June, 1997

LIDS-P 2393

Research Supported By:

NSF grant 9300494-DMI

A Note on Error Bounds for Convex and Nonconvex Programs

Bertsekas, D.P.

June 1997 **LIDS-P-2393**

A NOTE ON ERROR BOUNDS FOR[1] CONVEX AND NONCONVEX PROGRAMS

by

Dimitri P. Bertsekas[2]

Abstract

Given a single feasible solution x_F and a single infeasible solution x_I of a mathematical program, we provide an upper bound to the optimal dual value. We assume that x_F satisfies a weakened form of the Slater condition. We apply the bound to convex programs and we discuss its relation to Hoffman-like bounds. As a special case, we recover a bound due to Mangasarian [Man97] on the distance of a point to a convex set specified by inequalities.

[1] Research supported by NSF under Grant 9300494-DMI.

[2] Dept. of Electrical Engineering and Computer Science, M.I.T., Cambridge, Mass., 02139.

1. INTRODUCTION

We consider the problem

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; x \in X, \qquad g_j(x) \leq 0, \qquad j = 1, \ldots, r, \end{aligned} \tag{1}$$

where X is a nonempty subset of $\Re^n$, and $f : \Re^n \mapsto \Re$, $g_j : \Re^n \mapsto \Re$ are given functions. We denote by $g(x)$ be the vector of constraint functions

$$g(x) = \big(g_1(x), \ldots, g_r(x)\big),$$

and we write the constraints $g_j(x) \leq 0$ compactly as $g(x) \leq 0$. In our notation, all vectors are column vectors and a prime denotes transposition.

Let f^* and q^* be the optimal primal and dual value, respectively:

$$f^* = \inf_{\substack{x \in X \\ g_j(x) \leq 0,\, j=1,\ldots,r}} f(x), \tag{2}$$

$$q^* = \sup_{\mu \geq 0} q(\mu), \tag{3}$$

where $q : \Re^r \mapsto [-\infty, +\infty)$ is the dual function given by

$$q(\mu) = \inf_{x \in X} \big\{ f(x) + \mu' g(x) \big\}. \tag{4}$$

Throughout the paper, we assume the following:

Assumption 1: We have two vectors x_F and x_I from X such that:

(a) x_F is feasible, i.e., $g(x_F) \leq 0$.

(b) x_F is infeasible, i.e., $g_j(x_I) > 0$ for at least one j. Furthermore, its cost $f(x_I)$ is strictly smaller than the cost $f(x_F)$ of x_F.

We note that by weak duality, we have $q^* \leq f(x_F)$. We will show that the value $f(x_I)$ can be used to improve this upper bound. In particular, we prove the following result in Section 3:

Proposition 1: Under Assumption 1, there holds

$$\frac{q^* - f(x_I)}{f(x_F) - f(x_I)} \leq \frac{\Gamma}{\Gamma + 1}, \tag{5}$$

where

$$\Gamma = \inf\big\{ \gamma \geq 0 \mid g(x_I) \leq -\gamma g(x_F) \big\}. \tag{6}$$

If $\Gamma = \infty$ because there is no $\gamma \geq 0$ such that $g(x_I) \leq -\gamma g(x_F)$, the bound in Eq. (3) reduces to the trivial bound $q^* \leq f(x_F)$.

Note that we have $\Gamma < \infty$ if the Slater condition

$$g_j(x_F) < 0, \qquad \forall\, j = 1, \ldots, r, \tag{7}$$

holds. More generally, we have $\Gamma < \infty$ if and only if the following weakened form of the Slater condition holds:

$$g_j(x_I) \leq 0 \qquad \text{for all } j \text{ with } g_j(x_F) = 0. \tag{8}$$

If the above condition holds, we have

$$\Gamma = \max_{\{j \mid g_j(x_F) < 0\}} \frac{g_j(x_I)}{-g_j(x_F)}. \tag{9}$$

Figure 1 illustrates the idea underlying the bound (5), (6). In the case of a single constraint ($r = 1$) the bound reduces to

$$\frac{q^* - f(x_I)}{f(x_F) - f(x_I)} \leq \frac{\hat{f} - f(x_I)}{f(x_F) - f(x_I)} = \frac{g(x_I)}{g(x_I) - g(x_F)}, \tag{10}$$

where $\hat{f}$ is the point of intersection of the vertical axis of $\Re^2$ with the line segment connecting the vectors $\big(g(x_F), f(x_F)\big)$ and $\big(g(x_I), f(x_I)\big)$. When there are multiple constraints, this line segment can be projected on the 2-dimensional subspace spanned by the vertical axis $(0, 1)$ of $\Re^{r+1}$ and the vector $\big(g(x_I), 0\big)$. The inequality (10) can then be applied on this subspace in a suitably modified form (see the proof in the next section).

Figure 1 also suggests the following slightly stronger version of our bound:

$$\frac{\tilde{f} - f(x_I)}{f(x_F) - f(x_I)} \leq \frac{\Gamma}{\Gamma + 1}, \tag{11}$$

where

$$\tilde{f} = \inf\big\{w \mid (z, w) \in \text{Conv}(A)\big\}, \tag{12}$$

the subset A of $\Re^2$ is given by

$$A = \big\{(z, w) \mid \text{ there exists } x \in X \text{ such that } \; g(x) \leq z, \; f(x) \leq w\big\},$$

and Conv(A) denotes its convex hull. Indeed, we prove this bound in Section 3, and we also prove that

$$\tilde{f} \leq f^*.$$

Furthermore, in the case where X is convex, and f and g_j are convex over X, we have $\tilde{f} = f^*$. We state the corresponding bound as a proposition:

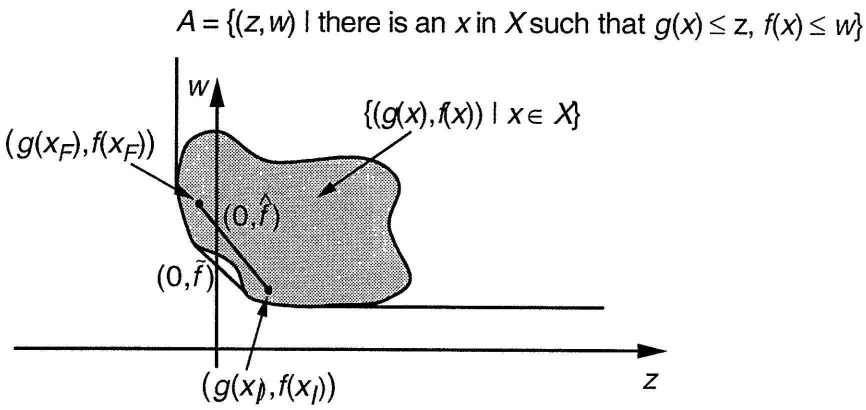

Figure 1: Geometrical interpretation of the bound (8) in the case where there is only one constraint. We consider the convex hull of the subset A of $\Re^2$ given by

$$A = \Big\{(z, w) \mid \text{ there exists } x \in X \text{ such that } \; g(x) \le z, \; f(x) \le w\Big\}.$$

Let $\hat{f}$ be the point of intersection of the vertical axis of $\Re^2$ with the line segment connecting the vectors $\big(g(x_F), f(x_F)\big)$ and $\big(g(x_I), f(x_I)\big)$. The vector $(0, \hat{f})$ belongs to Conv(A). Also, by Euclidean geometry, we have

$$\frac{\hat{f} - f(x_I)}{f(x_F) - f(x_I)} = \frac{g(x_I)}{g(x_I) - g(x_F)},$$

and by the definition of q^* we have

$$q^* \le \tilde{f} \le \hat{f} \le f^*,$$

where

$$\tilde{f} = \inf\Big\{w \mid (z, w) \in \text{Conv}(A)\Big\}.$$

Combining these two relations, the bound (5), (6) follows.

Thus under these convexity assumptions, we have the following:

Proposition 2: In addition to Assumption 1, assume that X is convex, and f and g_j are convex over X. Then, there holds

$$\frac{f^* - f(x_I)}{f(x_F) - f(x_I)} \le \frac{\Gamma}{\Gamma + 1}. \tag{13}$$

2. RELATIONS TO EXISTING BOUNDS

There are several analytical and algorithmic contexts where both feasible and infeasible solutions are known in optimization problems (e.g., in primal-dual algorithms), and in which our bound may prove useful. As an illustration of one such context, let us derive an error bound for the distance of a point to a convex set specified by inequality constraints. A similar error bound for this projection problem was derived by Mangasarian [Man97] using different methods, and was the inspiration for the present paper. In particular, let $y \in \Re^n$ be a given vector and consider the following projection problem

$$\begin{aligned}&\text{minimize} \;\; \|y - x\| \\ &\text{subject to} \;\; x \in X, \qquad g_j(x) \leq 0, \quad j = 1, \ldots, r.\end{aligned}$$

Let us assume that X is a convex set and g_j are convex over X. Furthermore:

(a) $y \in X$ and $g_j(y) > 0$ for at least one j.

(b) There exists a vector $x_F \in X$ such that

$$g_j(x_F) \leq 0, \qquad \forall\, j = 1, \ldots, r,$$

$$g_j(y) \leq 0 \qquad \text{for all } j \text{ with } g_j(x_F) = 0.$$

Then we can apply the error bound (13) with $f(x) = \|y - x\|$, f^* equal to the distance $d(y)$ of y from the convex set $X \cap \{x \mid g_j(x) \leq 0,\ j = 1, \ldots, r\}$, $x_I = y$, and $f(x_I) = 0$, $f(x_F) = \|y - x_F\|$. We have

$$d(y) \leq \frac{\Gamma}{\Gamma + 1}\|y - x_F\|, \tag{14}$$

where Γ is given by Eq. (9). It is easily seen that $\Gamma \leq \hat{\Gamma}$, where

$$\hat{\Gamma} = \frac{\max_{\{j \mid g_j(x_F) < 0\}} g_j(x_I)}{\min_{\{j \mid g_j(x_F) < 0\}} -g_j(x_F)}, \tag{15}$$

and the inequality (14) yields

$$d(y) \leq \frac{\hat{\Gamma}}{\hat{\Gamma} + 1}\|y - x_F\|, \tag{16}$$

or equivalently

$$d(y) \leq \frac{\max_{\{j \mid g_j(x_F) < 0\}} g_j(x_I)}{\max_{\{j \mid g_j(x_F) < 0\}} g_j(x_I) + \min_{\{j \mid g_j(x_F) < 0\}} -g_j(x_F)}\|y - x_F\|. \tag{17}$$

This bound coincides with the relative error bound derived by Mangasarian ([Man97], Theorem 2.2) under the assumption that $X = \Re^n$ and $g_j(x_F) < 0$ for all j. Note, however, that using Γ from Eq. (9) in place of $\hat{\Gamma}$ as in Eq. (16) yields a stronger bound.

For a generalization of the bound (14), let us consider replacing the distance $\|y - x\|$ with a more general metric. In particular, consider the problem

$$\begin{aligned} &\text{minimize} \;\; f(x, y) \\ &\text{subject to} \;\; x \in X, \qquad g_j(x) \leq 0, \quad j = 1, \ldots, r, \end{aligned}$$

where X is a convex set, g_j are convex over X, and $f(\cdot, y)$ is convex over X and satisfies

$$f(y, y) = 0, \qquad f(x_F, y) > 0.$$

Then, if $f^*(y)$ is the optimal cost of this problem, the preceding analysis can be used to show that [cf. Eq. (14)]

$$f^*(y) \leq \frac{\Gamma}{\Gamma + 1} f(x_F, y),$$

where Γ is given by

$$\Gamma = \max_{\{j \mid g_j(x_F) < 0\}} \frac{g_j(y)}{-g_j(x_F)}.$$

We finally note that the bound given here is fundamentally different from the well-known Hoffman's bound [Hof52] and its extensions (see e.g., [LuL94], [LuT92a], [LuT92b], [Man97], [PaL96], [Pan97], which give many additional references). To see this, we note that Propositions 1 and 2 do not explicitly require the existence of a Lagrange multiplier for problem (1). By contrast, as we will show shortly, Hoffman-like bounds essentially amount to assertions on the uniform boundedness of the Lagrange multipliers of some parametric convex program as the parameter vector ranges over some set.

Indeed let X be convex subset of $\Re^n$, let y be a parameter vector taking values in X, and consider the parametric program

$$\begin{aligned} &\text{minimize} \;\; f(x, y) \\ &\text{subject to} \;\; x \in X, \qquad g_j(x, y) \leq 0, \qquad j = 1, \ldots, r, \end{aligned} \tag{18}$$

where for each $y \in X$, $f(\cdot, y)$ and $g_j(\cdot, y)$ are convex over X. We assume that for each $y \in X$, the optimal value $f^*(y)$ of this program is finite and that when the constraints $g_j(x, y) \leq 0$ are dualized, there is no duality gap; that is, the optimal value $q^*(y)$ of the dual problem

$$\begin{aligned} &\text{maximize} \;\; q(\mu, y) \\ &\text{subject to} \;\; \mu \geq 0 \end{aligned} \tag{19}$$

is equal to $f^*(y)$, where $q(\mu, y)$ is the dual function

$$q(\mu, y) = \inf_{x \in X} \{ f(x, y) + \mu' g(x, y) \}.$$

Consider a penalty function $P : \Re^r \mapsto \Re$ that is convex and satisfies

$$P(u) = 0, \qquad \forall\, u \le 0,$$

$$P(u) > 0, \qquad \text{if } u_j > 0 \text{ for some } j = 1, \ldots, r.$$

Let $c > 0$ denote a penalty parameter. It is shown in [Ber95] [Prop. 5.4.1(a)] that we have

$$f^*(y) = \inf_{x \in X} \{ f(x, y) + cP\big(g(x, y)\big) \}$$

if and only if

$$u' \mu^*(y) \le cP(u), \qquad \forall\, u \in \Re^r,$$

for some dual optimal solution $\mu^*(y)$ [an optimal solution of the dual problem (14), which is also referred to as a Lagrange multiplier].

Thus, a bound of the form

$$f^*(y) \le f(y, y) + \bar{c}P\big(g(y, y)\big), \qquad \forall\, y \in X \tag{20}$$

holds if and only if there exists a uniform bounding constant $\bar{c} > 0$ such that

$$u' \mu^*(y) \le \bar{c}P(u), \qquad \forall\, u \in \Re^r,\ y \in X. \tag{21}$$

For the above relation to hold, it is essential that the penalty function P be nondifferentiable, such as for example

$$P(u) = \|u^+\|_1, \qquad P(u) = \|u^+\|_2, \qquad P(u) = \|u^+\|_\infty,$$

where u^+ is the vector with components $\max\{0, u_j\}$, $j = 1, \ldots, r$. Given any of these choices, it is seen that Eq. (21), and the equivalent bound (20), hold if and only if for every $y \in Y$, it is possible to select a Lagrange multiplier $\mu^*(y)$ of the parametric problem (18) such that the set $\{\mu^*(y) \mid y \in X\}$ is bounded.

If we now specialize the preceding discussion to the parametric program

$$\begin{aligned} &\text{minimize} \quad \|y - x\| \\ &\text{subject to} \quad x \in X, \qquad g_j(x) \le 0, \qquad j = 1, \ldots, r, \end{aligned} \tag{22}$$

we see that a bound of the form

$$d(y) \le \bar{c}P\big(g(y)\big), \qquad \forall\, y \in X \tag{23}$$

holds if and only if $\overline{c}$ satisfies Eq. (20). Thus, the Hoffman-like bound (23) holds if and only if the projection problem (22) has a Lagrange multiplier $\mu^*(y)$ such that the set $\{\mu^*(y) \mid y \in X\}$ is bounded. The latter assertion can be made under a number of conditions, such as the following two:

(a) $X = \Re^n$ and g_j are linear (this is the original Hoffman's bound [Hof52]). For a simple way to prove this, let $g_j(x) = a_j'x - b_j$, where a_j is a vector in $\Re^n$ and b_j is a scalar. Then the projection program (22) can be shown to have at least one Lagrange multiplier $\mu(y)$, satisfying

$$\frac{y - \hat{y}}{\|y - \hat{y}\|} = \sum_{j \in J(y)} \mu_j^*(y) a_j,$$

where $\hat{y}$ is the unique projection of y, and $J(y)$ is a subset of indices such that the set of vectors $\{a_j \mid j \in J(y)\}$ is linearly independent. Since the vector in the right-hand side of the above equation has norm 1, it follows that the set $\{\mu^*(y) \mid y \in \Re^n\}$ is bounded. This line of argument can be generalized to the case where, instead of $\|y - x\|$ in Eq. (22), we have a cost function $f(x, y)$, which is differentiable and convex with respect to x, and is such that the set $\{\nabla_x f(y, y) \mid y \in \Re^n\}$ is bounded.

(b) For each $y \in X$, a Slater condition holds; that is there exists a vector $\overline{x}(y) \in X$ such that $g_j\big(\overline{x}(y)\big) < 0$ for all $j = 1, \ldots, r$. Furthermore, there is a constant γ such that

$$\frac{\|y - \overline{x}(y)\| - d(y)}{\min_{j=1,\ldots,r}\big\{-g_j\big(\overline{x}(y)\big)\big\}} \leq \gamma, \qquad \forall\, y \in X. \tag{24}$$

This is Mangasarian's principal result [Man97], who assumed that $X = \Re^n$. For a proof of this result, note that the Slater condition implies (see e.g., [Ber95], p. 450) that for each $y \in X$ there exists a Lagrange multiplier $\mu^*(y)$ with

$$\sum_{j=1}^{r} \mu_j^*(y) \leq \frac{\|y - \overline{x}(y)\| - d(y)}{\min_{j=1,\ldots,r}\big\{-g_j\big(\overline{x}(y)\big)\big\}}. \tag{25}$$

Thus, Eq. (24) implies the boundedness of the set $\{\mu^*(y) \mid y \in X\}$, and hence the existence of a uniform bounding constant $\overline{c}$ in Eq. (23).

3. PROOF OF PROPOSITIONS 1 AND 2

We consider the subset of $\Re^{r+1}$

$$A = \{(z,w) \mid \text{ there exists } x \in X \text{ such that } g(x) \le z,\ f(x) \le w\},$$

and its convex hull $\mathrm{Conv}(A)$. The vectors $\big(g(x_F), f(x_F)\big)$ and $\big(g(x_I), f(x_I)\big)$ belong to A. In addition, the vector $(0, \tilde{f})$, where

$$\tilde{f} = \inf\{w \mid (z,w) \in \mathrm{Conv}(A)\},$$

is in the closure of $\mathrm{Conv}(A)$. Let us now show that $q^* \le \tilde{f}$, as indicated by Fig. 1.

Indeed, for each $(z,w) \in \mathrm{Conv}(A)$, there exist $\xi_1 \ge 0$ and $\xi_2 \ge 0$ with $\xi_1 + \xi_2 = 1$, and $x_1 \in X$, $x_2 \in X$ such that

$$\xi_1 g(x_1) + \xi_2 g(x_2) \le z,$$

$$\xi_1 f(x_1) + \xi_2 f(x_2) \le w.$$

Furthermore, by the definition of the dual function q, we have for all $\mu \in \Re^r$,

$$q(\mu) \le f(x_1) + \mu' g(x_1),$$

$$q(\mu) \le f(x_2) + \mu' g(x_2).$$

Combining the preceding four inequalities, we obtain

$$q(\mu) \le w + \mu' z, \qquad \forall\ (z,w) \in \mathrm{Conv}(A),\ \mu \ge 0.$$

The above inequality holds also for all (z,w) that are in the closure of $\mathrm{Conv}(A)$, and in particular, for $(z,w) = (0,\tilde{f})$. It follows that

$$q(\mu) \le \tilde{f}, \qquad \forall\ \mu \ge 0,$$

from which, by taking the maximum over $\mu \ge 0$, we obtain $q^* \le \tilde{f}$.

Let γ be any nonnegative scalar such that $g(x_I) \le -\gamma g(x_F)$, and consider the vector

$$\Delta = -\gamma g(x_F) - g(x_I).$$

Since $\Delta \ge 0$, it follows that the vector

$$\big(-\gamma g(x_F), f(x_I)\big) = \big(g(x_I) + \Delta, f(x_I)\big)$$

also belongs to the set A. Thus the three vectors

$$\big(g(x_F), f(x_F)\big), \quad (0, \tilde{f}), \quad \big(-\gamma g(x_F), f(x_I)\big)$$

belong to the closure of $\mathrm{Conv}(A)$, and form a triangle in the plane spanned by the "vertical" vector $(0,1)$ and the "horizontal" vector $\big(g(x_F), 0\big)$.

Let $(0, \hat{f})$ be the intersection of the vertical axis with the line segment connecting the vectors $\big(g(x_F), f(x_F)\big)$ and $\big(-\gamma g(x_F), f(x_I)\big)$ (there is a point of intersection because $\gamma \geq 0$). We have by Euclidean triangle geometry (cf. Fig. 1)

$$\frac{\hat{f} - f(x_I)}{f(x_F) - f(x_I)} = \frac{\gamma}{\gamma + 1}. \tag{26}$$

Since the vectors $\big(g(x_F), f(x_F)\big)$ and $\big(-\gamma g(x_F), f(x_I)\big)$ both belong to $\mathrm{Conv}(A)$, we also have $(0, \hat{f}) \in \mathrm{Conv}(A)$. Therefore, there exist vectors $x_1, \ldots, x_m \in X$ and nonnegative scalars $\xi_1, \ldots, \xi_m$ with $\sum_{i=1}^m \xi_i = 1$, satisfying

$$\sum_{i=1}^m \xi_i g(x_i) \leq 0, \qquad \sum_{i=1}^m \xi_i f(x_i) \leq \hat{f}.$$

Thus, if $\hat{f} < \tilde{f}$, we must have $\sum_{i=1}^m \xi_i f(x_i) < \tilde{f}$, contradicting the definition of $\tilde{f}$. It follows that $\tilde{f} \leq \hat{f}$ and since $q^* \leq \tilde{f}$, as shown earlier, from Eq. (26) we obtain

$$\frac{q^* - f(x_I)}{f(x_F) - f(x_I)} \leq \frac{\tilde{f} - f(x_I)}{f(x_F) - f(x_I)} \leq \frac{\gamma}{\gamma + 1}. \tag{27}$$

Taking the infimum over $\gamma \geq 0$, the error bound (5), (6) follows.

Assume now that X is convex, and f and g_j are convex over X. Then the set A is known to be convex under these assumptions (see e.g., [Ber95], Prop. 5.3.1, p. 446), and we have $\tilde{f} = f^*$. Proposition 2 then follows from Eq. (27). **Q.E.D.**

REFERENCES

[Ber95] Bertsekas, D. P., 1995. "Nonlinear Programming," Athena Scientific, Belmont, MA.

[Man97] Mangasarian, O. L., 1997. "Error Bounds for Nondifferentiable Convex Inequalities under a Strong Slater Constraint Qualification," Mathematical Programming Technical Report No. 96-04, Univ. of Wisconsin-Madison.

[Hof52] A. J. Hoffman, 1952. "On Approximate Solutions of Systems of Linear Inequalities," Natl. Bur. Standards, Vol. 49, pp. 263-265.

[LuL94] X.-D. Luo, and Z.-Q. Luo, 1994. "Extension of Hoffman's Error Bound to Polynomial Systems," SIAM J. on Optimization, Vol. 4, pp. 383-392.

[LuT92a] X.-D. Luo, and P. Tseng, 1992. "Error Bound and Convergence Analysis of Matrix Splitting Algorithms for the Affine Variational Inequality Problem," SIAM J. on Optimization, Vol. 2, pp. 43-54.

[LuT92b] X.-D. Luo, and P. Tseng, 1992. "On the Linear Convergence of Convex Essentially Smooth Minimization," SIAM J. on Control and Optimization, Vol. 30, pp. 408-425.

[PaL96] A. S. Lewis and J.-S. Pang, 1996. "Error Bounds for Convex Inequality Systems," working apper, Dept. of Math. Sciences, The Johns Hopkins University, Baltimore, MD; to appear in J.P. Crouzeix, ed., Proceedings of the Fifth International Symposium on Generalized Convexity.

[Pan97] J.-S. Pang, 1997. "Error Bounds in Mathematical Programming," manuscript, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD.