SIAM J. OPTIMIZATION
Vol. 6, No. 2, pp. 265–282, May 1996

001

A UNIFIED ANALYSIS OF HOFFMAN'S BOUND VIA FENCHEL DUALITY*

JAMES V. BURKE† AND PAUL TSENG†

Abstract. In 1952, Hoffman showed that the distance from any point to the solution set of a linear system is bounded above by a constant times the norm of the residual. Subsequently, this bound has been studied extensively and has found many practical uses. In this paper, we consider an extension of Hoffman's bound to a partially infinite-dimensional setting in which the norm defining the distance is replaced by a positively homogeneous convex function and the linear system is replaced by a convex inclusion of the form $Ax - a \in K$, where A is a continuous linear operator from some normed linear space into R^m, a is an element of the range space of A, and K is a nonempty closed convex cone in R^m. When specialized to the finite-dimensional case, we unify and extend many existing results on Hoffman's bound. Our analysis is based on the use of Fenchel duality to express the distance as the supremum of a certain concave function over a bounded subset of the polar of K. Much of our analysis also extends to the case where K is a nonempty closed convex set, not only a convex cone.

Key words. Fenchel duality, error bound, linear system, positively homogeneous convex function, convex cone, semidefinite programming

AMS subject classifications. 49M39, 90C25, 90C31, 90C34, 90C48, 49K40

1. Introduction.

In 1952, A. Hoffman [14] published a very interesting result showing that the distance, measured in some l_p-norm ($p \geq 1$), from any point x to the solution set of a linear system

$$A\xi \leq a,$$

where A is a real matrix and a is a real vector, is bounded above by some constant (depending on the matrix A only) times the norm of the residual

$$[Ax - a]_+,$$

where $[x]_+$ denotes the positive part of x. A local version of this bound was discovered slightly earlier by Rosenbloom [36]. Why is Hoffman's bound interesting? First, many iterative methods for solving linear systems have the property that they decrease the residual to zero. Hoffman's bound guarantees that the iterates generated by these methods are approaching the solution set. Second, Hoffman's bound is a key to the sensitivity analysis of linear/integer programs (see [9], [31], [33]), the computation of local error bounds (see [25], [31]), and the convergence analysis of descent methods for linearly constrained minimization (see [10], [11], [16], [25], [37]–[39]).

In the original work of Hoffman [14], an explicit formula for the constant was also given for the cases $p = 1, 2, \infty$. In [36], this was done for $p = 2$. Subsequently, much research has focused on sharpening this constant and on extending the bound to l_p spaces for other values of p. First Robinson [31], using LP duality, derived a new estimate of the constant for the case $p = 2$. Mangasarian [26] and Mangasarian and Shiau [29], also using LP duality, derived simpler estimates for the case $p = \infty$ (although their argument readily extends to the case $p = 1$). An estimate similar to that of [29] was also obtained by Cook et al. [9] in the context of integer programming

* Received by the editors March 26, 1993; accepted for publication (in revised form) November 1, 1994. This research was supported by National Science Foundation grants DMS-9303772 and CCR-9103804.

† Department of Mathematics, University of Washington, Seattle, WA 98195.

although, as was pointed out in [29, Remark 2.5], the latter estimate is weaker than that given in [29] (also see [18]). More recently, Hoffman's bound has been refined and extended to the case of arbitrary p by Li [22] and by Güler, Hoffman, and Rothblum [13], to the case of an infinite system of linear inequalities by Hu and Wang [15], and to the infinite-dimensional case under the l_∞-norm by Bergthaller and Singer [3]. In particular, a sharp estimate of the constant was given in [22] which, as was remarked in [22], can also be inferred from the analysis in [3]. A simplified proof of Hoffman's original bound was given by Güler [12] and, very recently, Klatte and Thiere [20] showed that the sharp estimate of the constant given in [22] is equal to the estimate given in [19] and in [31]. Finally, we note that Hoffman's bound has also been extended to convex programs and systems of convex inequalities [2], [7], [27], [32], to systems of analytic equalities/inequalities [24], and to the extremal solutions of linear programs [21]. However, such extensions are beyond the scope of the present paper and will not be treated any further.

In this paper, we consider the general setting in which A is a continuous linear operator from some real normed linear space X to another real normed linear space Y, a is an element of Y, and "$\leq$" denotes a partial ordering on Y induced by some nonempty closed convex cone K in Y. (Extensions of our results to the case where K is a convex set are discussed in §6.) Special attention will be given to the partially infinite case in which Y is finite-dimensional (see §§3–5). Here, we identify the normed space Y with the pair $(\mathbf{R}^m, \|\cdot\|)$, where $\|\cdot\|$ is a given norm on $\mathbf{R}^m$. The partially infinite case arises in a number of emerging applications, including constrained interpolation, spectral estimation, and semi-infinite linear programming (see [5] and references therein). Applications to the case in which the cone K is not polyhedral arise in the study of *positive semidefinite programming* [1], [6], [30].

Our two main contributions are (1) to unify and extend previous results on Hoffman's bound (for a linear system) to the partially infinite case and (2) to extend the theory of Hoffman bounds to closed convex sets and, in particular, the case of nonpolyhedral convex cones such as the cone of real positive semidefinite matrices. In particular, we derive general conditions under which the distance, measured in the norm on X, from any point x in X to the solution set of the convex cone inclusion

$$A\xi - a \in K$$

is bounded above by some constant (depending on A and K only) times the distance, measured in the norm on Y, from $Ax - a$ to K. We give an estimate of the constant in this bound which, in the case where X is an l_p (L_p) space for some $1 < p < \infty$, is as sharp as existing estimates. We also show that, in many cases, the norm used to measure the distance on X can be replaced by a gauge function [34, §15], i.e., a nonnegative positively homogeneous convex function. These results unify and extend existing results on Hoffman's bound (see [3], [9], [13], [22], [29]) which consider only the case in which K is the Cartesian product of closed intervals (see §5).

There are two key elements to our analysis. First, we use a version of Fenchel's duality theorem for convex programs with generalized constraints to express the distance from x to the solution set of $A\xi - a \in K$ as the supremum of a certain concave function over K°, the polar cone of K (see (1)). Second, we construct a cone $W \subset K^\circ$ that is large enough so that the preceding supremum is unchanged when K° is replaced by W (see (2)) and yet is small enough so that $W \cap (A^*)^{-1}\mathrm{B}^\circ$ is bounded, where B° denotes the dual ball in Y^*. Then, by observing that this supremum is unchanged when W is further intersected with $(A^*)^{-1}\mathrm{B}^\circ$, the desired bound readily follows (see §2 for a detailed argument). We note that there are typically several possible choices for the cone W and that different choices of W yield bounds that may be very different in nature. Moreover, for some choices of W, the set $W \cap (A^*)^{-1}\mathrm{B}^\circ$ is intimately related to the extreme points of $K^\circ \cap (A^*)^{-1}\mathrm{B}^\circ$ (see Lemma 5).

Following [34] and [35], we adopt the following notation throughout. We denote by R^m the m-dimensional real space. All vectors in R^m are column vectors and the superscript T denotes transpose. For any subspace S of R^m, we denote by $S^\perp$ the orthogonal complement of S. For any real normed linear space X, we denote by $\|\cdot\|$ the norm on X and by X^* the vector space of continuous linear functionals on X. For each $x \in X$ and $x^* \in X^*$, we denote by $\langle x^*, x\rangle$ the value of the function x^* at x (so the bilinear form $\langle \cdot, \cdot \rangle$ is a pairing of X^* and X). It is well known that X^* is a Banach space when endowed with the dual norm $\|x^*\|_* = \sup_{\|x\|\le 1} \langle x^*, x\rangle$ defined for all $x^* \in X^*$. We will use B to denote the unit ball on X (i.e., $\mathrm{B} = \{x \in X : \|x\| \le 1\}$) and use B° to denote the dual unit ball on X^* (i.e., $\mathrm{B}^\circ = \{x^* \in X^* : \|x^*\|_* \le 1\}$). For each nonempty closed convex set C in X, we let ψ_C denote the indicator function for C, i.e.,

$$\psi_C(x) := \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{else,} \end{cases}$$

and let

$$\operatorname{dist}(x \mid C) := \inf_{\xi \in C} \|x - \xi\|$$

for all $x \in X$. The recession cone of C, denoted $\operatorname{rec}(C)$, is the set of directions y such that $C + y \subset C$. When X is finite-dimensional, we denote by $\operatorname{ri} C$ the relative interior of C.

In this paper we simultaneously deal with two normed linear spaces X and Y and, for simplicity, we use $\|\cdot\|$ to denote the norm on either X or Y and use $\|\cdot\|_*$ to denote the dual norm on either X^* or Y^*. Analogously, we use $\langle \cdot, \cdot \rangle$ to denote the bilinear form on either $X \times X^*$ or $Y \times Y^*$, use B to denote the unit ball on either X or Y, and use B° to denote the dual unit ball on either X^* or Y^*, etc. For any convex function $f\colon X \mapsto \mathrm{R} \cup \{+\infty\}$, we denote by f^* the conjugate of f, i.e.,

$$f^*(x^*) = \sup_x \{\langle x^*, x\rangle - f(x)\}$$

for all $x^* \in X^*$. The effective domain of the function f, denoted $\operatorname{dom}(f)$, is the set of points $x \in X$ for which $f(x) < +\infty$. For any nonempty closed convex cone K in Y, we denote by $\operatorname{lin}(K) = K \cap (-K)$ the lineality of K, by $\operatorname{span}(K)$ the linear span of the elements of K, and by K° the polar of K, i.e.,

$$K^\circ = \{\, y^* \in Y^* : \langle y^*, y\rangle \le 0 \text{ for all } y \in K \,\}.$$

For any linear operator A from X to Y, we denote by $A^*\colon Y^* \mapsto X^*$ the adjoint of A. We also denote by $\operatorname{ran}(A)$ the range of A and by $\ker A^*$ the kernal of A^*.

2. Extensions of Hoffman's bound. Let X and Y be two real normed linear spaces. Let f be a real-valued convex function on X whose conjugate f^* is nonnegative everywhere, i.e., $f^*(x^*) \ge 0$ for all $x^* \in X^*$. (We can think of f as a generalization of the norm function on X.) Let K be a nonempty closed convex cone in Y and let A be a linear operator from X to Y. Our goal is to determine conditions under which the "generalized distance" from a point $x \in X$ to the solution set of the convex cone inclusion $A\xi - a \in K$, given by

$$\inf_{A\xi - a \in K} f(x-\xi),$$

is bounded above by some constant times

$$\operatorname{dist}(Ax - a \mid K)$$

uniformly in x and in a. In the case where $X = \mathrm{R}^n$, $Y = \mathrm{R}^m$ for some m and n, f is the norm function on X, and K is the nonpositive orthant in Y, this bound reduces to Hoffman's original bound.

In our analysis, we use a version of Fenchel's duality theorem and certain properties of convex sets to express the above generalized distance as the supremum of the dual norm over a bounded subset of K°. More precisely, we first use a version of Fenchel's duality theorem to express the "generalized distance" from x to the solution set of $A\xi - a \in K$ as the supremum of a concave function over K°:

$$\inf_{A\xi - a \in K} f(x-\xi) = \sup_{y^* \in K^\circ} \{\langle y^*, Ax - a\rangle - f^*(A^* y^*)\}\ . \tag{1}$$

Then we find a cone $W \subset K^\circ$ (independent of x) such that the above supremum is unchanged when K° is replaced by W, i.e.,

$$\sup_{y^* \in K^\circ} \{\langle y^*, Ax - a\rangle - f^*(A^* y^*)\} = \sup_{y^* \in W} \{\langle y^*, Ax - a\rangle - f^*(A^* y^*)\}. \tag{2}$$

Finally, we use the nonnegativity of f^* and a basic property of the dual norm to bound the right-hand side of (2) by $\operatorname{dist}(Ax - a \mid K)$ times the constant

$$\sup_{y^* \in W \cap (A^*)^{-1} D} \|y^*\|_*\, ,$$

where D is the "generalized dual unit ball" given by $D = \{\ x^* \in X^* : f^*(x^*) \le 1\ \}$. Thus the two relations (1) and (2) are key to extending Hoffman's bound to our problem setting. To ensure that the bound is not trivial (i.e., the above constant is not equal to ∞), we will further choose W so that

$$W \cap (A^*)^{-1} D \text{ is bounded.}$$

(Thus, if $K^\circ \cap (A^*)^{-1} D$ is bounded, we can simply choose K° to be W.) We defer the discussion of when the strong duality relation (1) holds to §3 and the discussion of when the desired cone W exists to §§4 and 5.

We show below that if (1) and (2) hold, we immediately obtain a nontrivial extension of Hoffman's bound to our problem setting (also see [12, p. 3] for a related derivation for linear systems). We will refine this extended bound shortly (see Theorem 2).

THEOREM 1. *Let X and Y be real normed linear spaces. Let f be a real-valued convex function on X whose conjugate f^* is nonnegative everywhere, let K be a nonempty closed convex cone in Y, let A be a continuous linear operator from X to Y, let $a \in \operatorname{ran}(A) - K$, and let $x \in X$. Suppose that* (1) *holds and there is a cone $W \subset K^\circ$ satisfying* (2). *Then*

$$\inf_{A\xi - a \in K} f(x-\xi) \le \mu_1 \operatorname{dist}(Ax - a \mid K)\ , \tag{3}$$

where $\mu_1 \in \{-\infty\} \cup [0, \infty]$ *is given by*

$$\mu_1 := \sup_{\substack{y^* \in W \\ A^* y^* \in \operatorname{dom}(f^*)}} \|y^*\|_* \; . \tag{4}$$

(We use the convention that $\infty \cdot 0 = (-\infty) \cdot 0 = 0$ *and* $\mu_1 = -\infty$ *if the supremum in* (4) *is taken over an empty set.)*

Proof. Using (1) and (2), we obtain

$$\begin{aligned} \inf_{A\xi - a \in K} f(x - \xi) &= \sup_{y^* \in W} \{\langle y^*, Ax - a\rangle - f^*(A^* y^*)\} \\ &\le \sup_{\substack{y^* \in W \\ A^* y^* \in \operatorname{dom}(f^*)}} \langle y^*, Ax - a\rangle \\ &\le \sup_{\substack{y^* \in W \\ A^* y^* \in \operatorname{dom}(f^*)}} \|y^*\|_* \operatorname{dist}(Ax - a \mid K) \\ &= \mu_1 \operatorname{dist}(Ax - a \mid K), \end{aligned}$$

where the first inequality follows from the nonnegativity of f^* and the second inequality follows from the inequality

$$\langle y^*, y\rangle \le \|y^*\|_* \operatorname{dist}(y \mid K) \qquad \forall y \in Y, y^* \in K^\circ, \tag{5}$$

which in turns follows from the observation (cf. the definition of the dual norm and the polar cone) that $\langle y^*, y\rangle \le \langle y^*, y - y'\rangle \le \|y^*\|_* \|y - y'\|$ for all $y \in Y$, $y^* \in K^\circ$, $y' \in K$. (In fact, it follows from the minimum norm duality theorem of functional analysis (see [23, §5.13]) that equality in (5) is attained for some $y^* \in K^\circ$ or, equivalently, that $\operatorname{dist}(y \mid K) = (\psi_{B^\circ \cap K^\circ})^*(y)$ (see [8, Thm. 3.1]).) $\square$

In some literature, the bound (3) is stated with the right-hand quantity $\operatorname{dist}(Ax - a \mid K)$ replaced by $\operatorname{dist}(\bar{a} - a \mid K)$ or by $\|\bar{a} - a\|$, where $\bar{a}$ is any element of $Ax - K$. However, these modified bounds are equivalent to (3) in the sense that one holds if and only if the others hold. To see this, note that for any $\bar{a} \in Ax - K$ we have

$$\begin{aligned} \operatorname{dist}(Ax - a \mid K) &\le \operatorname{dist}(Ax - \bar{a} \mid K) + \operatorname{dist}(\bar{a} - a \mid K) \\ &= \operatorname{dist}(\bar{a} - a \mid K) \le \|\bar{a} - a\|, \end{aligned}$$

with equality holding throughout when $\bar{a}$ is an element of $Ax - K$ nearest to a in the norm on Y.

In many cases, we can choose the cone W so that in addition

$$A^* W \ne \{0\} \quad \text{whenever} \quad W \ne \{0\} \tag{6}$$

(see Lemma 5). This has the advantage that when f is positively homogeneous, the bound given in Theorem 1 can be further refined, as we show in the next two theorems.

First, we consider the case where f is real-valued positively homogeneous convex of degree 1 and positive except at the origin. We refine the bound given in Theorem 1 by applying the theorem to this case and then showing that the supremum in (4) can be taken over a smaller set. In what follows, we denote by f° the polar of f on X^* [34, p. 128], i.e.,

$$f^\circ(x^*) := \sup_{x \ne 0} \frac{\langle x^*, x\rangle}{f(x)}.$$

THEOREM 2. *Let X, Y, K, A, a, and x be as in Theorem 1 and let f be a real-valued positively homogeneous convex function on X of degree 1 that is positive except at the origin. Suppose that* (1) *holds and there is a cone $W \subset K^\circ$ satisfying* (2)*. Then*

$$\inf_{A\xi - a \in K} f(x-\xi) \le \mu_2 \operatorname{dist}(Ax - a \mid K), \tag{7}$$

where $\mu_2 \in \{-\infty\} \cup [0,\infty]$ is given by

$$\mu_2 := \sup_{\substack{y^* \in W \\ f^\circ(A^* y^*) \le 1}} \|y^*\|_* \ . \tag{8}$$

(We use the convention that $\infty \cdot 0 = (-\infty) \cdot 0 = 0$ and $\mu_2 = -\infty$ if the supremum in (8) *is taken over an empty set.) If in addition W satisfies* (6)*, then the inequality sign in* (8) *can be replaced by an equality sign.*

Proof. Since $f(0) = 0$ so that $f^*(x^*) \ge 0$ for all $x^* \in X^*$, we have, upon applying Theorem 1, that (3) holds with μ_1 given by (4). Since f is positively homogeneous of degree 1 so that $f^*(x^*) < \infty$ if and only if $f^\circ(x^*) \le 1$, we see that $\mu_2 = \mu_1$ and hence (7) holds (with μ_2 given by (8)). Now, suppose that W also satisfies (6), and we will show that the inequality sign in (8) can be replaced by an equality sign.

First, consider the case where $W \neq \{0\}$ and $A^*y^* \neq 0$ for all nonzero $y^* \in W$. Then the supremum in (8) is unchanged when y^* is further restricted to satisfy $A^*y^* \neq 0$. Since W is a cone and, as is easily checked, f° is positively homogeneous of degree 1 and positive except at the origin, multiplying any such y^* by $1/f^\circ(A^*y^*)$ would yield a new y^* in W that satisfies $f^\circ(A^*y^*) = 1$ and whose dual norm is no less than before. This shows that the supremum in (8) is unchanged when the inequality sign is replaced by an equality sign.

Second, consider the case where $W \neq \{0\}$ and $A^*y^* = 0$ for some nonzero $y^* \in W$. Then the supremum in (8) is taken over an unbounded set, so the supremum equals ∞. Suppose that we replace the inequality sign in (8) by an equality sign. Then, since $A^*W \neq \{0\}$ by (6), the supremum in (8) would be taken over a nonempty set which, by the existence of y^*, is unbounded. Thus, the supremum in (8) would still equal ∞.

Third, consider the case where $W = \{0\}$. Then, since f is positive except at the origin so that $f^*(0) = 0$, we have that the right-hand side of (2) equals zero, and so (1) and (2) imply that the left-hand side of (7) equals zero. Since f is zero only at the origin, we have $Ax - a \in K$, implying $\operatorname{dist}(Ax - a \mid K) = 0$, so both sides of (7) equal zero (under the convention $(-\infty) \cdot 0 = 0$) regardless of whether the inequality sign in (8) is replaced by an equality sign. □

Next, we consider the case where f is real-valued positively homogeneous convex of degree $1 < p < \infty$ and positive except at the origin. We will refine the bound given in Theorem 1 by suitably modifying the proof of this theorem and of Theorem 2. This refined bound is closely related to the bound given by (7) when X is an ℓ_p space and will be used in §4 to unify existing results on Hoffman's bound.

THEOREM 3. *Let X, Y, K, A, a, and x be as in Theorem 1 and let f be a real-valued positively homogeneous convex function on X of degree $p \in (1,\infty)$ that is positive except at the origin. Suppose that* (1) *holds and there is a cone $W \subset K^\circ$ satisfying* (2) *and* (6)*. Then*

$$\inf_{A\xi - a \in K} f(x-\xi)^{\frac{1}{p}} \le \mu_3 \operatorname{dist}(Ax - a \mid K), \tag{9}$$

where $\mu_3 \in \{-\infty\} \cup [0, \infty]$ is given by

$$\mu_3 := \left(\frac{1}{p}\right)^{\frac{1}{p}} \left(\frac{1}{q}\right)^{\frac{1}{q}} \sup_{\substack{y^* \in W \\ f^*(A^*y^*)=1}} \|y^*\|_* \tag{10}$$

with $q \in (1, \infty)$ satisfying $1/p + 1/q = 1$. (We use the convention that $\infty \cdot 0 = (-\infty) \cdot 0 = 0$ and $\mu_3 = -\infty$ if the supremum in (10) is taken over an empty set.)

Proof. We divide the proof into three cases as in the proof of Theorem 2. First, suppose that $W \neq \{0\}$ and $A^*y^* \neq 0$ for all nonzero $y^* \in W$. Since f is real-valued and positively homogeneous convex of degree $p > 1$, we also have that $f^*(x^*) > 0$ for all $x^* \neq 0$. (To see this, note that $x^* \neq 0$ implies the existence of an $x \in X$ with $\langle x^*, x\rangle > 0$. Then $\langle x^*, \theta x\rangle - f(\theta x) = \theta \langle x^*, x\rangle - \theta^p f(x) > 0$ for all $\theta > 0$ sufficiently small.) Then, it follows that $f^*(A^*y^*) > 0$ for all nonzero $y^* \in W$. By combining (1) with (2) and then using the preceding observation, we obtain the following chain of equalities and inequalities:

$$\begin{aligned}
\inf_{A\xi - a \in K} f(x - \xi) &= \sup_{y^* \in W} \{\langle y^*, Ax - a\rangle - f^*(A^*y^*)\} \\
&\leq \sup_{\substack{y^* \in W \\ f^*(A^*y^*)<\infty}} \{\|y^*\|_* \operatorname{dist}(Ax - a \mid K) - f^*(A^*y^*)\} \\
&= \sup_{\substack{y' \in W \\ f^*(A^*y')=1}} \left\{ \sup_{\substack{\theta \geq 0 \\ y^* = \theta y'}} \{\|y^*\|_* \operatorname{dist}(Ax - a \mid K) - f^*(A^*y^*)\} \right\} \\
&= \sup_{\substack{y' \in W \\ f^*(A^*y')=1}} \left\{ \sup_{\theta \geq 0} \{ \theta \|y'\|_* \operatorname{dist}(Ax - a \mid K) - \theta^q \} \right\},
\end{aligned}$$

where the inequality follows from (5). The second equality follows from the observations that W is a cone which is not simply the origin and that $0 < f^*(A^*y^*)$ for all nonzero $y^* \in W$. The final equality follows from the fact that f^* is positively homogeneous of degree q. Now, for any positive number t, the supremum of the function $\theta \mapsto \theta t - \theta^q$ (also using $q > 1$) is attained at $\theta = (t/q)^{1/(q-1)}$ with a value of

$$\frac{q-1}{q} \left(\frac{1}{q}\right)^{1/(q-1)} t^{q/(q-1)} = \frac{1}{p} \left(\frac{1}{q}\right)^{p/q} t^p ,$$

where the equality follows from the fact that $p = q/(q-1)$. Using this in the above relation and then taking the pth root of both sides yields (9) with μ_3 given by (10).

Second, suppose that $W \neq \{0\}$ and $A^*y^* = 0$ for some nonzero $y^* \in W$. Then, since $A^*W \neq \{0\}$ by (6), the supremum in (10) is taken over a nonempty set which, by the existence of y^*, is unbounded. Thus, $\mu_3 = \infty$ and so (9) holds trivially.

Third, suppose that $W = \{0\}$. Then since f^* is positively homogeneous so that $f^*(0) = 0$, we have that the right-hand side of (2) equals zero and so (1) and (2) imply that the left-hand side of (9) equals zero. Since f is zero only at the origin, we have $Ax - a \in K$, implying that $\operatorname{dist}(Ax - a \mid K) = 0$ and hence both sides of (9) equal zero (under the convention $(-\infty) \cdot 0 = 0$). □

In the case where X is an ℓ_p (L_p) space with $1 < p < \infty$, we can apply (7) with μ_2 given by (8) and f taken to be the norm function on X, for which we have $f^\circ(x^*) = \|x^*\|_*$. We can also apply (9) with μ_3 given by (10) and f taken to be the norm function on X raised to the pth power, for which we have

$$f^*(x^*) = \left(\frac{1}{p}\right)^{\frac{q}{p}} \left(\frac{1}{q}\right) \|x^*\|_*^q .$$

Then note that $\mu_2 = \mu_3$, so the two bounds (7) and (9) are in fact equivalent in this case.

Although the bounds given in Theorems 1, 2, and 3 are attractive for their generality and simplicity, their utility is still in question since we have yet to identify conditions under which the strong duality relation (1) holds and to find a cone $W \subset K^\circ$ such that (2) and (6) hold with the corresponding constants μ_1, μ_2, μ_3 finite. We will address these and related issues in the remaining sections.

3. Checking for strong duality. In this section we assume that $Y = \mathrm{R}^m$ for some m and we identify a constraint qualification under which the strong duality relation (1) holds.

LEMMA 4. *Let f be a real-valued convex function on a real normed linear space X, let A be a continuous linear operator from X to R^m for some m, and let K be a nonempty closed convex cone in R^m of the form $K = K_1 \cap K_2$ for some closed convex polyhedral set K_1 and some closed convex set K_2. Then, for every $a \in \operatorname{ran}(A) - (K_1 \cap \operatorname{ri} K_2)$ and every $x \in X$, the strong duality relation* (1) *holds.*

Proof. Fix any $a \in \operatorname{ran}(A) - (K_1 \cap \operatorname{ri} K_2)$ and any $x \in X$. Upon applying a recent result of Borwein and Lewis [4, Cor. 4.6] to the primal problem

$$\inf_{A\xi \in K+a} \{f(x-\xi) + \psi_{K+a}(A\xi)\}$$

and using the observation that $(\psi_{K+a})^*$ is the pointwise sum of $(\psi_K)^* = \psi_{K^\circ}$ and the linear function $y^* \mapsto \langle y^*, a\rangle$, it readily follows that the strong duality relation (1) holds. $\square$

The above duality result and its relatives are extremely powerful tools in convex analysis. Interested readers are referred to [4, 17, 34, 35] for further discussions of their applications.

4. Choosing the cone W: General case. In this section we assume that $Y = \mathrm{R}^m$ for some m, and we propose a choice for the cone $W \subset K^\circ$ and identify conditions under which (2) and (6) hold with the corresponding constants μ_2, μ_3 finite. Then, by applying the results of previous sections with this choice of W, we obtain the desired extension of Hoffman's bound to the cone inclusion case.

Let K be a nonempty closed convex cone in R^m and let A be a continuous linear operator from the normed linear space X to R^m. Let $S \subset \mathrm{R}^m$ be the subspace

$$S := \ker A^* \cap \operatorname{lin}(K^\circ)\ , \tag{11}$$

define the cone

$$\hat{K} := K + S, \tag{12}$$

and consider the following subset W_1 of $\hat{K}^\circ$ given by

$$\begin{aligned} W_1 &:= \left\{ y^* \in \hat{K}^\circ : \begin{array}{c} \text{There does not exist nonzero } z^* \in \hat{K}^\circ \\ \text{with } A^*z^* = 0 \text{ and } y^* - z^* \in \hat{K}^\circ. \end{array} \right\} \\ &= \hat{K}^\circ \backslash [\hat{K}^\circ + [(\ker A^* \cap \hat{K}^\circ)\backslash\{0\}]], \end{aligned} \tag{13}$$

where we adopt the convention that the Minkowski sum of any set with the empty set is itself the empty set. Roughly speaking, W_1 comprises the minimal elements of $\hat{K}^\circ$ under the partial ordering: $y^* \prec y'$ if $y' \neq y^*$, $A^*y' = A^*y^*$, and $y' - y^* \in \hat{K}^\circ$. It is readily seen that W_1 is a cone (although in general nonconvex) and, in the case where $\ker A^* \cap \hat{K}^\circ \neq \{0\}$, is contained in the boundary of $\hat{K}^\circ$. Also, notice that (see [34, Cor. 16.4.2])

$$\hat{K}^\circ = K^\circ \cap S^\perp,$$

and so $W_1 \subset \hat{K}^\circ \subset S^\perp$. The purpose of restricting W_1 to the subspace $S^\perp$ is to ensure that its intersection with $\ker A^*$ contains no line (since it can be seen that $S^\perp = \operatorname{ran}(A) + \operatorname{span}(K)$), which in turn is needed to ensure that the set over which the supremum in (8) or (10) (with W set to W_1) is taken is bounded.

We state and prove the first main result of this section, showing that the choice $W = W_1$ indeed satisfies (2) and (6), among other things.

LEMMA 5. *Let A be a continuous linear operator from the normed linear space X to R^m. Let K be a nonempty closed convex cone in R^m. Let S, $\hat{K}$, and W_1 be given by, respectively,* (11), (12), *and* (13). *Then the following hold.*

(a) *Let f be a real-valued convex function on X. For every $a \in \operatorname{ran}(A) - K$ and every $x \in X$,*

$$\sup_{y^* \in K^\circ} \{\langle y^*, Ax - a\rangle - f^*(A^*y^*)\} = \sup_{y^* \in W_1} \{\langle y^*, Ax - a\rangle - f^*(A^*y^*)\}. \tag{14}$$

(b) *Let D be any convex set in X^* for which $(A^*)^{-1}D$ is closed and has $\ker A^*$ as its recession cone. Then $W_1 \cap (A^*)^{-1}D$ is contained in the convex hull of the extreme points of the closed convex set*

$$G := \hat{K}^\circ \cap (A^*)^{-1}D.$$

(Thus, $W_1 \cap (A^*)^{-1}D$ is bounded whenever the extreme points of G are bounded, which in particular holds when G is polyhedral or when G is bounded.)

(c) *$W_1 = \{0\}$ if and only if $A^*W_1 = \{0\}$ if and only if $\operatorname{ran}(A) \subset K$.*

Proof. For every $x^* \in A^*\hat{K}^\circ$, let us define the nonempty closed convex set

$$C(x^*) := \{\, y^* \in \hat{K}^\circ : A^*y^* = x^* \,\}.$$

Since

$$\operatorname{lin}(\operatorname{rec}(C(x^*))) = \ker A^* \cap \operatorname{lin}(K^\circ) \cap S^\perp = S \cap S^\perp = \{0\},$$

where $\operatorname{rec}(C(x^*))$ denotes the recession cone of $C(x^*)$, we conclude that $C(x^*)$ contains at least one extreme point [34, Cor. 18.5.3].

We claim that for every $x^* \in A^*\hat{K}^\circ$, the extreme points of $C(x^*)$ are in W_1. To see this, fix any extreme point y of $C(x^*)$. If y were not in W_1, there would exist a nonzero $z \in \hat{K}^\circ$ such that $y - z \in \hat{K}^\circ$ and $A^*z = 0$. Then both $y + z$ and $y - z$ would be in $\hat{K}^\circ$ with $x^* = A^*(y - z) = A^*(y + z)$, implying that both $y + z$ and $y - z$ are in $C(x^*)$. But since $y = (1/2)(y + z) + (1/2)(y - z)$, this would show that y is not an extreme point of $C(x^*)$, which is a contradiction.

(a) Fix any $a \in \operatorname{ran}(A) - K$ and any $x \in X$. Let $\bar{x}$ be any element of X satisfying $A\bar{x} - a \in K$. (Such an $\bar{x}$ exists since $a \in \operatorname{ran}(A) - K$.)

First, we claim that

$$\sup_{y^* \in K^\circ} \{\langle y^*, Ax - a\rangle - f^*(A^*y^*)\} = \sup_{y^* \in \hat{K}^\circ} \{\langle y^*, Ax - a\rangle - f^*(A^*y^*)\}. \tag{15}$$

To see this, consider any $y^* \in K^\circ$. Since $S \subset \mathrm{lin}\,(K^\circ)$ so that $K^\circ = S + [K^\circ \cap S^\perp]$, we can write y^* as $y^* = y_1^* + y_2^*$ for some $y_1^* \in S$ and some $y_2^* \in K^\circ \cap S^\perp$. Consequently,

$$\begin{aligned}\langle y^*, Ax - a\rangle - f^*(A^*y^*) &= \langle y_2^*, Ax - a\rangle - f^*(A^*y_2^*) + \langle y_1^*, -a\rangle \\ &= \langle y_2^*, Ax - a\rangle - f^*(A^*y_2^*)\ ,\end{aligned}$$

where the second equality follows from $-a \in \mathrm{ran}\,(A) + K \subset \mathrm{ran}\,(A) + \mathrm{span}\,(K) = S^\perp$.

Second, we claim that for every $x^* \in A^*\hat{K}^\circ$,

$$\sup_{y^* \in C(x^*)} \langle y^*, Ax - a\rangle = \sup_{y^* \in C(x^*) \cap W_1} \langle y^*, Ax - a\rangle. \tag{16}$$

To see this, observe that for every $y^* \in C(x^*)$ we have

$$\begin{aligned}\langle y^*, Ax - a\rangle &= \langle y^*, A(x - \bar{x})\rangle + \langle y^*, A\bar{x} - a\rangle \\ &\le \langle y^*, A(x - \bar{x})\rangle \\ &= \langle x^*, x - \bar{x}\rangle,\end{aligned}$$

where the inequality follows from the fact that $A\bar{x} - a \in K$ and $y^* \in K^\circ$ (cf. $C(x^*) \subset \hat{K}^\circ \subset K^\circ$). Thus, the supremum on the left-hand side of (16) is finite. Since $y^* \mapsto \langle y^*, Ax - a\rangle$ is linear, this supremum is unchanged when we further restrict y^* to the extreme points of $C(x^*)$. By the claim shown earlier, all extreme points of $C(x^*)$ are in W_1.

Upon using (15) and (16) and the observation that

$$\hat{K}^\circ = \bigcup_{x^* \in A^*\hat{K}^\circ} C(x^*),$$

we obtain

$$\begin{aligned}\sup_{y^* \in K^\circ} \{\langle y^*, Ax - a\rangle - f^*(A^*y^*)\} &= \sup_{y^* \in \hat{K}^\circ} \{\langle y^*, Ax - a\rangle - f^*(A^*y^*)\} \\ &= \sup_{x^* \in A^*\hat{K}^\circ} \left\{ \sup_{y^* \in C(x^*)} \langle y^*, Ax - a\rangle - f^*(x^*) \right\} \\ &= \sup_{x^* \in A^*\hat{K}^\circ} \left\{ \sup_{y^* \in C(x^*) \cap W_1} \langle y^*, Ax - a\rangle - f^*(x^*) \right\} \\ &= \sup_{y^* \in \hat{K}^\circ \cap W_1} \{\langle y^*, Ax - a\rangle - f^*(A^*y^*)\} \\ &= \sup_{y^* \in W_1} \{\langle y^*, Ax - a\rangle - f^*(A^*y^*)\},\end{aligned}$$

where the last equality follows trivially from the fact that $W_1 \subset \hat{K}^\circ$.

(b) If G is the empty set, then so is $W_1 \cap (A^*)^{-1}D$ (since $W_1 \subset \hat{K}^\circ$) and the claim follows trivially. Suppose that $G \neq \emptyset$. Since the lineality of G is the same as that of $C(0)$, which we have already shown to be $\{0\}$, it follows that G is a closed convex set containing no lines. Then, by a fundamental representation theorem for a closed convex set in R^m (see [34, Thm. 18.5]), any point y^* in G may be represented as

$$y^* = y' + z^*$$

for some y' in the convex hull of the extreme points of G and some z^* in the recession cone of G. Since the recession cone of G is precisely $\ker A^* \cap \hat{K}^\circ$, then, whenever y^* is not in the convex hull of the extreme points of G, the z^* in the previous representation of y^* must be nonzero and satisfy $z^* \in \ker A^* \cap \hat{K}^\circ$ and $y^* - z^* \in G \subset \hat{K}^\circ$. This in turn shows that $y^* \notin W_1$.

(c) Clearly if $A^*W_1 \neq \{0\}$, then $W_1 \neq \{0\}$. Conversely, if $A^*W_1 = \{0\}$, it is easily checked that $W_1 = \{0\}$. (Since for any nonzero $y^* \in \hat{K}^\circ$ with $A^*y^* = 0$, we have that $z^* := y^*$ is in $\hat{K}^\circ$ and satisfies $A^*z^* = 0$ and $y^* - z^* \in \hat{K}^\circ$, so $y^* \notin W_1$.)

Since $\hat{K} = K + [\ker A^* \cap \operatorname{lin}(K^\circ)]$ (see (11), (12)), we have

$$\begin{aligned} \operatorname{ran}(A) \not\subset K &\iff \operatorname{ran}(A) \not\subset \hat{K} \\ &\iff A^*\hat{K}^\circ \neq \{0\}. \end{aligned}$$

Hence, if $\operatorname{ran}(A) \not\subset K$, there is a nonzero x^* in $A^*\hat{K}^\circ$ so, by an earlier claim, $C(x^*)$ contains at least one extreme point, say y^*, which is in W_1. Since $A^*y^* = x^* \neq 0$, this shows that $A^*W_1 \neq \{0\}$. Conversely, suppose that $\operatorname{ran}(A) \subset K$ or, equivalently, $A^*\hat{K}^\circ = \{0\}$. Since $W_1 \subset \hat{K}^\circ$, it follows that $A^*W_1 = \{0\}$. □

Remarks.

1. We remark that the converse of Lemma 5(b) also holds in the sense that every extreme point of G is an element of $W_1 \cap (A^*)^{-1}D$. This follows from the observation that every extreme point y^* of G is an extreme point of $C(A^*y^*)$, which in turn, as was argued in the preceding proof, is an element of W_1. Thus, maximizing a concave function over $W_1 \cap (A^*)^{-1}D$ is equivalent to maximizing the same function over the extreme points of G.

2. Part (c) of Lemma 5 provides a simple check of when (6) with $W = W_1$ holds, namely, $\operatorname{ran}(A) \not\subset K$ or, equivalently (since $\operatorname{ran}(A)$ is a subspace), $\operatorname{ran}(A) \not\subset \operatorname{lin}(K)$. Thus whether (6) with $W = W_1$ holds depends on K through $\operatorname{lin}(K)$ only. And if (6) with $W = W_1$ does not hold or, equivalently, $\operatorname{ran}(A) \subset K$, then the inclusion $Ax - a \in K$ has a solution if and only if $-a \in K$, in which case the distance $\operatorname{dist}(Ax - a \mid K)$ equals zero for all $x \in X$.

Using Lemmas 4 and 5, we are now in a position to state conditions on A, K, a, and x under which the hypotheses of Theorems 2 and 3 with $W = W_1$ hold. We also give conditions on A and K under which the associated constants μ_2 and μ_3 are finite. We summarize the results in the next theorem.

THEOREM 6. *Let X, A, K, $\mathbf{R}^m$ be as in Lemma 5 and let W be the cone W_1 given by (11), (12), and (13). Assume that $K = K_1 \cap K_2$ for some polyhedral cone K_1 and some closed convex cone K_2 in $\mathbf{R}^m$. Then the following hold.*

(a) *Let f be a real-valued positively homogeneous convex function on X of degree 1 that is positive except at the origin. For every $a \in \operatorname{ran}(A) - (K_1 \cap \operatorname{ri} K_2)$ and every $x \in X$, the bound (7) with μ_2 given by (8) holds. Moreover, if $\operatorname{ran}(A) \not\subset K$, then the inequality sign in (8) can be replaced by an equality sign, and if either K is polyhedral or $\operatorname{ran}(A) + K = \mathbf{R}^m$, then μ_2 is finite.*

(b) *Let f be a real-valued positively homogeneous convex function on X of degree $p \in (1, \infty)$ that is positive except at the origin. For every $a \in \operatorname{ran}(A) - (K_1 \cap \operatorname{ri} K_2)$ and every $x \in X$, the bound (9) with μ_3 given by (10) holds. Moreover, if $\operatorname{ran}(A) \not\subset K$ and either K is polyhedral or $\operatorname{ran}(A) + K = \mathbf{R}^m$, then μ_3 is finite.*

Proof. We prove part (b) only. Part (a) can be similarly proved by replacing f^*, μ_3 and Theorem 3 in the argument below with, respectively, f°, μ_2 and Theorem 2. Fix any $a \in \operatorname{ran}(A) - (K_1 \cap \operatorname{ri} K_2)$ and any $x \in X$. If $\operatorname{ran}(A) \subset K$, then $-a \in K$ and it follows that $Ax - a \in K$ for all $x \in X$, and so the bound (9) holds trivially (since both sides of (9) equal zero). If $\operatorname{ran}(A) \not\subset K$, then parts (a) and (c) of Lemma 5 yield that (2) and (6) with $W = W_1$ hold. Since (1) holds by Lemma 4, it follows from Theorem 3 that the bound (9) also holds.

Suppose that $\operatorname{ran}(A) \not\subset K$ and either K is polyhedral or $\operatorname{ran}(A) + K = \mathrm{R}^m$. We will show that μ_3 is finite. Since $\operatorname{ran}(A) \not\subset K$, Lemma 5(c) yields that (6) with $W = W_1$ holds and so either μ_3 is finite or $\mu_3 = \infty$. By the choice of f, there exists a scalar $\alpha > 0$ such that

$$\{x^* \in X^* \,:\, f^*(x^*) \le 1\} \subset \alpha \mathrm{B}^\circ. \tag{17}$$

Choose any $x^1, \ldots, x^k$ in X such that $Ax^1, \ldots, Ax^k$ span $\operatorname{ran}(A)$ and normalized so that $\|Ax^1\| = \cdots = \|Ax^k\| = 1$. Let

$$D := \{x^* \in X^* \,:\, -\alpha \le \langle x^*, x^i \rangle \le \alpha,\ i = 1, \ldots, k\}.$$

Then D contains $\alpha\mathrm{B}^\circ$ and the set

$$(A^*)^{-1}D = \{y^* \in \mathrm{R}^m \,:\, -\alpha \le \langle y^*, Ax^i \rangle \le \alpha,\ i = 1, \ldots, k\}$$

is polyhedral with recession cone $\ker A^*$. Let G be as defined in Lemma 5(b). If K is polyhedral, then G is also polyhedral. If $\operatorname{ran}(A) + K = \mathrm{R}^m$ so that $S^\perp = \mathrm{R}^m$, then

$$G = K^\circ \cap (A^*)^{-1}D,$$

which must be bounded since its recession cone is $\ker A^* \cap K^\circ = (\operatorname{ran}(A) + K)^\circ = \{0\}$. Thus, in either case the extreme points of G form a bounded set so, by Lemma 5(b), $W_1 \cap (A^*)^{-1}D$ is bounded. Since D contains $\alpha\mathrm{B}^\circ$, it follows from (17) that $\mu_3 \ne \infty$. $\square$

A few remarks about the condition $\operatorname{ran}(A) + K = \mathrm{R}^m$ are in order. First, this condition implies $S^\perp = \mathrm{R}^m$ and hence $\hat{K} = K + S = K$. Second, this condition is equivalent to the solvability of the inclusion $Ax - a \in K$ for all choices of the vector a. Third, although this condition (and assuming $\operatorname{ran}(A) \not\subset K$) suffices to guarantee the finiteness of μ_2 and of μ_3, it is far from necessary since K being polyhedral also suffices.

5. Choosing the cone W: Cartesian product case. As we had remarked in §1, most previous extensions of Hoffman's bound consider only the case in which $Y = \mathrm{R}^m$ for some m and K is the Cartesian product of closed intervals. Moreover, the constants are typically expressed as the supremum of the dual norm taken over all y^* in $K^\circ \cap (A^*)^{-1}\mathrm{B}^\circ$ such that the the "rows" of A corresponding to the nonzero components of y^* are linearly independent. In this section we consider this Cartesian product case and show that these extensions can alternatively be deduced from Theorem 6, so that Theorem 6 unifies and further extends previous extensions of Hoffman's bound. To facilitate the comparison, we need the following key lemma showing that the aforementioned supremum is unchanged if it is instead taken over all y^* in $W_1 \cap (A^*)^{-1}\mathrm{B}^\circ$.

LEMMA 7. *Let X, A, R^m be as in Lemma 5. Let $A_1, \ldots, A_m$ be the unique elements of X^* satisfying*

$$Ax = (\langle A_1, x \rangle, \ldots, \langle A_m, x \rangle)^T \quad \forall x \in X. \tag{18}$$

Let K be a nonempty closed convex cone in R^m of the form

$$K = I_1 \times \cdots \times I_m, \tag{19}$$

where $I_1, \ldots, I_m$ are closed intervals, each of which is either $(-\infty, 0]$ or $\{0\}$ or $[0, \infty)$. Let $\mathcal{I} = \{ i \in \{1, \ldots, m\} : I_i = \{0\} \}$ and assume that A_i, $i \in \mathcal{I}$, are linearly independent. Let S, $\hat{K}$, and W_1 be given by, respectively, (11), (12), and (13), and let W_2 be the cone given by

$$W_2 := \left\{ y^* \in K^\circ : \begin{array}{c} \textit{The elements } A_i, \textit{ with either } i \in \mathcal{I} \textit{ or the ith} \\ \textit{component of } y^* \textit{ nonzero, are linearly independent} \end{array} \right\}.$$

Then, for any convex set D in X^* such that the right-hand supremum below is finite, we have

$$\sup_{\substack{y^* \in W_2 \\ A^* y^* \in D}} \|y^*\|_* = \sup_{\substack{y^* \in W_1 \\ A^* y^* \in D}} \|y^*\|_* . \tag{20}$$

Proof. Since A_i, $i \in \mathcal{I}$, are linearly independent, it is readily seen that $S^\perp$, which equals $\operatorname{ran}(A) + \operatorname{span}(K)$, is all of R^m. Thus, $\hat{K}^\circ = K^\circ$. Consider any $y^* \in W_1$ with $A^* y^* \in D$. Suppose that y^* is not in W_2, so there exists a nonzero vector $z^* \in \mathrm{R}^m$ with $A^* z^* = 0$ and whose support (i.e., $\{i \in \{1, \ldots, m\} : z_i^* \neq 0\}$) is contained in the union of $\mathcal{I}$ with the support of y^*. Let L be the line segment

$$L := \{y^* + \theta z^* \; : \; \theta \in \mathrm{R}\} \cap K^\circ.$$

Then, for any $y' \in L$, the support of y' is contained in the union of $\mathcal{I}$ with the support of y^*, so y' is in W_1 (since y^* is in W_1 and $\hat{K}^\circ = K^\circ$ is the Cartesian product of closed intervals); moreover, y' satisfies $A^* y' = A^* y^* \in D$. Hence we conclude that L is a subset of $W_1 \cap (A^*)^{-1} D$ which, by the assumption that the right-hand supremum in (20) is finite, is bounded. Then the convex function $y' \mapsto \|y'\|_*$ attains its maximum over L at an endpoint of L. Let y' be such an endpoint. Then $y' \in W_1$, $A^* y' \in D$, $\|y'\|_* \geq \|y^*\|_*$, and, since y' is an endpoint of L, the union of $\mathcal{I}$ with the support of y' is strictly contained in the union of $\mathcal{I}$ with the support of y^*. If y' is not in W_2, then we repeat the above reduction procedure with y' in place of y^*, etc. After at most m repetitions, we would obtain a y^* satisfying in addition $y^* \in W_2$. Since the choice of y^* was arbitrary, the equality (20) must hold. □

Remarks.

1. The cone W_2 has been employed in the literature to replace the cone K° in the reduction (2). However, such a reduction is applicable only in the case where K is the Cartesian product of closed intervals (so that W_2 is defined).

2. When A is expressed in the form (18), its adjoint can be expressed analogously in the form

$$A^* y^* = y_1^* A_1 + \cdots + y_m^* A_m \quad \forall y^* = (y_1^*, \ldots, y_m^*)^T \in \mathrm{R}^m.$$

In the case where $X = \mathrm{R}^n$ for some n and A is represented as a matrix, the A_i's are simply the columns of A^T.

3. Under the preceding assumption that A_i, with $i \in \mathcal{I}$, are linearly independent, we in fact have

$$W_2 \subset W_1.$$

To see this, recall that $S^\perp = \mathrm{R}^m$ in this case, so if there were some $y^* \in W_2$ not in W_1, there would exist a nonzero $z^* \in K^\circ$ with $y^* - z^* \in K^\circ$ and $A^* z^* = 0$. Then the first two relations would imply that the union of $\mathcal{I}$ with the support of z^* is contained in the union of $\mathcal{I}$ with the support of y^*, which together with $A^*z^* = 0$ would show that A_i, with either $i \in \mathcal{I}$ or the ith component of y^* nonzero, are linearly dependent, thus contradicting the hypothesis that y^* is in W_2.

By applying Theorem 6(a) with $f(\cdot) = \|\cdot\|$ and Lemma 7, we have the following extension of Hoffman's bound, which can be readily compared with previous extensions of this bound.

THEOREM 8. *Let X, A, K, R^m be as in Lemma 7. Assume that A and K have the form* (18) *and* (19), *respectively, and $\{A_i : I_i = \{0\}\}$ are linearly independent. Let W_2 be as in Lemma 7. Then, for every $a \in \operatorname{ran}(A) - (K_1 \cap \operatorname{ri} K_2)$ and every $x \in X$, we have*

$$\inf_{A\xi - a \in K} \|x - \xi\| \le \mu_4 \operatorname{dist}(Ax - a \mid K), \tag{21}$$

where $\mu_4 \in \{-\infty\} \cup [0, \infty]$ is given by

$$\mu_4 := \sup_{\substack{y^* \in W_2 \\ \|A^*y^*\|_* = 1}} \|y^*\|_* . \tag{22}$$

Moreover, if $\operatorname{ran}(A) \not\subset K$, then μ_4 is finite.

We note that in the case where A and K have the form (18) and (19), respectively, the assumption that $\{A_i : I_i = \{0\}\}$ be linearly independent can be made without loss of generality. This follows from the observation that an $a = (a_1, \ldots, a_m)^T$ and an $x \in X$ satisfy $Ax - a \in K$ if and only if they satisfy $\langle A_i, x\rangle - a_i \in I_i$ for all $i \notin \mathcal{I}\backslash\mathcal{I}'$, where $\mathcal{I} = \{i \in \{1, \ldots, m\} : I_i = \{0\}\}$ and $\mathcal{I}'$ is any subset of $\mathcal{I}$ such that $\{A_i : i \in \mathcal{I}'\}$ are linearly independent and span the space spanned by $\{A_i : i \in \mathcal{I}\}$.

Below we relate Theorem 8 to existing extensions of Hoffman's bound. First, consider the special case in which

$$X = \mathrm{R}^n, \quad K = (-\infty, 0]^m ,$$

and the norms on X and R^m are, respectively, l_p- and l_r-norms for some $p \in [1, \infty]$ and some $r \in [1, \infty]$. Then the set G defined in Lemma 5(b), with D taken to be the l_∞-ball in R^n, is precisely the set $\sigma(A)$ defined by Güler, Hoffman, and Rothblum [13], and it follows from Lemma 5(b) and the remark following Lemma 5 that μ_4 given by (22) is equal to the constant $K(A)$ given in [13]. Hence the bound (21) is equivalent to that given in [13] (see Theorem 2.1 and Corollary 2.3 therein) for $p = r = \infty$ and is at least as sharp as the latter for all other values of p and r. Second, consider the special case in which

$$X = \mathrm{R}^n, \quad A = \begin{pmatrix} B \\ C \end{pmatrix}, \quad K = (-\infty, 0]^k \times \{0\}^l$$

for some $k \times n$ matrix B and some $l \times n$ matrix C of full row rank (with $k + l = m$). Since K has the form (19) and C has full row rank, we can replace W_1 in the definition of μ_4 (cf. (22)) with W_2 and (21) would still hold. Then it is readily seen that μ_4 is equal to the constant $\alpha_{\mu,\nu}(A, C)$ given by Li [22] for every choice of the norm on X. Hence (21) is equivalent to the bound given in [22] (see Theorem 2.4 therein). (The bound in [22] is stated with $\operatorname{dist}(Ax - a \mid K)$ replaced by $\|\bar{a} - a\|$, where $\bar{a}$ is any element of $Ax - K$. However, as we noted in §2, this change does not affect the sharpness of the bound.) Also, when the norm on X is the l_∞-norm and the norm on R^m is the l_r-norm for some $r \in [1, \infty]$, it may be seen that (21) is equivalent to the bound given by Mangasarian and Shiau [29, Thm. 2.2]. If the norm on R^m is monotone, further simplification of the right-hand side in the bounds is possible (see [26] and [29]). In general, the constant in Hoffman's bound (and its extensions) is quite difficult to compute. Studies of the computational issues are given in [19], [28].

6. Extension to convex set inclusion. For simplicity, we have restricted the preceding discussion to the case of a convex cone inclusion. However, our approach readily extends to the more general case of a convex set inclusion of the form

$$B\xi - b \in C, \tag{23}$$

where B is a continuous linear operator from some real normed linear space X to another real normed linear space Y, b is an element of Y, and C is some nonempty closed convex set in Y. The idea of the extension is to apply a standard lifting trick to recast the set inclusion (23) as a cone inclusion of the kind studied in previous sections, and then to adapt the results of previous sections to the cone inclusion. We illustrate this idea below.

The following two convex cones associated with the set C will play significant roles in our analysis: the cone generated by $\{-1\} \times C$ given by

$$\mathrm{cone}\,(\{-1\} \times C) := \cup_{\lambda \geq 0}(\{-1\} \times C) \cdot \lambda$$

and the recession cone of C denoted by $\mathrm{rec}\,(C)$. A key property of $\mathrm{rec}\,(C)$ is that $\mathrm{rec}\,(C)^\circ$ is the closure of $\mathrm{dom}\,(\psi_C^*)$ (called the *barrier cone* of C) [34, Cor. 14.2.1].

Observe that the set inclusion (23) has the same solution set as the following cone inclusion:

$$A\xi - a \in K,$$

where we let

$$A = \begin{bmatrix} 0 \\ B \end{bmatrix}, \qquad a = \begin{bmatrix} 1 \\ b \end{bmatrix}, \qquad K = \mathrm{cl}\;\mathrm{cone}\,(\{-1\} \times C), \tag{24}$$

where cl denotes the closure. Let the norm on $\mathrm{R} \times Y$ be defined by $\|(\mu, y)\| = |\mu| + \|y\|$. Now, we can apply Theorem 6 directly to the cone inclusion, but the resulting bound turns out to be far from sharp. Instead, we will suitably modify the proof of Theorem 6 to take into account the structures of A, a, and K and thus obtain a much sharper bound.

THEOREM 9. *Let X, B, C be as given above with $Y = \mathrm{R}^m$ for some m. Assume that $C = C_1 \cap C_2$ for some closed convex polyhedral set C_1 and some closed convex set C_2 in R^m. Then for every $b \in \mathrm{ran}\,(B) - (C_1 \cap \mathrm{ri}\, C_2)$ and every $x \in X$, we have*

$$\inf_{B\xi - b \in C} \|x - \xi\| \leq \mu_5 \mathrm{dist}\,(Bx - b \mid C), \tag{25}$$

where $\mu_5 \in \{-\infty\} \cup [0, \infty]$ is given by

$$\mu_5 := \sup_{\substack{(\alpha, y^*) \in W_1 \\ \|B^* y^*\|_* \leq 1}} \|y^*\|_* \tag{26}$$

and W_1 is given by (11), (12), and (13) with A and K defined as in (24). (We use the convention that $\infty \cdot 0 = (-\infty) \cdot 0 = 0$ and $\mu_5 = -\infty$ if the supremum in (26) is taken *over an empty set.) Moreover, if* $\operatorname{ran}(B) \not\subset \operatorname{rec}(C)$, *then the inequality sign in* (26) *can be replaced by an equality sign, and if either C is polyhedral or* $\operatorname{ran}(B)+\operatorname{rec}(C) = \mathrm{R}^m$, *then μ_5 is finite.*

Proof. Fix any $b \in \operatorname{ran}(B) - (C_1 \cap \operatorname{ri} C_2)$ and any $x \in X$. Let a be given by (24) and let $K_1 = \operatorname{cl}\operatorname{cone}(\{-1\} \times C_1)$, $K_2 = \operatorname{cl}\operatorname{cone}(\{-1\} \times C_2)$. It is straightforward to verify that $a \in \operatorname{ran}(A) - (K_1 \cap \operatorname{ri} K_2)$ and hence, by Lemmas 4 and 5(a) (with $f(\cdot) = \|\cdot\|$), we have

$$\begin{aligned}
\inf_{B\xi - b \in C} \|x - \xi\| &= \inf_{A\xi - a \in K} \|x - \xi\| \\
&= \sup_{z^* \in W_1} \{\langle z^*, Ax - a\rangle - \psi_{\mathrm{B}^\circ}(A^* z^*)\} \\
&= \sup_{(\alpha, y^*) \in W_1} \left\{ \left\langle \begin{pmatrix} \alpha \\ y^* \end{pmatrix}, \begin{pmatrix} -1 \\ Bx - b \end{pmatrix} \right\rangle - \psi_{\mathrm{B}^\circ}(B^* y^*) \right\} \\
&= \sup_{\substack{(\alpha, y^*) \in W_1 \\ \|B^* y^*\|_* \le 1}} \{\langle y^*, Bx - b\rangle - \alpha\} \\
&\le \sup_{\substack{(\alpha, y^*) \in W_1 \\ \|B^* y^*\|_* \le 1}} \{\langle y^*, Bx - b\rangle - \psi_C^*(y^*)\} \\
&\le \sup_{\substack{(\alpha, y^*) \in W_1 \\ \|B^* y^*\|_* \le 1}} \|y^*\|_* \operatorname{dist}(Bx - b \mid C)\ ,
\end{aligned}$$

where the first inequality follows from the fact that $W_1 \subset K^\circ = \operatorname{epi}(\psi_C^*)$ so that $(\alpha, y^*) \in W_1$ implies $\alpha \ge \psi_C^*(y^*)$. The last inequality follows from the observation that

$$\begin{aligned}
\psi_C^*(y^*) &= \sup_{y \in C} \langle y^*, y\rangle \\
&= \langle y^*, Bx - b\rangle - \inf_{y \in C} \langle y^*, Bx - b - y\rangle \\
&\ge \langle y^*, Bx - b\rangle - \|y^*\|_* \operatorname{dist}(Bx - b \mid C)
\end{aligned}$$

for all $y^* \in \mathrm{R}^m$.

If $\operatorname{ran}(B) \not\subset \operatorname{rec}(C)$, then $\ker B^* \not\supset \operatorname{rec}(C)^\circ$ and there would exist a $y^* \in \operatorname{rec}(C)^\circ$ such that $B^* y^* \neq 0$. Then, by an argument analogous to the proof of Theorem 2, we obtain that the inequality sign in (26) can be replaced by an equality sign.

If C is polyhedral, then so is K and the finiteness of μ_5 follows from an argument analogous to the proof of Theorem 6. If $\operatorname{ran}(B) + \operatorname{rec}(C) = \mathrm{R}^m$, then $(B^*)^{-1}\mathrm{B}^\circ \cap \operatorname{rec}(C)^\circ$ would be bounded since its recession cone is $\ker B^* \cap \operatorname{rec}(C)^\circ = (\operatorname{ran}(B) + \operatorname{rec}(C))^\circ = \{0\}$. Since $W_1 \subset K^\circ = \operatorname{epi}(\psi_C^*)$ so that $(\alpha, y^*) \in W_1$ implies $y^* \in \operatorname{dom}(\psi_C^*)$, we have that

$$\mu_5 = \sup_{\substack{(\alpha, y^*) \in W_1 \\ \|B^* y^*\|_* \le 1}} \|y^*\|_* \le \sup_{\substack{y^* \in \operatorname{dom}(\psi_C^*) \\ \|B^* y^*\|_* \le 1}} \|y^*\|_* = \sup_{\substack{y^* \in \operatorname{rec}(C)^\circ \\ y^* \in (B^*)^{-1}\mathrm{B}^\circ}} \|y^*\|_* < \infty, \tag{27}$$

where the last equality follows from the fact that $\operatorname{cl}\operatorname{dom}(\psi_C^*) = \operatorname{rec}(C)^\circ$ [34, Cor. 14.2.1]. □

Remarks.

1. Note that if $\operatorname{ran}(B) \subset \operatorname{rec}(C)$, then the inclusion $B\xi - b \in C$ has a solution if and only if $-b \in C$, in which case every $\xi \in X$ is a solution.

2. Under the condition $\operatorname{ran}(B) + \operatorname{rec}(C) = \mathrm{R}^m$, we have

$$\operatorname{ran}(A) + \operatorname{span}(K) = \mathrm{R}^{m+1},$$

in which case S given by (11) reduces to the origin and μ_5 can be more simply written as

$$\mu_5 = \sup_{\substack{y^* \in W_3 \\ \|B^* y^*\|_* \leq 1}} \|y^*\|_* \ ,$$

where

$$W_3 := \left\{ y^* \in \operatorname{rec}(C)^\circ : \begin{array}{c} \text{There does not exist a nonzero } z^* \in \operatorname{rec}(C)^\circ \\ \text{with } B^* z^* = 0 \text{ and } y^* - z^* \in \operatorname{rec}(C)^\circ \end{array} \right\}.$$

Also, under this condition, the right-hand side of (27) is finite and hence can be used to estimate μ_5. In general, the bound (25) still holds if we replace W_1 in (26) by any subset of $\mathrm{R} \times \operatorname{rec}(C)^\circ$ that contains W_1.

3. In the special case where C, C_1, and C_2 are all cones, Theorem 9 reduces to Theorem 6(a) with $f(\cdot) = \|\cdot\|$.

REFERENCES

[1] F. Alizadeh, *Combinatorial Optimization with Interior Point Methods and Semidefinite Matrices*, Ph.D. thesis, Department of Computer Science, University of Minnesota, Minneapolis, MN, 1991.

[2] A. A. Auslender and J.-P. Crouzeix, *Global regularity theorems*, Math. Oper. Res., 13 (1988), pp. 1–11.

[3] C. Bergthaller and I. Singer, *The distance to a polyhedron*, Linear Algebra Appl., 169 (1992), pp. 111–129.

[4] J. M. Borwein and A. S. Lewis, *Partially finite convex programming, Part* I: *Quasi relative interiors and duality theory*, Math. Programming, 57 (1992), pp. 15–48.

[5] ———, *Partially finite convex programming, Part* II: *Explicit lattice models*, Math. Programming, 57 (1992), pp. 15–48.

[6] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in Systems and Control*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1994.

[7] L. H. Brady, *Condition Constants for Solutions of Convex Inequalities*, Ph.D. thesis, Department of Computer Sciences, University of Wisconsin, Madison, WI, 1988.

[8] J. V. Burke and S.-P. Han, *A Gauss–Newton approach to solving generalized inequalities*, Math. Oper. Res., 11 (1986), pp. 632–643.

[9] W. Cook, A. M. H. Gerards, A. Schrijver, and É. Tardos, *Sensitivity results in integer linear programming*, Math. Programming, 34 (1986), pp. 251–264.

[10] J.-L. Goffin, *The relaxation method for solving systems of linear inequalities*, Math. Oper. Res., 5 (1980), pp. 388–414.

[11] O. Güler, *Augmented Lagrangian algorithms for linear programming*, J. Optim. Theory Appl., 75 (1992), pp. 445–470.

[12] ———, *Distance to a convex polyhedron and Hoffman's lemma*, Department of Management Sciences, University of Iowa, Iowa City, IA, July 1991, manuscript.

[13] O. Güler, A. J. Hoffman, and U. R. Rothblum, *Approximations to solutions to systems of linear inequalities*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 688–696.

[14] A. J. Hoffman, *On approximate solutions of systems of linear inequalities*, J. Res. Natl. Bur. Standards, 49 (1952), pp. 263–265.

[15] H. Hu and Q. Wang, *On approximate solutions of infinite systems of linear inequalities*, Linear Algebra Appl., 114/115 (1989), pp. 429–438.

[16] A. N. Iusem and A. R. De Pierro, *On the convergence properties of Hildreth's quadratic programming algorithm*, Math. Programming, 47 (1990), pp. 37–51.

[17] V. Jeyakumar, *Duality and infinite-dimensional optimization*, Nonlinear Anal., 15 (1990), pp. 1111–1122.

[18] D. Klatte, *Eine Bemerkung zur parametrischen quadratischen Optimierung*, in Seminarbericht Nr. 50, Sektion Mathematik, Humboldt-Universität zu Berlin, Berlin, 1983, pp. 174–185.

[19] D. KLATTE AND G. THIERE, *Error bounds for solutions of linear equations and inequalities*, Z. Oper. Res., 41 (1995), pp. 191–214.

[20] ———, *A note on Lipschitz constants for solutions of linear inequalities and equations*, Institut für Operations Research, Universität Zürich, to appear in Linear Algebra Appl., 1996.

[21] W. LI, *Sharp Lipschitz constants for basic optimal solutions and basic feasible solutions of linear programs,* SIAM J. Control Optim., 32 (1994), pp. 140–153.

[22] ———, *The sharp Lipschitz constants for feasible and optimal solutions of a perturbed linear program,* Linear Algebra Appl., 187 (1993), pp. 15–40.

[23] D. G. LUENBERGER, *Optimization by Vector Space Methods*, John Wiley and Sons, New York, 1969.

[24] Z.-Q. LUO AND J.-S. PANG, *Error bounds for analytic systems and their applications*, Math. Programming, 67 (1994), pp. 1–28.

[25] Z.-Q. LUO AND P. TSENG, *On the linear convergence of descent methods for convex essentially smooth minimization,* SIAM J. Control Optim., 30 (1992), pp. 408–425.

[26] O. L. MANGASARIAN, *A condition number for linear inequalities and linear systems*, in Methods of Operations Research 43, Proc. 6th Symposium über Operations Research, Universität Augsburg, September 7–9, 1981, G. Bamberg and O. Opitz, eds., Verlagsgruppe Athenäum/Hain/Scriptor/Hanstein, Konigstein, 1981, pp. 3–15.

[27] ———, *A condition number for differentiable convex inequalities*, Math. Oper. Res., 10 (1985), pp. 175–179.

[28] O. L. MANGASARIAN AND T.-H. SHIAU, *A variable-complexity norm maximization problem,* SIAM J. Alg. Discrete Methods, 7 (1986), pp. 455–461.

[29] ———, *Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems,* SIAM J. Control Optim., 25 (1987), pp. 583–595.

[30] Y. NESTEROV AND A. NEMIROVSKII, *Interior-Point Algorithms in Convex Programming*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1994.

[31] S. M. ROBINSON, *Bounds for error in the solution set of a perturbed linear program,* Linear Algebra Appl., 6 (1973), pp. 69–81.

[32] ———, *An application of error bounds for convex programming in a linear space*, SIAM J. Control, 13 (1975), pp. 271–273.

[33] ———, *A characterization of stability in linear programming,* Oper. Res., 25 (1977), pp. 435–447.

[34] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[35] ———, *Conjugate Duality and Optimization*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1974.

[36] P. C. ROSENBLOOM, *Quelques classes de problèmes extrémaux*, Bull. Soc. Math. France, 79 (1951), pp. 1–58.

[37] R. A. TAPIA, Y. ZHANG, AND Y. YE, *On the convergence of the iteration sequence in primal-dual interior point methods*, Math. Programming, 68 (1995), pp. 141–154.

[38] P. TSENG AND D. P. BERTSEKAS, *On the convergence of the exponential multiplier method for convex programming*, Math. Programming, 60 (1993), pp. 1–19.

[39] P. TSENG AND Z.-Q. LUO, *On the convergence of the affine scaling algorithm*, Math. Programming, 56 (1992), pp. 301–319.