Math. Program. 85: 181–192 (1999) / DOI 10.1007/s10107980036a

William W. Hager · M. Seetharama Gowda

Stability in the presence of degeneracy and error estimation

Received February 10, 1997 / Revised version received June 6, 1998
Published online October 9, 1998

Abstract. Given an approximation to a local minimizer to a nonlinear optimization problem and to associated multipliers, we obtain a tight error estimate in terms of the violation of the first-order conditions. Our results apply to degenerate optimization problems where independence of the active constraint gradients and strict complementarity can be violated.

Key words. stability analysis – perturbation theory – degenerate optimization – error estimation – quadratic program stability – merit functions

1. Introduction

We obtain estimates for the error in an approximation to the solution to an optimization problem. One of our main objectives is to establish error estimates that apply in situations where the Mangasarian-Fromovitz constraint qualification (MFCQ) does not necessarily hold, or where strict complementarity is violated. For a system of equalities and inequalities

$$g(z) \le 0 \quad \text{and} \quad h(z) = 0,$$

where $g : \mathbf{R}^n \to \mathbf{R}^l$ and $h : \mathbf{R}^n \to \mathbf{R}^m$, the Mangasarian-Fromovitz constraint qualification [9] holds at a solution z_* if the constraint gradients $\nabla h_i(z_*)$ are linearly independent and there exists $y \in \mathbf{R}^n$ such that $\nabla g(z_*)y < 0$ and $\nabla h(z_*)y = 0$. It is well known (see [10] or [11]) that when the MFCQ is not satisfied by a linear inequality system, then arbitrary small perturbations can yield an infeasible system. Using a Lipschitz stability result for the solutions of quadratic programming problems, based on Robinson's continuity properties [13] (also see [6] and [7]) for polyhedral multifunctions, we estimate the distance from a given point to the solution of a nonquadratic problem, without imposing the MFCQ. We note that estimates for the error in an approximate solution to a programming problem play an integral role in Wright's [16] recent sequential quadratic programming algorithm for degenerate optimization problems.

In a related paper [4], Facchinei, Fischer, and Kanzow obtain similar upper bounds for the error in an approximate solution to an optimization problem by applying results

W.W. Hager: Department of Mathematics, University of Florida, Gainesville, FL 32611, USA,
e-mail: `hager@math.ufl.edu`, `http://www.math.ufl.edu/~hager`

M. Seetharama Gowda: Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21250, USA,
e-mail: `gowda@math.umbc.edu`, `http://www.math.umbc.edu/~gowda`

Mathematics Subject Classification (1991): 90C31, 90C30, 65G

of Klatte [8], however, their analysis assumes that the MFCQ holds. Also, a result [14, Theorem 4.2] of Robinson can be used to obtain an analogous estimate, assuming the MFCQ. In the approach of [4], one assumes that the usual second-order sufficient optimality condition holds for all multipliers associated with a given local optimum, and one obtains an error estimate at all points in a neighborhood of the optimal solution and the associated multipliers. In our approach, we assume that the second-order condition holds at a given solution/multiplier pair, and we obtain an error estimate in a neighborhood of the given solution/multiplier pair. For polyhedral constraints, the estimate for the solution error is valid in a neighborhood of the solution, without constraints on the choice of the multipliers. In contrast to our work, the lack of smoothness in problem functions is taken into account in [4] by working with generalized Jacobians, while here we assume that the functions are twice differentiable at the point under consideration. In another paper [16], Wright obtains error estimates in the case where the multipliers at the reference point are bounded from zero, again assuming the MFCQ.

Our approach is in the spirit of a result [1, Theorem 3.2] of Dontchev. There, in a nondegenerate setting, he estimates the distance from a given point to a solution of an inclusion by perturbing the inclusion so that the given point becomes a solution. Properties of an associated linearization are used to analyze the effect of the perturbation. When the MFCQ does not hold, special care is needed in the analysis of the linearized problem since it may be infeasible for certain perturbations, and continuity arguments cannot be applied. Our paper concludes by demonstrating the tightness of a particular merit function which is essentially the same one studied in [4]. That is, we show that this merit function yields not only an upper bound for the distance to the solution set, but a lower bound as well.

2. Local uniqueness and quadratic program stability

Our error estimates are based on a result concerning the stability of stationary points for a quadratic program. Initially, let us consider the local uniqueness of stationary points for the following optimization problem:

$$\text{minimize} \quad f(z) \quad \text{subject to} \quad Az \le r, \quad Bz = s, \quad z \in \mathbf{R}^n, \tag{1}$$

where A is $l \times n$ and B is $m \times n$. The first-order optimality conditions associated with (1) at a point z where f is differentiable can be expressed: There exist λ and μ such that

$$\nabla f(z) + \lambda^{\mathsf{T}} A + \mu^{\mathsf{T}} B = 0, \quad Az - r \in N(\lambda), \quad Bz - s = 0, \tag{2}$$

where N is the normal cone:

$$N(\lambda) = \begin{cases} \{\pi : \pi \le 0, \ \ \pi^{\mathsf{T}}\lambda = 0\} \ \text{ if } \ \lambda \ge 0, \\ \ \emptyset \text{ otherwise.} \end{cases}$$

Throughout this paper $\|\cdot\|$ denotes the Euclidean norm, and β denotes a generic positive constant that has different values in different equations.

Proposition 1. *Suppose that f is twice continuously differentiable at a solution (z_*, λ_*, μ_*) to (2) and that there exists a scalar $\alpha > 0$ such that*

$$w^{\mathsf{T}}\nabla^2 f(z_*)w \geq \alpha\|w\|^2 \tag{3}$$

for each w such that $Bw = 0$, $(Aw)_+ = 0$, and $(Aw)_0 \leq 0$, where the $+$ and 0 subscripts are used to denote the subvectors associated with those indices i for which $(\lambda_*)_i > 0$ and $(\lambda_*)_i = 0 = (Az_* - r)_i$, respectively. Then there exists a neighborhood $\mathcal{N}_1$ of z_* with the property that if $z \in \mathcal{N}_1$ and (z, λ, μ) is a solution to (2), then $z = z_*$.

Proof. Our proof exploits the technique of Robinson in [14, Theorem 2.4]. There he considers a nonlinear optimization problem that satisfies the Mangasarian-Fromovitz constraint qualification. Using bounds on the multipliers, he extracts limits and proves a similar local uniqueness result. In our case, we do not assume the Mangasarian-Fromovitz constraint qualification, and there are no bounds on the multipliers.

nstead, the polyhedral structure of the constraints leads to local uniqueness.

The proof is by contradiction. Suppose that there exists a sequence $\{z_k\}$ approaching z_* and associated multipliers such that (z_k, λ_k, μ_k) satisfies (2) for each k, and $(z_k - z_*)/\|z_k - z_*\|$ approaches some unit vector w. Multiplying the first equation in (2), evaluated at (z_k, λ_k, μ_k), by $(z_k - z_*)$, we conclude that

$$\begin{aligned} 0 &= \nabla f(z_k)(z_k - z_*) + \lambda_k^{\mathsf{T}} A(z_k - z_*) \\ &= \nabla f(z_k)(z_k - z_*) + \lambda_k^{\mathsf{T}}(Az_k - a) + \lambda_k^{\mathsf{T}}(a - Az_*). \end{aligned} \tag{4}$$

For any vector y, let y_- denote the subvector associated with those indices i for which $(Az_* - a)_i < 0$. By complementary slackness, $\lambda_k^{\mathsf{T}}(Az_k - a) = 0$, while for k sufficiently large, $(\lambda_k)_- = 0$ since $(Az_* - a)_- < 0$, z_k converges to z_*, and $(\lambda_k)_-^{\mathsf{T}}(Az_k - a)_- = 0$. Consequently,

$$\lambda_k^{\mathsf{T}}(Az_k - a) = 0 = \lambda_k^{\mathsf{T}}(a - Az_*), \tag{5}$$

for k sufficiently large, and (4) implies that $\nabla f(z_k)(z_k - z_*) = 0$ for k sufficiently large. By taking limits, we conclude that $\nabla f(z_*)w = 0$. Since $Az_k \leq a$ and $(Az_* - a)_+ = (Az_* - a)_0 = 0$, it follows, by taking limits, that $(Aw)_+ \leq 0$ and $(Aw)_0 \leq 0$. In a similar manner, $Bw = 0$. Moreover, multiplying (2), evaluated at (z_*, λ_*, μ_*) by w, we deduce that $(Aw)_+ = 0$ since $\nabla f(z_*)w = 0$ and $\lambda_+ > 0$. Hence, w satisfies the constraints associated with (3).

Now, as in [14, Theorem 2.4], we consider the scalar function s defined by

$$s(t) = (\nabla f(z_t) + \lambda_t^{\mathsf{T}} A)(z_k - z_*) + (\lambda_k - \lambda_*)^{\mathsf{T}}(a - Az_t) \tag{6}$$

where $(z_t, \lambda_t) = (1 - t)(z_*, \lambda_*) + t(z_k, \lambda_k)$. The first term in (6) vanishes for $t = 0$ and $t = 1$ by the first-order condition (2). The observation (5) implies that

$$(\lambda_k - \lambda_*)^{\mathsf{T}}(a - Az_*) = 0,$$

for k sufficiently large, while

$$(\lambda_k - \lambda_*)^{\mathsf{T}}(a - Az_k) = -\lambda_*^{\mathsf{T}}(a - Az_k) \leq 0$$

since $Az_k \le a$ and $\lambda_* \ge 0$. Hence, $s(0) = 0 \ge s(1)$, and by the mean value theorem, there exists $t_k \in (0, 1)$ such that

$$0 \ge s'(t_k) = (z_k - z_*)^{\mathsf{T}} \nabla^2 f(z_{t_k})(z_k - z_*).$$

Dividing by $\|z_k - z_*\|^2$ and taking limits, we obtain a contradiction to (3).

□

Now let us consider a quadratic programming problem

$$\text{minimize} \quad \frac{1}{2} z^{\mathsf{T}} Q z - z^{\mathsf{T}} \varphi \quad \text{subject to} \quad Az \le r, \quad Bz = s, \quad z \in \mathbf{R}^n, \tag{7}$$

where Q is an $n \times n$ symmetric matrix and $\varphi \in \mathbf{R}^n$. The first-order conditions can be expressed in the following way:

$$\psi = \begin{pmatrix} \varphi \\ r \\ s \end{pmatrix} \in \mathcal{F}(z, \lambda, \mu), \tag{8}$$

where

$$\mathcal{F}(z, \lambda, \mu) = \begin{bmatrix} Qz + A^{\mathsf{T}}\lambda + B^{\mathsf{T}}\mu \\ Az \\ Bz \end{bmatrix} - \begin{bmatrix} 0 \\ N(\lambda) \\ 0 \end{bmatrix} \quad \text{if} \quad \lambda \ge 0, \tag{9}$$

and $\mathcal{F}(z, \lambda, \mu) = \emptyset$ otherwise. Using Proposition 1, we have

Lemma 1. *Suppose that $(z, \lambda, \mu) = (z_*, \lambda_*, \mu_*)$ is a solution to (8) corresponding to $\psi = \psi_*$, and that there exists a scalar $\alpha > 0$ such that*

$$w^{\mathsf{T}} Q w \ge \alpha \|w\|^2 \tag{10}$$

for each w such that $Bw = 0$, $(Aw)_+ = 0$, and $(Aw)_0 \le 0$, where the $+$ and 0 subscripts are used to denote the subvectors associated with those indices i for which $(\lambda_*)_i > 0$ and $(\lambda_*)_i = 0 = (Az_* - r)_i$, respectively. Then there exist a scalar $\beta > 0$ and neighborhoods $\mathcal{N}_1$ of z_* and $\mathcal{N}_*$ of ψ_* with the property that if (z, λ, μ) is a solution to (8) corresponding to $\psi \in \mathcal{N}_*$ and $z \in \mathcal{N}_1$, then

$$\|z - z_*\| \le \beta \|\psi - \psi_*\|. \tag{11}$$

Proof. The function $\mathcal{F}$ defined in (9) is a polyhedral multifunction since its graph, defined by

$$\operatorname{gr} \mathcal{F} = \{(w, \psi) : \psi \in \mathcal{F}(w)\},$$

is the union of finitely many polyhedral convex sets. Robinson notes in [13] that the set-valued inverse of a polyhedral multifunction is a polyhedral multifunction, and that polyhedral multifunctions are closed under (finite) composition. Letting $\mathcal{P}_1$ be the projection defined by $\mathcal{P}_1(z, \lambda, \mu) = z$, it follows that $\mathcal{P}_1 \circ \mathcal{F}^{-1}$ is a polyhedral multifunction. Using a theorem of Walkup and Wets [15] concerning a Lipschitzian characterization of convex polyhedra, Robinson proves that a polyhedral multifunction is locally upper Lipschitzian at every point, and the Lipschitz constant is independent of the point. Hence, there exists a constant β and an associated neighborhood $\mathcal{N}_*$ of ψ_* such that

$$\mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi) \subset \mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi_*) + \beta\|\psi - \psi_*\|\mathcal{B} \quad \text{for all} \quad \psi \in \mathcal{N}_*, \tag{12}$$

where $\mathcal{B}$ is the unit ball in $\mathbf{R}^n$. Letting $\mathcal{N}_1$ be as in Proposition 1, it follows that since $\mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi)$ is the set of z-components of solutions to (8), we have

$$\mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi_*) \cap \mathcal{N}_1 = \{z_*\}. \tag{13}$$

Choose $\mathcal{N}_1$ smaller if necessary so that it is contained within the ball of center z_* and radius $\rho/3$ where ρ is the distance from z_* to $(\mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi_*)) \setminus \{z_*\}$. Shrink $\mathcal{N}_*$ further if necessary so that

$$\{z_*\} + \beta\|\psi - \psi_*\|\mathcal{B} \subset \mathcal{N}_1 \quad \text{for all} \quad \psi \in \mathcal{N}_*.$$

Now if $\psi \in \mathcal{N}_*$ and $z \in \mathcal{N}_1 \cap \mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi)$, then by (12), (11) holds.

□

Under the hypotheses of Lemma 1, an example given in [12] shows that $\mathcal{N}_1 \cap \mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi)$ may be multivalued for ψ near ψ_*, even though (13) holds; moreover, as ψ approaches ψ_*, some of the elements of $\mathcal{N}_1 \cap \mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi)$ may not be local minimizers of (7).

3. Error estimates

Let us consider the following inequality constrained optimization problem:

$$\text{minimize} \quad f(z) \quad \text{subject to} \quad g(z) \leq 0, \quad h(z) = 0, \quad z \in \mathbf{R}^n, \tag{14}$$

where f is real-valued, $g : \mathbf{R}^n \to \mathbf{R}^l$, and $h : \mathbf{R}^n \to \mathbf{R}^m$. Given $\lambda \in \mathbf{R}^l$ and $\mu \in \mathbf{R}^m$, the Lagrangian $\mathcal{L}$ is defined by

$$\mathcal{L}(z, \lambda, \mu) = f(z) + \lambda^{\mathsf{T}} g(z) + \mu^{\mathsf{T}} h(z).$$

The first-order necessary conditions associated with (14) can be expressed:

$$T(z, \lambda, \mu) \in F(\lambda), \tag{15}$$

where

$$T(z, \lambda, \mu) = \begin{pmatrix} \nabla_z \mathcal{L}(z, \lambda, \mu) \\ g(z) \\ h(z) \end{pmatrix}, \quad F(\lambda) = \begin{pmatrix} 0 \\ N(\lambda) \\ 0 \end{pmatrix}. \tag{16}$$

Let $\mathcal{M}$ denote the set of all multipliers associated with a local minimizer z_* for (14). That is, letting Λ denote the pair (λ, μ), then $\Lambda \in \mathcal{M}$ if and only if $\nabla_z \mathcal{L}(z_*, \Lambda) = 0$ with $\lambda \geq 0$ and $\lambda^{\mathsf{T}} g(z_*) = 0$. Then we have

Lemma 2. *Suppose that f, g, and h are twice differentiable at a local minimizer z_* of (14), $\Lambda_* \in \mathcal{M} \neq \emptyset$, and*

$$w^{\mathsf{T}} \nabla_z^2 \mathcal{L}(z_*, \Lambda_*) w \geq \alpha \|w\|^2 \tag{17}$$

whenever $\nabla h(z_*)w = 0 = \nabla g_+(z_*)w$ and $\nabla g_0(z_*)w \leq 0$, where g_+ and g_0 denote the components of g associated with indices i for which $(\lambda_*)_i > 0$ and $(\lambda_*)_i = 0 = g_i(z_*)$, respectively. Then there exists a neighborhood $\mathcal{N}$ of (z_*, Λ_*) and constants γ and $\delta > 0$ with the property that for every $(z, \Lambda) \in \mathcal{N}$ and for each y with $\|y\| \leq \delta$ and

$$T(z, \Lambda) + y \in F(\Lambda), \tag{18}$$

we have

$$\|z - z_*\| + \|\Lambda - \hat{\Lambda}\| \leq \gamma \|y\|, \tag{19}$$

where $\hat{\Lambda}$ is the closest element of $\mathcal{M}$ to Λ.

Proof. Throughout this proof, the generic constant β is uniformly bounded when $\mathcal{N}$ is sufficiently small. Let us consider the quadratic program (7) in the case that

$$Q = \nabla_z^2 \mathcal{L}(z_*, \Lambda_*), \quad A = \nabla g(z_*), \quad B = \nabla h(z_*).$$

We make two choices for ψ. Our first choice is

$$\psi = \psi_1 = L(z, \Lambda) - T(z, \Lambda) - y,$$

where y satisfies (18), and

$$L(z, \Lambda) = \begin{pmatrix} Qz + A^{\mathsf{T}}\lambda + B^{\mathsf{T}}\mu \\ Az \\ Bz \end{pmatrix}.$$

In this case, $(z, \Lambda) \in \mathcal{F}^{-1}(\psi_1)$. Our second choice is

$$\psi = \psi_2 = L(z_*, \Lambda_2) - T(z_*, \Lambda_2) = \begin{pmatrix} Qz_* - \nabla f(z_*) \\ Az_* - g(z_*) \\ Bz_* - h(z_*) \end{pmatrix},$$

where Λ_2 is arbitrary since the terms containing it cancel. Taking $\Lambda_2 = \Lambda_*$, we see that $(z_*, \Lambda_*) \in \mathcal{F}^{-1}(\psi_2)$. By the differentiability assumption, ψ_1 is close to ψ_2 when (z, Λ) is close to (z_*, Λ_*), and y is close to 0. Consequently, by choosing $\mathcal{N}$ and δ sufficiently small, and by taking $\Lambda_2 = \Lambda$, Lemma 1 gives us the estimate

$$\|z - z_*\| \leq \beta \|T(z, \Lambda) - T(z_*, \Lambda) - (L(z, \Lambda) - L(z_*, \Lambda)) + y\| \tag{20}$$

for all $(z, \Lambda) \in \mathcal{N}$ and $\|y\| \leq \delta$.

Given any $\epsilon > 0$, it follows from the differentiability assumption that for $\mathcal{N}$ sufficiently small,

$$\|T(z, \Lambda) - T(z_*, \Lambda) - L(z - z_*, 0)\| \leq \epsilon \|z - z_*\| \tag{21}$$

for all $(z, \Lambda) \in \mathcal{N}$. Combining this with (20), we have

$$\|z - z_*\| \leq \beta \|y\|. \tag{22}$$

An estimate for the error in the multipliers is gotten from Hoffman's result [5] for the distance to the boundary of a polyhedron. That is, if a linear system of inequalities is feasible, then the distance from any given point to the feasible set is bounded by a constant times the norm of the constraint violation at the given point. Since $\mathcal{M}$ is the solution set for the linear inequality system

$$\nabla_z \mathcal{L}(z_*, \Lambda) = 0, \quad \Lambda = (\lambda, \mu), \quad \lambda \geq 0, \quad \lambda_- = 0,$$

where λ_- denotes the subvector of λ associated with those indices i such that $g_i(z_*) < 0$, the norm of the constraint violation at any given Λ is $\|\nabla_z \mathcal{L}(z_*, \Lambda)\| + \|\lambda_-\|$ (the constraint $\lambda \geq 0$ is satisfied due to (18)). Let us take $\mathcal{N}$ and δ small enough that $g_-(z) + y_- < 0$ when $(z, \Lambda) \in \mathcal{N}$ and $\|y\| \leq \delta$. In this case the complementary slackness condition associated with (18) implies that $\lambda_- = 0$, and Hoffman's result yields

$$\|\Lambda - \hat{\Lambda}\| \leq \beta \|\nabla_z \mathcal{L}(z_*, \Lambda)\|. \tag{23}$$

By the definition of y, by the differentiability assumption, and by choosing $\mathcal{N}$ smaller if necessary, we have

$$\|\nabla_z \mathcal{L}(z_*, \Lambda)\| \leq \|\nabla_z \mathcal{L}(z, \Lambda)\| + \|\nabla_z \mathcal{L}(z_*, \Lambda) - \nabla_z \mathcal{L}(z, \Lambda)\| \leq \|y\| + \beta \|z - z_*\|. \tag{24}$$

Combining (22)–(24), the proof is complete.

□

Since the normal cone $N(\lambda)$ changes discontinuously when a component of λ becomes zero, sharper estimates for the error in (z, λ, μ) may be gotten by estimating the error in an intermediate point. This idea is the motivation for the following result.

Theorem 1. *Suppose the assumptions of Lemma 2 are in effect. Then there exists a neighborhood $\mathcal{N}$ of (z_*, Λ_*) and constants γ and $\delta > 0$ with the property that for every (z, Λ), for each $(\bar{z}, \bar{\Lambda}) \in \mathcal{N}$, and for each y with $\|y\| \leq \delta$ and*

$$T(\bar{z}, \bar{\Lambda}) + y \in F(\bar{\Lambda}), \tag{25}$$

we have

$$\|z - z_*\| + \|\Lambda - \hat{\Lambda}\| \leq \|z - \bar{z}\| + \|\Lambda - \bar{\Lambda}\| + \gamma \|y\|, \tag{26}$$

where $\hat{\Lambda}$ is the closest element of $\mathcal{M}$ to Λ.

Proof. By the triangle inequality and the analysis of Lemma 2, we have

$$\|z - z_*\| \le \|z - \bar{z}\| + \|\bar{z} - z_*\| \le \|z - \bar{z}\| + \gamma\|y\|.$$

In a similar fashion, Lemma 2 yields

$$\begin{aligned}\|\Lambda - \hat{\Lambda}\| &= \inf_{\tilde{\Lambda}\in\mathcal{M}} \|\Lambda - \tilde{\Lambda}\| \\ &\le \|\Lambda - \bar{\Lambda}\| + \inf_{\tilde{\Lambda}\in\mathcal{M}} \|\bar{\Lambda} - \tilde{\Lambda}\| \\ &\le \|\Lambda - \bar{\Lambda}\| + \gamma\|y\|.\end{aligned}$$

This completes the proof.

□

When the functions g and h are affine, the multipliers in (20) and (21) cancel leading to the following special form of Lemma 2 and Theorem 1.

Corollary 1. *Suppose that f is twice differentiable at a local minimizer z_* of (14), g and h are affine, $\Lambda_* \in \mathcal{M}$, and (17) holds whenever $\nabla h(z_*)w = 0 = \nabla g_+(z_*)w$ and $\nabla g_0(z_*)w \le 0$, where g_+ and g_0 denote the components of g associated with indices i for which $(\lambda_*)_i > 0$ and $(\lambda_*)_i = 0 = g_i(z_*)$, respectively. Then there exists a neighborhood $\mathcal{N}_1$ of z_* and constants γ and $\delta > 0$ with the property that for every $z \in \mathcal{N}_1$ and for each Λ and y satisfying (18) with $\|y\| \le \delta$, we have*

$$\|z - z_*\| + \|\Lambda - \hat{\Lambda}\| \le \gamma\|y\|,$$

where $\hat{\Lambda}$ is the closest element of $\mathcal{M}$ to Λ. Moreover, if z is arbitrary, but $\bar{z} \in \mathcal{N}_1$ and $(\bar{z}, \bar{\Lambda})$ satisfies (25), then we have

$$\|z - z_*\| + \|\Lambda - \hat{\Lambda}\| \le \|z - \bar{z}\| + \|\Lambda - \bar{\Lambda}\| + \gamma\|y\|.$$

Proof. As noted above, restrictions on the choice of the multipliers only arise when we make the transition from (20) to (21) in the proof of Lemma 2. However, when g and h are affine, the multiplier terms cancel, and hence, the conclusion of the Lemma 2 and Theorem 1 are valid without constraints on the multipliers.

□

In Corollary 1, we restrict the norm of the perturbation y. Since there are no conditions on Λ in Corollary 1, one approach for removing the constraint $\|y\| \le \delta$ is to choose Λ to minimize the distance from $T(z, \Lambda)$ to $F(\Lambda)$. When z is close to z_*, $T(z, \Lambda_*)$ is close to $F(\Lambda_*)$, and hence, if Λ is chosen to minimize the distance from $T(z, \Lambda)$ to $F(\Lambda)$, then the constraint $\|y\| \le \delta$ is satisfied automatically for z in a neighborhood of z_*. A second approach for removing this restriction on y is to strengthen the form of second-order sufficient condition in order to obtain a quadratic program stability result analogous to Lemma 1, but with the perturbations ψ unrestricted. The following result removes the neighborhoods of Lemma 1. In the proof of [13, Proposition 4], Robinson notes this property in the context of a real valued function, while Klatte in [7, Corollary 5.1] proves this, in a more general context, for the solution to a quadratic programming problem.

Lemma 3. *Suppose that there exists a scalar $\alpha > 0$ such that*

$$w^{\mathsf{T}} Q w \ge \alpha \|w\|^2 \tag{27}$$

for each w such that $Bw = 0$. Then for each $\psi = (\varphi, r, s)$ where r and s are chosen so that the system

$$Az \le r \quad \text{and} \quad Bz = s, \tag{28}$$

is feasible, (7) has a unique minimizer $z(\psi)$. Moreover, there exists a scalar $\beta > 0$ with the property that for every $\psi_1 = (\varphi_1, r_1, s_1)$ and $\psi_2 = (\varphi_2, r_2, s_2)$, where the associated linear systems (28) corresponding to $(r, s) = (r_i, s_i)$, $i = 1,\ 2$, are feasible, we have

$$\|z(\psi_2) - z(\psi_1)\| \le \beta \|\psi_2 - \psi_1\|. \tag{29}$$

Proof. For completeness, we give a short proof of the Klatte/Robinson observation. By [3, Section 4] and (27), there exists a unique solution $z = z(\psi)$ to (7) when (28) is feasible. Moreover, any z that satisfies the first-order optimality conditions for (7) is a strict local minimizer. By Robinson's result [13, Proposition 1], a polyhedral multifunction is locally upper Lipschitzian at every point with Lipschitz constant β independent of the point. That is, for any ψ_0, there exists an associated neighborhood $\mathcal{N}$ for which

$$\mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi) \subset \mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi_0) + \beta \|\psi - \psi_0\| \mathcal{B} \quad \text{for all} \quad \psi \in \mathcal{N}, \tag{30}$$

where $\mathcal{B}$ is the unit ball in $\mathbf{R}^n$.

Now suppose that for $\psi_1 = (\varphi_1, r_1, s_1)$ and $\psi_2 = (\varphi_2, r_2, s_2)$, the associated linear systems (28) corresponding to $(r, s) = (r_i, s_i)$, $i = 1,\ 2$, are feasible. Then for all $\psi \in [\psi_1, \psi_2]$, the line segment connecting ψ_1 and ψ_2, the associated linear systems (28) remain feasible, and hence, the corresponding unique solution $z(\psi)$ exists. Since $\mathcal{P}_1 \circ \mathcal{F}^{-1}(\psi) = z(\psi)$, it follows from (30) that if $\psi_0 \in [\psi_1, \psi_2]$, then for all $\psi \in [\psi_1, \psi_2]$ near ψ_0, we have

$$\|z(\psi) - z(\psi_0)\| \le \beta \|\psi - \psi_0\|. \tag{31}$$

Let $\bar{\psi}$ be the point on the line segment $[\psi_1, \psi_2]$ that is closest to ψ_2 and that satisfies

$$\|z(\bar{\psi}) - z(\psi_1)\| \le \beta \|\bar{\psi} - \psi_1\|. \tag{32}$$

Since $z(\cdot)$ is continuous on $[\psi_1, \psi_2]$ by (31), the point $\bar{\psi}$ exists. Since (31) holds for $\psi_0 = \psi_1$ and for ψ near ψ_1, we see that $\bar{\psi} \ne \psi_1$. We wish to conclude that $\bar{\psi} = \psi_2$. If $\bar{\psi} \ne \psi_2$, then applying (31) with $\psi_0 = \bar{\psi}$, we deduce that

$$\|z(\psi) - z(\bar{\psi})\| \le \beta \|\psi - \bar{\psi}\|$$

for ψ between $\bar{\psi}$ and ψ_2 with ψ near $\bar{\psi}$. Since the triangle inequality becomes an equality for three successive colinear points, we have

$$\begin{aligned} \|z(\psi) - z(\psi_1)\| &\le \|z(\psi) - z(\bar{\psi})\| + \|z(\bar{\psi}) - z(\psi_1)\| \\ &\le \beta \|\psi - \bar{\psi}\| + \beta \|\bar{\psi} - \psi_1\| \\ &= \beta \|\psi - \psi_1\|. \end{aligned}$$

Since ψ can be chosen closer to ψ_2 than $\bar{\psi}$, we have a contradiction. Hence, $\bar{\psi} = \psi_2$ in (32), and the proof is complete.

□

Using Lemma 3 instead of Lemma 1, we have the following analogue of Theorem 1:

Theorem 2. *Suppose that f is twice differentiable at a local minimizer z_* of (14), g and h are affine, $\Lambda_* \in \mathcal{M}$, and for some $\alpha > 0$, (17) holds whenever $\nabla h(z_*)w = 0$. Then there exists a neighborhood $\mathcal{N}_1$ of z_* and a constant γ with the property that for every (z, Λ), for every $(\bar{z}, \bar{\Lambda})$ with $\bar{z} \in \mathcal{N}_1$, and for each y satisfying (25), we have*

$$\|z - z_*\| \le \|z - \bar{z}\| + \gamma\|y\|.$$

Proof. Again, the proof of this result is contained in the proofs of Lemma 2 and Theorem 1, except that Lemma 3 is used in place of Lemma 1.

□

4. An application

As a specific application of Theorem 1, we consider the following choices for the variables: $\bar{z} = z$ and $\bar{\mu} = \mu$, while $\bar{\lambda}$ is defined by

$$\bar{\lambda}_i = \begin{cases} 0 & \text{if } \lambda_i < -g_i(z) \text{ or } \lambda_i < 0, \\ \lambda_i & \text{otherwise.} \end{cases} \tag{33}$$

And $y = (q, r, s)$ where

$$q = -\nabla_z \mathcal{L}(z, \bar{\lambda}, \mu), \quad s = -h(z), \quad r_i = \begin{cases} -g_i(z) & \text{if } \bar{\lambda}_i > 0 \text{ or } g_i(z) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

With these definitions, the inclusion (25) of Theorem 1 is satisfied. This is, $\bar{\lambda} \ge 0$ and if $\bar{\lambda}_i > 0$, then $g_i(z) + r_i = 0$, while q and s are the negatives of the corresponding components of $T(\bar{w})$. The variables r and $\bar{\lambda}$ were chosen in the following way: The multiplier was changed so that the nonnegativity constraint $\bar{\lambda} \ge 0$ was satisfied, and then we perturbed further either the inequality constraint $g(z) \le 0$ or the multiplier making the smallest change so that the complementarity condition $g(z) + r \in N(\bar{\lambda})$ was satisfied. For this special choice of $\bar{w}$ and y, we now show that the right side of (26) not only provides an upper bound for the error in $w = (z, \lambda, \mu)$, but a lower bound as well.

Theorem 3. *If f, g, and h are twice differentiable at a local minimizer z_* of (14) and $\Lambda_* \in \mathcal{M} \ne \emptyset$, then there exists a neighborhood $\mathcal{N}$ of (z_*, Λ_*) and a constant γ with the property that for each $(z, \lambda, \mu) \in \mathcal{N}$, we have*

$$\|\lambda - \bar{\lambda}\| + \|y\| \le \gamma(\|z - z_*\| + \|\Lambda - \hat{\Lambda}\|), \tag{35}$$

where $\bar{\lambda}$ and y are given in (33) and (34), and $\hat{\Lambda}$ is the closest element of $\mathcal{M}$ to Λ.

Proof. The generic constant β used in this proof is uniformly bounded when the neighborhood $\mathcal{N}$ is sufficiently small. Observe that

$$\begin{aligned} \|y\| &\le \|q\| + \|r\| + \|s\| \\ &= \|\nabla_z \mathcal{L}(z, \bar{\lambda}, \mu)\| + \|r\| + \|h(z)\| \\ &\le \beta\|\lambda - \bar{\lambda}\| + \|\nabla_z \mathcal{L}(z, \lambda, \mu)\| + \|r\| + \|h(z)\|. \end{aligned} \tag{36}$$

Since $h(z_*) = 0$, we have $\|h(z)\| \le \beta\|z - z_*\|$. Also, by the differentiability assumption, we have

$$\|\nabla_z \mathcal{L}(z, \Lambda) - \nabla_z \mathcal{L}(z^*, \Lambda)\| \le \beta\|z - z^*\|.$$

Since $\nabla_z \mathcal{L}(z_*, \hat{\Lambda}) = 0$, we have

$$\|\nabla_z \mathcal{L}(z^*, \Lambda)\| = \|\nabla_z \mathcal{L}(z^*, \lambda) - \nabla_z \mathcal{L}(z^*, \hat{\Lambda})\| \le \beta\|\Lambda - \hat{\Lambda}\|.$$

Choose $\mathcal{N}$ small enough that for all $(z, \lambda, \mu) \in \mathcal{N}$, we have $g_i(z) < 0$ and $|\lambda_i| < |g_i(z)|$ whenever $g_i(z_*) < 0$ (and hence, $\lambda_i^* = 0$). By its definition, the components of r are either 0 or $-g_i(z)$. If $r_i = -g_i(z)$, then by (34), either $g_i(z) > 0$ or $\bar{\lambda}_i > 0$. In the former case, $g_i(z_*) = 0$ by the choice of $\mathcal{N}$. That is, $g_i(z) < 0$ for all $(z, \lambda, \mu) \in \mathcal{N}$ when $g_i(z_*) < 0$. Hence, if $g_i(z) > 0$, we must have $g_i(z_*) = 0$. In the latter case ($g_i(z) \le 0$ and $\bar{\lambda}_i > 0$), it follows from (33) that $\lambda_i \ge -g_i(z) \ge 0$. Since $|\lambda_i| \ge |g_i(z)|$, it follows from the choice of $\mathcal{N}$ that $g_i(z_*) = 0$. In summary, if $r_i \ne 0$, then $r_i = g_i(z)$ and $g_i(z_i^*) = 0$. Since g is differentiable, for these nonzero components of r, we have

$$|r_i| = |g_i(z)| \le \beta\|z - z_*\|.$$

Hence, $\|r\| \le \beta\|z - z^*\|$. Combining these estimates for the terms in (36) gives

$$\|y\| \le \beta(\|z - z_*\| + \|\Lambda - \hat{\Lambda}\| + \|\lambda - \bar{\lambda}\|). \tag{37}$$

Also, by the definition of $\bar{\lambda}$, each component of $\lambda - \bar{\lambda}$ is either zero or λ_i, and $(\lambda - \bar{\lambda})_i = \lambda_i \ne 0$ only when either $\lambda_i < 0$, or $0 \le \lambda_i < -g_i(z)$. In the former case,

$$|\lambda_i - \bar{\lambda}_i| = |\lambda_i| \le |\lambda_i - \hat{\lambda}_i|$$

since $\hat{\lambda} \ge 0$. In the latter case, if $g_i(z_*) < 0$, then $\hat{\lambda}_i = 0$ for all $(\hat{\lambda}, \hat{\mu}) \in \mathcal{M}$, and we have

$$|\lambda_i - \bar{\lambda}_i| = |\lambda_i| = |\lambda_i - \hat{\lambda}_i|.$$

On the other hand, if $g_i(z_*) = 0$, then the relation $0 \le \lambda_i < -g_i(z)$ implies that

$$|\lambda_i - \bar{\lambda}_i| = |\lambda_i| \le |g_i(z)| \le \beta\|z - z_*\|.$$

Combining these observations gives

$$\|\lambda - \bar{\lambda}\| \le \|\lambda - \hat{\lambda}\| + \beta\|z - z_*\|.$$

Taking into account (37), we have

$$\|\lambda - \bar{\lambda}\| + \|y\| \le \beta\|z - z_*\| + \|\Lambda - \hat{\Lambda}\|,$$

which completes the proof.

□

By Theorems 1 and 3, the expression

$$\|\lambda - \bar{\lambda}\| + \|r\| + \|h(z)\| + \|\nabla_z \mathcal{L}(z, \bar{\lambda}, \mu)\|$$

where $\bar{\lambda}$ and r are defined in (33) and (34), respectively, tightly measures the error in an approximation (z, λ, μ) to the solution to the optimization problem (14) in the sense that it is bounded from above and below by constants times the true error.

Acknowledgements. This work was supported by the National Science Foundation. We wish to thank the referees along with the participants of the Nineteenth Symposium on Mathematical Programming with Data Perturbations, George Washington University, Washington, DC, May 22-23, 1997, for their comments and suggestions concerning this paper.

References

1. Dontchev, A.L. (1995): Characterizations of Lipschitz stability in optimization. In: Lucchetti, L., Revalski, R., eds., Well-posedness and Stability of Optimization Problems and Related Topics, pp. 95–115. Kluwer
2. Hager, W.W. (1997): Convergence of Wright's stabilized SQP algorithm. Mathematics Department, University of Florida, Gainesville, FL 32611, January, 1997
3. Dontchev, A.L., Hager, W.W. (1993): Lipschitzian stability in nonlinear control and optimization. SIAM J. Control Optim. **31**, 569–603
4. Facchinei, F., Fischer, A., Kanzow, C. (1996): On the accurate identification of active constraints. Technische Universität Dresden, June, 1996
5. Hoffman, A.J. (1952): On approximate solutions of systems of linear inequalities. J. Res. Natl. Bureau Standards **49**, 263–265
6. Klatte, D. (1985): On the Lipschitz behavior of optimal solutions in parametric problems of quadratic optimization and linear complementarity. Optimization **16**, 819–831
7. Klatte, D. (1987): Lipschitz continuity of infima and optimal solutions in parametric optimization: The polyhedral case. In: Guddat, J., Jongen, H.T., Kummer, B., Nožika, F., eds., Parametric Optimization and Related Topics. Adademie-Verlag, Berlin
8. Klatte, D. (1992): Nonlinear optimization problems under data perturbations. In: Krabs, W., Zowe, J., eds., Modern Methods of Optimization, pp. 204–235. Springer, Berlin
9. Mangasarian, O.L., Fromovitz, S. (1967): The Fritz-John necessary optimality conditions in the presence of equality and inequality constraints. J. Math. Anal. Appl. **17**, 37–47
10. Robinson, S.M. (1973): Perturbations in finite-dimensional systems of linear inequalities and equations. Mathematics Research Center, University of Wisconsin-Madison, August, 1973
11. Robinson, S.M. (1975): Stability theory for systems of inequalities, Part 1: Linear systems. SIAM J. Numer. Anal. **12**, 754–769
12. Robinson, S.M. (1980): Strongly regular generalized equations. Math. Oper. Res. **5**, 43–62
13. Robinson, S.M. (1981): Some continuity properties of polyhedral multifunctions. Math. Program. Study **14**, 206–214
14. Robinson, S.M. (1982): Generalized equations and their solution, part II: Applications to nonlinear programming. Math. Program. Study **19**, 200–221
15. Walkup, D.W., Wets, R.J.-B. (1969): A Lipschitzian characterization of convex polyhedra. Proc. Am. Math. Soc. **20**, 167–173
16. Wright, S.J. (1998): Superlinear convergence of a stabilized SQP method to a degenerate equation. Argonne National Laboratory, Mathematics and Computer Science Division, Argonne, Illinois, November, 1996. Comput. Optim. Appl. **11**(3)