normal cone operator, N_C (recall (5)), is paramonotone as it coincides with the subdifferential of the indicator function of C .

Corollary 3.11: *Under the assumptions of Proposition 3.9, one has:*

- (i) *If for some $j_0 \in \{1, \dots, m\}$, the function f_{j_0} is differentiable at $\bar{x} \in \mathcal{A}$, then $\tilde{T}_{j_0}(x) = \{\nabla f_{j_0}(\bar{x})\}$, for all $x \in \mathcal{A}$.*
- (ii) *If for some $j_0 \in \{1, \dots, m\}$, the function f_{j_0} is differentiable on $\mathcal{A}$, then ∇f_{j_0} is constant on $\mathcal{A}$.*

Proof: (i) follows straightforwardly from Proposition 3.9, taking into account that if f_{j_0} is differentiable at $\bar{x}$, then $\tilde{T}_{j_0}(\bar{x}) = \partial f_{j_0}(\bar{x}) = \{\nabla f_{j_0}(\bar{x})\}$ (since $\emptyset \neq \tilde{T}_{j_0}(\bar{x}) \subset \{\nabla f_{j_0}(\bar{x})\}$), which entails that $\tilde{T}_{j_0}(x) = \tilde{T}_{j_0}(\bar{x}) = \{\nabla f_{j_0}(\bar{x})\}$ whenever $x \in \mathcal{A}$.

(ii) comes from (i) since for every $\bar{x}, x \in \mathcal{A}$ we have

$$\{\nabla f_{j_0}(\bar{x})\} = \tilde{T}_{j_0}(x) \subset \partial f_{j_0}(x) = \{\nabla f_{j_0}(x)\};$$

hence $\nabla f_{j_0}(\bar{x}) = \nabla f_{j_0}(x)$. ■

4. Simultaneous projections and displacement mappings

This section is mainly devoted to study the minimal weighted distance to two disjoint non-empty closed and convex subsets S_1 and S_2 of a Hilbert space Y . We

will denote by $d : Y \times Y \rightarrow \mathbb{R}$ the *distance function* on Y , i.e. $d(x, y) := \|x - y\|$, and by $d_{S_i} : Y \rightarrow \mathbb{R}$ the distance function to S_i , $i = 1, 2$. We set

$$d(S_1, S_2) := \inf_{s_1 \in S_1, s_2 \in S_2} d(s_1, s_2).$$

For arbitrary real numbers $\alpha_1, \alpha_2 > 0$, with $\alpha_1 + \alpha_2 = 1$, and $p \geq 1$, we define

$$v(\alpha_1, \alpha_2, p) := \inf_{x \in Y} [\alpha_1 d(x, S_1)^p + \alpha_2 d(x, S_2)^p],$$

$$\mathcal{A}(\alpha_1, \alpha_2, p) := \arg \min [\alpha_1 d_{S_1}^p + \alpha_2 d_{S_2}^p].$$

Observe that $v(\alpha_1, \alpha_2, p)$ and $\mathcal{A}(\alpha_1, \alpha_2, p)$ are the optimal value and the set of optimal solutions of problem (2) for the case of two sets. Notice that $\mathcal{A}(\alpha_1, \alpha_2, p)$ may be empty; consider, e.g. the case when $Y := \mathbb{R}^2$, S_1 is the convex hull of a branch of a hyperbola and S_2 is one of its asymptotes; in this case $v(\alpha_1, \alpha_2, p) = 0$ is not attained.

We denote by $P_1 := P_{S_1}$ and $P_2 := P_{S_2}$ the metric projections over S_1 and S_2 , respectively.

We are going to consider three cases depending on the values of the power p and parameters α_1 and α_2 . At this moment we advance that in the case when $d_{S_1}^p$ and $d_{S_2}^p$ are differentiable we are able to apply Corollary 3.11 to derive information about $\mathcal{A}(\alpha_1, \alpha_2, p)$. Going further, Proposition 2.3(i) establishes the differentiability of $d_{S_i}^p$ on the whole space Y when $p \geq 2$, which allows us to tackle the case of a finite amount of sets.

Case 1. $p = 1$, $\alpha_1 \neq \alpha_2$. Without loss of generality, we assume that $\alpha_1 > \alpha_2$. The following result has a clear geometrical meaning according to Proposition 1.1. In it $\Pi_1 : S_1 \times S_2 \rightarrow S_1$ denotes the projection mapping, defined by $\Pi_1(s_1, s_2) = s_1$. Observe that, appealing to this projection mapping, $\mathcal{A}(\alpha_1, \alpha_2, 1)$ can be seen as the set of the first components $s_1 \in S_1$ of pairs of closest points $(s_1, s_2) \in S_1 \times S_2$.

Proposition 4.1: *If $\alpha_1 > \alpha_2$, then:*

- (i) $v(\alpha_1, \alpha_2, 1) = \alpha_2 d(S_1, S_2)$,
- (ii) $\mathcal{A}(\alpha_1, \alpha_2, 1) = \arg \min_{S_1} d_{S_2} = \Pi_1(\arg \min_{S_1 \times S_2} d)$.

Proof: (i) To prove inequality ' $\geq$ ', just observe that, for each $x \in Y$, we have

$$\begin{aligned} \alpha_1 d(x, S_1) + \alpha_2 d(x, S_2) &\geq \alpha_2 (d(x, S_1) + d(x, S_2)) \\ &= \alpha_2 (\|P_1(x) - x\| + d(x, S_2)) \\ &\geq \alpha_2 d(P_1(x), S_2) \\ &\geq \alpha_2 d(S_1, S_2). \end{aligned} \tag{12}$$

Let us show the opposite inequality ' $\leq$ '. We have

$$\begin{aligned}\alpha_2 d(S_1, S_2) &= \alpha_2 \inf_{x \in S_1} \inf_{y \in S_2} d(x, y) = \alpha_2 \inf_{x \in S_1} d(x, S_2) \\ &= \inf_{x \in S_1} [\alpha_1 d(x, S_1) + \alpha_2 d(x, S_2)] \geq v(\alpha_1, \alpha_2, 1).\end{aligned}\quad (13)$$

(ii) First, we have that $\mathcal{A}(\alpha_1, \alpha_2, 1) \subset S_1$ since for any $x \notin S_1$ the first inequality of (12) is strict. Hence

$$\begin{aligned}\mathcal{A}(\alpha_1, \alpha_2, 1) &\subset \arg \min_{x \in S_1} [\alpha_1 d(x, S_1) + \alpha_2 d(x, S_2)] \\ &= \arg \min_{S_1} \alpha_2 d_{S_2}.\end{aligned}$$

To show the opposite inclusion, take $\bar{x} \in \arg \min_{S_1} d_{S_2}$ and observe that

$$\begin{aligned}\alpha_1 d(\bar{x}, S_1) + \alpha_2 d(\bar{x}, S_2) &= \alpha_2 d(\bar{x}, S_2) = \alpha_2 \inf_{x \in S_1} d(x, S_2) \\ &= \alpha_2 d(S_1, S_2) = v(\alpha_1, \alpha_2, 1),\end{aligned}$$

where we have used the fact that, for $x := \bar{x}$, the inequality of (13) is indeed an equality, appealing to statement (i).

Now let us show that $\arg \min_{S_1} d_{S_2} = \Pi_1(\arg \min_{S_1 \times S_2} d)$. To prove the inclusion $\subset$, let $\bar{s}_1 \in \arg \min_{S_1} d_{S_2}$ and $s_1 \in S_1$. Then, for every $s_2 \in S_2$, we have

$$d(\bar{s}_1, P_2(\bar{s}_1)) = d(\bar{s}_1, S_2) \leq d(s_1, S_2) \leq d(s_1, s_2);$$

hence $(\bar{s}_1, P_2(\bar{s}_1)) \in \arg \min_{S_1 \times S_2} d$, implying that $\bar{s}_1 \in \Pi_1(\arg \min_{S_1 \times S_2} d)$, thus proving the desired inclusion. We now proceed to prove the opposite inclusion. Let $\bar{s}_1 \in \Pi_1(\arg \min_{S_1 \times S_2} d)$ and $s_1 \in S_1$. There exists $\bar{s}_2 \in S_2$ such that $(\bar{s}_1, \bar{s}_2) \in \arg \min_{S_1 \times S_2} d$, and for every $s_2 \in S_2$ we have

$$d(\bar{s}_1, S_2) \leq d(\bar{s}_1, \bar{s}_2) \leq d(s_1, s_2);$$

taking the infimum over $s_2 \in S_2$, this yields $d(\bar{s}_1, S_2) \leq d(s_1, S_2)$, which implies that $\bar{s}_1 \in \arg \min_{S_1} d_{S_2}$. Thus $\Pi_1(\arg \min_{S_1 \times S_2} d) \subset \arg \min_{S_1} d_{S_2}$, and the proof is complete. ■

Case 2. $p = 1, \alpha_1 = \alpha_2 = \frac{1}{2}$.

From now on $]P_1(x), P_2(x)[$ represents the segment of points between $P_1(x)$ and $P_2(x)$, except these two ones; i.e. $]P_1(x), P_2(x)[:= \{(1 - \lambda)P_1(x) + \lambda P_2(x) : 0 < \lambda < 1\}$.

According to Proposition 4.1, one has $\mathcal{A}(\alpha_1, \alpha_2, 1) \subset S_i$ when $\alpha_i > \alpha_{3-i}$. Our next proposition shows that the situation is quite different when $\alpha_1 = \alpha_2$: indeed, $\mathcal{A}(\frac{1}{2}, \frac{1}{2}, 1) \cap S_1 \neq \emptyset \neq \mathcal{A}(\frac{1}{2}, \frac{1}{2}, 1) \cap S_2$ whenever $d(S_1, S_2) = \|\bar{s}_1 - \bar{s}_2\|$ for some $(\bar{s}_1, \bar{s}_2) \in S_1 \times S_2$, in which case one can easily prove, using statement (ii), that $\frac{1}{2}(\bar{s}_1 + \bar{s}_2) \in \mathcal{A}(\frac{1}{2}, \frac{1}{2}, 1)$, so that $\mathcal{A}(\frac{1}{2}, \frac{1}{2}, 1) \not\subset S_1 \cup S_2$.

Proposition 4.2: *One has:*

- (i) $v(\frac{1}{2}, \frac{1}{2}, 1) = \frac{1}{2}d(S_1, S_2)$,
- (ii) $\mathcal{A}(\frac{1}{2}, \frac{1}{2}, 1) = \{x \in Y : x \in]P_1(x), P_2(x)[\} \cup \arg \min_{S_1} d_{S_2} \cup \arg \min_{S_2} d_{S_1}$.

Proof: (i) For $x \in Y$, we have

$$\begin{aligned} d(x, S_1) + d(x, S_2) &= \|x - P_1(x)\| + \|x - P_2(x)\| \geq \|P_1(x) - P_2(x)\| \\ &\geq d(S_1, S_2), \end{aligned}$$

which proves the inequality $\geq$. To prove the opposite inequality, it suffices to observe that, for $s_1 \in S_1$ and $s_2 \in S_2$, we have

$$\begin{aligned} \|s_1 - s_2\| &= d(s_1, S_1) + \|s_1 - s_2\| \geq d(s_1, S_1) + d(s_1, S_2) \\ &\geq 2v(\frac{1}{2}, \frac{1}{2}, 1). \end{aligned}$$

(ii) Let $x \in \mathcal{A}(\frac{1}{2}, \frac{1}{2}, 1)$. If $x \notin S_1 \cup S_2$, then $x \in]P_1(x), P_2(x)[$, since otherwise we would have

$$\begin{aligned} d(x, S_1) + d(x, S_2) &= \|x - P_1(x)\| + \|x - P_2(x)\| > \|P_1(x) - P_2(x)\| \\ &\geq d(S_1, S_2), \end{aligned}$$

a contradiction with (i). If $x \in S_1$, then, for any $s_1 \in S_1$ we have

$$d(x, S_2) = d(x, S_1) + d(x, S_2) \leq d(s_1, S_1) + d(s_1, S_2) = d(s_1, S_2),$$

which shows that $x \in \arg \min_{S_1} d_{S_2}$. In the same way, if $x \in S_2$, then $x \in \arg \min_{S_2} d_{S_1}$. We have thus proved the inclusion $\subset$. To prove the opposite inclusion, let $x \in Y$ be such that $x \in]P_1(x), P_2(x)[$ and take $\lambda \in]0, 1[$ such that $x = (1 - \lambda)P_1(x) + \lambda P_2(x)$. Combining this equality with the inequalities $\langle s_i - P_i(x), x - P_i(x) \rangle \leq 0$, which hold for every $s_i \in S_i$, $i = 1, 2$, we obtain $\langle s_1 - P_1(x), P_2(x) - P_1(x) \rangle \leq 0$ and $\langle s_2 - P_2(x), P_1(x) - P_2(x) \rangle \leq 0$. Adding the latter inequalities, we get $\langle s_2 - s_1 + P_1(x) - P_2(x), P_1(x) - P_2(x) \rangle \leq 0$; hence

$$\|P_1(x) - P_2(x)\|^2 \leq \langle s_1 - s_2, P_1(x) - P_2(x) \rangle \leq \|s_1 - s_2\| \|P_1(x) - P_2(x)\|.$$

Therefore, $\|s_1 - s_2\| \geq \|P_1(x) - P_2(x)\|$, and we deduce that

$$\begin{aligned} d(x, S_1) + d(x, S_2) &= \|x - P_1(x)\| + \|x - P_2(x)\| = \|P_1(x) - P_2(x)\| \leq \|s_1 - s_2\|. \end{aligned}$$

Since $s_i \in S_i$, $i = 1, 2$, are arbitrarily chosen, we conclude

$$d(x, S_1) + d(x, S_2) \leq d(S_1, S_2),$$

which, by (i), says that $x \in \mathcal{A}(\frac{1}{2}, \frac{1}{2}, 1)$. It remains to prove that

$$\arg \min_{S_1} d_{S_2} \cup \arg \min_{S_2} d_{S_1} \subset \mathcal{A}(\frac{1}{2}, \frac{1}{2}, 1).$$

For symmetry reasons, it suffices to prove that $\arg \min_{S_1} d_{S_2} \subset \mathcal{A}(\frac{1}{2}, \frac{1}{2}, 1)$, but this inclusion follows from the fact that, for $x \in \arg \min_{S_1} d_{S_2}$, we have

$$\begin{aligned} d(x, S_1) + d(x, S_2) &= d(x, S_2) = \min_{s_1 \in S_1} d(s_1, S_2) = \min_{s_1 \in S_1} \min_{s_2 \in S_2} d(s_1, s_2) \\ &= \min_{s_1 \in S_1, s_2 \in S_2} d(s_1, s_2) = d(S_1, S_2). \end{aligned} \quad \blacksquare$$

Case 3. $p > 1$. In our current setting, it is known that the function $d_{S_i} : Y \rightarrow \mathbb{R}$ is convex and differentiable outside S_i and for every $x \in Y \setminus S_i$ one has (recall Proposition 2.3(ii))

$$\nabla d_{S_i}(x) = (x - P_i(x)) / \|x - P_i(x)\|. \quad (14)$$

The following theorem shows that the case $p > 1$ is quite different from the case $p = 1$. Indeed, statement (i) is in sharp contrast with Propositions 4.1 and 4.2, which show that $\mathcal{A}(\alpha_1, \alpha_2, 1) \cap (S_1 \cup S_2) \neq \emptyset$ whenever $d(S_1, S_2)$ is attained. On the other hand, statement (iii) shows that $\mathcal{A}(\alpha_1, \alpha_2, p)$ coincides with the set of fixed points of a certain convex combination (depending on α_1 and α_2) of the projection mappings P_1 and P_2 .

Theorem 4.3: *If $p > 1$, then:*

- (i) $\mathcal{A}(\alpha_1, \alpha_2, p) \cap (S_1 \cup S_2) = \emptyset$.
- (ii) *For each $i = 1, 2$, the function $d_{S_i}^p$ is differentiable on $\mathcal{A}(\alpha_1, \alpha_2, p)$ and*

$$\nabla d_{S_i}^p(x) = p \|x - P_i(x)\|^{p-2} (x - P_i(x)), \quad x \in \mathcal{A}(\alpha_1, \alpha_2, p).$$

- (iii) $\mathcal{A}(\alpha_1, \alpha_2, p)$ *coincides with the set of fixed points of*

$$W_{\alpha_1, \alpha_2, p} := \frac{\alpha_1^{\frac{1}{p-1}}}{\alpha_1^{\frac{1}{p-1}} + \alpha_2^{\frac{1}{p-1}}} P_1 + \frac{\alpha_2^{\frac{1}{p-1}}}{\alpha_1^{\frac{1}{p-1}} + \alpha_2^{\frac{1}{p-1}}} P_2.$$

- (iv) $v(\alpha_1, \alpha_2, p) = \gamma_{\alpha_1, \alpha_2, p} d(S_1, S_2)^p$, *where*

$$\gamma_{\alpha_1, \alpha_2, p} := \alpha_1 \left(\frac{\alpha_2^{\frac{1}{p-1}}}{\alpha_1^{\frac{1}{p-1}} + \alpha_2^{\frac{1}{p-1}}} \right)^p + \alpha_2 \left(\frac{\alpha_1^{\frac{1}{p-1}}}{\alpha_1^{\frac{1}{p-1}} + \alpha_2^{\frac{1}{p-1}}} \right)^p.$$

Proof: (i) It will suffice to prove that $\mathcal{A}(\alpha_1, \alpha_2, p) \cap S_1 = \emptyset$. Let $x \in S_1$ and pick $\lambda > 0$ such that $\frac{\lambda^p}{1-(1-\lambda)^p} < \frac{\alpha_2}{\alpha_1}$ (this is possible, since $\lim_{\lambda \rightarrow 0^+} \frac{\lambda^p}{1-(1-\lambda)^p} = 0$). Since

$$d((1-\lambda)x + \lambda P_2(x), S_1) \leq d((1-\lambda)x + \lambda P_2(x), x) = \lambda \|P_2(x) - x\|$$

and

$$\begin{aligned} d((1-\lambda)x + \lambda P_2(x), S_2) \\ \leq d((1-\lambda)x + \lambda P_2(x), P_2(x)) = (1-\lambda) \|x - P_2(x)\|, \end{aligned}$$

we have

$$\begin{aligned} \alpha_1 d((1-\lambda)x + \lambda P_2(x), S_1)^p + \alpha_2 d((1-\lambda)x + \lambda P_2(x), S_2)^p \\ \leq \alpha_1 \lambda^p \|P_2(x) - x\|^p + \alpha_2 (1-\lambda)^p \|x - P_2(x)\|^p \\ = (\alpha_1 \lambda^p + \alpha_2 (1-\lambda)^p) \|x - P_2(x)\|^p < \alpha_2 \|x - P_2(x)\|^p \\ = \alpha_1 d(x, S_1)^p + \alpha_2 d(x, S_2)^p, \end{aligned}$$

which shows that $x \notin \mathcal{A}(\alpha_1, \alpha_2, p)$, thus proving that $\mathcal{A}(\alpha_1, \alpha_2, p)$ and S_1 are disjoint.

(ii) is a consequence of (i) taking (14) into account.

(iii) For simplicity of notation, for $x \in \mathcal{A}(\alpha_1, \alpha_2, p)$ and $i = 1, 2$ we will denote

$$D_i^p(x) := \alpha_i \nabla d_{S_i}^p(x).$$

Let $x \in \mathcal{A}(\alpha_1, \alpha_2, p)$. The equality $D_1^p(x) + D_2^p(x) = 0$ yields

$$\alpha_1 \|x - P_1(x)\|^{p-2} (x - P_1(x)) + \alpha_2 \|x - P_2(x)\|^{p-2} (x - P_2(x)) = 0, \quad (15)$$

from which we deduce that

$$\begin{aligned} x &= \frac{\alpha_1 \|x - P_1(x)\|^{p-2}}{\alpha_1 \|x - P_1(x)\|^{p-2} + \alpha_2 \|x - P_2(x)\|^{p-2}} P_1(x) \\ &\quad + \frac{\alpha_2 \|x - P_2(x)\|^{p-2}}{\alpha_1 \|x - P_1(x)\|^{p-2} + \alpha_2 \|x - P_2(x)\|^{p-2}} P_2(x) \\ &= \frac{\alpha_1}{\alpha_1 + \alpha_2 \left(\frac{\|x - P_2(x)\|}{\|x - P_1(x)\|} \right)^{p-2}} P_1(x) + \frac{\alpha_2}{\alpha_1 \left(\frac{\|x - P_1(x)\|}{\|x - P_2(x)\|} \right)^{p-2} + \alpha_2} P_2(x). \end{aligned}$$

Since the condition $D_1^p(x) + D_2^p(x) = 0$ implies that $\|D_1^p(x)\| = \|D_2^p(x)\|$, that is, $\alpha_1 p \|x - P_1(x)\|^{p-1} = \alpha_2 p \|x - P_2(x)\|^{p-1}$, which is equivalent to

the equality

$$\frac{\alpha_1}{\alpha_2} = \left(\frac{\|x - P_2(x)\|}{\|x - P_1(x)\|} \right)^{p-1}, \quad (16)$$

we obtain

$$x = \frac{\alpha_1}{\alpha_1 + \alpha_2 \left(\frac{\alpha_1}{\alpha_2} \right)^{\frac{p-2}{p-1}}} P_1(x) + \frac{\alpha_2}{\alpha_1 \left(\frac{\alpha_2}{\alpha_1} \right)^{\frac{p-2}{p-1}} + \alpha_2} P_2(x) = W_{\alpha_1, \alpha_2, p}(x).$$

This shows that x is a fixed point of $W_{\alpha_1, \alpha_2, p}$.

Conversely, if $x \in Y$ is a fixed point of $W_{\alpha_1, \alpha_2, p}$, then $x \notin S_1 \cup S_2$. Indeed, otherwise, if, say, $x \in S_1$, then, from the equalities $x = W_{\alpha_1, \alpha_2, p}(x)$ and $P_1(x) = x$ we would obtain $x = P_2(x) \in S_2$, thus contradicting the assumption that $S_1 \cap S_2 = \emptyset$. Therefore, the functions d_{S_i} , $i = 1, 2$, are differentiable at x . From $x = W_{\alpha_1, \alpha_2, p}(x)$, it follows that

$$\alpha_1^{\frac{1}{p-1}} (x - P_1(x)) + \alpha_2^{\frac{1}{p-1}} (x - P_2(x)) = 0, \quad (17)$$

from which we deduce (16). Now, using (16), we can rewrite (17) as (15) to obtain the equality $D_1^p(x) + D_2^p(x) = 0$, which shows that $x \in \mathcal{A}(\alpha_1, \alpha_2, p)$.

(iv) To prove this statement we distinguish two cases.

Case 1. $\mathcal{A}(\alpha_1, \alpha_2, p) \neq \emptyset$. In this case, take $\bar{x} \in \mathcal{A}(\alpha_1, \alpha_2, p)$ and, appealing to statement (iii), write

$$\begin{aligned} v(\alpha_1, \alpha_2, p) &= \alpha_1 d(\bar{x}, S_1)^p + \alpha_2 d(\bar{x}, S_2)^p \\ &= \alpha_1 \left\| \frac{\alpha_2^{\frac{1}{p-1}}}{\alpha_1^{\frac{1}{p-1}} + \alpha_2^{\frac{1}{p-1}}} (P_1(\bar{x}) - P_2(\bar{x})) \right\|^p \\ &\quad + \alpha_2 \left\| \frac{\alpha_1^{\frac{1}{p-1}}}{\alpha_1^{\frac{1}{p-1}} + \alpha_2^{\frac{1}{p-1}}} (P_1(\bar{x}) - P_2(\bar{x})) \right\|^p \\ &= \gamma_{\alpha_1, \alpha_2, p} \|P_1(\bar{x}) - P_2(\bar{x})\|^p \\ &= \gamma_{\alpha_1, \alpha_2, p} d(S_1, S_2)^p, \end{aligned}$$

where the last equality comes by taking into account that, from (iii) and Proposition 4.2, we have $\bar{x} \in]P_1(\bar{x}), P_2(\bar{x})[\subset \mathcal{A}(\frac{1}{2}, \frac{1}{2}, 1)$ and, hence,

$$\|P_1(\bar{x}) - P_2(\bar{x})\| = d(\bar{x}, S_1) + d(\bar{x}, S_2) = 2v\left(\frac{1}{2}, \frac{1}{2}, 1\right) = d(S_1, S_2).$$

Case 2. $\mathcal{A}(\alpha_1, \alpha_2, p) = \emptyset$. In this case, write

$$\begin{aligned} v(\alpha_1, \alpha_2, p) &= \lim_r \left[\alpha_1 d(x^r, S_1)^p + \alpha_2 d(x^r, S_2)^p \right] \\ &= \lim_r \left[\alpha_1 \|x^r - P_1(x^r)\|^p + \alpha_2 \|x^r - P_2(x^r)\|^p \right], \end{aligned}$$

for a certain sequence $\{x^r\}_{r \in \mathbb{N}} \subset Y$. For each r , let

$$\tilde{x}^r := W_{\alpha_1, \alpha_2, p}(x^r).$$

Next, for each r , we are going to apply statement (iii) to the particular case $\tilde{S}_1 := \{P_1(x^r)\}$ and $\tilde{S}_2 := \{P_2(x^r)\}$, denoting as $\tilde{W}_{\alpha_1, \alpha_2, p}$ the counterpart of $W_{\alpha_1, \alpha_2, p}$ for $\tilde{S}_1$ and $\tilde{S}_2$ playing the role of S_1 and S_2 , respectively. In this way, let us see that $\tilde{x}^r$ is a fixed point of $\tilde{W}_{\alpha_1, \alpha_2, p}$. We have

$$\begin{aligned} \tilde{W}_{\alpha_1, \alpha_2, p}(\tilde{x}^r) &= \frac{\alpha_1^{\frac{1}{p-1}}}{\alpha_1^{\frac{1}{p-1}} + \alpha_2^{\frac{1}{p-1}}} P_{\tilde{S}_1}(\tilde{x}^r) + \frac{\alpha_2^{\frac{1}{p-1}}}{\alpha_1^{\frac{1}{p-1}} + \alpha_2^{\frac{1}{p-1}}} P_{\tilde{S}_2}(\tilde{x}^r) \\ &= \frac{\alpha_1^{\frac{1}{p-1}}}{\alpha_1^{\frac{1}{p-1}} + \alpha_2^{\frac{1}{p-1}}} P_1(x^r) + \frac{\alpha_2^{\frac{1}{p-1}}}{\alpha_1^{\frac{1}{p-1}} + \alpha_2^{\frac{1}{p-1}}} P_2(x^r) \\ &= W_{\alpha_1, \alpha_2, p}(x^r) = \tilde{x}^r. \end{aligned}$$

Then, according to (iii),

$$\tilde{x}^r \in \arg \min_{x \in Y} [\alpha_1 d(x, P_1(x^r))^p + \alpha_2 d(x, P_2(x^r))^p].$$

Hence, appealing to Case 1 above (again, in the particular case $\tilde{S}_1 := \{P_1(x^r)\}$ and $\tilde{S}_2 := \{P_2(x^r)\}$) we have that

$$\alpha_1 d(\tilde{x}^r, P_1(x^r))^p + \alpha_2 d(\tilde{x}^r, P_2(x^r))^p = \gamma_{\alpha_1, \alpha_2, p} \|P_1(x^r) - P_2(x^r)\|^p.$$

(Observe that $\gamma_{\alpha_1, \alpha_2, p}$ does not depend on subsets S_1 and S_2 as it only involves α_1, α_2 and p). Therefore,

$$\begin{aligned} v(\alpha_1, \alpha_2, p) &= \lim_r [\alpha_1 d(x^r, P_1(x^r))^p + \alpha_2 d(x^r, P_2(x^r))^p] \\ &\geq \limsup_r [\alpha_1 d(\tilde{x}^r, P_1(x^r))^p + \alpha_2 d(\tilde{x}^r, P_2(x^r))^p] \\ &= \gamma_{\alpha_1, \alpha_2, p} \limsup_r \|P_1(x^r) - P_2(x^r)\|^p \geq \gamma_{\alpha_1, \alpha_2, p} d(S_1, S_2)^p. \end{aligned}$$

So, we have established the inequality $v(\alpha_1, \alpha_2, p) \geq \gamma_{\alpha_1, \alpha_2, p} d(S_1, S_2)^p$. In order to prove the opposite inequality, write

$$d(S_1, S_2) = \lim_r d(x^r, y^r)$$

for some sequences $\{x^r\} \subset S_1$, $\{y^r\} \subset S_2$. For each r , we have

$$\begin{aligned} v(\alpha_1, \alpha_2, p) &= \inf_{x \in Y} [\alpha_1 d(x, S_1)^p + \alpha_2 d(x, S_2)^p] \\ &\leq \inf_{x \in Y} [\alpha_1 \|x - x^r\|^p + \alpha_2 \|x - y^r\|^p] \\ &= \gamma_{\alpha_1, \alpha_2, p} d(x^r, y^r)^p, \end{aligned}$$

where again in the equality we have appealed to Case 1 to the particular case $\widehat{S}_1 := \{x^r\}$ and $\widehat{S}_2 := \{y^r\}$. Then we have

$$v(\alpha_1, \alpha_2, p) \leq \lim_r \gamma_{\alpha_1, \alpha_2, p} d(x^r, y^r)^p = \gamma_{\alpha_1, \alpha_2, p} d(S_1, S_2)^p,$$

which finishes the proof. ■

Notice that the set $\mathcal{A}(\frac{1}{2}, \frac{1}{2}, p)$ does not depend on p , since, by Theorem 4.3(iii), it coincides with the set of fixed points of $\frac{1}{2}(P_1 + P_2)$. Also notice that $\mathcal{A}(\alpha_1, \alpha_2, 2)$ coincides with the set of fixed points of $\alpha_1 P_1 + \alpha_2 P_2$.

The following lemma provides the counterpart of Theorem 4.3(ii) for the case $p \geq 2$. The stronger condition $p \geq 2$ guarantees that d_S^p is differentiable everywhere, not only on the optimal solution set $\mathcal{A}(\alpha_1, \alpha_2, p)$.

Lemma 4.4: *Take $p \geq 2$, and let $\emptyset \neq S \subset Y$ be a closed convex set. The function d_S^p is differentiable on Y and we have*

$$\nabla d_S^p(x) = p d_S^{p-2}(x) (x - P_S(x)), \quad \text{for } x \in Y.$$

Proof: Just write $d_S^p(x)$ as $(d_S^2(x))^{p/2}$ and apply Proposition 2.3(i). ■

The fact that the function d_S^p is differentiable on the whole space X enables us to tackle the case of a finite amount of subsets $S_1, \dots, S_m$, with $\bigcap_{i=1}^m S_i = \emptyset$, $m \in \mathbb{N}$. For simplicity, we use the notation

$$\mathcal{A}(\alpha, p) := \arg \min \sum_{i=1}^m \alpha_i d_{S_i}^p,$$

where $\alpha := (\alpha_1, \alpha_2, \dots, \alpha_m)$, with $\alpha_i > 0$, $i = 1, \dots, m$, and $\sum_{i=1}^m \alpha_i = 1$. The following theorem gathers the announced application of Corollary 3.11(ii). It has as an immediate consequence that the projections of a point $x \in \mathcal{A}(\alpha, p)$ determine those of every other point $y \in \mathcal{A}(\alpha, p)$, since one has $P_i(y) = y + P_i(x) - x$.

Theorem 4.5: *If $p > 1$ and $m = 2$, or $p \geq 2$, the displacement mappings $I - P_i$, $i = 1, \dots, m$ are constant on $\mathcal{A}(\alpha, p)$.*

Proof: From Theorem 4.3(ii) and Lemma 4.4 if any of the current cases occurs we have that $d_{S_i}^p$ is differentiable on $\mathcal{A}(\alpha, p)$, for each $i = 1, \dots, m$. Hence, by

Corollary 3.11(ii),

$$\nabla d_{S_i}^p(x) = p \|x - P_i(x)\|^{p-2} (x - P_i(x)) \quad (18)$$

is constant on $\mathcal{A}(\alpha, p)$, $i = 1, \dots, m$ (again, $P_i := P_{S_i}$, $i = 1, 2, \dots, m$). So,

$$\|\nabla d_{S_i}^p(x)\| = p \|x - P_i(x)\|^{p-1}$$

is constant on $\mathcal{A}(\alpha, p)$, too, and hence so is $\|x - P_i(x)\|$. Therefore, from (18), we conclude that $I - P_i$ is constant on $\mathcal{A}(\alpha, p)$, $i = 1, \dots, m$. ■

As a consequence of the previous theorem, taking Proposition 1.1 into account, we derive the following corollary. Roughly speaking, under the current assumptions, the corollary says that the smallest translations of the sets S_i that achieve a nonempty intersection are unique. To obtain such translations, one just has to pick $x \in \mathcal{A}(\alpha, p)$ and set $u_i = x - P_i(x)$; then, the required translated sets are $S_i + u_i$. The constancy of $I - P_i$ on $\mathcal{A}(\alpha, p)$ shows that this procedure does not depend on the choice of x .

Corollary 4.6: *If $p > 1$ and $m = 2$, or $p \geq 2$, problem (3) has a unique optimal solution, provided that problem (2) is solvable.*

To summarize our results on the solution set $\mathcal{A}(\alpha_1, \alpha_2, p)$, let us mention that they show essential differences between the cases $p = 1$ and $p > 1$. Indeed, Propositions 4.1 and 4.2 show that computing $\mathcal{A}(\alpha_1, \alpha_2, 1)$ essentially reduces to finding projections onto the sets S_i , whereas for $p > 1$ the situation is substantially more complicated, since computing $\mathcal{A}(\alpha_1, \alpha_2, p)$ involves finding fixed points of a certain convex combination of the projection mappings P_1 and P_2 .

5. Distance to feasibility

This section is focussed on the distance to feasibility for convex inequality systems in $\mathbb{R}^n$ under right-hand side perturbations. In this framework, lower and upper estimates for such a distance are provided in terms of some elements whose existence is guaranteed from Corollary 3.11. Both estimates coincide when confined to linear systems.

Let us consider the parameterized system,

$$\sigma(b) := [g_i(x) \leq b_i, i = 1, \dots, m],$$

where $x \in \mathbb{R}^n$, $(b_i)_{i=1, \dots, m} \equiv b \in \mathbb{R}^m$, and $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function, $i = 1, 2, \dots, m$. To start with, the space of variables, $\mathbb{R}^n$, is endowed with an arbitrary norm, $\|\cdot\|$, with dual norm $\|\cdot\|_*$ and the associated distances denoted by d and d_* , respectively. From Corollary 5.6 on we consider $\mathbb{R}^n$ equipped with the

Euclidean norm, $\|\cdot\|_2$. The space of parameters, $\mathbb{R}^m$, is endowed with any p -norm, $\|\cdot\|_p$, provided that $p \geq 2$, and the associated distance is denoted by d_p . We denote by Θ_c the set of consistent parameters; i.e.

$$\Theta_c := \{b \in \mathbb{R}^m \mid \sigma(b) \text{ is consistent}\}.$$

Throughout this section we consider a fixed $\bar{b} \in \mathbb{R}^m \setminus \Theta_c$ and our aim is to estimate

$$d_p(\bar{b}, \Theta_c) = \inf \left\{ \|\bar{b} - b\|_p : b \in \Theta_c \right\},$$

called the distance from $\bar{b}$ to feasibility.

Our first result in this section provides a formula to compute the distance to feasibility by solving a simple unconstrained optimization problem in the space of variables.

Proposition 5.1: *Let $\bar{b} \in \mathbb{R}^m \setminus \Theta_c$, then*

$$d_p(\bar{b}, \Theta_c)^p = \inf_{x \in \mathbb{R}^n} \sum_{i=1}^m [g_i(x) - \bar{b}_i]_+^p.$$

Proof: To establish the inequality ' $\leq$ ', take any $x \in \mathbb{R}^n$ and define

$$b_i := \bar{b}_i + [g_i(x) - \bar{b}_i]_+, \quad i = 1, \dots, m.$$

One can easily check that $b = (b_i)_{i=1, \dots, m} \in \Theta_c$ and, hence,

$$d_p(\bar{b}, \Theta_c)^p \leq d_p(\bar{b}, b)^p = \sum_{i=1}^m [g_i(x) - \bar{b}_i]_+^p.$$

Since $x \in \mathbb{R}^n$ has been arbitrarily chosen, then

$$d_p(\bar{b}, \Theta_c)^p \leq \inf_{x \in \mathbb{R}^n} \sum_{i=1}^m [g_i(x) - \bar{b}_i]_+^p.$$

Let us prove the converse inequality. Take any $b \in \Theta_c$, i.e. such that, for some $\bar{x} \in \mathbb{R}^n$, $g_i(\bar{x}) \leq b_i$, $i = 1, \dots, m$; then, $g_i(\bar{x}) - \bar{b}_i \leq b_i - \bar{b}_i$, $i = 1, \dots, m$, and so

$$[g_i(\bar{x}) - \bar{b}_i]_+ \leq [b_i - \bar{b}_i]_+ \leq |b_i - \bar{b}_i|, \quad i = 1, \dots, m.$$

Hence

$$\inf_{x \in \mathbb{R}^n} \sum_{i=1}^m [g_i(x) - \bar{b}_i]_+^p \leq \sum_{i=1}^m [g_i(\bar{x}) - \bar{b}_i]_+^p \leq \|\bar{b} - b\|_p^p.$$

Since the previous inequality is held for all $b \in \Theta_c$, then $\inf_{x \in \mathbb{R}^n} \sum_{i=1}^m [g_i(x) - \bar{b}_i]_+^p \leq d_p(\bar{b}, \Theta_c)^p$. ■

The well-known Ascoli formula establishes that the distance from a point $x \in \mathbb{R}^n$ to a half-space $H := \{x \in \mathbb{R}^n \mid \langle a, x \rangle \leq b\}$, with $0_n \neq a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, is given by

$$d_H(x) = \frac{[\langle a, x \rangle - b]_+}{\|a\|_*}. \quad (19)$$

The following result is focussed on the extension of (19) to the convex case, where a convex inequality of the form ' $g(x) \leq b$ ' is considered. In this context, the distance from $x \in \mathbb{R}^n$ to the nonempty closed convex set $S := \{x \in \mathbb{R}^n \mid g(x) \leq b\}$, denoted by $d_S(x)$, is lower and upper bounded by quotients involving the residual $[g(x) - b]_+$ and the minimum norm of some subgradients of g . Regarding these quotients, we use the convention $\frac{0}{0} := 0$.

Proposition 5.2: *Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function and $b \in \mathbb{R}$ be such that the corresponding sublevel set, S , is nonempty. Then we have:*

(i) *For any $x \in \mathbb{R}^n$,*

$$d_S(x) \geq \frac{[g(x) - b]_+}{d_*(0_n, \partial g(x))};$$

(ii) *Assume that there exists $\hat{x} \in \mathbb{R}^n$ (called a Slater point) such that $g(\hat{x}) < b$. Then, for any $x \in \mathbb{R}^n$,*

$$d_S(x) \leq \frac{[g(x) - b]_+}{d_*(0_n, \partial g(P_S(x)))},$$

where $P_S(x)$ is the metric projection set of x on S with respect to the norm $\|\cdot\|$.

Proof: (i) Inequality $g(x) \leq b$ turns out to be equivalent (same solution set, S) to its standard linearization by means of the Fenchel conjugate, g^* , (see, e.g. [27, Formula (3)]), namely system

$$\{\langle u, x \rangle \leq g^*(u) + b, \ u \in \partial g(\mathbb{R}^n)\}.$$

The distance $d_S(x)$ may be computed by means of [28, Lemma 1], yielding (with the convention $\frac{0}{0} := 0$)

$$d_S(x) = \sup \left\{ \frac{[\langle v, x \rangle - \alpha]_+}{\|v\|_*} \mid (v, \alpha) \in \text{conv} \left\{ (u, g^*(u) + b), \ u \in \partial g(\mathbb{R}^n) \right\} \right\}$$

$$\begin{aligned}
&\geq \sup \left\{ \left| \frac{[\langle u, x \rangle - (g^*(u) + b)]_+}{\|u\|_*} \right| \mid u \in \partial g(\mathbb{R}^n) \right\} \\
&\geq \sup \left\{ \left| \frac{[g(x) - b]_+}{\|u\|_*} \right| \mid u \in \partial g(x) \right\} \\
&= \frac{[g(x) - b]_+}{\inf \{ \|u\|_* \mid u \in \partial g(x) \}} = \frac{[g(x) - b]_+}{d_*(0_n, \partial g(x))},
\end{aligned}$$

where in the third step we have appealed to the fact that

$$g(x) = g^{**}(x) = \langle u, x \rangle - g^*(u) \Leftrightarrow u \in \partial g(x).$$

- (ii) It follows from [29, Lemma 2(ii)]. Observe that for $x \in S$ we apply the convention $\frac{0}{0} := 0$, whereas for $x \notin S$ the existence of a Slater point entails that points of $P_S(x)$ are not minimizers of g (since $P_S(x) \subset g^{-1}(b)$), and then $d_*(0_n, \partial g(P_S(x))) > 0$. $\blacksquare$

Remark 5.3: If $x \in \mathbb{R}^n \setminus S$, item (ii) in the previous lemma entails $d_S(x) \leq [g(x) - b]_+ / \delta(b)$, with δ being the function defined by

$$b \mapsto \delta(b) := d_*(0_n, \partial g(g^{-1}(b)));$$

here $\partial g(g^{-1}(b)) = \bigcup_{g(y)=b} \partial g(y)$. For instance, if $g(x_1, x_2) := e^{x_1} + e^{x_2}$, with the Euclidean norm in $\mathbb{R}^2$, then $\delta(b) = b/\sqrt{2}$ for $b > 0$.

The bounds provided by Proposition 5.2 may be difficult to compute, as they involve subgradients. By appealing to this proposition and bounding (from above and below) the norm of such subgradients we derive the next corollary.

Corollary 5.4: Let $\bar{b} \in \mathbb{R}^m \setminus \Theta_c$ and assume that $S_i := \{x \in \mathbb{R}^n \mid g_i(x) \leq \bar{b}_i\} \neq \emptyset$, $i = 1, \dots, m$. Then, the following statements hold:

- (i) Let $\emptyset \neq C \subset \mathbb{R}^n$ be a closed convex set such that, for each $i \in \{1, \dots, m\}$, there exists an upper bound $u_i \geq d_*(0_n, \partial g_i(x))$ for all $x \in C$. Then,

$$d_p(\bar{b}, \Theta_c)^p \leq \inf_{x \in C} \sum_{i=1}^m (u_i)^p d_{S_i}^p(x) = \inf_{x \in \mathbb{R}^n} \left[\sum_{i=1}^m (u_i)^p d_{S_i}^p(x) + I_C(x) \right], \quad (20)$$

where I_C is the indicator function of C ; i.e. $I_C(x) = 0$ if $x \in C$ and $I_C(x) = +\infty$ if $x \in \mathbb{R}^n \setminus C$.

- (ii) Assume that for each $i \in \{1, \dots, m\}$ there exists a lower bound $0 < l_i \leq d_*(0_n, \partial g_i(P_{S_i}(x)))$ for all $x \in \mathbb{R}^n \setminus S_i$. Then,

$$d_p(\bar{b}, \Theta_c)^p \geq \inf_{x \in \mathbb{R}^n} \sum_{i=1}^m (l_i)^p d_{S_i}^p(x). \quad (21)$$

Proof: (i) comes straightforwardly from Propositions 5.1 and 5.2 (i), taking into account the obvious fact that $\inf_{x \in \mathbb{R}^n} \sum_{i=1}^m [g_i(x) - \bar{b}_i]_+^p \leq \inf_{x \in C} \sum_{i=1}^m [g_i(x) - \bar{b}_i]_+^p$.
(ii) follows immediately from Propositions 5.1 and 5.2 (ii). ■

The following simple example is intended to illustrate the computation of the upper and lower bounds for the distance to feasibility provided in the previous corollary, as well as the role played by set C .

Example 5.5: Let us consider the convex system, in $\mathbb{R}^2$ with the Euclidean norm ($p = 2$),

$$[(x_1 + 1)^2 - x_2 \leq -1, (x_1 - 1)^2 + x_2 \leq -1];$$

i.e. with $\bar{b} = (-1, -1)$, $g_1(x) = (x_1 + 1)^2 - x_2$ and $g_2(x) = (x_1 - 1)^2 + x_2$. The reader can easily check that the infimum in Proposition 5.1 is attained at $(0, 0)$, hence the distance to feasibility equals $2\sqrt{2}$. With this information, we can take $C = \{(0, 0)\}$ in order to get $u_i = \sqrt{5}$, $i = 1, 2$, whereas $l_i = 1$, $i = 1, 2$, independently of C . Nevertheless, a more natural choice for C without this a priori information would be $[-1, 1] \times \mathbb{R}$, in which case we can take $u_i = \sqrt{17}$, $i = 1, 2$; in any case, in this example, the first coordinate must be bounded on C in order to get finite u_i 's.

On the one hand, taking $l_i = 1$, $i = 1, 2$, condition (ii) in the previous corollary yields

$$d_2(\bar{b}, \Theta_c)^2 \geq \inf_{x \in \mathbb{R}^2} [d_{S_1}^2(x) + d_{S_2}^2(x)] = \frac{1}{2} d(S_1, S_2)^2,$$

where we have applied Theorem 4.3 (iv) with $\alpha_1 = \alpha_2 = \frac{1}{2}$. On the other hand, if we take $C := [-1, 1] \times \mathbb{R}$ and $u_i = \sqrt{17}$, $i = 1, 2$, then Corollary 5.4 (i) yields

$$\frac{1}{2} d_2(\bar{b}, \Theta_c)^2 \leq 17 \inf_{x \in C} [d_{S_1}^2(x) + d_{S_2}^2(x)] = \frac{17}{2} \inf_{x \in \mathbb{R}^2} [d_{S_1}^2(x) + d_{S_2}^2(x)],$$

where the last equality is obtained by taking into account that an easy argument based on convexity of the objective function and the symmetry with respect to the origin shows that the infimum over $\mathbb{R}^2$ is attained at the origin. Consequently, we obtain the bounds

$$d(S_1, S_2)^2 \leq d_2(\bar{b}, \Theta_c)^2 \leq 17 d(S_1, S_2)^2.$$

Finally, we can obtain specific lower and upper estimates for the aimed distance to feasibility by using the estimation $d(S_1, S_2)^2 \simeq 6.71$, which is obtained by a standard computation of nonlinear programming. In this case the upper bound is quite rough, but in other examples we could get even the exact distance to

feasibility if we have $l_i = u_i$ as for instance $[x_1^2 - x_2 \leq -1, x_1^2 + x_2 \leq -1]$ and taking $C = \{0\} \times \mathbb{R}$.

Provided that $C, u = (u_i)_{i=1,\dots,m}, l = (l_i)_{i=1,\dots,m}$ satisfy the conditions of the previous corollary, we consider the argmin sets coming from (20) and (21):

$$\mathcal{A}(C, u) := \arg \min \left[\sum_{i=1}^m (u_i)^p d_{S_i}^p(x) + I_C(x) \right],$$

$$\mathcal{A}(l) := \arg \min \sum_{i=1}^m (l_i)^p d_{S_i}^p(x).$$

Then we can state another corollary of Proposition 5.2, appealing also to Corollary 3.11. Indeed, it brings to light the advantages of appealing to $\mathcal{A}(C, u)$ and $\mathcal{A}(l)$, instead of working directly with $\arg \min \sum_{i=1}^m [g_i(x) - \bar{b}_i]_+^p$. The key point is that, in the current case in which $p \geq 2$, each function $d_{S_i}^p$ is differentiable on $\mathbb{R}^n$ (see Lemma 4.4), which allows to appeal to Corollary 3.11, while this is not the case of $[g_i(\cdot) - \bar{b}_i]_+^p$.

The following corollary uses Theorem 4.5 to refine the bounds provided by Corollary 5.4. Hereafter in this section we consider that $\mathbb{R}^n$ is endowed with the Euclidean norm $\|\cdot\|_2$ and $P_S(x)$ will denote the unique projection point of x on a closed convex set S .

Corollary 5.6: *Keeping the previous notation, assume that $\mathcal{A}(C, u)$ and $\mathcal{A}(l)$ are nonempty. Then, we have that:*

- (i) d_{S_i} is constant on both $\mathcal{A}(C, u)$ and $\mathcal{A}(l)$, for each $i = 1, \dots, m$;
- (ii) For each $i = 1, \dots, m$, let us denote by d_i^+ and d_i^- the constant values of $u_i d_{S_i}(\cdot)$ and $l_i d_{S_i}(\cdot)$ on $\mathcal{A}(C, u)$ and $\mathcal{A}(l)$, respectively, and let $d^+ = (d_i^+)_{i=1,\dots,m}$ and $d^- = (d_i^-)_{i=1,\dots,m}$. Then,

$$d_p(\bar{b}, \Theta_c) \leq \|d^+\|_p.$$

If, in addition, for each $i = 1, \dots, m$ there exists $\hat{x}_i \in \mathbb{R}^n$ such that $g_i(\hat{x}_i) < \bar{b}_i$, then

$$d_p(\bar{b}, \Theta_c) \geq \|d^-\|_p.$$

Proof: (i) Regarding $\mathcal{A}(l)$, the statement coincides with the one of Theorem 4.5 (in the case when $p \geq 2$) just replacing each α_i with $(l_i)^p$. With respect to $\mathcal{A}(C, u)$, the statement comes from an analogous argument to the one of that theorem, just by adding the nondifferentiable mapping I_C . For completeness, we include here a sketch of the proof. Observe that all functions $x \mapsto (u_i)^p d_{S_i}^p(x)$ are convex and differentiable on $\mathbb{R}^n$, and $x \mapsto I_C(x)$ is a

proper lower semicontinuous convex function from $\mathbb{R}^n$ to $] -\infty, +\infty]$. Hence, the regularity condition (10) is satisfied, yielding

$$\mathcal{A}(C, u) = \left\{ x \in \mathbb{R}^n \mid 0_n \in \sum_{i=1}^m (u_i)^p \nabla d_{S_i}^p(x) + \partial I_C(x) \right\} (\neq \emptyset).$$

From Corollary 3.11, for each $i = 1, \dots, m$, we have that $\nabla d_{S_i}^p$ is constant on $\mathcal{A}(C, u)$, hence d_{S_i} is also constant on $\mathcal{A}(C, u)$ since taking norms we have

$$\begin{aligned} \|\nabla d_{S_i}^p(x)\| &= \|p d_{S_i}^{p-2}(x)(x - P_i(x))\| \\ &= p d_{S_i}^{p-1}(x), \quad \text{for each } x \in \mathcal{A}(C, u), \end{aligned}$$

where $P_i(x)$ denotes the projection of x on S_i (recall again Lemma 4.4).

(ii) follows immediately from (i) and Corollary 5.4. ■

5.1. Linear systems

This subsection is devoted to the linear case, i.e. where $g_i(x) = \langle a_i, x \rangle$, for some $a_i \in \mathbb{R}^n$, $i = 1, \dots, m$. In this particular case, obviously $\partial g_i(x) = \{a_i\}$ for all $x \in \mathbb{R}^n$, $i = 1, \dots, m$. Let us consider $\bar{b}$ such that

$$\sigma(\bar{b}) = [\langle a_i, x \rangle \leq \bar{b}_i, i = 1, \dots, m]$$

is inconsistent and for each i there exists $\hat{x}_i \in \mathbb{R}^n$ such that $\langle a_i, \hat{x}_i \rangle < \bar{b}_i$ (observe that it is always held when $a_i \neq 0_n$ or $\bar{b}_i > 0$). According to the notation of Corollary 5.4, we can choose:

$$C = \mathbb{R}^n, l_i = u_i = \|a_i\|_*, i = 1, \dots, m.$$

Hence $\mathcal{A}(C, u) = \mathcal{A}(l)$, and $d_i^+ = d_i^-$ for all i . Let us denote by $\mathcal{A} := \mathcal{A}(C, u)$ and $\bar{d} := (d_i^+)_{i=1, \dots, m}$.

The following corollary follows straightforwardly from Corollary 5.6. It gives a simple formula for the distance to feasibility in the linear case.

Corollary 5.7: *Under the current assumptions, we have*

$$d_p(\bar{b}, \Theta_c) = \|\bar{d}\|_p,$$

where $d_i^+ = \|a_i\|_* d_{S_i}(x) = [\langle a_i, x \rangle - \bar{b}_i]_+$, for all $x \in \mathcal{A}$. Moreover $\sigma(\bar{b} + \bar{d})$ is a consistent system nearest to $\sigma(\bar{b})$.

The next result is devoted to provide an operative expression for determining $\bar{d}$ with the Euclidean norm in both the space of variables and the space of parameters. For simplicity all norms are denoted by $\|\cdot\|$, A represents the matrix

whose rows are a'_i , $i = 1, \dots, m$, A' denotes its transpose and, for any $y \in \mathbb{R}^m$, $[y]_+$ denotes positive part coordinate by coordinate; i.e.

$$[y]_+ := ([y_i]_+)_{i=1, \dots, m}.$$

Theorem 5.8: *The following conditions are equivalent:*

- (i) $(x^0, h^0) \in \mathcal{A} \times \{\bar{d}\}$;
- (ii) (x^0, h^0) is a solution of the system, in the variable (x, h) ,

$$\begin{cases} [Ax - \bar{b}]_+ = h, \\ A'h = 0_n. \end{cases} \quad (22)$$

- (iii) (x^0, h^0) is an optimal solution of the quadratic problem, in the variable (x, h) ,

$$\begin{aligned} \min \quad & \langle h, h \rangle \\ \text{s.t.} \quad & Ax \leq \bar{b} + h, \\ & h \geq 0_m. \end{aligned} \quad (23)$$

Proof: Let us see (i) $\Rightarrow$ (ii). Let $(x^0, h^0) \in \mathcal{A} \times \{\bar{d}\}$, i.e. $x^0 \in \mathcal{A}$ and $h^0 = \bar{d}$. By Corollary 5.7, $h_i^0 (= d_i^+) = [a'_i x^0 - \bar{b}_i]_+$, for all i . Moreover, the optimality condition

$$x^0 \in \mathcal{A} = \arg \min \sum_{i=1}^m \|a_i\|^2 d_{S_i}^2(x)$$

is equivalent to

$$0_n = \sum_{i=1}^m \|a_i\|^2 \nabla d_{S_i}^2(x^0) = 2 \sum_{i=1}^m \|a_i\| d_{S_i}(x^0) a_i = 2 \sum_{i=1}^m [a'_i x^0 - \bar{b}_i]_+ a_i; \quad (24)$$

in other words

$$0_n = \sum_{i=1}^m h_i^0 a_i = A'h^0.$$

So, (x^0, h^0) is a solution of system (22).

- (ii) $\Rightarrow$ (i) Let (x^0, h^0) be a solution of (22); i.e. $h^0 = [Ax^0 - \bar{b}]_+$ and

$$0_n = \sum_{i=1}^m h_i^0 a_i = A'h^0.$$

Then, by repeating the previous argument of (24), we have

$$0_n = \sum_{i=1}^m \|a_i\|^2 \nabla d_{S_i}^2(x^0),$$

which means that $x^0 \in \mathcal{A}$. Then, appealing again to Corollary 5.7, we deduce $h^0 = \bar{d}$.

Now, let us prove (ii) $\Leftrightarrow$ (iii). By the Karush-Kuhn-Tucker (KKT, in brief) conditions, (x^0, h^0) is an optimal solution of (23) if and only if there exist $\lambda, \mu \in \mathbb{R}_+^m$ such that

$$\begin{cases} -\begin{pmatrix} 0_n \\ 2h^0 \end{pmatrix} = \begin{pmatrix} A' \\ -I_m \end{pmatrix} \lambda + \begin{pmatrix} 0_{n \times m} \\ -I_m \end{pmatrix} \mu, \\ \begin{pmatrix} Ax^0 - \bar{b} - h^0 \end{pmatrix} \lambda = 0, & -(h^0)' \mu = 0, \\ Ax^0 - \bar{b} - h^0 \leq 0_m, & h^0 \geq 0_m. \end{cases} \quad (25)$$

So, $A'\lambda = 0_n$, and $h^0 = \frac{\lambda + \mu}{2}$. Moreover, $h_i^0 \mu_i = 0$ for all i . Let us see that $\mu = 0_m$. If $h_i^0 = 0$, then $\lambda_i + \mu_i = 0$, which entails $\lambda_i = \mu_i = 0$, while, if $h_i^0 > 0$, then $\mu_i = 0$. Therefore

$$h^0 = \frac{\lambda}{2} \quad (26)$$

and, so,

$$A'h^0 = 0_n.$$

Let us see that $[Ax^0 - \bar{b}]_+ = h^0$. Observe that $(a'_i x^0 - \bar{b}_i - h_i^0) \lambda_i = 0$ for all i . If $a'_i x^0 - \bar{b}_i < 0$, then $a'_i x^0 - \bar{b}_i - h_i^0 < 0$, thus we have $\lambda_i = 0$ and

$$h_i^0 = \frac{\lambda_i}{2} = 0.$$

If $a'_i x^0 - \bar{b}_i > 0$, then $h_i^0 > 0$ and $\lambda_i > 0$, yielding $a'_i x^0 - \bar{b}_i - h_i^0 = 0$. Finally, if $a'_i x^0 - \bar{b}_i = 0$, then $h_i^0 \lambda_i = 0$, and from (26) we have $h_i^0 = 0$. So,

$$\left[a'_i x^0 - \bar{b}_i \right]_+ = h_i^0, \quad \text{for all } i,$$

and consequently (x^0, h^0) is a solution of (22).

Reciprocally, if (x^0, h^0) is a solution of (22) and we consider

$$\lambda = 2h^0 \quad \text{and} \quad \mu = 0_m,$$

it can be easily seen that x^0, h^0, λ and μ satisfy the KKT conditions (25), and then (x^0, h^0) is an optimal solution for problem (23). ■

6. Conclusion

We have found new properties of paramonotone operators. In view of their applications to convex feasibility problems and convex inequality systems, we have studied operators that are both paramonotone and bimonotone, proving that they are constant on their domains. As an application to convex feasibility problems,

this result implies the uniqueness of the smallest translation of mutually disjoint convex sets such that the corresponding translated sets have a nonempty intersection. We have also obtained lower and upper bounds on the smallest right-hand side perturbations of inconsistent convex inequality systems that make them consistent, which turn out to be the exact value in the case of linear systems.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research has been partially supported by Grants PGC2018-097960-B-C21 and PID2022-136399NB-C22 from MICINN, Spain, and ERDF, ‘A way to make Europe’, European Union, and Grant PROMETEO/2021/063 from Generalitat Valenciana, Spain.

ORCID

J. Camacho  <http://orcid.org/0000-0002-2527-7351>

M.J. Cánovas  <http://orcid.org/0000-0002-1320-485X>

J.E. Martínez-Legaz  <http://orcid.org/0000-0002-6845-6202>

J. Parra  <http://orcid.org/0000-0001-8864-975X>

References

- [1] Bauschke HH, Combettes PL. Convex analysis and monotone operator theory in Hilbert spaces. New York: Springer; 2011.
- [2] Borwein JM, Vanderwerff JD. Convex functions. Cambridge, MA: Cambridge University Press; 2010.
- [3] Burachik RS, Iusem AN. Set-valued mappings and enlargement of monotone operators. New York: Springer; 2009.
- [4] Phelps RR. Convex functions, monotone operators and differentiability. 2nd ed. Berlin: Springer-Verlag; 1993.
- [5] Rockafellar RT, Wets RJ-B. Variational analysis. 3rd ed. Berlin: Springer-Verlag; 2009.
- [6] Simons S. From Hahn-Banach to Monotonicity. Berlin: Springer-Verlag; 2008.
- [7] Zălinescu C. Convex analysis in general vector spaces. Singapore: World Scientific Publishing; 2002.
- [8] Censor Y, Iusem AN, Zenios SA. An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. Math Program (Ser A). 1998;81:373–400. doi: [10.1007/BF01580089](https://doi.org/10.1007/BF01580089)
- [9] Bruck RE. An iterative solution of a variational inequality for certain monotone operators in Hilbert space. Bull AMS. 1975;81:890–892. doi: [10.1090/bull/1975-81-05](https://doi.org/10.1090/bull/1975-81-05)
- [10] Iusem AN. On some properties of paramonotone operators. J Convex Anal. 1998;5:269–278.
- [11] Bausche HH, Wang X, Yao L. On Borwein–Wiersma decompositions of monotone linear relations. SIAM J Optim. 2010;20:2636–2652. doi: [10.1137/09078016X](https://doi.org/10.1137/09078016X)
- [12] Bauschke HH, Combettes PL, Luke DR. Finding best approximation pairs relative to closed convex sets in Hilbert spaces. J Approx Theory. 2004;127:178–192. doi: [10.1016/j.jat.2004.02.006](https://doi.org/10.1016/j.jat.2004.02.006)

- [13] Ioffe AD. On stability of solutions to systems of convex inequalities. *J Convex Anal.* **2012**;19:1017–1032.
- [14] Cánovas MJ, López MA, Parra J, et al. Distance to ill-posedness for linear inequality systems under block perturbations: convex and infinite-dimensional cases. *Optimization.* **2011**;60:925–946. doi: [10.1080/02331934.2011.606624](https://doi.org/10.1080/02331934.2011.606624)
- [15] Peña J. Conditioning of convex programs from a primal-dual perspective. *Math Oper Res.* **2001**;26:206–220. doi: [10.1287/moor.26.2.206.10551](https://doi.org/10.1287/moor.26.2.206.10551)
- [16] Renegar J. Incorporating condition measures into the complexity theory of linear programming. *SIAM J Optim.* **1995**;5:506–524. doi: [10.1137/0805026](https://doi.org/10.1137/0805026)
- [17] Freund RM, Vera JR. Some characterizations and properties of the “ distance to ill-posedness” and the condition measure of a conic linear system. *Math Program.* **1999**;86:225–260. doi: [10.1007/s10107990063a](https://doi.org/10.1007/s10107990063a)
- [18] Freund RM, Vera JR. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM J Optim.* **1999**;10:155–176. doi: [10.1137/S105262349732829X](https://doi.org/10.1137/S105262349732829X)
- [19] Lucchetti R. Convexity and Well-Posed problems. New York: Springer-Verlag; **2006**.
- [20] Mordukhovich BS, Nam NM. Convex analysis and beyond. Volume I. Basic theory. Cham: Springer-Verlag; **2022**. (Springer series in operations research and financial engineering).
- [21] Rockafellar RT. Extension of Fenchel’ duality theorem for convex functions. *Duke Math J.* **1966**;33:81–89. doi: [10.1215/S0012-7094-66-03312-6](https://doi.org/10.1215/S0012-7094-66-03312-6)
- [22] Martínez-Legaz JE, Svaiter BF. Monotone operators representable by l.s.c. convex functions. *Set-Valued Anal.* **2005**;13:21–46. doi: [10.1007/s11228-004-4170-4](https://doi.org/10.1007/s11228-004-4170-4)
- [23] Simons S. Positive sets and monotone sets. *J Convex Anal.* **2007**;14:297–317.
- [24] Holmes RB. Smoothness of certain metric projections on Hilbert space. *Trans Am Math Soc.* **1973**;184:87–100. doi: [10.1090/tran/1973-184-00](https://doi.org/10.1090/tran/1973-184-00)
- [25] Bauschke HH, Wang X, Yao L. Rectangularity and paramonotonicity of maximally monotone operators. *Optimization.* **2014**;63:487–504. doi: [10.1080/02331934.2012.707653](https://doi.org/10.1080/02331934.2012.707653)
- [26] Burke JV, Ferris MC. Characterization of solution sets of convex programs. *Oper Res Lett.* **1991**;10:57–60. doi: [10.1016/0167-6377\(91\)90087-6](https://doi.org/10.1016/0167-6377(91)90087-6)
- [27] Beer G, Cánovas MJ, López MA, et al. Lipschitz modulus of linear and convex inequality systems with the Hausdorff metric. *Math Program.* **2021**;189:75–98. doi: [10.1007/s10107-020-01543-9](https://doi.org/10.1007/s10107-020-01543-9)
- [28] Cánovas MJ, Gómez-Senent FJ, Parra J. Regularity modulus of arbitrarily perturbed linear inequality systems. *J Math Anal Appl.* **2008**;343:315–327. doi: [10.1016/j.jmaa.2008.01.021](https://doi.org/10.1016/j.jmaa.2008.01.021)
- [29] Cánovas MJ, Gómez-Senent FJ, Parra J. Linear regularity, equirregularity, and intersection mappings for convex semi-infinite inequality systems. *Math Program Ser B.* **2010**;123:33–60. doi: [10.1007/s10107-009-0312-7](https://doi.org/10.1007/s10107-009-0312-7)