Yokohama Publishers

ISSN 2189-3764   *ONLINE JOURNAL*

© Copyright 2020

STABILITY OF THE OPTIMAL VALUES UNDER SMALL PERTURBATIONS OF THE CONSTRAINT SET

DANIEL REEM, SIMEON REICH, AND ALVARO DE PIERRO

Abstract. This paper discusses a general and useful stability principle which, roughly speaking, says that given a uniformly continuous function defined on an arbitrary metric space, if the function is bounded on the constraint set and we slightly change this set, then its optimal (extreme) values on this set vary slightly, and, moreover, they are actually uniformly continuous as a function of the constraint set. The principle holds in a much more general setting than a metric space, since the distance function may be asymmetric, may attain negative and even infinite values, and so on. This stability principle leads to applications in parametric optimization, mixed linear-nonlinear programming and analysis of Lipschitz continuity, as well as to a general scheme for tackling a wide class of non-convex and non-smooth optimization problems. We also discuss the issue of stability when the objective function is merely continuous. As a byproduct of our analysis we obtain a significant generalization of the concept of a generalized inverse of a linear operator and a very general variant of the so-called "Hoffman's Lemma".

1. Introduction

1.1. **Background:** The issue of stability in optimization problems has both theoretical and practical importance. More precisely, suppose that we are given an optimization problem consisting of a space X, an objective function (a target function) $f : X \to [-\infty, \infty]$, a constraint set $\emptyset \neq A \subseteq X$, and various parameters which influence the problem (for example, the parameters may influence the constraint set A, namely $A = A(t)$ for some fixed parameter t in a parameter space; another example: parameters which define f, namely there is some function g of two variables such that $f(x) = g(x,t)$, $x \in A$, where t is a parameter in a parameter space; in both cases one minimizes with respect to x). A natural question is what happens to the optimal values of the objective function f, as well as to its sets of minimizers and maximizers, when we slightly perturb some of the elements involved in the formulation of the problem. If the optimal values (and perhaps also the sets of minimizers and maximizers) change slightly as a result of the change in the elements of the problem, then the problem exhibits a certain kind of stability.

Of course, the type of stability depends on the way in which we measure all the pertinent changes.

Stability is a desired property which ensures that if one is able to control various types of imprecision which are inherent in many optimization problems (such as noise, inexact measurements, representation errors due to real parameters/expressions which are approximated by a finite decimal representation of them, and so on), then the resulting optimal values will not change a lot.

There is a very long chain of research works that contain results related to the issue of stability in optimization problems, as can be seen by looking at the following (far from being exhaustive) list of references and the references therein: [2,8,12–14,21–23,26,28–30,39,45,46,48,49,55,56,58,71,73,74,78,79,81,87,89,91,95]. These works consider a variety of optimization problems, in various settings (often Euclidean spaces, but sometimes also in infinite-dimensional Banach spaces or other spaces). They establish certain properties related to either the optimal values of the objective function or its sets of minimizers and maximizers, under suitable assumptions on the objective function and the structure of the set of constraints, such as convexity, linear or piecewise linear structure, quadratic structure, compactness, and so on. Frequently the established properties are semicontinuity (lower or upper) and closedness; occasionally (usually in finite-dimensional Euclidean spaces) stronger properties are established, such as continuity, Lipschitz continuity, convexity, and differentiability. The aforementioned results are interpreted as being stability results.

In the above-mentioned cases the perturbations occur in the space of parameters which define either the set of constraints or the objective function. This is one of the reasons why the corresponding research field to which these works belong is traditionally called "parametric optimization", although they can also be regarded as belonging to "variational analysis" [54,75]. The parameters themselves are frequently real numbers or vectors in a linear space endowed with a norm or a topology, but sometimes they belong to a different entity. For instance, in [4,76] the parameter is a probability measure; in [72] the parameter is a pair consisting of a point in a topological space and a probability measure; in [84,85] the parameter is a sample belonging to a probability space and one discusses various probabilistic types of convergence, as well as convergence in the sense of inner or outer limits (limits which are related to the Kuratowski-Painlevé notion of convergence). For related but somewhat different types of stability, see, for example, [5,7,25,60,61,67,82,92,93].

1.2. **Our stability principle:** In this paper we discuss a different type of stability, which, in our opinion, is not less natural than other types of stability discussed in the literature. More precisely, after some preliminaries (Section 2) we present in Section 3 our stability principle which, roughly speaking, says the following: given a uniformly continuous function defined on an arbitrary metric space, if the function is bounded on the constraint set and we slightly change this set, where the change is measured with respect to the Hausdorff distance, then its optimal (extreme) values on this set vary slightly.

One can think of the above-mentioned stability principle as a continuous dependence result, where the resulting continuous functions are the infimum and the

supremum of the objective function over the constraint set. This is how we actually formulate the stability principle (Theorem 3.1). The advantage of this point of view is that it enables one to establish other properties of the infimum and supremum functions, such as their uniform continuity and, sometimes, their Lipschitz continuity. Such properties may not have a very intuitive interpretation. Theorem 3.1, which is new to the best of our knowledge (but see Remark 3.5(e) below for some predecessors and other related results), has a potential to extend, and perhaps to re-prove, some of the stability results mentioned earlier (for an illustration, see Section 5). It seems to be especially promising for finite-dimensional normed spaces, manifolds, and the like, since, as is well known, a continuous function defined on a closed and bounded (hence compact) subset of such spaces is uniformly continuous on this subset. Another relevant example is continuous linear functionals on arbitrary normed spaces (since, as is well known, they are automatically Lipschitz continuous).

Interestingly, our stability result holds in a much more general setting than metric spaces, since we assume essentially nothing on the distance function: in particular, it can be negative, infinite, asymmetric, and so on; this simple observation is, in fact, a key contribution. As a result, our "topology-free" stability principle holds a promise to be relevant to many scenarios in the literature in which the distance function is not a metric, as we illustrate below and in Section 5.

First example: the distance function is the well-known Bregman divergence/ distance [15] (see also [68] for a recent semi-survey and an extensive re-examination) or one of its generalizations [63]; second example: distortion measures, divergences, and distance functions used in information theory, data analysis, data processing, machine learning and the like [3, 9, 31, 40, 41]; third example: Finsler quasi-metric spaces [80]; fourth example: distance functions which appear in fixed point theory [35, 37, 38, 65]; fifth example: distances induced by quasi-norms [36]; sixth example: distances induced by asymmetric norms (such as asymmetric Minkowski functionals) [19, 51]; seventh example: distances induced by star bodies [18, 33, 50, 53]; eighth example: numerous distance functions which appear in many scientific and technological areas [24]; ninth example: the setting is the energy levels of an atom, namely the space is $X := \mathbb{N}$, and $d(x, y)$ denotes the amount of energy that one needs to invest in order to bring an electron from energy level $x \in X$ to energy level $y \in X$ (in this case $d(x, y) < 0$ means that one gains energy when the electron jumps from energy level x to energy level y).

The assumptions made in the formulation of Theorem 3.1 are essential, as we illustrate in Section 3 by using some counterexamples. Despite this, we are able to formulate a counterpart to Theorem 3.1, namely Theorem 3.2, in which the uniform continuity can be replaced by mere continuity, provided the optimal values are infinite.

1.3. **Additional major contributions:** Sections 4–7 of the paper present additional major contributions, among them several applications of Theorem 3.1, as well as new and seemingly unrelated concepts and results. One application (Section 4) is in parametric optimization, where we prove a general continuity result related to the so-called "optimal value function". A second application appears in Section

5, where we apply the result of Section 4 to mixed linear-nonlinear programming problems. As a byproduct of our analysis we obtain a significant generalization of the concept of a generalized inverse of a linear operator and a very general variant of the so-called "Hoffman's Lemma". A third application is presented in Section 6, where we obtain a sequence of Lipschitz constants related to certain functions. This latter application has recently been used in [69] in the context of estimating the rate of convergence of a certain first order proximal gradient method. A fourth application (Section 7) is a general scheme for tackling a wide class of non-convex and non-smooth optimization problems.

2. PRELIMINARIES

Our setting is an arbitrary nonempty set X endowed with an arbitrary function $d : X^2 \to [-\infty, \infty]$. We refer to the set X as the "space", to d as the "distance function" or the "pseudo-distance", and to (X, d) as the "pseudo-distance space". Note that d is, in general, not a metric, for example because it can attain negative and even infinite values, it may be asymmetric, and so on, and hence we use the notion "pseudo-distance space". We denote by 2^X the set of all subsets of X. Given $\psi : X \to [-\infty, \infty]$, its effective domain is the set $\mathrm{dom}(\psi) := \{x \in X : \psi(x) \in \mathbb{R}\}$. Given an arbitrary function $f : X \to \mathbb{R}$, we denote by $\mathrm{SUP}_f : 2^X \to [-\infty, \infty]$ and $\mathrm{INF}_f : 2^X \to [-\infty, \infty]$ the supremum and infimum operators, respectively, associated with f. These operators are naturally defined by $\mathrm{SUP}_f(A) := \sup\{f(a) : a \in A\}$ and $\mathrm{INF}_f(A) := \inf\{f(a) : a \in A\}$ for each $A \in 2^X$, respectively, where we use the usual convention that $\sup \emptyset := -\infty$ and $\inf \emptyset := \infty$. We have $\mathrm{dom}(\mathrm{SUP}_f) \neq \emptyset$ and $\mathrm{dom}(\mathrm{INF}_f) \neq \emptyset$ because they contain all the singletons $\{x\}$, $x \in X$, by the assumption that $f(x) \in \mathbb{R}$ for each $x \in X$.

Given a point $x \in X$ and a nonempty subset A of X, the distance between them is defined by $d(x, A) := \inf\{d(x, a) : a \in A\}$. Given $r > 0$, we denote $B[A, r] := \{x \in X : d(x, A) \leq r\}$ and refer to this set as the "closed ball of radius r around A" (this "ball" always contains A if d is a metric, but in general it may be empty). Given another $\emptyset \neq A' \subseteq X$, we denote the standard distance between A and A' by

$$(2.1) \qquad\qquad d(A, A') := \inf\{d(a, a') : a \in A, a' \in A'\}.$$

We define the asymmetric Hausdorff distance between A and A' by

$$(2.2) \qquad\qquad D_{asyH}(A, A') := \sup\{d(a, A') : a \in A\},$$

and the (symmetric) Hausdorff distance (also called the "Pompeiu-Hausdorff distance") between A and A' by

$$(2.3) \qquad\qquad D_H(A, A') := \max\{D_{asyH}(A, A'), D_{asyH}(A', A)\}.$$

It is a simple consequence of (2.1), (2.2) and (2.3) that $d(A, A')$, $D_{asyH}(A, A')$ and $D_H(A, A')$ belong to $[-\infty, \infty]$ and $d(A, A') \leq D_{asyH}(A, A') \leq D_H(A, A')$.

Roughly speaking, the Hausdorff distance quantifies the "degree of similarity" between A and A'. More precisely, if (as it happens, for example, in the real world where d is the usual Euclidean distance) there is a resolution parameter $r > 0$, namely a positive number r with the property that we cannot distinguish, by using

distance measurements, between two points x and y in X if $d(x, y) < r$, and if $D_H(A, A') < r$, then we cannot distinguish between A and A' by using distance measurements. Indeed, the inequality $D_H(A, A') < r$ implies that for each $a \in A$, there is $a' \in A'$ such that $d(a, a') < r$ (namely, no point in A can be distinguished from a point in A') and for each $a' \in A'$, there is $a \in A$ such that $d(a', a) < r$ (namely, no point in A' can be distinguished from a point in A). We note that in general (if, say, $d(x, y) = \infty$ for all $(x, y) \in X^2$), given $\delta > 0$ and $\emptyset \neq A \subseteq X$, there may not be any $\emptyset \neq A' \subseteq X$ such that $D_H(A, A') < \delta$; if, however, d is a metric, then we always have $0 = D_H(A, A) < \delta$.

In the sequel we consider various continuity notions of real functions. These notions are word for word as in the usual (metric) case, although in the general case their interpretation is less intuitive. More precisely, let $f : X \to \mathbb{R}$ and $\emptyset \neq A \subseteq X$ be given. We say that f is continuous at $x \in X$ if for every $\epsilon > 0$, there exists $\delta > 0$ such that for each $y \in X$ satisfying $d(x, y) < \delta$, we have $|f(x) - f(y)| < \epsilon$. We say that f is continuous on A if it is continuous at each $x \in A$. We say that f is uniformly continuous on A if for each $\epsilon > 0$, there exists $\delta > 0$ such that for all $(x, y) \in A^2$ satisfying $d(x, y) < \delta$, we have $|f(x) - f(y)| < \epsilon$. Given another pseudo-distance space (I, d_I) and $\psi : X \to I$, we say that ψ is Lipschitz continuous on $A \subseteq X$ with a constant $0 \neq \Lambda \in \mathbb{R}$ (or, briefly, that ψ is Λ-Lipschitz continuous on A) if $d_I(\psi(x), \psi(y)) \leq \Lambda d(x, y)$ for all $(x, y) \in A^2$ (here and elsewhere, if $d(x, y) \in \mathbb{R}$ for all $(x, y) \in A^2$, then the case $\Lambda = 0$ is legitimate too).

Given a family $(A_t)_{t \in I}$ of nonempty subsets of X and a point $t_0 \in I$, we say that $\limsup_{t \to t_0} D_{H,X}(A_{t_0}, A_t) \leq 0$ if for every $\epsilon > 0$, there exists $\delta > 0$ such for all $t \in I$ which satisfy $d_I(t_0, t) < \delta$, we have $D_{H,X}(A_{t_0}, A_t) < \epsilon$; here, for the sake of clarity, we denoted D_H by $D_{H,X}$, to emphasize that it is induced by d and not by d_I.

We finish this section with a few remarks. First, the fact that the pseudo-distance d may be asymmetric, means that the above-mentioned types of continuity, as well as other notions (such as the distance $d(x, A)$ between some $x \in X$ and $\emptyset \neq A \subseteq X$, as well as the "ball" $B[A, r]$, $r > 0$) will be different if, in the corresponding definitions, we change the order in which we measure the distance between two given points. For example, in the definition of continuity, instead of assuming the inequality $d(x, y) < \delta$ we could have assumed that $d(y, x) < \delta$. Nevertheless, in either case the relevant proofs follow essentially the same reasoning. Second, we could extend our setting and some of the results even further (for example, by assuming that $d : X^2 \to L$, where $L \neq \emptyset$ is an arbitrary linearly/totally/simply ordered set), but we refrain from doing this here.

3. Stability of the optimal values

In this section we present the stability principle in both its finite and infinite versions. Despite the general setting, the proofs are rather simple. We present them for the sake of completeness and in order to eliminate any suspicion that possibly a few subtle points have been missed due to the general setting. The counterexamples which come afterward show that the assumptions imposed in the formulation of the principle are essential. At the end of the section we make a few relevant comments regarding this stability principle.

Theorem 3.1 (Stability of the optimal values: the finite case). *Let (X, d) be a pseudo-distance space and let $f : X \to \mathbb{R}$ be given. If f is uniformly continuous on (X, d) in the sense of Section 2, then SUP_f is uniformly continuous on the pseudo-distance space $(\mathrm{dom}(\mathrm{SUP}_f), D_H)$, and INF_f is uniformly continuous on the pseudo-distance space $(\mathrm{dom}(\mathrm{INF}_f), D_H)$. If f is Λ-Lipschitz continuous for some $\Lambda > 0$ (in the sense of Section 2), then SUP_f is Λ-Lipschitz continuous on $(\mathrm{dom}(\mathrm{SUP}_f), D_H)$ and INF_f is Λ-Lipschitz continuous on $(\mathrm{dom}(\mathrm{INF}_f), D_H)$.*

Proof. We only consider the assertions regarding SUP_f, because the assertions regarding INF_f can be proved in a similar manner, or can be deduced from the assertions regarding SUP_f by working with $-f$ instead of f.

Fix some arbitrary $\epsilon > 0$. Assume first that f is uniformly continuous. Then there is $\delta > 0$ such that for all $(x, y) \in X^2$ which satisfy $d(x, y) < \delta$, we have $|f(x) - f(y)| < 0.5\epsilon$. We first prove that SUP_f is uniformly continuous on its effective domain, where we actually show that the δ from the previous lines can be associated, in the definition of the uniform continuity of SUP_f, with the given ϵ. Let $A, A' \in \mathrm{dom}(\mathrm{SUP}_f)$ be arbitrary such that $D_H(A, A') < \delta$ (if no such sets exist, then the proof is complete, vacuously). Since $\mathrm{SUP}_f(A) \in \mathbb{R}$, there is some $a \in A$ such that

$$(3.1) \qquad\qquad \mathrm{SUP}_f(A) - 0.5\epsilon < f(a).$$

From (2.3) we have $d(a, A') \leq D_H(A, A') < \delta$. Hence there is some $a' \in A'$ such that $d(a, a') < \delta$. The uniform continuity of f and (3.1) imply that $\mathrm{SUP}_f(A) - 0.5\epsilon < f(a) < f(a') + 0.5\epsilon$. Since $-\infty < f(a') \leq \mathrm{SUP}_f(A')$, we obtain $\mathrm{SUP}_f(A) - \mathrm{SUP}_f(A') < \epsilon$. By interchanging the roles of A and A' we obtain, using an argument similar to the above one, that $-\epsilon < \mathrm{SUP}_f(A) - \mathrm{SUP}_f(A')$.

Now assume that f is Λ-Lipschitz continuous for some $\Lambda > 0$. Fix some arbitrary $A, A' \in \mathrm{dom}(\mathrm{SUP}_f)$. We need to show that $|\mathrm{SUP}_f(A) - \mathrm{SUP}_f(A')| \leq \Lambda D_H(A, A')$. The assertion is obvious if $D_H(A, A') = \infty$, and so from now on we may assume that $D_H(A, A') < \infty$. We observe that the case $D_H(A, A') = -\infty$ is impossible since the fact that f is Λ-Lipschitz continuous implies that $0 \leq |f(x) - f(y)| \leq \Lambda d(x, y)$ for all $(x, y) \in X^2$, and so the assumption that and $\Lambda > 0$ implies that d and hence D_H are nonnegative. Thus we can actually assume that $D_H(A, A') \in [0, \infty)$. The definition of $\mathrm{SUP}_f(A)$ implies that (3.1) holds for some $a \in A$. Since from (2.3) we have $d(a, A') \leq D_H(A, A') < D_H(A, A') + 0.5\epsilon$, there is some $a' \in A'$ such that $d(a, a') < D_H(A, A') + 0.5\epsilon$. We conclude from (3.1) and the Λ-Lipschitz continuity of f that $\mathrm{SUP}_f(A) - 0.5\epsilon < f(a) \leq f(a') + \Lambda d(a, a') \leq f(a') + \Lambda D_H(A, A') + 0.5\Lambda\epsilon$. As obviously $f(a') \leq \mathrm{SUP}_f(A')$, it follows that $\mathrm{SUP}_f(A) - \mathrm{SUP}_f(A') < \Lambda D_H(A, A') + 0.5\epsilon(1 + \Lambda)$. Since ϵ can be arbitrarily small, we have $\mathrm{SUP}_f(A) - \mathrm{SUP}_f(A') \leq \Lambda D_H(A, A')$. By interchanging the roles of A and A' we obtain, using an argument similar to the above one, that $-\Lambda D_H(A, A') \leq \mathrm{SUP}_f(A) - \mathrm{SUP}_f(A')$. $\qquad\square$

Theorem 3.2 (Stability of the optimal values: the infinite case). *Let (X, d) be a pseudo-distance space. Suppose that $f : X \to \mathbb{R}$ is continuous in the sense of Section 2. Given an arbitrary $\emptyset \neq A \subseteq X$, if $\mathrm{SUP}_f(A) = \infty$, then for all $\mu \in \mathbb{R}$, there exists $\delta > 0$ such that for each nonempty subset $A' \subseteq X$ satisfying $D_{asyH}(A, A') < \delta$ (in particular, for each $\emptyset \neq A' \subseteq X$ satisfying $D_H(A, A') < \delta$), the following inequality*

holds:

(3.2) $\mu < \text{SUP}_f(A')$.

Similarly, if $\text{INF}_f(A) = -\infty$, *then for all* $\mu \in \mathbb{R}$, *there exists* $\delta > 0$ *such that for each nonempty subset* $A' \subseteq X$ *satisfying* $D_{asyH}(A, A') < \delta$ *(in particular, for each* $\emptyset \neq A' \subseteq X$ *satisfying* $D_H(A, A') < \delta$*), one has*

(3.3) $\text{INF}_f(A') < \mu$.

Proof. We only consider the case where $\text{SUP}_f(A) = \infty$, because the proof in the case where $\text{INF}_f(A) = -\infty$ employs similar arguments (or, alternatively, can be deduced from the first case by taking $-f$ instead of f). Let $\mu \in \mathbb{R}$ and $\epsilon > 0$ be arbitrary. The definition of $\text{SUP}_f(A)$ and the fact that $\text{SUP}_f(A) = \infty$ imply that there exists a point $a \in A$ such that

(3.4) $\mu + \epsilon < f(a)$.

Since f is continuous on X, it is continuous at a. Thus for the given ϵ, there exists $\delta > 0$ such that for every $x \in X$ satisfying $d(a, x) < \delta$, we have

(3.5) $f(a) < f(x) + \epsilon$.

Let $\emptyset \neq A' \subseteq X$ be arbitrary such that $D_{asyH}(A, A') < \delta$. If no such subset A' exists, then the proof is complete (the assertion holds vacuously). Otherwise, the inequality $D_{asyH}(A, A') < \delta$ and (2.2) imply that $d(a, A') \leq D_{asyH}(A, A') < \delta$. Therefore there is a point $a' \in A'$ such that $d(a, a') < \delta$. We conclude from (3.4) and (3.5) (with $x := a'$) that $\mu + \epsilon < f(a) < f(a') + \epsilon$, namely $\mu < f(a')$. Since obviously $f(a') \leq \text{SUP}_f(A')$, we have $\mu < \text{SUP}_f(A')$, that is, (3.2) holds.   □

Counterexample 3.3. The uniform continuity assumption on f in Theorem 3.1 is essential. Indeed, let $X := \mathbb{R}$ and let d be the standard absolute value metric; let $A := \cup_{k=2}^{\infty}[2k, 2k+1]$. For each $1 < j \in \mathbb{N}$ and $1 < k \in \mathbb{N}$, let $J_{j,k}$ be defined by $J_{j,k} := [2k, 2k + 1 + (1/k)]$ if $j = k$ and $J_{j,k} := [2k, 2k + 1]$ if $j \neq k$. For every $1 < j \in \mathbb{N}$, let $A_j := \cup_{k=2}^{\infty} J_{j,k}$. Let $\mathbb{N}_1 := \mathbb{N} \backslash \{1\}$, and let $f : X \to \mathbb{R}$ be defined as follows for each $t \in X$:

$$f(t) := \begin{cases} 0, & \text{if } t \in (-\infty, 4], \\ 0, & \text{if } t \in [2k, 2k+1] \text{ for some } k \in \mathbb{N}_1, \\ -2kt + 2k(2k+1), & \text{if } t \in [2k+1, 2k+1+\frac{1}{2k}] \text{ for some } k \in \mathbb{N}_1, \\ 4kt - 4k(2k+1) - 3, & \text{if } t \in [2k+1+\frac{1}{2k}, 2k+1+\frac{1}{k}] \text{ for some } k \in \mathbb{N}_1, \\ \dfrac{kt}{1-k} - \dfrac{k(2k+2)}{1-k}, & \text{if } t \in [2k+1+\frac{1}{k}, 2k+2] \text{ for some } k \in \mathbb{N}_1. \end{cases}$$

Then $f(t)$ is defined for all $t \in X$ and f is continuous, but not uniformly continuous on X. In addition, for each $1 < j \in \mathbb{N}$, we have $\text{INF}_f(A_j) = -1$ and $\text{SUP}_f(A_j) = 1$. Since $\text{INF}_f(A) = 0 = \text{SUP}_f(A)$ and $\lim_{j \to \infty} D_H(A, A_j) = 0$, we conclude that neither INF_f nor SUP_f are continuous at A.

Counterexample 3.4. Here is another counterexample related to Theorem 3.1, where now X is bounded. For each $k \in \mathbb{N}$, let $e_k \in \ell_2$ be the vector having 1 in its k-th component and 0 in the other components, and let $[0, e_k]$ be the line segment which connects the origin 0 of ℓ_2 with e_k. For each $j, k \in \mathbb{N}$, define $J_{j,k} := [0, ((k+1)/k)e_k)$ (half-open line segment) if $j = k$ and $J_{j,k} := [0, e_k]$ if $j \neq k$.

Let X be defined by $X := \cup_{k=1}^{\infty} [0, ((k+1)/k)e_k)$ and d be the metric induced by the ℓ_2 norm. Let $A := \cup_{k=1}^{\infty} [0, e_k]$ and for all $j \in \mathbb{N}$, let $A_j := \cup_{k=1}^{\infty} J_{j,k}$. Then X, A and A_j, $j \in \mathbb{N}$, are bounded, and we have $A \subset X$ and also $A_j \subset X$ for each $j \in \mathbb{N}$. For each $k \in \mathbb{N}$, let $f_k : [0, (k+1)/k) \to \mathbb{R}$ be defined by $f_k(t) := 0$ for every $t \in [0,1]$ and $f_k(t) := \sin(2\pi/(1 + (1/k) - t))$ for all $t \in [1, (k+1)/k)$. Now let $f : X \to \mathbb{R}$ be defined as follows: given $0 \neq x = (x_i)_{i=1}^{\infty} \in X$, there exists a unique $k \in \mathbb{N}$ such that $x \in [0, ((k+1)/k)e_k)$, namely $x_i = 0$ if $i \neq k$ and $x_i \in [0, (k+1)/k)$ if $i = k$, where $i, k \in \mathbb{N}$; in this case, define $f(x) := f_k(x_k)$; for $x = 0$, let $f(x) := 0$. Then f is continuous on X, but it is not uniformly continuous there. Moreover, $\mathrm{INF}_f(A) = 0 = \mathrm{SUP}_f(A)$. However, for each $j \in \mathbb{N}$, we have $\mathrm{INF}_f(A_j) = -1$ and $\mathrm{SUP}_f(A_j) = 1$, and, in addition, $\lim_{j \to \infty} D_H(A, A_j) = 0$. Thus neither INF_f nor SUP_f are continuous at A.

Remark 3.5. (a) **The level of generality of d:** Both Theorem 3.1 and Theorem 3.2 hold even though we assume essentially nothing regarding d. In particular, d may attain infinite values, may be negative, may not satisfy the triangle inequality and so on.

(b) **A redundant assumption:** In the proof of Theorem 3.1 we required both A and A' to be in $\mathrm{dom}(\mathrm{SUP}_f)$. This requirement is actually redundant, since if $A \in \mathrm{dom}(\mathrm{SUP}_f)$ and $\emptyset \neq A' \subseteq X$ satisfies $D_H(A, A') < \delta$ (where δ is associated with some given $\epsilon > 0$ in the definition of uniform continuity of f), then for a fixed $a' \in A'$ we can find $a \in A$ such that $d(a', a) < \delta$, and so the uniform continuity of f implies that $-\infty < f(a') < f(a) + \epsilon \leq \mathrm{SUP}_f(A) + \epsilon$; in other words, $A' \in \mathrm{dom}(\mathrm{SUP}_f)$.

(c) **An intuitive interpretation:** The previous part implies an intuitive interpretation of Theorem 3.1. Indeed, suppose that we are given a real function f which is defined on an arbitrary pseudo-distance space (X, d), and suppose that we know that f is uniformly continuous on X; given a nonempty subset A of X on which f is bounded from above, suppose that we perturb A slightly, where the perturbation is measured with respect to the Hausdorff distance; let A' be the new subset which is obtained from the original subset A; then f is bounded from above on A' and its supremum over A' is almost equal to the supremum of f over A. Similarly, if f is bounded from below, then the infimum of f over A' is almost equal to the infimum of f over A. Moreover, if f is known to be Lipschitz continuous, then the perturbation of the optimal values "behaves better".

(d) **Stability of the sets of minimizers and maximizers:** Theorem 3.1 raises the corresponding question regarding stability of the sets of minimizers and maximizers of the function under consideration, that is, not only the stability of its optimal values. In general, stability does not hold as is shown in the counterexample in the next paragraph. However, under further assumptions a weak stability principle related to the minimizers and maximizers can be formulated: Roughly speaking, given a continuous function $f : X \to \mathbb{R}$, where (X, d) is a compact metric space, and given some nonempty and closed subset A of X, if one slightly perturbs A to a new nonempty and closed subset A', then the set of minimizers of f over A is "slightly" perturbed too in the sense that the asymmetric Hausdorff distance (and hence the standard distance) between

the set of minimizers of f over A' and the set of minimizers of f over A is small; an analogous assertion holds regarding the perturbed and original sets of maximizers. This result is essentially known [43, Theorem 4.5], and its proof is rather simple.

As for the promised counterexample, consider the case where $X := [-20, 20]$ with the absolute value metric. Let $A := [0, \pi]$ and $A'_\epsilon := [-\epsilon, \pi - \epsilon]$ for every $\epsilon \in (0, 1/2)$. In addition, let $f : X \to \mathbb{R}$ be defined by $f(t) := |\sin(t)|$, $t \in X$. Then f is uniformly continuous on X and its set of minimizers over A is $\{0, \pi\}$. However, no matter how small ϵ is, the set of minimizers of f over A'_ϵ is $\{0\}$. Thus the Hausdorff distance between these sets is π, despite the fact that $D_H(A'_\epsilon, A) = \epsilon \to 0$ as ϵ tends to 0. On the other hand, the stability result stated in the previous paragraph does hold because the asymmetric Hausdorff distance between $\{0\}$ and $\{0, \pi\}$ is 0.

(e) **Predecessors of our stability principle:** Here we discuss a few predecessors of Theorem 3.1, and also additional related results. We mention them in the next paragraphs, but before doing so we note that there are several significant differences between these results and our ones, for instance with respect to the setting, with respect to the proofs, the fact that we present explicit estimates and more; we also note that we have not been aware of these results when we derived Theorem 3.1.

The first relevant result is a sequential continuity (not uniform continuity or Lipschitz continuity as in our Theorem 3.1) result regarding the maximal value function in the setting of metric spaces, which seems to be implicit in Kummer [43]. Interestingly, it seems that this fact has not been observed so far in the literature, not even in [43] (the explicit relevant results there are for compact metric spaces [43, Theorem 4.5], finite-dimensional Euclidean spaces with a quasiconvex objective function [43, Theorem 4.10], and Banach spaces [43, Theorem 4.11] with quite demanding and technical assumptions).

Another relevant result appears in [44, Satz 2.1 and Section 4]. It is a sequential continuity result related to the maximal value function of a quadratic function over a constraint set which is the sum of a compact set K and a polyhedral cone U, both located in a finite-dimensional Euclidean space. A related result appears in [42, Satz 2.3] (infimal value of a rather specific functional in a normed space setting). The notion of convergence in all of the above-mentioned cases is either convergence with respect to the Hausdorff distance or a slightly more general notion.

Yet another result appears in [47, Corollary 4]. It essentially says that the infimum of a certain functional, the variable of which is a compact subset of a finite-dimensional Euclidean space (up to slight change of notation, this is essentially the γ functional mentioned in [47, Relation (9)]), is sequentially continuous with respect to the Kuratowski-Painlevé notion of convergence of sets.

Finally, we also note that Corollary 4.1 below, regarding the continuity, and Lipschitz continuity, of the so-called "optimal value function" or "marginal function", has predecessors, and we mention them in the beginning of Section 4.

4. APPLICATION 1: CONTINUITY OF THE OPTIMAL VALUE FUNCTION FROM PARAMETRIC OPTIMIZATION

In this section we show how Theorem 3.1 can be used to prove that the so-called "optimal value function" (or "marginal function", or "inf-projection") from parametric optimization is continuous under certain assumptions. Corollary 4.1 below extends, to the setting of pseudo-distance spaces, related results formulated in a metric space setting, such as [49, Lemma 3.18, Parts 3 and 4], [39, Lemma 1] and [58, Theorem 3.1.22 and Proposition 3.3.10].

Corollary 4.1. *Let (X, d_X) and (I, d_I) be two arbitrary pseudo-distance spaces. Suppose that $\mathscr{C}$ is a nonempty set of nonempty subsets of X. Assume that $(A_t)_{t \in I}$ is a family of subsets in $\mathscr{C}$. Given $f : X \to \mathbb{R}$, define $\phi^* : I \to [-\infty, \infty]$ and $\phi_* : I \to [-\infty, \infty]$ by $\phi^*(t) := \mathrm{SUP}_f(A_t) := \sup\{f(x) : x \in A_t\}$ and $\phi_*(t) := \mathrm{INF}_f(A_t) := \inf\{f(x) : x \in A_t\}$, respectively, for each $t \in I$. Then the following two statements hold:*

(a) *Assume that $\mathscr{C}$ has the property that for each $A \in \mathscr{C}$, there is $r_A > 0$ such that $B[A, r_A] \in \mathscr{C}$. Assume also that f is uniformly continuous on each $A \in \mathscr{C}$, and for each $t_0 \in I$, one has $\limsup_{t \to t_0} D_{H,X}(A_{t_0}, A_t) \leq 0$ in the sense of Section 2 (here $D_{H,X}$ is the Hausdorff distance induced by d_X and not by d_I). If $A_t \in \mathrm{dom}(\mathrm{SUP}_f)$ for all $t \in I$, then ϕ^* is a continuous function from I to $\mathbb{R}$, and if $A_t \in \mathrm{dom}(\mathrm{INF}_f)$ for every $t \in I$, then ϕ_* is a continuous function from I to $\mathbb{R}$.*

(b) *Assume that $\mathscr{C}$ has the following two properties: first, that for each $C_1, C_2 \in \mathscr{C}$, there is $C_{1,2} \in \mathscr{C}$ such that $C_1 \cup C_2 \subseteq C_{1,2}$, and second, that f is $\Lambda(C)$-Lipschitz continuous on each $C \in \mathscr{C}$ for some $\Lambda(C) > 0$. Assume also that for all $t, s \in I$ there is some $\alpha_{t,s} > 0$ such that $D_{H,X}(A_t, A_s) \leq \alpha_{t,s} d_I(t, s)$.*

    If $A_t \in \mathrm{dom}(\mathrm{SUP}_f)$ for all $t \in I$, then for all $t, s \in I$ there is some $\Lambda_{t,s} > 0$ such that $|\phi^*(t) - \phi^*(s)| \leq \alpha_{t,s} \Lambda_{t,s} d_I(t, s)$. In particular, if $A_t \in \mathrm{dom}(\mathrm{SUP}_f)$ for all $t \in I$, and f is Λ-Lipschitz continuous on X and there is some $\alpha > 0$ such that $D_{H,X}(A_t, A_s) \leq \alpha d_I(t, s)$ for all $t, s \in I$, then ϕ^* is $\alpha\Lambda$-Lipschitz continuous on I. Similarly, if $A_t \in \mathrm{dom}(\mathrm{INF}_f)$ for every $t \in I$, then for all $t, s \in I$ there is some $\Lambda_{t,s} > 0$ such that $|\phi_*(t) - \phi_*(s)| \leq \alpha_{t,s} \Lambda_{t,s} d_I(t, s)$. In particular, if $A_t \in \mathrm{dom}(\mathrm{INF}_f)$ for every $t \in I$, and f is Λ-Lipschitz continuous on X and there is some $\alpha > 0$ such that $D_{H,X}(A_t, A_s) \leq \alpha d_I(t, s)$ for all $t, s \in I$, then ϕ_* is $\alpha\Lambda$-Lipschitz continuous on I.

Proof. We start with Part (a) regarding ϕ^*. Let $t_0 \in I$ and $\epsilon > 0$ be arbitrary. Since we assume that $B[A_{t_0}, r_0] \in \mathscr{C}$ for some $r_0 > 0$, it follows, in particular, that $B[A_{t_0}, r_0] \neq \emptyset$. Moreover, if $\emptyset \neq A' \subseteq X$ satisfies $D_{H,X}(A_{t_0}, A') < r_0$, then $A' \subseteq B[A_{t_0}, r_0]$, as follows from the definition of $B[A_{t_0}, r_0]$ and (2.3). Since f is uniformly continuous on any subset which belongs to $\mathscr{C}$, it is uniformly continuous on $B[A_{t_0}, r_0]$. Moreover, by our assumption, $A_{t_0} \in \mathrm{dom}(\mathrm{SUP}_f)$. Thus, we can apply Theorem 3.1, where the space there is $B[A_{t_0}, r_0]$ and the pseudo-distance is the restriction of d_X to $B[A_{t_0}, r_0]^2$. This theorem implies, in particular, that for our ϵ, there exists $\delta_0 \in (0, r_0]$ such that for all $\emptyset \neq A' \subseteq B[A_{t_0}, r_0]$ satisfying

$A' \in \mathrm{dom}(\mathrm{SUP}_f)$ and $D_{H,X}(A_{t_0}, A') < \delta_0$, we have

(4.1)
$$|\phi^*(t_0) - \mathrm{SUP}_f(A')| < \epsilon.$$

Since we assume that $\limsup_{t \to t_0} D_{H,X}(A_{t_0}, A_t) \le 0$, it follows that for the positive number δ_0, there exists $\delta > 0$ such that for each $t \in I$ which satisfies $d_I(t_0, t) < \delta$, we have $D_{H,X}(A_{t_0}, A_t) < \delta_0$. Since $A_t \in \mathrm{dom}(\mathrm{SUP}_f)$, we can substitute $A' := A_t$ in (4.1). In other words, $|\phi^*(t_0) - \phi^*(t)| < \epsilon$ for each $t \in I$ satisfying $d_I(t_0, t) < \delta$. Therefore ϕ^* is continuous at t_0. The proof of the claim regarding ϕ_* is similar.

Now we prove Part (b) regarding ϕ^*. Let $t, s \in I$ be given. According to our assumptions, there is some $A_{t,s} \in \mathscr{C}$ and $\Lambda_{t,s} > 0$ such that $A_t \cup A_s \subseteq A_{t,s}$ and f is $\Lambda_{t,s}$-Lipschitz continuous on $A_{t,s}$. Since we also assume that there is some $\alpha_{t,s} > 0$ such that $D_{H,X}(A_t, A_s) \le \alpha_{t,s} d_I(t, s)$, Theorem 3.1 (in which the space is $A_{t,s}$ and the pseudo-distance is the restriction of d_X to $A_{t,s}^2$) implies that
(4.2)
$$|\phi^*(t) - \phi^*(s)| = |\mathrm{SUP}_f(A_t) - \mathrm{SUP}_f(A_s)| \le \Lambda_{t,s} D_{H,X}(A_t, A_s) \le \alpha_{t,s} \Lambda_{t,s} d_I(t, s).$$

In particular, if f is Λ-Lipschitz continuous on X for some $\Lambda > 0$ and there is some $\alpha > 0$ such that $D_{H,X}(A_t, A_s) \le \alpha d_I(t, s)$ for all $t, s \in I$, then we conclude from Theorem 3.1 (as in (4.2)) that ϕ^* is $\alpha\Lambda$-Lipschitz continuous on I. The proof of the claim regarding ϕ_* is similar. $\square$

Remark 4.2. An example of a set $\mathscr{C}$ having the property mentioned in the formulation of Corollary 4.1(a) is provided by the set of all nonempty and bounded subsets of a metric space: in Corollary 6.1 below we use this example. A second example for $\mathscr{C}$ is the set of all nonempty subsets of a metric space, or, more generally, the set of all nonempty subsets of a pseudo-distance space (X, d), such that (X, d) has the property that $0 = d(x, x) \le d(x, y)$ for every $(x, y) \in X^2$ (for instance, this happens if X is a real or complex vector space and d is the distance induced by a Minkowski functional of a convex subset of X which contains the origin: in Section 5 below we use this set; another example: X is the zone of a Bregman function and d is the associated Bregman divergence [15, 68]). A third example is as follows: again, we consider a pseudo-distance space (X, d) having the property that $0 = d(x, x) \le d(x, y)$ for every $(x, y) \in X^2$, and are also given a uniformly continuous function $f : X \to \mathbb{R}$; then Theorem 3.1 and Remark (3.5)b imply that we can take $\mathscr{C}$ to be either $\mathrm{dom}(\mathrm{SUP}_f)$ or $\mathrm{dom}(\mathrm{INF}_f)$.

Similarly, if the space X is a metric space and f is Lipschitz continuous on every nonempty and bounded subset of the space, then the set of all nonempty and bounded subsets of the space is an example for a set $\mathscr{C}$ having the property mentioned in the formulation of Corollary 4.1(b). If f is Lipschitz continuous on the whole space, then the set of all nonempty subsets of the space is an example for $\mathscr{C}$. Two additional examples for $\mathscr{C}$ are the set of all nonempty convex subsets of a normed space, assuming that f is Lipschitz continuous on every nonempty convex subset of the space, and the set of all nonempty, convex and bounded subsets of the space, assuming that f is Lipschitz continuous on every nonempty, convex and bounded subset of the space.

As a final remark, we note that in Corollary 4.1(b), if $d_I(t, s) \in \mathbb{R}$ for all $t, s \in I$, then both $\alpha_{t,s}$ and α can be assigned the value 0.

5. Application 2: mixed linear-nonlinear programming, extended generalized inverses, a general variant of Hoffman's lemma

In this section we consider two mixed linear-nonlinear programming problems and establish continuity properties of the corresponding optimal value functions. In the first case (Example 5.9) the objective function is, in general, nonlinear but the constraints are linear, and in the second case (Example 5.11) the function is nonlinear and the constraints are partly linear and partly nonlinear. Our results extend partly, but significantly, a theory which was developed in previous works: see Remark 5.12 below for more details. Along the way we present a very general variant of the so-called Hoffman's Lemma (see Lemma 5.6) and generalize the so-called "generalized inverse" of a linear operator (Definition 5.2, Remarks 5.3–5.5). We present our results in Subsection 5.2. Before presenting them, we need some background, in the form of a few definitions and remarks, which are presented in Subsection 5.1.

5.1. **Background.** In this subsection we discuss some concepts which we use later. We start by recalling the concept of a Minkowski functional.

Definition 5.1. *Suppose that X is a vector space over $\mathbb{R}$ or $\mathbb{C}$. The Minkowski functional associated with a convex subset $C \subseteq X$ which contains 0 is the function $\mathscr{M}_C : X \to [0, \infty]$ defined by*

$$(5.1) \qquad \mathscr{M}_C(x) := \inf\{\mu : \mu \geq 0 \text{ and } x \in \mu C\}, \quad x \in X,$$

where, of course, $\mu C = \{\mu c : c \in C\}$ and $\inf \emptyset := \infty$.

It is well known that $\mathscr{M}_C$ might be a norm, but unless C satisfies certain properties, $\mathscr{M}_C$ is not a norm in general (for instance, it may be asymmetric and may attain the value $+\infty$). Nevertheless, $\mathscr{M}_C$ enjoys several properties similar to those of a norm, for example it is positively homogeneous (that is, $\mathscr{M}_C(\lambda x) = \lambda \mathscr{M}_C(x)$ for all $x \in X$ and $\lambda > 0$) and subadditive (namely, $\mathscr{M}_C(x+y) \leq \mathscr{M}_C(x) + \mathscr{M}_C(y)$ for all $x, y \in X$). See [83, p. 26] for more details (note that there additional assumptions are imposed on C, but the proofs hold when one merely assumes that C is convex and $0 \in C$, and the assertions hold also in the cases where $\mathscr{M}_C$ attains the value $+\infty$; in fact, one can impose even weaker assumptions for the assertions to hold).

Now we introduce an extension of the concept of a generalized inverse of a linear operator.

Definition 5.2. *Let X and Y be nonempty sets. Assume that on both sets a binary operation $+$ is defined. Here we abuse our notation and denote both operations by "$+$". Assume further that the $+$ operation on Y has a right-neutral element 0, that is, $y + 0 = y$ for each $y \in Y$. Given $L : X \to Y$, suppose that its kernel $\{x \in X : Lx = 0\}$ is nonempty. Assume that $\widetilde{Y}$ is a subset of Y which contains the image $L(X)$ of X by L. The operator L is said to have an "extended generalized inverse" (EGI) with domain $\widetilde{Y}$ if there exists a (not necessarily additive) operator $\widetilde{L} : \widetilde{Y} \to X$ which has the following property: for all $t \in L(X)$ there are $v \in X$ and $a_{0,v}$ in the kernel of L such that $t = Lv$ and $\widetilde{L}(t) = v + a_{0,v}$.*

Remark 5.3. Definition 5.2 significantly generalizes the concept of a "generalized inverse" of a linear operator (also called the "generalized reciprocal", or the "Moore-Penrose generalized inverse", or the "Moore-Penrose pseudo-inverse", or the "Moore-Penrose inverse", in honor of the contributions of Moore [52] and Penrose [59] to this theory). Indeed, consider first the case of a standard finite-dimensional generalized inverse. Here one starts with a linear operator $L : X \to Y$ which acts between two real or complex finite-dimensional vector spaces X and Y, and its generalized inverse is the linear operator $L^\dagger : Y \to X$, which uniquely exists, and satisfies the following four relations: $LL^\dagger L = L$, $L^\dagger LL^\dagger = L^\dagger$, $(LL^\dagger)^* = LL^\dagger$ and $(L^\dagger L)^* = L^\dagger L$, where $*$ denotes the linear conjugation. Since the generalized inverse $L^\dagger$ automatically satisfies $L^\dagger L = id_X - P_{A_0} = P_{L^*(Y)}$ (this follows from [11, Ex. 58, p. 80] and [11, Theorem 1, p. 12]), where $L(X)$ is the image/range of L, A_0 is kernel of L, P_{A_0} is the projection operator onto A_0 along $L^*(Y)$, and id_X is the identity operator on X, our extended generalized inverse $\widetilde{L}$ does generalize the concept of (standard) generalized inverse.

More generally, suppose that L is a bounded linear operator which acts between two Banach spaces X and Y and has the following two properties: first, its kernel A_0 satisfies $X = A_0 \oplus A_0'$ for some closed linear subspace A_0' of X, and second, its image $L(X)$ satisfies $Y = L(X) \oplus M'$ for some closed linear subspace M' of Y (namely, A_0 is topologically complemented in X and $L(X)$ is topologically complemented in Y). A (Moore-Penrose) generalized inverse to L is a bounded linear operator $L^\dagger : Y \to X$ which satisfies the following four relations: $LL^\dagger L = L$, $L^\dagger LL^\dagger = L^\dagger$, $LL^\dagger = P_{L(X)}$ and $L^\dagger L = id_X - P_{A_0}$, where P_{A_0} is the projection operator onto A_0 along A_0' and $P_{L(X)}$ is the projection operator onto $L(X)$ along M'. Such an operator sometimes exists. It can be seen that in Definition 5.2 we required only a relation weaker than the fourth relation, and we did not require $\widetilde{L}$ to be linear, L to be bounded, and the spaces to be Banach spaces (or even vector spaces).

We note that the concept of generalized inverse of a linear operator has been extensively studied during the last 60 years or so and has found various applications: see, for instance, the books [11, 17, 27, 32, 88] and the semi-survey [57] (and the references therein); see also the online bibliographic list [10] which is composed of 1670 items, and last updated in June 2001. Hence we believe that our extension of this concept has a promising potential to yield diverse applications too. We also note that other generalizations of this concept exist (see, for example, [11, 90] and the references therein), but they seem to have a somewhat different nature than our generalization.

Remark 5.4. Consider the setting of Definition 5.2. Assume further that X has a right-neutral element 0 (that is, $x + 0 = x$ for each $x \in X$). If $L(0) = 0$ (as, in particular, happens if L is additive), then L has an EGI. Indeed, given $t \in L(x)$, there is at least one $v_t \in X$ such that $Lv_t = t$. Now we use the Axiom of Choice to define, for each $t \in \widetilde{Y} := L(X)$, an operator $\widetilde{L} : \widetilde{Y} \to X$ by $\widetilde{L}t := v_t$. Obviously $\widetilde{L}t = v_t + 0$, and since $L(0) = 0$, one can see that the required conditions mentioned in Definition 5.2 are satisfied.

Remark 5.5. For applications, one usually needs to require more from an extended generalized inverse $\widetilde{L}$ of L, such as its Lipschitz continuity. Here we want to mention

two cases in which one can provide a linear and bounded EGI in spaces which are not necessarily normed spaces.

(i) In this case we show that $\widetilde{L}$ can be taken to be a standard (Moore-Penrose) generalized inverse $L^{\dagger}$, where the main technical work is to show that $L^{\dagger}$ is Lipschitz continuous with respect to the given pseudo-distances. Our basic setting is two Banach spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ and a bounded linear operator $L : X \to Y$ which is known to have a standard generalized inverse $L^{\dagger} : Y \to X$ which is a bounded linear operator. We assume further that $\mathscr{M}_{C_X} : X \to [0, \infty]$ is a Minkowski functional induced by a convex subset $C_X \subseteq X$ which contains 0 in its interior (interior with respect to $\|\cdot\|_X$), and $\mathscr{M}_{C_Y} : Y \to [0, \infty]$ is a Minkowski functional induced by a convex subset $C_Y \subseteq Y$ which is bounded with respect to the norm $\|\cdot\|_Y$ and contains 0. It can be checked immediately that $\mathscr{M}_{C_X}(x) \leq \alpha\|x\|_X$ for all $x \in X$, where $\alpha := 1/r$ and $r > 0$ is the radius of some open ball (with respect to $\|\cdot\|_X$) which is centered at 0 and is contained in C_X, and $\|y\|_Y \leq \beta\mathscr{M}_{C_Y}(y)$ for all $y \in Y$, where $\beta := \sup\{\|c\|_Y : c \in C_Y\}$. Hence for all $y, z \in Y$,

$$\mathscr{M}_{C_X}(L^{\dagger}y - L^{\dagger}z) \leq \alpha\|L^{\dagger}y - L^{\dagger}z\|_X \leq \alpha\|L^{\dagger}\|\|y - z\|_Y \leq \alpha\|L^{\dagger}\|\beta\mathscr{M}_{C_Y}(y - z),$$

and so $\widetilde{L} := L^{\dagger}$ is $\alpha\|L^{\dagger}\|\beta$-Lipschitz continuous as a function from the pseudo-distance space $(X, \mathscr{M}_{C_X})$ to the pseudo-distance space $(Y, \mathscr{M}_{C_Y})$.

(ii) As a second example, suppose that $(X, \|\cdot\|_X)$ is a real or complex normed space. Assume that $(Y, \|\cdot\|_Y)$ is a normed space (over $F = \mathbb{R}$ or $F = \mathbb{C}$). Assume that $\mathscr{S}_X : X \to [0, \infty]$ has the property that for some $\kappa > 0$ and all scalars α and β (real or complex),

$$(5.2) \qquad \mathscr{S}_X(\alpha u + \beta v) \leq \kappa\left(|\alpha|\mathscr{S}_X(u) + |\beta|\mathscr{S}_X(v)\right), \quad \forall u, v \in X,$$

where we regard $0 \cdot \infty$ as 0. For instance, the previously mentioned inequality holds if $\mathscr{S}_X$ is a norm or a quasi-norm; it also holds if $\mathscr{S}_X$ is the Minkowski functional induced by a convex subset C_X of X which contains 0 and has the property that $C_X \subseteq \sigma(-C_X)$ for some $\sigma > 0$ if the field associated with X is real (if the field associated with X is complex, then (5.2) holds when $C_X \subseteq \sigma(iC_X)$ for some $\sigma > 0$; this assumption implies, in particular, that $C_X \subseteq \sigma^2(-C_X)$).

Suppose that $L : X \to Y$ is a linear operator having the property that its range $\widetilde{Y} := L(X)$ is finite-dimensional. In what follows we assume that $\widetilde{Y} \neq \{0\}$, since otherwise the analysis becomes rather trivial. Since $\widetilde{Y}$ is finite-dimensional, it has a basis $(w_j)_{j=1}^n$, where $n \in \mathbb{N}$ and for all $j \in \{1, \ldots, n\}$ there is some $v_j \in X$ such that $w_j = L(v_j)$. Let $V := \text{span}\{v_j : j \in \{1, \ldots, n\}\}$ and $L_V : V \to \widetilde{Y}$ be the restriction of L to V. Then L_V is invertible because $(w_j)_{j=1}^n$ is a basis of $\widetilde{Y}$, and L_V^{-1} is bounded since it is a linear mapping which acts between the finite-dimensional spaces $\widetilde{Y}$ and V. Denote $\widetilde{L} := L_V^{-1}$.

We assume that $\mathscr{S}_X$ has the property that $\mathscr{S}_X(\widetilde{L}w_j) < \infty$ for all $j \in \{1, \ldots, n\}$ (this happens, if, for example, $\mathscr{S}_X(x) < \infty$ for all $x \in X$, but it can happen in other cases as well). Assume further that $\mathscr{S}_Y : Y \to [0, \infty]$ has the properties that the set $S_{\widetilde{Y}} := \{y \in \widetilde{Y} : \mathscr{S}_Y(y) = 1\}$ is bounded with

respect to $\|\cdot\|_Y$, that $\mathscr{S}_Y$ is positively homogeneous on $\widetilde{Y}$, and $\mathscr{S}_Y(y) > 0$ for all $0 \neq y \in \widetilde{Y}$. These properties hold if, for example, $\mathscr{S}_Y$ is the Minkowski functional induced by a convex subset E of Y which contains 0 and is closed and bounded with respect to $\|\cdot\|_Y$ (in this case one also has $E = \{y \in Y : \mathscr{S}_Y(y) \leq 1\}$, as follows, for instance, essentially word for word from the proof of [66, Lemma 3.1(a)]). Another example for $\mathscr{S}_Y$ are functions which induce certain star bodies; such functions appear in the theory of geometry of numbers [18, 33, 50, 53]; for instance, if $Y = \mathbb{R}^3$, then $\mathscr{S}_Y$, defined for all $y = (y_1, y_2, y_3) \in Y$ by

$$\mathscr{S}_Y(y) := (|y_1||y_2||y_3|)^{1/3} + (|y_2|^2|y_3|^3)^{1/5} + \sqrt{|y_3^2 - y_1^2|} + (\sqrt{|y_1|} + \sqrt{|y_2|} + \sqrt{|y_3|})^2,$$

has the required properties on the whole space Y.

We claim that under the above-mentioned assumptions, $\widetilde{L}$ is Lipschitz continuous as a linear operator from $(\widetilde{Y}, \mathscr{S}_Y)$ to $(X, \mathscr{S}_X)$. Indeed, any $y \in \widetilde{Y}$ can be written uniquely as $y = \sum_{j=1}^n \lambda_j w_j$, where each λ_j, $j \in \{1, \ldots, n\}$, is a scalar in F. Let $T : F^n \to \widetilde{Y}$ be the linear operator defined by $T(\mu_j)_{j=1}^n := \sum_{j=1}^n \mu_j w_j$ for every $(\mu_j)_{j=1}^n \in F^n$. Then T is invertible, and since T^{-1} acts between two finite-dimensional spaces, T^{-1} is bounded (where the norm on F^n is the max norm $\|\cdot\|_\infty$). Since $S_{\widetilde{Y}}$ is nonempty (because $y/\mathscr{S}_Y(y) \in S_{\widetilde{Y}}$ for all $0 \neq y \in \widetilde{Y}$) and bounded with respect to $\|\cdot\|_Y$, it follows that $T^{-1}(S_{\widetilde{Y}})$ is bounded in F^n, namely $\eta := \sup\{\|T^{-1}(s)\|_\infty : s \in S_{\widetilde{Y}}\} < \infty$. This inequality, (5.2), induction (on n) and the assumption that $\mathscr{S}_X(\widetilde{L}w_j) < \infty$ for each $j \in \{1, \ldots, n\}$, imply that for all $y \in S_{\widetilde{Y}}$,

$$(5.3) \quad \mathscr{S}_X(\widetilde{L}y) = \mathscr{S}_X(\widetilde{L}(\sum_{j=1}^n \lambda_j w_j)) = \mathscr{S}_X(\sum_{j=1}^n \lambda_j \widetilde{L}w_j)$$

$$\leq \tau \sum_{j=1}^n |\lambda_j| \mathscr{S}_X(\widetilde{L}w_j) \leq \tau\eta \sum_{j=1}^n \mathscr{S}_X(\widetilde{L}w_j) =: \sigma < \infty,$$

where $\tau := \max\{\kappa, \ldots, \kappa^{n-1}\}$. Since $S_{\widetilde{Y}} = \{y \in \widetilde{Y} : \mathscr{S}_{\widetilde{Y}}(y) = 1\}$, our assumptions on $\mathscr{S}_Y$ imply that for all $y, z \in \widetilde{Y}$ which satisfy $y \neq z$, we have $\mathscr{S}_Y(y - z) > 0$ and $(y - z)/\mathscr{S}_Y(y - z) \in S_{\widetilde{Y}}$. Hence, using (5.2), it follows from (5.3) that

$$\mathscr{S}_X(\widetilde{L}y - \widetilde{L}z) \leq \kappa \mathscr{S}_Y(y - z)\mathscr{S}_X\left(\widetilde{L}\left(\frac{y - z}{\mathscr{S}_Y(y - z)}\right)\right) \leq \kappa\sigma\mathscr{S}_Y(y - z).$$

In other words, $\widetilde{L}$ is $\kappa\sigma$-Lipschitz continuous, as claimed. We note that when $\mathscr{S}_X$ and $\mathscr{S}_Y$ are norms, then we can, of course, take also $\|\widetilde{L}\|$ as a Lipschitz constant of $\widetilde{L}$. Finally, since, given $t \in \widetilde{Y}$, there is $v \in V$ such that $Lv = t$, and since $\widetilde{L}t = L_V^{-1}L_V v = v + 0$, it follows that $\widetilde{L}$ is an EGI.

5.2. **The application.** In this subsection we present the application of our stability results to parametric optimization. We do this in Example 5.9 and Example 5.11, which are also based on Lemma 5.6 below. This lemma is related to, but different

from, the so called "Hoffman's Lemma" and several variants of it: see Remark 5.8 below for more details.

Lemma 5.6 (A general variant of Hoffman's Lemma)**.** *Let X be a commutative additive group and let $\mathscr{S}_X : X \to [-\infty, \infty]$ be given. Let $d : X^2 \to [-\infty, \infty]$ be the pseudo-distance induced by $\mathscr{S}_X$, namely $d(u, v) := \mathscr{S}_X(u-v)$ for all $u, v \in X$. Let Y be an additive group and assume that $L : X \to Y$ is an additive operator which has an extended generalized inverse $\widetilde{L} : \widetilde{Y} \to X$ with domain $\widetilde{Y}$, where $\widetilde{Y}$ is an additive subgroup of Y which satisfies $I := L(X) \subseteq \widetilde{Y}$. Given $\mathscr{S}_{\widetilde{Y}} : \widetilde{Y} \to [-\infty, \infty]$, suppose that $\widetilde{L}$ is α-Lipschitz continuous from $(\widetilde{Y}, \mathscr{S}_{\widetilde{Y}})$ to $(X, \mathscr{S}_X)$ for some $0 \neq \alpha \in \mathbb{R}$ (the case $\alpha = 0$ is allowed too if $\mathscr{S}_{\widetilde{Y}}(y) \in \mathbb{R}$ for all $y \in \widetilde{Y}$). Denote $A_t := \{x \in X : Lx = t\}$ for all $t \in I$. Then $D_H(A_s, A_t) \leq \max\{\alpha \mathscr{S}_{\widetilde{Y}}(s-t), \alpha \mathscr{S}_{\widetilde{Y}}(t-s)\}$ for all $s, t \in I$.*

Proof. Let $t \in I$ be given. Then, according to the assumption that $\widetilde{L}$ is an EGI of L (Definition 5.2), there are $v \in X$ and $a_{0,v} \in A_0$ such that $t = Lv$ and $\widetilde{L}t = v + a_{0,v}$. Hence $v = \widetilde{L}t - a_{0,v}$. The definition of A_t implies that $v \in A_t$, and therefore $\widetilde{L}t - a_{0,v}$ is in A_t. But, as is well known and can be proved immediately, A_t, which is the set of solutions $z \in X$ to the inhomogeneous additive equation $Lz = t$, is the sum of A_0 and an arbitrary particular solution to the inhomogeneous equation. Since, as shown above, $\widetilde{L}t - a_{0,v}$ solves the inhomogeneous equation $Lz = t$, we have $A_t = A_0 + (\widetilde{L}t - a_{0,v})$, and so $A_t = A_0 + \widetilde{L}t$ since X is commutative and A_0 is invariant under translation. Similarly, $A_s = A_0 + \widetilde{L}s$ for all $s \in I$. Fix some $s \in I$ and let $x \in A_s$. Then $x = a_{0,x} + \widetilde{L}s$ for some $a_{0,x} \in A_0$. Let $y := a_{0,x} + \widetilde{L}t$. Then $y \in A_0 + \widetilde{L}t$, namely, $y \in A_t$. In addition, since $\widetilde{L}$ is α-Lipschitz continuous on $\widetilde{Y}$ with respect to $\mathscr{S}_{\widetilde{Y}}$ and $\mathscr{S}_X$, we have $\mathscr{S}_X(\widetilde{L}(\widetilde{x}) - \widetilde{L}(\widetilde{y})) \leq \alpha \mathscr{S}_{\widetilde{Y}}(\widetilde{x} - \widetilde{y})$ for all $\widetilde{x}$ and $\widetilde{y}$ in $\widetilde{Y}$. Since $I \subseteq \widetilde{Y}$, we conclude from the previous lines and the fact that X is commutative that $\mathscr{S}_X(x - y) = \mathscr{S}_X(\widetilde{L}(s) - \widetilde{L}(t)) \leq \alpha \mathscr{S}_{\widetilde{Y}}(s - t)$. Since $y \in A_t$, we have $d(x, A_t) \leq \mathscr{S}_X(x - y)$, and so $d(x, A_t) \leq \alpha \mathscr{S}_{\widetilde{Y}}(s - t)$. But x was an arbitrary element in A_s, and so $D_{asyH}(A_s, A_t) = \sup_{x \in A_s} d(x, A_t) \leq \alpha \mathscr{S}_{\widetilde{Y}}(s - t)$. Similarly, $D_{asyH}(A_t, A_s) \leq \alpha \mathscr{S}_{\widetilde{Y}}(t - s)$. Thus $D_H(A_s, A_t) \leq \max\{\alpha \mathscr{S}_{\widetilde{Y}}(s-t), \alpha \mathscr{S}_{\widetilde{Y}}(t-s)\}$. $\square$

Remark 5.7 (An even more general variant of Hoffman's Lemma)**.** It is possible to generalize Lemma 5.6 even further. For instance, suppose that $\widetilde{L}$ has the property that for some function $\nu : I^2 \to [-\infty, \infty]$, we have $\mathscr{S}_X(\widetilde{L}s - \widetilde{L}t) \leq \nu(s, t)$ for all $(s, t) \in I^2$. Then an analysis similar to the one used in the proof of Lemma 5.6 shows that $D_H(A_s, A_t) \leq \max\{\nu(s, t), \nu(t, s)\}$ for all $s, t \in I$.

Remark 5.8. Lemma 5.6 (and Remark 5.7) is related to, but different from, some known results, such as [20, Theorem 4.15, Theorem 4.16], [86, Theorem 1] and [94, Theorem 2.1]. All of these results are variants of the so-called Hoffman's Lemma [34, the Theorem in Section 2].

The differences between Lemma 5.6 and the above-mentioned results are both in the formulations and in the methods of proof. For instance, the nature of the above-mentioned results is finite-dimensional (either the spaces are finite-dimensional or there are finite systems of equalities/inequalities) and no EGI appears (in fact, in the above-mentioned works, only in [20, Theorem 4.15, Theorem 4.16] one can

see the appearance of the standard generalized inverse of a matrix, but its use, in [20, Lemma 4.1], is significantly different from our use of the EGI in Lemma 5.6); in addition, in these works the spaces are always assumed to be normed spaces, although in Hoffman's paper [34] one allows a more general magnitude function (which, implicitly, has to satisfy certain relations with respect to the Euclidean norm); on the other hand, the nature of Lemma 5.6 is rather general (possibly infinite-dimensional vector spaces or even spaces which are not vector spaces, magnitude functions which are much more general than a norm, and so on) and we use (and introduce) the concept of an extended generalized inverse.

We note that [70, Theorem 7] is also a general variant of Hoffman's Lemma, but its nature is different from Lemma 5.6. For instance, it is for convex processes in Banach spaces, that, when restricted to the linear case, some restrictive assumptions are imposed on the operator, unless the spaces are finite-dimensional. In addition, the method of proof in [70, Theorem 7] is significantly different from the one of Lemma 5.6, and, in particular, it does not use the notion of an EGI (however, interestingly, it does use a certain inverse, namely an inverse of a set-valued operator).

Anyway, since Lemma 5.6 is related to the above-mentioned results, one can regard it as a very general variant of Hoffman's Lemma.

Example 5.9 (Nonlinear function, linear constraints)**.** Consider the setting of Lemma 5.6. Assume that $f : X \to \mathbb{R}$ is bounded from below and is also uniformly continuous with respect to the distance induced by the "pseudo-magnitude" $\mathscr{S}_X : X \to [0, \infty]$ (not necessarily a Minkowski functional). Denote by d_I the distance on I induced by $\mathscr{S}_{\widetilde{Y}}$. Suppose that $\mathscr{S}_{\widetilde{Y}}(0) = 0$. Suppose also that the conjugate pseudo-magnitude $\overline{\mathscr{S}_{\widetilde{Y}}} : X \to [0, \infty]$, which is defined by $\overline{\mathscr{S}_{\widetilde{Y}}}(x) := \mathscr{S}_{\widetilde{Y}}(-x)$ for each $x \in X$, is continuous at 0 (the continuity is with respect to $\mathscr{S}_{\widetilde{Y}}$). Let $\phi_* : I \to \mathbb{R}$ be the optimal value function defined by $\phi_*(t) := \mathrm{INF}_f(A_t)$, $t \in I$, namely

$$\phi_*(t) = \inf\{f(x) : x \in X, \, Lx = t\}, \quad t \in L(X).$$

Let $\mathscr{C}$ be the set of all nonempty subsets of X. Since f is bounded from below and $A_t \neq \emptyset$ for every $t \in I$, it follows that $A_t \in \mathrm{dom}(\mathrm{INF}_f)$ for each $t \in I$. Moreover, as we have shown in Lemma 5.6, there exists some real number α, which must be nonnegative since $\mathscr{S}_{\widetilde{Y}}$ is nonnegative (unless $\mathscr{S}_{\widetilde{Y}}$ vanishes identically, but in this case we can obviously replace α by a nonnegative number), such that $D_H(A_s, A_t) \leq \max\{\alpha\mathscr{S}_{\widetilde{Y}}(s - t), \alpha\overline{\mathscr{S}_{\widetilde{Y}}}(s - t)\}$ for all $s, t \in I$. Since $\overline{\mathscr{S}_{\widetilde{Y}}}$ is continuous at 0 and $\overline{\mathscr{S}_{\widetilde{Y}}}(0) = 0 = \mathscr{S}_{\widetilde{Y}}(0)$, we have $\lim_{t \to s} D_H(A_s, A_t) = 0$ for all $s \in I$. Since the assumptions on $\mathscr{S}_{\widetilde{Y}}$ and hence on d_I imply that $\mathscr{C}$ satisfies the conditions needed in Corollary 4.1(a), this corollary implies that ϕ_* is continuous. If, in addition, f is Λ-Lipschitz continuous on X for some $\Lambda > 0$ and $\overline{\mathscr{S}_{\widetilde{Y}}}$ is β-Lipschitz continuous with respect to $\mathscr{S}_{\widetilde{Y}}$ for some $\beta > 0$, then ϕ_* is $\alpha \cdot \max\{1, \beta\} \cdot \Lambda$-Lipschitz continuous, as follows from Corollary 4.1(b).

Remark 5.10. In connection with Example 5.9, it could be of interest to note that functions which are uniformly continuous with respect to a pseudo-magnitude (even with respect to a Minkowski functional) may not be uniformly continuous with respect to a norm $\|\cdot\|$. Indeed, let $X := \mathbb{R}^2$ with the usual Euclidean norm

$\|\cdot\|$, and let $C := [-2, 1] \times \{0\}$. Consider the Minkowski functional $\mathscr{M}_C$ induced by C. A simple calculation shows that $\mathscr{M}_C(x_1, x_2) = x_1$ if $x = (x_1, x_2) \in [0, \infty) \times \{0\}$, $\mathscr{M}_C(x_1, x_2) = -0.5x_1$ if $(x_1, x_2) \in (-\infty, 0] \times \{0\}$ and $\mathscr{M}_C(x_1, x_2) = \infty$ otherwise. In particular, $\mathscr{M}_C(0) = 0$ and $0 \leq \overline{\mathscr{M}_C}(x) \leq 2\mathscr{M}_C(x)$ at any $x \in X$, and hence the conjugate $\overline{\mathscr{M}_C}$ (see Example 5.9) is continuous at 0 with respect to $\mathscr{M}_C$.

Now let $f : X \to \mathbb{R}$ be defined for all $(x_1, x_2) \in \mathbb{R}^2$ by $f(x_1, x_2) := g(x_2)$, where $g : \mathbb{R} \to \mathbb{R}$ is some discontinuous function (with respect to the usual absolute value metric). It follows that f is not uniformly continuous with respect to the Euclidean norm, since it is not even continuous with respect to this norm. To see that f is uniformly continuous with respect to $\mathscr{M}_C$, let $\epsilon > 0$ be given and let $\delta := \epsilon$. Given any (x_1, x_2) and (y_1, y_2) in X which satisfy $\mathscr{M}_C((x_1, x_2) - (y_1, y_2)) < \delta$, one has, in particular, $\mathscr{M}_C((x_1 - y_1, x_2 - y_2)) < \infty$. Hence $x_2 = y_2$. Since $f(x_1, x_2) = g(x_2)$ and $f(y_1, y_2) = g(y_2)$, it follows that $|f(x_1, x_2) - f(y_1, y_2)| = 0 < \epsilon$, and so indeed f is uniformly continuous.

Example 5.11 (Nonlinear function, mixed linear-nonlinear constraints)**.** Given $n \in \mathbb{N}$, endow $\mathbb{R}^n$ with some norm $\|\cdot\|$ and suppose that C is compact and convex, and has a nonempty interior $\mathrm{Int}(C)$. Assume that $f : C \to \mathbb{R}$ is continuous. Given some $m \in \mathbb{N}$, suppose that $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator, where the norm on $\mathbb{R}^m$ is $\|\cdot\|_{\mathbb{R}^m}$ (just an arbitrary norm). For all t in the image $L(\mathbb{R}^n)$ of $\mathbb{R}^n$ by L, denote $E_t := \{x \in \mathbb{R}^n : Lx = t\}$. Denote by I the set of all $t \in L(\mathbb{R}^n)$ for which the following regularity condition holds: the affine subspace E_t intersects $\mathrm{Int}(C)$, namely there is some $x \in \mathrm{Int}(C)$ such that $Lx = t$. Assume that $I \neq \emptyset$ (as shown in Remark 5.12(ii) below, this latter assumption actually implies that I is convex and open in $L(\mathbb{R}^n)$). For all $t \in I$, let $A_t := E_t \cap C$ and let $\phi_* : I \to \mathbb{R}$ be the optimal value function defined by $\phi_*(t) := \mathrm{INF}_f(A_t)$, $t \in I$, that is,

$$\phi_*(t) = \inf\{f(x) : x \in C, \, Lx = t\}, \quad t \in I.$$

Given $t \in I$, since $E_t \cap \mathrm{Int}(C) \neq \emptyset$, we have $A_t \neq \emptyset$, and since E_t is closed and C is compact, A_t is compact. Thus the Extreme Value Theorem implies that $A_t \in \mathrm{dom}(\mathrm{INF}_f)$ for each $t \in I$. We prove below that $\lim_{t \to s} D_H(A_s, A_t) = 0$ for all $s \in I$. This fact, when combined with the fact that the continuous function f is actually uniformly continuous (since C is compact), implies that we may use Corollary 4.1(a) (in which the first space is C with the restriction of $\|\cdot\|$ to C as the distance function, d_I is the distance induced on I by $\|\cdot\|_{\mathbb{R}^m}$, and $\mathscr{C}$ is the set of all nonempty subsets of C), from which we conclude that ϕ_* is continuous.

To see that indeed $\lim_{t \to s} D_H(A_s, A_t) = 0$ for all $s \in I$, suppose to the contrary that this is not true. Then there are $s \in I$, $\epsilon > 0$ and a sequence $(t_k)_{k=1}^\infty$ of elements in I such that $\lim_{k \to \infty} t_k = s$ and $D_H(A_s, A_{t_k}) \geq \epsilon$ for each $k \in \mathbb{N}$. Therefore either $D_{asyH}(E_{t_k} \cap C, E_s \cap C) \geq \epsilon$ for all $k \in N_1$, where $N_1 \subseteq \mathbb{N}$ is an infinite set, or $D_{asyH}(E_s \cap C, E_{t_k} \cap C) \geq \epsilon$ for all $k \in N_2$, where $N_2 \subseteq \mathbb{N}$ is an infinite set.

Consider the first case. It implies that for each $k \in N_1$, there is some $x_k \in E_{t_k} \cap C$ such that $d(x_k, E_s \cap C) > 0.5\epsilon$. Since the sequence $(x_k)_{k \in N_1}$ is contained in the compact set C, there is an infinite set $N_{11} \subseteq N_1$ and $x_\infty \in C$ such that $\lim_{k \in N_{11}} x_k = x_\infty$. Lemma 5.6, Remark 5.5(ii) and the assumption that $\lim_{k \to \infty} t_k = s$ imply that $\lim_{k \to \infty} D_H(E_{t_k}, E_s) = 0$, and so from (2.2) we conclude

that $\lim_{k\to\infty,k\in N_{11}} d(x_k,E_s)=0$. This fact and the continuity of the distance function imply that $d(x_\infty,E_s)=0$. Thus (since E_s is closed) $x_\infty\in E_s$. We conclude that $x_\infty\in E_s\cap C$. However, since $d(x_k,E_s\cap C)>0.5\epsilon$ for each $k\in N_1$, when we pass to the limit $k\to\infty$, $k\in N_{11}$ and use the continuity of the distance function, we have $d(x_\infty,E_s\cap C)\geq 0.5\epsilon$, a contradiction.

Now consider the second case which was mentioned two paragraphs earlier. It implies that for each $k\in N_2$, there is some $y_k\in E_s\cap C$ such that $d(y_k,E_{t_k}\cap C)>0.5\epsilon$. Since the sequence $(y_k)_{k\in N_2}$ is contained in the compact set $E_s\cap C$, there is an infinite set $N_{22}\subseteq N_2$ and $y_\infty\in E_s\cap C$ such that $\lim_{k\in N_{22}} y_k=y_\infty$. According to our assumption, $E_s\cap\mathrm{Int}(C)\neq\emptyset$. Let $z\in E_s\cap\mathrm{Int}(C)$ and $r:=\min\{0.1\epsilon,0.5\|y_\infty-z\|\}$. If $y_\infty=z$, then $r=0$. We let $w:=z$ and observe that $w\in\mathrm{Int}(C)$. Otherwise, let $w:=y_\infty+r((z-y_\infty)/\|z-y_\infty\|)$. In this latter case w belongs to the half-open line segment $(y_\infty,z]$. As is well known [73, Theorem 6.1, p. 45], since $z\in\mathrm{Int}(C)$ and $y_\infty\in C$, the convexity of C implies that $(y_\infty,z]$ is contained in $\mathrm{Int}(C)$. Thus $w\in\mathrm{Int}(C)$ again. Hence there is some $\rho\in(0,r)$ such that the ball of radius ρ with center w is contained in C. Moreover, $w\in E_s$ since E_s is convex and $[y_\infty,z]\subseteq E_s$.

Lemma 5.6 and Remark 5.5(ii) imply that $\lim_{k\to\infty} D_H(E_{t_k},E_s)=0$. As a result, it follows that $\lim_{k\to\infty} d(w,E_{t_k})=0$. Thus for all $k\in\mathbb{N}$ sufficiently large there is some $u_k\in E_{t_k}$ such that $\|w-u_k\|<\rho$. Our choice of ρ implies that $u_k\in C$ and the triangle inequality implies that $\|y_\infty-u_k\|\leq\|y_\infty-w\|+\|w-u_k\|<r+\rho<2r\leq 0.2\epsilon$. Therefore

$$(5.4) \qquad\qquad d(y_\infty,E_{t_k}\cap C)\leq\|y_\infty-u_k\|<0.25\epsilon$$

for all $k\in\mathbb{N}$ sufficiently large and, in particular, for all $k\in N_{22}$ sufficiently large. On the other hand, the fact that $\lim_{k\in N_{22}} y_k=y_\infty$ implies that $\|y_\infty-y_k\|<0.25\epsilon$ for all $k\in N_{22}$ sufficiently large. Since $d(y_k,E_{t_k}\cap C)>0.5\epsilon$ for each $k\in N_2$, the triangle inequality implies that $0.5\epsilon<d(y_k,E_{t_k}\cap C)\leq\|y_k-y_\infty\|+d(y_\infty,E_{t_k}\cap C)<0.25\epsilon+d(y_\infty,E_{t_k}\cap C)$. Thus $0.25\epsilon<d(y_\infty,E_{t_k}\cap C)$ for all $k\in N_{22}$ sufficiently large. This inequality contradicts (5.4) and proves that the second case mentioned several paragraphs above cannot hold too. Thus we indeed have $\lim_{t\to s} D_H(A_s,A_t)=0$ for all $s\in I$, as asserted.

Below we collect a few remarks regarding Examples 5.9 and 5.11.

Remark 5.12.   (i) Examples 5.9 and 5.11 extend partly, but significantly, the stability theory developed in [89, pp. 279–280] and [87, Theorem 2] (for a related theory, see [89, p. 281] and [77, Lemma 4.1]). This theory has been applied to analyzing stochastic programs [87, Section 4], [77, Section 4]. The setting in [89, pp. 279–280] and [87, Theorem 2] is a finite-dimensional Euclidean space, a polyhedral set C, and an objective function f which is Lipschitz continuous on a set which contains C; another requirement in [89, pp. 279–280] is that either the level-sets of f are bounded or $C^\infty\cap A_0=\{0\}$, where A_0 is the kernel of the linear operator L and C^∞ is the horizontal cone associated with C. It is proved in [89] that the associated optimal value function ϕ_* is Lipschitz continuous under these assumptions (note: the constraint set there is written as $\{x\in C: Lx=b-t\}$, where b is a given vector and t is the parameter; thus by a simple change of variable we can arrive at this formulation). It can be seen that Examples 5.9 and 5.11 extend this theory to the

case of spaces which are not necessarily normed spaces (and not necessarily finite-dimensional), a constraint set C which is either the entire space or a (usually non-polyhedral) convex body, a linear operator L which should have a Lipschitz continuous EGI, and an objective function f which is either Lipschitz continuous or merely uniformly continuous. We are still able to derive the continuity of ϕ_* under these conditions, and sometimes (Example 5.9) its Lipschitz continuity.

 We believe that the theory developed in this section can be extended further, and, in particular, that it is possible to remove (at least in some interesting cases) the compactness and finite-dimensionality assumptions from Example 5.11.

(ii) The set I mentioned in Example 5.11 is actually open and convex whenever it is nonempty. Indeed, given $s \in I$, let $x \in E_s \cap \text{Int}(C)$. In particular, $x \in \text{Int}(C)$ and hence there exists some $r > 0$ such that the ball of radius r about x is contained in $\text{Int}(C)$. Remark 5.5(ii) and Lemma 5.6 imply that there is some $\alpha > 0$ such that $D_H(E_s, E_t) \leq \alpha\|s - t\|_{\mathbb{R}^m}$ for all $t \in L(\mathbb{R}^n)$. Thus, if $\delta := r/\alpha$, then for each $t \in L(\mathbb{R}^n)$ which satisfies $\|s - t\|_{\mathbb{R}^m} < \delta$, we have $D_H(E_s, E_t) < r$. This inequality and the fact that x also belongs to E_s imply that there is some $y \in E_t$ such that $\|x - y\| < r$. Thus $y \in \text{Int}(C) \cap E_t$ and so $t \in I$ for all t in the ball of radius δ about s. To see that I is convex, let $s, t \in I$ and $\lambda \in [0, 1]$ be given. Then $Lx_s = s$ and $Lx_t = t$ for some $x_s, x_t \in \text{Int}(C)$. We have $\lambda x_s + (1 - \lambda)x_t \in \text{Int}(C)$ since $\text{Int}(C)$ is convex. In addition, $L(\lambda x_s + (1 - \lambda)x_t) = \lambda L(x_s) + (1 - \lambda)L(x_t) = \lambda s + (1 - \lambda)t$. Consequently, $\lambda s + (1 - \lambda)t \in I$, as required.

6. Application 3: a sequence of Lipschitz constants

 In this section we use Corollary 4.1 in order to show that, under some assumptions, a rather general sequence of positive numbers can be a sequence of Lipschitz constants associated with a given function (each Lipschitz constant corresponds to a certain subset on which one measures the Lipschitz continuity of the function). Corollary 6.1 below has recently been applied in the analysis of a telescopic proximal gradient method [69].

Corollary 6.1. *Suppose that $f : U \to \mathbb{R}$ is a twice continuously (Fréchet) differentiable function defined on an open and convex subset U of some real normed space $(X, \|\cdot\|)$, $X \neq \{0\}$. Suppose that C is a convex subset of X which has the property that $C \cap U \neq \emptyset$. Assume that f'' is bounded and uniformly continuous on bounded subsets of $C \cap U$. Fix an arbitrary $y_0 \in C \cap U$, and let $s := \sup\{\|f''(x)\| : x \in C \cap U\}$ and $s_0 := \|f''(y_0)\|$. If $s = \infty$, then for each strictly increasing sequence $(\lambda_k)_{k=1}^\infty$ of positive numbers which satisfies $\lambda_1 > s_0$ and $\lim_{k \to \infty} \lambda_k = \infty$, there exists an increasing sequence $(S_k)_{k=1}^\infty$ of bounded and convex subsets of C (and also closed if C is closed) which satisfies the following properties: first, $S_k \cap U \neq \emptyset$ for all $k \in \mathbb{N}$, second, $\cup_{k=1}^\infty S_k = C$, third, for each $k \in \mathbb{N}$, the function f' is Lipschitz continuous on $S_k \cap U$ with λ_k as a Lipschitz constant; moreover, if C contains more than one point, then also $S_k \cap U$ contains more than one point for each $k \in \mathbb{N}$. Finally, if $s < \infty$, then f' is Lipschitz continuous on $C \cap U$ with s as a Lipschitz constant.*

Proof. Suppose first that $s = \infty$. Let $I := [0, \infty)$ and let d_I be the standard absolute value metric. For each $t \in I$, define B_t to be the intersection of C with the closed ball of radius t and center y_0 (here $B_0 := \{y_0\}$). Then B_t is a bounded and convex subset of C for each $t \in I$, and it is also closed if C is closed. Let $A_t := B_t \cap U$ for each $t \in I$. Then $A_t \neq \emptyset$ (it contains y_0), convex and bounded for all $t \in I$. In addition, $\cup_{t \in I} B_t = C$ and $\cup_{t \in I} A_t = C \cap U$. An immediate verification shows that $D_H(A_t, A_{t'}) \leq |t - t'|$ for all $t, t' \in I$. Since f'' exists and is bounded on bounded subsets of $C \cap U$, the function h, which is defined by $h(x) := \|f''(x)\|$, $x \in C \cap U$, is finite at each point, and it is also bounded on each bounded subset of $C \cap U$. This implies that the function $\phi^* : I \to (-\infty, \infty]$, which is defined by $\phi^*(t) := \mathrm{SUP}_h(A_t)$ for each $t \in I$, satisfies $\phi^*(t) \in [0, \infty)$ for each $t \in I$. In addition, since f'' is uniformly continuous on bounded subsets of $C \cap U$, the triangle inequality shows that the function h, too, is uniformly continuous on bounded subsets of $C \cap U$. We conclude from the previous lines that the conditions needed in Corollary 4.1(a) hold (the first pseudo-distance space there is $C \cap U$, where the pseudo-distance is the metric which is induced by the restriction of the norm of X to $C \cap U$; in addition, the set $\mathscr{C}$ in Corollary 4.1(a) is the set of all nonempty and bounded subsets of $C \cap U$), and consequently, ϕ^* is a continuous function on I.

Since $s = \sup\{\|f''(x)\| : x \in C \cap U\} = \infty$, for each $\rho \geq \|f''(y_0)\| = s_0$, there exists $x \in C \cap U$ such that $\|f''(x)\| > \rho$. Since $\cup_{t \in I} A_t = C \cap U$, there exists $t(x) \in I$ such that $x \in A_{t(x)}$. As a result, from the definition of ϕ^* we see that $\phi^*(t(x)) \geq \|f''(x)\| > \rho$. By applying the classical Intermediate Value Theorem to the continuous function ϕ^* on the interval $[0, t(x)]$, we conclude that each value between $\phi^*(0) = s_0$ and $\phi^*(t(x))$ is attained. In particular, ρ is attained. Since ρ was an arbitrary number which is greater than or equal to s_0 and since ϕ^* is increasing, it follows that the image of $I = [0, \infty)$ under ϕ^* is the interval $[s_0, \infty)$. Therefore, given $k \in \mathbb{N}$, since $\lambda_k \geq \lambda_1 > s_0$, there exists $t_k \in [0, \infty)$ such that $\phi^*(t_k) = \lambda_k$, and this t_k must be positive, otherwise $t_k = 0$ and hence $s_0 = \phi^*(0) = \phi^*(t_k) = \lambda_k$, a contradiction.

Let $S_k := B_{t_k}$ for each $k \in \mathbb{N}$. Then S_k is bounded and convex for each $k \in \mathbb{N}$, and it is also closed if C is closed. In addition, $S_k \cap U$ is nonempty (it contains y_0), bounded and convex for every $k \in \mathbb{N}$. Since $\|f''(x)\| \leq \sup\{\|f''(y)\| : y \in S_k \cap U\} = \phi^*(t_k)$ for all $x \in S_k \cap U$, and since f' is continuously differentiable on U and hence on $S_k \cap U$, the (generalized) Mean Value Theorem applied to f' (see [1, Theorem 1.8, p. 13, and also p. 23]; this theorem is formulated for Gâteaux differentiable functions acting between real Banach spaces, but it holds as well for Fréchet differentiable functions acting between real normed spaces, because no completeness assumption is needed in the proof, and the Fréchet and Gâteaux derivatives coincide in our case) implies that f' is Lipschitz continuous on $S_k \cap U$ with $\phi^*(t_k)$ as a Lipschitz constant, namely with λ_k as a Lipschitz constant.

Now we show that $\cup_{k=1}^{\infty} S_k = C$. Indeed, since ϕ^* is increasing and $(\lambda_k)_{k=1}^{\infty}$ is strictly increasing, it follows that $(t_k)_{k=1}^{\infty}$ is increasing. Hence $\ell := \lim_{k \to \infty} t_k$ exists and it must be that $\ell = \infty$, otherwise $\lambda_k = \phi^*(t_k) \leq \phi^*(\ell) < \infty$ for all $k \in \mathbb{N}$, a contradiction to the assumption that $\lim_{k \to \infty} \lambda_k = \infty$. Hence the union of the closed balls with common center y_0 and radii t_k, $k \in \mathbb{N}$, is X. Thus the intersection

of this union with C is C itself. On the other hand, this intersection is $\cup_{k=1}^{\infty} S_k$, as follows from the definition of the subsets S_k, $k \in \mathbb{N}$. In other words, $\cup_{k=1}^{\infty} S_k = C$.

It remains to show that if C contains more than one point, then $S_k \cap U$ also contains more than one point for every $k \in \mathbb{N}$. Indeed, take some arbitrary $w_0 \in C$ which satisfies $w_0 \neq y_0$. The line segment $[y_0, w_0]$ is contained in C because C is convex. Since U is open and $y_0 \in U$, there is a sufficiently small closed ball B of center y_0 and positive radius $r < \min\{\|y_0 - w_0\|, t_k\}$ such that $B \subset U$. The intersection of B with C contains the segment $[y_0, y_0 + r\theta]$, where $\theta := (w_0 - y_0)/\|w_0 - y_0\|$. Since $r < t_k$, it follows from the definition of S_k that $B \cap C \subset B_{t_k} = S_k$. Hence S_k contains the nondegenerate segment $[y_0, y_0 + r\theta]$, namely it contains more than one point.

Finally, we need to consider the case where $s < \infty$. In this case $\|f''(x)\| \leq s < \infty$ for every $x \in C \cap U$. Since $C \cap U$ is convex and f' is Fréchet (hence Gâteaux) differentiable on U, the Mean Value Theorem applied to f' implies that f' is Lipschitz continuous on $C \cap U$ with s as a Lipschitz constant. $\square$

7. Application 4: a general scheme for tackling a wide class of nonconvex and nonsmooth optimization problems

7.1. **The method:** Given a pseudo-distance space (X, d), consider the general optimization problem of minimizing (or maximizing) a given uniformly continuous function $f : X \to \mathbb{R}$ over a nonempty subset $A \subseteq X$. Theorem 3.1 suggests a general scheme for approximating both $\mathrm{INF}_f(A)$ and $\mathrm{SUP}_f(A)$. Indeed, consider the case of approximating $\mathrm{INF}_f(A)$ (the case of approximating $\mathrm{SUP}_f(A)$ follows a similar reasoning) and assume that it is known that $\mathrm{INF}_f(A) \in \mathbb{R}$. Assume also that we are able to approximate A by a sequence $(A_k)_{k=1}^{\infty}$ of subsets of X such that $\lim_{k \to \infty} D_H(A, A_k) = 0$ and $\mathrm{INF}_f(A_k) \in \mathbb{R}$ for all $k \in \mathbb{N}$. Furthermore, assume that we are also able to compute an approximation $\tilde{\sigma}_k$ to $\mathrm{INF}_f(A_k)$ so that $\lim_{k \to \infty} |\tilde{\sigma}_k - \mathrm{INF}_f(A_k)| = 0$. Then Theorem 3.1 ensures that the following equality holds: $\lim_{k \to \infty} \tilde{\sigma}_k = \lim_{k \to \infty} [\tilde{\sigma}_k - \mathrm{INF}_f(A_k)] + \lim_{k \to \infty} \mathrm{INF}_f(A_k) = \mathrm{INF}_f(A)$. Consequently, the general scheme is nothing but computing $\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\sigma}_3, \ldots$.

7.2. **A few remarks:** The above-mentioned method seems to be useful in cases where A itself does not have a "finite representation" or is not easily computable. For example, suppose that A is a component (or the union of the components) of a double zone diagram induced by finitely many sites contained in a convex body in a finite-dimensional strictly convex normed space, or, more generally, in a compact geodesic metric space X which has the geodesic inclusion property [64, Definition 3.1]. A double zone diagram is an exotic geometric object which is defined to be a fixed point of a certain operator which acts on tuples of sets. While its existence is known in general [65, Theorem 5.5] and one can even represent explicitly one of the double zone diagrams, this representation is not finite in the sense that it based on an infinite increasing union of known "inner tuples" of sets: see [64, Theorem 5.2] (this representation was observed before in [6, Lemma 5.1] in a simpler setting). Thus, if we want to estimate the distance from a given point $q \in X$ to A, namely to estimate $\mathrm{INF}_f(A)$ for $f(x) := d(q, x)$, $x \in X$ (this is a uniformly continuous function), then we can use the above-mentioned method since it is known [64, Corollary 5.3] that if

A_k denotes a component of the "inner tuple" in iteration number k corresponding to A, $k \in \mathbb{N}$ (or the union of the components if A itself is the union of components of the double zone diagram), then $\lim_{k \to \infty} D_H(A, A_k) = 0$.

As a second example, consider the problem of minimizing a continuous function f over a finite-dimensional Euclidean space, where the constraint set A is induced by a finite system of convex inequalities. Assume further that it is known that A is contained in some known (and possibly large) closed ball. In this case f is automatically uniformly continuous on the ball. Estimating A is not always a simple task, but in [16] one can find a method which does exactly this. More precisely, it produces, in finitely many steps, an inner and outer polytopial approximations $Q'_\epsilon \subseteq A \subseteq Q''_\epsilon$ to A having Hausdorff distance from each other (and hence also from A) which is not greater than a known tolerance parameter ϵ. In particular, by letting $\epsilon_k := 1/k$ and $A_k := Q'_{\epsilon_k}$ for each $k \in \mathbb{N}$, we have $\lim_{k \to \infty} D_H(A_k, A) = 0$, and hence we can use the above-mentioned method in order to estimate $\mathrm{INF}_f(A)$.

As a final remark in this section, we note that the idea of estimating the optimal value of some function over a given constraint set A by estimating it over an approximating set A_k and taking the limit $k \to \infty$ appears in other works, such as [43, p. 367] (in a very intuitive and brief form), in [47] (the setting there is a finite-dimensional Euclidean space and the approximation is with respect to inner and outer limits of compact sets), and in [62, Subsections 1.1 and 1.2] (in the setting of an interval/box, or, more generally, in the setting of a compact metric space).

Acknowledgments

References

[1] A. Ambrosetti and G. Prodi, *A Primer of Nonlinear Analysis*, Cambridge University Press, New York, USA, 1993.

[2] D. T. V. An and N. D. Yen, *Subdifferential stability analysis for convex optimization problems via multiplier sets*, Vietnam J. Math. **46** (2018), 365–379.

[3] M. Anthony and J. Ratsaby, *Large-width bounds for learning half-spaces on distance spaces*, Discrete Appl. Math. **243** (2018), 73–89.

[4] Z. Artstein, *Sensitivity with respect to the underlying information in stochastic programs*, J. Comput. Appl. Math. **56** (1994), 127–136.

[5] Z. Artstein and R. J.-B. Wets, *Stability results for stochastic programs and sensors, allowing for discontinuous objective functions*, SIAM J. Optim. **4** (1994), 537–550.

[6] T. Asano, J. Matoušek and T. Tokuyama, *Zone diagrams: existence, uniqueness, and algorithmic challenge*, SIAM J. Comput. **37** (2007), 1182–1198.

[7] H. Attouch and R. J.-B. Wets, *Quantitative stability of variational systems. III. ε-approximate solutions*, Math. Program. (Ser. A) **61** (1993), 197–214.

[8] B. Bank J. Guddat, D. Klatte, B. Kummer and K. Tammer, *Non-Linear Parametric Optimization*, Birkhäuser Verlag, Basel-Boston, Mass., 1983.

[9] M. Basseville, *Divergence measures for statistical data processing – an annotated bibliography*, Signal Processing **93** (2013), 621–633.

[10] A. Ben-Israel, Adi Ben-Israel's Bibliography on Generalized Inverses, 2001. This is an online list: `https://www.math.technion.ac.il/iic/GI.html` . Last updated: June 20, 2001, with 1670 bibliographic items (Retrieved date: May 29, 2019).

[11] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, 2 ed., CMS Books in Mathematics, Springer, New York, 2003.

[12] C. Berge, *Topological Spaces: Including a Treatment of Multi-Valued Functions, Vector Spaces, and Convexity*, Macmillan, New York, 1963. Translation of: "Espaces Topologiques: Fonctions Multivoques", 1959.

[13] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research. Springer-Verlag, New York, 2000.

[14] J. M. Borwein, *Stability and regular points of inequality systems*, J. Optim. Theory Appl. **48** (1986), 9–52.

[15] L. M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, Comput. Math. Math. Phys. **7** (1967), 200–217.

[16] D. Butnariu and Y. Censor, *A method for approximating the solution set of a system of convex inequalities by polytopes*, J. Comput. Appl. Math. **36** (1991), 289–304.

[17] S. L. Campbell and C. D. Meyer, Jr., *Generalized inverses of linear transformations*, Dover Publications, Inc., New York, 1991.

[18] J. W. C. Cassels, *An Introduction to the Geometry of Numbers*, Classics in mathematics. Springer, Berlin-New York, 1997 (reprint of the 1971 edition).

[19] C. Cobzaş, *Functional Analysis in Asymmetric Normed Spaces*, Frontiers in Mathematics. Birkhäuser/Springer Basel AG, Basel, 2013.

[20] J. W. Daniel, *On perturbations in systems of linear inequalities*, SIAM J. Numer. Anal. **10** (1973), 299–307.

[21] A. Daniilidis, M. A. Goberna, M.A. López and R. Lucchetti, *Stability in linear optimization under perturbations of the left-hand side coefficients*, Set-Valued Var. Anal. **23** (2015), 737–758.

[22] G. B. Dantzig, J. Folkman and N. Shapiro, *On the continuity of the minimum sets of a continuous function*, J. Math. Anal. Appl. **17** (1967), 519–548.

[23] S. Dempe and P. Mehlitz, *Lipschitz continuity of the optimal value function in parametric optimization*, J. Global Optim. **61** (2015), 363–377.

[24] M. M. Deza and E. Deza, *Encyclopedia of Distances*, fourth ed, Springer, Berlin, 2016.

[25] N. Dinh, M. A. Goberna and M. A. López, *On the stability of the feasible set in optimization problems*, SIAM J. Optim. **20** (2010), 2254–2280.

[26] N. Dinh, M. A. Goberna and M. A. López, *On the stability of the optimal value and the optimal set in optimization problems*, J. Convex Anal. **19** (2012), 927–953.

[27] D. S. Djordjević and V. Rakočević, *Lectures on Generalized Inverses*, University of Niš, Faculty of Sciences and Mathematics, Niš, 2008.

[28] S. Dolecki, *Semicontinuity in constrained optimization. II*, Control Cybernet. **7** (1978), 51–68.

[29] A. V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Mathematics in Science and Engineering, vol. 165, Academic Press, Inc., Orlando, FL, 1983.

[30] A. V. Fiacco and Y. Ishizuka, *Sensitivity and stability analysis for nonlinear programming*, Ann. Oper. Res. **27** (1990), 215–235.

[31] R. M. Gray, *Entropy and Information Theory: First Edition, Corrected*, Springer-Verlag, New York, NY, USA, 2013.

[32] C. W. Groetsch, *Generalized Inverses of Linear Operators: Representation and Approximation*, Marcel Dekker, Inc., New York-Basel, 1977.

[33] P. M. Gruber and C. G. Lekkerkerker, *Geometry of Numbers*, second ed, North Holland, 1987.

[34] A. J. Hoffman, *On approximate solutions of systems of linear inequalities*, J. Research National Bureau Standards **49** (1952), 263–265.

[35] Z. Kadelburg and S. Radenović, *On generalized metric spaces: a survey*, TWMS J. Pure Appl. Math. **5** (2014), 3–13.

[36] N. Kalton, Quasi-Banach spaces, in: Handbook of the Geometry of Banach Spaces, vol. 2, W. B. Johnson and J. Lindenstrauss, Eds. North-Holland, Amsterdam, 2003, pp. 1099–1130.

[37] M. A. Khamsi, *Generalized metric spaces: a survey*, J. Fixed Point Theory Appl. **17** (2015), 455–475.

[38] W. Kirk and N. Shahzad, *Fixed Point Theory in Distance Spaces*, Springer, Cham, 2014.

[39] D. Klatte and B. Kummer, *Stability properties of infima and optimal solutions of parametric optimization problems*, in: Nondifferentiable optimization: motivations and applications (Sopron, 1984), Lecture Notes in Econom. and Math. Systems, vol. 255, Springer, Berlin, 1985, pp. 215–229.

[40] J. Kogan, *Introduction to Clustering Large and High–Dimensional Data*, Cambridge University Press, New York, NY, USA, 2007.

[41] S. Köknar-Tezel and L. J. Latecki, *Improving SVM classification on imbalanced data sets in distance spaces*, in: 2009 Ninth IEEE International Conference on Data Mining (ICDM 2009), 2009, pp. 259–267.

[42] W. Krabs, *k Zur stetigen abhängigkeit des extremalwertes eines konvexen optimierungsproblems von einer stetigen änderung des problems*, ZAMM - J. Appl. Math. Mech./ Zeitschrift für Angewandte Mathematik und Mechanik **52** (1972), 359–368.

[43] B. Kummer, *Global stability of optimization problems*, Mathematische Operationsforschung und Statistik. Series Optimization **8** (1977), 367–383.

[44] B. Kummer, *Globale stabilität quadratischer optimierungsprobleme*, Wiss. Z. Humboldt-Univ., Math.-Nat. R. **XXVI** (1977), 565–569.

[45] G. M. Lee and S. Phạm, *Stability and genericity for semi-algebraic compact programs*, J. Optim. Theory Appl. **169** (2016), 473–495.

[46] G. M. Lee, N. N. Tam and N. D. Yen, *Stability of linear-quadratic minimization over Euclidean balls*, SIAM J. Optim. **22** (2012), 936–952.

[47] A. Lent and Y. Censor, *The primal-dual algorithm as a constraint-set-manipulation device*, Math. Program. (Ser. A) **50** (1991), 343–357.

[48] D. T. Luc and M. Volle, *On epsilon-stability in optimization*, Vietnam J. Math. **46** (2018), 149–167.

[49] B. Luderer, L. Minchenko and T. Satsura, *Multivalued Analysis and Nonlinear Programming Problems with Perturbations*, Nonconvex Optimization and its Applications, vol. 66., Kluwer Academic Publishers, Dordrecht, 2002.

[50] K. Mahler, *On lattice points in n-dimensional star bodies I. Existence theorems*, Proc. Roy. Soc. Lond. A **187** (1946), 151–187.

[51] H. Minkowski, *Gesammelte Abhandlungen / von Hermann Minkowski ; unter Mitwirkung von Andreas Speiser und Hermann Weyl ; hrsg. von David Hilbert.* New York : Chelsea, 1967.

[52] E. H. Moore, *On the reciprocal of the general algebraic matrix*, Bull. Amer. Math. Soc. **26** (1920), 394–395.

[53] L. J. Mordell, *On the geometry of numbers in some non-convex regions*, Proc. London Math. Soc. **48** (1945), 339–390.

[54] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation I: Basic theory*, vol. 330 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2006.

[55] B. S. Mordukhovich, N. M. Nam and N. D. Yen, *Subgradients of marginal functions in parametric mathematical programming*, Math. Program. (Ser. B) **116** (2009), 369–396.

[56] M. Moussaoui and A. Seeger, *Sensitivity analysis of optimal value functions of convex parametric programs with possibly empty solution sets*, SIAM J. Optim. **4** (1994), 659–675.

[57] M. Z. Nashed, *Inner, outer, and generalized inverses in Banach and Hilbert spaces*, Numer. Func. Anal. Optim. **9** (1987), 261–325.

[58] D. Pallaschke and S. Rolewicz, *Foundations of Mathematical Optimization: Convex Analysis Without Linearity*, Mathematics and its Applications, vol. 388, Kluwer Academic Publishers Group, Dordrecht, 1997.

[59] R. Penrose, *A generalized inverse for matrices*, Proc. Cambridge Philos. Soc. **51** (1955), 406–413.

[60] H. X. Phu and N. D. Yen, *On the stability of solutions to quadratic programming problems*, Math. Program. (Ser. A) **89** (2001), 385–394.

[61] I. Pinelis, *A necessary and sufficient condition on the stability of the infimum of convex functions*, J. Convex Anal. **26** (2019), 77–87.

[62] D. Reem, *Fun with "Analysis I": basic theorems in calculus revisited*, arXiv:0709.4492 [math.HO] (2007; current version: [v8], 30 Jan 2019).

[63] D. Reem, *The Bregman distance without the Bregman function II*, in: Optimization Theory and Related Topics, Contemp. Math. (Amer. Math. Soc., Providence, RI), S. Reich and A. J. Zaslavski, Eds., vol. 568. 2012, pp. 213–223.

[64] D. Reem, *On the computation of zone and double zone diagrams*, Discrete & Computational Geometry **59** (2018), 253–292.

[65] D. Reem and S. Reich, *Zone and double zone diagrams in abstract spaces*, Colloquium Mathematicum **115** (2009), 129–145.

[66] D. Reem and S. Reich, *Fixed points of polarity type operators*, J. Math. Anal. Appl. **467** (2018), 1208–1232.

[67] D. Reem and S. Reich, *Solutions to inexact resolvent inclusion problems with applications to nonlinear analysis and optimization*, Rend. Circ. Mat. Palermo (2) **67** (2018), 337–371.

[68] D. Reem, S. Reich and A. De Pierro, *Re-examination of Bregman functions and new properties of their divergences*, Optimization **68** (2019), 279–348.

[69] D. Reem, S. Reich and A. De Pierro, *A telescopic Bregmanian proximal gradient method without the global Lipschitz continuity assumption*, J. Optim. Theory. Appl. **182** (2019), 851–884.

[70] S. M. Robinson, *Normed convex processes*, Trans. Amer. Math. Soc. **174** (1972), 127–140.

[71] S. M. Robinson, *Local epi-continuity and local optimization*, Math. Programming **37** (1987), 208–222.

[72] S. M. Robinson and R. J.-B. Wets, *Stability in two-stage stochastic programming*, SIAM J. Control Optim. **25** (1987), 1409–1416.

[73] R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, NJ, USA, 1970.

[74] R. T. Rockafellar, *Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming*, Math. Programming Stud. **17** (1982), 28–66.

[75] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 317, Springer-Verlag, Berlin, 1998.

[76] W. Römisch, *Stability of stochastic programming problems*, in: Stochastic Programming, A. Ruszczyński and A. Shapiro (Eds.), Handbooks Oper. Res. Management Sci., vol. 10, Elsevier Sci. B. V., Amsterdam, 2003, pp. 483–554.

[77] W. Römisch and R. J.-B. Wets, *Stability of ϵ-approximate solutions to convex stochastic programs*, SIAM J. Optim. **18** (2007), 961–979.

[78] S.-Q. Shan Y. Han and N.-J. Huang, *Upper semicontinuity of solution mappings to parametric generalized vector quasiequilibrium problems*, J. Funct. Spaces (2015), Art. ID 764187, 6.

[79] N. N. Tam and T. V. Nghi, *On the solution existence and stability of quadratically constrained nonconvex quadratic programs*, Optim. Lett. **12** (2018), 1045–1063.

[80] L. Tamássy, *Relation between metric spaces and Finsler spaces* Differential Geom. Appl. **26** (2008), 483–494.

[81] L. Thibault, *On subdifferentials of optimal value functions*, SIAM J. Control Optim. **29** (1991), 1019–1036.

[82] A. N. Tihonov, *Stability of a problem of optimization of functionals*, Ž. Vyčisl. Mat. i Mat. Fiz. **6** (1966), 631–634.

[83] J. van Tiel, *Convex Analysis: An Introductory Text*, John Wiley and Sons, Universities Press, Belfast, Northern Ireland, 1984.

[84] S. Vogel, *A stochastic approach to stability in stochastic programming*, J. Comput. Appl. Math. **56** (1994), 65–96.

[85] S. Vogel, *Qualitative stability of stochastic programs with applications in asymptotic statistics*, Statist. Decisions **23** (2005), 219–248.

[86] D. W. Walkup and R. J.-B. Wets, *A Lipschitzian characterization of convex polyhedra*, Proc. Amer. Math. Soc. **23** (1969), 167–173.

[87] D. W. Walkup and R. J.-B. Wets, *Some practical regularity conditions for nonlinear programs*, SIAM J. Control **7** (1969), 430–436.

[88] G. Wang, Y. Wei and S. Qiao, *Generalized Inverses: Theory and Computations*, Developments in Mathematics, second ed., vol. 53, Springer, Singapore; Science Press Beijing, Beijing, 2018.

[89] R. J.-B. Wets, *Lipschitz continuity of inf-projections*, Comput. Optim. Appl. **25** (2003), 269–282.

[90] Y. Xue and J. Cao, *Perturbation analysis of bounded homogeneous generalized inverses on Banach spaces*, Acta Mathematica Universitatis Comenianae **83** (2014), 181–194.

[91] N. D. Yen, Stability of the solution set of perturbed nonsmooth inequality systems and application, J. Optim. Theory Appl. **93** (1997), 199–225.

[92] A. J. Zaslavski, *Optimization on Metric and Normed Spaces*, Springer Optimization and Its Applications, vol. 44, Springer, New York, 2010.

[93] A. J. Zaslavski, *Nonconvex Optimal Control and Variational Problems*, Springer Optimization and Its Applications, vol. 82, Springer, New York, 2013.

[94] X. Y. Zheng and K. F. Ng, Hoffman's least error bounds for systems of linear inequalities. J. Global Optim. **30** (2004), 391–403.

[95] C. Zlobec, R. Gardner and A. Ben-Israel, *Regions of stability for arbitrarily perturbed convex programs*, in: Mathematical Programming with Data Perturbations, I (Washington, D.C., 1979), Lecture Notes in Pure and Appl. Math., vol. 73, Dekker, New York, 1982, pp. 69–89.

DANIEL REEM

Department of Mathematics, The Technion – Israel Institute of Technology, 3200003 Haifa, Israel.
E-mail address: `dream@technion.ac.il`

SIMEON REICH

Department of Mathematics, The Technion – Israel Institute of Technology, 3200003 Haifa, Israel.
E-mail address: `sreich@technion.ac.il`

ALVARO DE PIERRO

CNPq, Brazil
E-mail address: `depierro.alvaro@gmail.com`