Computational Optimization and Applications 12, 253–273 (1999)

Stabilized Sequential Quadratic Programming*

WILLIAM W. HAGER hager@math.ufl.edu
Department of Mathematics, University of Florida, Gainesville, FL 32611

Received January, 1997; Revised January 19, 1998

Abstract. Recently, Wright proposed a stabilized sequential quadratic programming algorithm for inequality constrained optimization. Assuming the Mangasarian-Fromovitz constraint qualification and the existence of a strictly positive multiplier (but possibly dependent constraint gradients), he proved a local quadratic convergence result. In this paper, we establish quadratic convergence in cases where both strict complementarity and the Mangasarian-Fromovitz constraint qualification do not hold. The constraints on the stabilization parameter are relaxed, and linear convergence is demonstrated when the parameter is kept fixed. We show that the analysis of this method can be carried out using recent results for the stability of variational problems.

Keywords: sequential quadratic programming, quadratic convergence, superlinear convergence, degenerate optimization, stabilized SQP, error estimation

Dedication: This paper is dedicated to Olvi L. Mangasarian on the occasion of his 65th birthday.

1. Introduction

Let us consider the following inequality constrained optimization problem:

$$\text{minimize} \quad f(z) \qquad \text{subject to} \quad c(z) \le 0, \quad z \in \mathbf{R}^n, \tag{1}$$

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

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

Let (z_k, λ_k) denote the current approximation to a local minimizer z_* and an associated multiplier λ_* for (1). In the sequential quadratic programming (SQP) algorithm, the new approximation z_{k+1} to z_* is given by $z_{k+1} = z_k + \Delta z$ where Δz is a local minimizer of the following quadratic problem:

$$\begin{aligned} &\text{minimize} \quad \nabla f(z_k)\Delta z + \frac{1}{2}\Delta z^{\mathsf{T}} \nabla_z^2 \mathcal{L}(z_k, \lambda_k)\Delta z \\ &\text{subject to} \quad c(z_k) + \nabla c(z_k)\Delta z \le 0 \end{aligned} \tag{2}$$

*This work was supported by the National Science Foundation.

There are various ways to specify the new multiplier. Often λ_{k+1} is a multiplier associated with the constraint in the quadratic problem (2).

The typical convergence theorem for (2) (for example, see Robinson's paper [16]) states that for (z_0, λ_0) in a neighborhood of a solution/multiplier pair (z_*, λ_*) associated with (1), the iteration is quadratically convergent when the following conditions hold:

(R1) The gradients of the active constraints are linearly independent.
(R2) The multipliers associated with the active constraints are strictly positive.
(R3) There exists a scalar $\alpha > 0$ such that

$$w^{\mathsf{T}} \nabla_z^2 \mathcal{L}(z_*, \lambda_*) w \geq \alpha \|w\|^2 \tag{3}$$

for each w satisfying $\nabla c_i(z_*) w = 0$ for every i such that $c_i(z_*) = 0$.

When the constraint gradients are linearly dependent, quadratic convergence in the SQP algorithm is lost in even the simplest cases. For example, consider the problem

$$\text{minimize} \quad z^2 \qquad \text{subject to} \quad z^2 \leq 0. \tag{4}$$

The unique solution is $z_* = 0$ while λ_* can be any nonnegative number. If the multiplier approximation λ_k is held fixed at $\underline{\lambda} \geq 0$ and if $z_0 > 0$, then for $0 \leq \underline{\lambda} \leq 1$, the iteration reduces to $z_{k+1} = \underline{\lambda} z_k/(1 + \underline{\lambda})$, while for $\underline{\lambda} > 1$, the iteration reduces to $z_{k+1} = z_k/2$. In either case, the convergence is linear.

Wright's stabilized sequential quadratic programming algorithm [19] is obtained by applying Rockafellar's augmented Lagrangian [18] to the quadratic program (2). If $\rho_k > 0$ is the penalty parameter at iteration k, then (z_{k+1}, λ_{k+1}) is a local minimax for the problem

$$\begin{aligned} \min_{z} \max_{\lambda \geq 0} \; & (z - z_k)^{\mathsf{T}} \nabla f(z_k) + \frac{1}{2}(z - z_k)^{\mathsf{T}} \nabla_z^2 \mathcal{L}(z_k, \lambda_k)(z - z_k) \\ & + \lambda^{\mathsf{T}}(c(z_k) + \nabla c(z_k)(z - z_k)) - \frac{1}{2}\rho_k \|\lambda - \lambda_k\|^2. \end{aligned} \tag{5}$$

Wright shows that this method is locally quadratically convergent if the following conditions hold:

(W1) The Mangasarian-Fromovitz [15] constraint qualification (MFCQ) holds. In the context of the inequality constrained problem (1), this means that there exists y such that $c(z_*) + \nabla c(z_*) y < 0$.
(W2) There exists a multiplier vector whose components associated with the active constraints are strictly positive.
(W3) For some fixed $\alpha > 0$, the coercivity condition (3) holds for all choices of λ_* satisfying the following first-order conditions:

$$\nabla_z \mathcal{L}(z_*, \lambda_*) = 0, \quad \lambda_* \geq 0, \quad c(z_*)^{\mathsf{T}} \lambda_* = 0. \tag{6}$$

(W4) The parameter ρ_k tends to zero proportional to the error in (z_k, λ_k).

Notice that (W1) is weaker than (R1) since there may exist y such that

$$c(z_*) + \nabla c(z_*)y < 0$$

even when the constraint gradients are linearly dependent. On the other hand, the MFCQ does not hold for the example (4), or in cases where an equality constraint is written as a pair of inequalities.

Let us consider the stabilized iteration (5) for the example (4) with z_0 near $z_* = 0$. There are two cases to consider, depending on the choice of $\lambda_k = \underline{\lambda}$. If $\underline{\lambda}$ is sufficiently large (for example, $\underline{\lambda} \geq z_k^2/(\rho_k + z_k^2)$), then at the solution of (5), the maximizing λ is positive and the successive iterates are given by

$$z_{k+1} = \frac{z_k^3}{(1+\underline{\lambda})\rho_k + 2z_k^2}. \tag{7}$$

Hence, if $\rho_k = z_k - z_* = z_k$ (the error at step k), then we have

$$z_{k+1} = \frac{z_k^2}{1+\underline{\lambda}+2z_k},$$

which implies local quadratic convergence to the solution $z_* = 0$.

The second case corresponds to the situation where the maximizing λ in (5) vanishes. For this to happen, we must have $\underline{\lambda} \leq z_k^2/(\rho_k + z_k^2)$, and the new iterate is expressed:

$$z_{k+1} = \frac{z_k\underline{\lambda}}{1+\underline{\lambda}}.$$

Again, if $\rho_k = z_k - z_* = z_k$, then

$$\underline{\lambda} \leq z_k^2\big/\big(\rho_k + z_k^2\big) = z_k/(1+z_k), \tag{8}$$

and we have

$$z_{k+1} = \frac{z_k\underline{\lambda}}{1+\underline{\lambda}} \leq \frac{z_k^2}{(1+\underline{\lambda})(1+z_k)}. \tag{9}$$

In each of these cases, the convergence to the solution $z_* = 0$ is locally quadratic.

Also notice in this example that choosing ρ_k much smaller than the error at step k can slow the convergence. In particular, if $\rho_k = 0$ and the max is changed to sup in (5), then the scheme (5) reduces to the usual SQP iteration (2) for which the convergence in the example (4) is linear. On the other hand, we still obtain fast convergence even when ρ_k is much larger than the error at step k. For example, if $\rho_k > 0$ is fixed, then (7) gives cubic convergence. Likewise, the initial inequality in (8) implies that $\underline{\lambda} \leq z_k^2/\rho_k$, which combines with (9) to give $z_{k+1} \leq z_k^3/\rho_k$. In either case, when $\rho_k > 0$ is fixed, we obtain cubic convergence near the solution $z_* = 0$. Hence, from an implementational viewpoint, a large ρ_k is safer than a small one.

In this example, quadratic convergence is preserved with the stabilized SQP scheme even though strict complementarity and the MFCQ are violated. In fact, when $\underline{\lambda} = 0$ and strict complementarity is violated, we have convergence in one step. In this paper, we show in general that Wright's stabilized scheme is locally, quadratically convergent even though both the MFCQ and strict complementarity are violated. In contrast to Wright's assumption (W3) that the second-order condition holds for all multipliers, we give in this paper a local analysis where a second-order condition is required to hold only at a given solution/multiplier pair (z_*, λ_*). When strict complementarity is violated, our second-order condition is slightly stronger than the usual second-order condition in that we assume

$$w^{\mathsf{T}} \nabla_z^2 \mathcal{L}(z_*, \lambda_*) w \geq \alpha \|w\|^2 \tag{10}$$

for all w satisfying $\nabla c_i(z_*) w = 0$ for every i such that $c_i(z_*) = 0$ and $(\lambda_*)_i > 0$. This strengthened form of the second-order sufficient condition first appears in Robinson's study [17] of Lipschitz stability of optimization problems. Dontchev and Rockafellar [7] show that this condition along with linear independence of the active constraint gradients are necessary and sufficient for Lipschitz stability of the solution and multipliers under canonical perturbations of the problem constraints and cost function.

The strong second-order sufficient condition is stable in the sense that it holds when $\nabla_z^2 \mathcal{L}(z_*, \lambda_*)$ and $\nabla c_i(z_*)$ are replaced by nearby matrices, while the usual second-order condition is unstable under problem perturbations. The usual second-order sufficient condition imposes on w in (10) the additional constraint $\nabla c_i(z_*) w \leq 0$ for every i such that $c_i(z_*) = 0 = (\lambda_*)_i$. That is, (10) must hold for all w in the set

$$\{w \in \mathbf{R}^n : \nabla c_i(z_*) w = 0 \;\; \forall i \in \mathcal{A}_+, \quad \nabla c_i(z_*) w \leq 0 \;\; \forall i \in \mathcal{A}_0\},$$

where

$$\mathcal{A}_+ = \{i : (\lambda_*)_i > 0\} \quad \text{and} \quad \mathcal{A}_0 = \{i : c_i(z_*) = 0 = (\lambda_*)_i\}.$$

If the usual second-order condition holds for some pair (z_*, λ_*), then we can perturb the constraint $c(z) \leq 0$ to $c(z) + \xi \leq 0$ where $\xi_i < 0$ if $(\lambda_*)_i = 0 = c_i(z_*)$, and $\xi_i = 0$ otherwise. For this perturbed problem, (z_*, λ_*) again satisfies the first-order conditions, however, the active constraints for the perturbed problem are precisely the constraints in the unperturbed problem with positive multipliers. Therefore, even though the usual second-order sufficient condition holds at (z_*, λ_*), small perturbations in the constraints can yield a problem whose stationary point does not satisfy this condition.

Our analysis of (5) is based on the application of tools from stability analysis. That is, we introduce parameters in the iteration map and we study how the map depends on the parameters using a stability result established in [6, Lemma 2.1]. Once we understand how the iteration map depends on the parameters, we can write down a convergence theorem. Other applications of stability theory to the convergence of algorithms and to the analysis of discretizations appear in [3–6], and [11]. Our analysis of (5) also leads to a new expression for the error in each iterate. In particular, we show that linear convergence is achieved when ρ_k is fixed, but small. This paper is a revised version of the report [12].

Another approach for dealing with degeneracy in nonlinear programming is developed by Fischer in [9]. In his approach, the original quadratic program (2) is retained, however, the multiplier estimate is gotten by solving a separate quadratic program. Fischer obtains quadratic convergence assuming the MFCQ, the second-order sufficient optimality condition, a constant rank condition for the active constraint gradients in a neighborhood of z_*, and a condition concerning the representation of the cost function gradient in terms of the constraint gradients. Although these assumptions seem more stringent than those used in our analysis of Wright's method, there are no parameters like ρ_k in Fischer's method that must be specified in each iteration.

2. Convergence theory

Let z_* denote a local minimizer for (1) and let λ_* be an associated multiplier satisfying the first-order conditions (6). To state our assumptions, we partition c and λ into (g, h) and (μ, π) where the components of h correspond to components of c associated with strictly positive components π_* of λ_*, while the components of g are the remaining components of c for which the associated components μ_* of λ_* could be zero. Let $\mathcal{M}$ denote the set of all multipliers associated with a local minimizer z_* for (1):

$$\lambda \in \mathcal{M} \quad \text{if and only if} \quad \nabla_z \mathcal{L}(z_*, \lambda) = 0, \quad \lambda \geq 0, \quad \text{and} \quad \lambda^{\mathsf{T}} c(z_*) = 0.$$

Letting $\mathcal{B}_\delta(z)$ denote the ball with center z and radius δ, our main result is the following:

Theorem 1. *Suppose that f and c are twice Lipschitz continuously differentiable in a neighborhood of a local minimizer z_* of* (1), *that $\lambda_* = (\mu_*, \pi_*)$ is an associated multiplier in $\mathcal{M}$ with $\pi_* > 0$, and that*

$$w^{\mathsf{T}} \nabla_z^2 \mathcal{L}(z_*, \lambda_*) w \geq \alpha \|w\|^2 \tag{11}$$

for each w such that $\nabla h(z_*) w = 0$. Then for any choice of the constant σ_0 sufficiently large, there exist constants σ_1, δ, and $\bar{\beta}$ with the property that $\sigma_0 \delta \leq \sigma_1$ and for each starting guess $(z_0, \lambda_0) \in \mathcal{B}_\delta(z_*, \lambda_*)$, there are iterates (z_k, λ_k) contained in $\mathcal{B}_\delta(z_*, \lambda_*)$, where each z_{k+1} is a strict local minimizer in the stabilized problem (5), *λ_{k+1} is the unique maximizer in* (5) *associated with $z = z_{k+1}$, and ρ_k is any scalar that satisfies the condition*

$$\sigma_0 \|z_k - z_*\| \leq \rho_k \leq \sigma_1. \tag{12}$$

Moreover, the following estimate holds:

$$\|z_{k+1} - z_*\| + \|\lambda_{k+1} - \hat{\lambda}_{k+1}\| \leq \bar{\beta}(\|z_k - z_*\|^2 + \|\lambda_k - \hat{\lambda}_k\|^2 + \rho_k \|\lambda_k - \hat{\lambda}_k\|), \tag{13}$$

where $\hat{\lambda}_k$ and $\hat{\lambda}_{k+1}$ are the closest elements of $\mathcal{M}$ to λ_k and λ_{k+1} respectively.

By Theorem 1, letting ρ_k go to zero proportional to the total error

$$\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\|$$

leads to local quadratic convergence. Techniques for estimating the error in the current iterate can be found in [13, 19]. Since Theorem 1 is a local convergence result, we assume (without loss of generality), that $c(z_*) = 0$. That is, if some constraint is inactive at z_*, we simply discard this constraint and apply Theorem 1 to the reduced problem, obtaining a neighborhood where the iterations converge and (13) holds. When this constraint is included in c, it can be shown that for (z_k, λ_k) near (z_*, λ_*), the associated component of the maximizing multiplier in (5) vanishes. Hence, the iterates obtained either with or without this inactive constraint included in c are identical.

Although an equality constraint does not appear explicitly in (1), we can include the equality constraint $e(z) = 0$ by writing it as a pair of inequalities: $e(z) \le 0$ and $-e(z) \le 0$. One of these constraint functions should be included in g and the other in h. There are an infinite number of multipliers associated with this pair of constraint functions with linearly dependent gradients, and it can always be arranged so that the associated component in π_* is strictly positive.

Throughout this paper, $\|\cdot\|$ denotes the Euclidean norm and β denotes a generic positive constant that has different values in different equations, and which can be bounded in terms of the derivatives through second order of f and c in a neighborhood of (z_*, λ_*) and in terms of fixed constants like α in (11). In order to prove Theorem 1, we recast (5) in the form of a perturbed variational inequality. Let T be the function defined by

$$T(z, \lambda, \underline{z}, \underline{\lambda}_1, \underline{\lambda}_2) = \begin{pmatrix} \nabla_z \mathcal{L}(\underline{z}, \lambda) + \nabla_z^2 \mathcal{L}(\underline{z}, \underline{\lambda}_1)(z - \underline{z}) \\ c(\underline{z}) + \nabla c(\underline{z})(z - \underline{z}) - \rho(\lambda - \underline{\lambda}_2) \end{pmatrix}, \tag{14}$$

where ρ and $p = (\underline{z}, \underline{\lambda}_1, \underline{\lambda}_2)$ are regarded parameters. Since we later impose a constraint on ρ in terms of p, as in (12), we do not make ρ an explicit argument of T. We study properties of solutions to the following inclusion relative to the parameters: Find (z, λ) such that

$$T(z, \lambda, p) \in \begin{pmatrix} 0 \\ N(\lambda) \end{pmatrix}, \quad \lambda \ge 0, \tag{15}$$

where N is the usual normal cone: If $\lambda \ge 0$, then $y \in N(\lambda)$ if and only if $y \le 0$ and $y^{\mathsf{T}}\lambda = 0$. By analyzing how the solutions to (15) depend on p, we will establish Theorem 1.

If (z_{k+1}, λ_{k+1}) is a local solution to (5), then for $p = p_k = (z_k, \lambda_k, \lambda_k)$, $(z, \lambda) = (z_{k+1}, \lambda_{k+1})$ is a solution to (15), and in this case, (15) represents the first-order optimality conditions associated with (5). More explicitly, (15) implies that

$$\nabla_z \mathcal{L}(z_k, \lambda_{k+1}) + \nabla_z^2 \mathcal{L}(z_k, \lambda_k)(z_{k+1} - z_k) = 0, \tag{16}$$

$$c(z_k) + \nabla c(z_k)(z_{k+1} - z_k) - \rho(\lambda_{k+1} - \lambda_k) \le 0, \quad \lambda_{k+1} \ge 0, \tag{17}$$

$$\lambda_{k+1}^{\mathsf{T}}(c(z_k) + \nabla c(z_k)(z_{k+1} - z_k) - \rho(\lambda_{k+1} - \lambda_k)) = 0. \tag{18}$$

Conditions (17) and (18) are equivalent to saying that λ_{k+1} achieves the maximum in (5) corresponding to $z = z_{k+1}$. By the standard rules for differentiating under a maximization (see [2]), the derivative of the extremand in (5) with respect to z is obtained by computing the partial derivative with respect to z and evaluating the resulting expression at that λ

where the extremand is maximized. Hence, (16) is equivalent to saying the derivative of the extremand with respect to z vanishes at $z = z_{k+1}$.

Observe that when $p = (z_*, \lambda_*, \tilde{\lambda})$, where $\tilde{\lambda}$ is an arbitrary element of $\mathcal{M}$, then $(z, \lambda) = (z_*, \tilde{\lambda})$ is a solution to (15). In this section, we apply the following stability result, describing how the solution to (15) changes as p changes, to obtain Theorem 1. The proof of this stability result is given in the next section.

Lemma 1. *Under the hypotheses of Theorem* 1, *for any choice of the constant σ_0 sufficiently large and for any $\sigma_1 > 0$, there exist constants β and δ such that $\sigma_0\delta \le \sigma_1$ and for each $p = (\underline{z}, \underline{\lambda}_1, \underline{\lambda}_2) \in \mathcal{B}_\delta(p_*)$ and for each ρ satisfying*

$$\Lambda(p) \le \rho \le \sigma_1 \quad \text{where } \Lambda(p) = \sigma_0 \|\underline{z} - z_*\|, \tag{19}$$

(15) *has a unique solution $(z, \lambda) = (z(p), \lambda(p)) \in \mathcal{N}(\rho)$ where*

$$\mathcal{N}(\rho) = \{(z, \lambda) : \|z - z_*\| + \rho\|\lambda - \lambda_*\| \le \rho\}.$$

Moreover, for every p_1 and $p_2 \in \mathcal{B}_\delta(p_*)$, and ρ satisfying (19) *for $p = p_1$ and $p = p_2$, if (z_1, λ_1) and (z_2, λ_2) are the associated solutions to* (15), *then we have*

$$\|z_1 - z_2\| + \rho\|\lambda_1 - \lambda_2\| \le \beta\|T(z_1, \lambda_1, p_1) - T(z_1, \lambda_1, p_2)\|. \tag{20}$$

There are three parts to the proof of Theorem 1. Initially, we show that the estimate (13) holds for each (z_k, λ_k) near (z_*, λ_*), where (z_{k+1}, λ_{k+1}) is a solution to (15) associated with $p = p_k = (z_k, \lambda_k, \lambda_k)$. Next, we show that for (z_0, λ_0) sufficiently close to (z_*, λ_*), we can construct a sequence $(z_1, \lambda_1), (z_2, \lambda_2), \ldots,$ contained in a fixed ball centered at (z_*, λ_*), where (z_{k+1}, λ_{k+1}) is the unique solution in $\mathcal{N}(\rho_k)$ to (15) for $p = (z_k, \lambda_k, \lambda_k)$. Finally, we show that for this unique solution (z_{k+1}, λ_{k+1}) to (15), z_{k+1} is a local minimizer of (5).

Part 1 (Error estimate). Let $\sigma_1 > 0$ be any fixed scalar (independent of k) and let σ_0 and δ be chosen in accordance with Lemma 1. By Lemma 1, there exists a ball $\mathcal{B}_\delta(p_*)$ with the property that for each $p_k = (z_k, \lambda_k, \lambda_k) \in \mathcal{B}_\delta(p_*)$, (15) has a unique solution $(z, \lambda) = (z_{k+1}, \lambda_{k+1})$ in $\mathcal{N}(\rho)$ where ρ is any scalar that satisfies the condition

$$\sigma_0\|z_k - z_*\| \le \rho \le \sigma_1. \tag{21}$$

We apply Lemma 1 taking

$$p_1 = (z_*, \lambda_*, \hat{\lambda}_k), \quad (z_1, \lambda_1) = (z_*, \hat{\lambda}_k),$$
$$p_2 = (z_k, \lambda_k, \lambda_k), \quad (z_2, \lambda_2) = (z_{k+1}, \lambda_{k+1}).$$

If λ_k is near λ_*, then $\hat{\lambda}_k$ is near λ_* since $\|\lambda_k - \hat{\lambda}_k\| \le \|\lambda_k - \lambda_*\|$ and

$$\|\hat{\lambda}_k - \lambda_*\| \le \|\hat{\lambda}_k - \lambda_k\| + \|\lambda_k - \lambda_*\| \le 2\|\lambda_k - \lambda_*\|.$$

Suppose that $p_2 = p_k \in \mathcal{B}_\delta(p_*)$ is close enough to p_* that $p_1 = (z_*, \lambda_*, \hat{\lambda}_k) \in \mathcal{B}_\delta(p_*)$ and $(z_1, \lambda_1) = (z_*, \hat{\lambda}_k) \in \mathcal{N}(\rho)$. Note that (19) holds for $p = p_1$ since $\Lambda(p_1) = 0$. Assuming that ρ is chosen so that (19) holds for $p = p_k = (z_k, \lambda_k, \lambda_k)$, it follows from (20) that

$$\|z_{k+1} - z_*\| + \rho\|\lambda_{k+1} - \hat{\lambda}_k\| \le \beta E_k, \tag{22}$$

where

$$E_k = \left\| \begin{pmatrix} \nabla_z \mathcal{L}(z_k, \hat{\lambda}_k) + \nabla_z^2 \mathcal{L}(z_k, \lambda_k)(z_* - z_k) \\ c(z_k) + \nabla c(z_k)(z_* - z_k) - \rho(\hat{\lambda}_k - \lambda_k) \end{pmatrix} \right\|. \tag{23}$$

Expanding E_k in a Taylor series around z_* gives

$$E_k \le \bar{E}_k := \beta(\|z_k - z_*\|^2 + \|\lambda_k - \hat{\lambda}_k\|\|z_k - z_*\| + \rho\|\lambda_k - \hat{\lambda}_k\|) \tag{24}$$

$$\le \beta(\|z_k - z_*\|^2 + \|\lambda_k - \hat{\lambda}_k\|^2 + \rho\|\lambda_k - \hat{\lambda}_k\|), \tag{25}$$

where β is a generic positive constant. The second inequality (25) is obtained using the relation $ab \le (a^2 + b^2)/2$. Combining (22), (23), and (25) establishes the estimate for z_{k+1} in Theorem 1.

Dividing (24) by ρ gives

$$\bar{E}_k/\rho \le \beta((\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\|)\|z_k - z_*\|/\rho + \|\lambda_k - \hat{\lambda}_k\|).$$

Utilizing the lower bound $\rho \ge \sigma_0\|z_k - z_*\|$, it follows that

$$\bar{E}_k/\rho \le \beta(\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\|). \tag{26}$$

Hence, dividing (22) by ρ and referring to (26), we deduce that

$$\|\lambda_{k+1} - \hat{\lambda}_k\| \le \beta(\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\|). \tag{27}$$

By the triangle inequality, we have

$$\|\lambda_{k+1} - \lambda_*\| \le \|\lambda_{k+1} - \hat{\lambda}_k\| + \|\hat{\lambda}_k - \lambda_k\| + \|\lambda_k - \lambda_*\|,$$

and combining this with (27) gives

$$\begin{aligned} \|\lambda_{k+1} - \lambda_*\| &\le \|\lambda_k - \lambda_*\| + \beta(\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\|) \\ &\le \|\lambda_k - \lambda_*\| + \beta(\|z_k - z_*\| + \|\lambda_k - \lambda_*\|). \end{aligned} \tag{28}$$

This shows that λ_{k+1} is near λ_* when (z_k, λ_k) is near (z_*, λ_*).

We now show that

$$\|\nabla_z \mathcal{L}(z_*, \lambda_{k+1})\| \le \beta\bar{E}_k. \tag{29}$$

In order to establish this, we exploit the Lipschitz continuity of $\nabla_z \mathcal{L}$, the bound (22), and our observation that λ_{k+1} is near λ_* to obtain

$$\|\nabla_z \mathcal{L}(z_*, \lambda_{k+1}) - \nabla_z \mathcal{L}(z_{k+1}, \lambda_{k+1})\| \le \beta \|z_{k+1} - z_*\| \le \beta E_k. \tag{30}$$

Expanding $\nabla_z \mathcal{L}(z_{k+1}, \lambda_{k+1})$ in a Taylor series around z_k and substituting from (16) gives

$$\begin{aligned}
&\|\nabla_z \mathcal{L}(z_{k+1}, \lambda_{k+1})\| \\
&\quad \le \left\|\nabla_z \mathcal{L}(z_k, \lambda_{k+1}) + \nabla_z^2 \mathcal{L}(z_k, \lambda_{k+1})(z_{k+1} - z_k)\right\| + \beta \|z_{k+1} - z_k\|^2 \\
&\quad = \left\|\left(\nabla_z^2 \mathcal{L}(z_k, \lambda_{k+1}) - \nabla_z^2 \mathcal{L}(z_k, \lambda_k)\right)(z_{k+1} - z_k)\right\| + \beta \|z_{k+1} - z_k\|^2 \\
&\quad \le \beta(\|z_{k+1} - z_k\| + \|\lambda_{k+1} - \lambda_k\|)\|z_{k+1} - z_k\| \\
&\quad \le \beta(\|z_{k+1} - z_k\|^2 + \|\lambda_{k+1} - \lambda_k\|^2).
\end{aligned} \tag{31}$$

By the triangle inequality, we have

$$\|z_{k+1} - z_k\| \le \|z_{k+1} - z_*\| + \|z_k - z_*\| \le \beta E_k + \|z_k - z_*\|.$$

Squaring this gives

$$\|z_{k+1} - z_k\|^2 \le \beta E_k^2 + 2\|z_k - z_*\|^2 \le \beta \bar{E}_k. \tag{32}$$

If it can be shown that

$$\|\lambda_{k+1} - \lambda_k\| \le \beta(\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\|), \tag{33}$$

then by squaring, we have

$$\|\lambda_{k+1} - \lambda_k\|^2 \le \beta(\|z_k - z_*\|^2 + \|\lambda_k - \hat{\lambda}_k\|^2) \le \beta \bar{E}_k. \tag{34}$$

Combining (31) with (32) and (34) gives

$$\|\nabla_z \mathcal{L}(z_{k+1}, \lambda_{k+1})\| \le \beta \bar{E}_k,$$

and combining this with (30) yields

$$\begin{aligned}
\|\nabla_z \mathcal{L}(z_*, \lambda_{k+1})\| \le{}& \|\nabla_z \mathcal{L}(z_{k+1}, \lambda_{k+1})\| \\
&+ \|\nabla_z \mathcal{L}(z_*, \lambda_{k+1}) - \nabla_z \mathcal{L}(z_{k+1}, \lambda_{k+1})\| \le \beta \bar{E}_k,
\end{aligned}$$

which completes the proof of (29).

To prove (33), we focus on the individual components of $\lambda_{k+1} - \lambda_k$ and establish the relation

$$|(\lambda_{k+1} - \lambda_k)_i| \le \beta(\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\|) \tag{35}$$

for each i. There are three cases to consider:

(C1) $(\lambda_{k+1})_i = 0 = (\lambda_k)_i$. For these components, (35) is a triviality.

(C2) $(\lambda_{k+1})_i > 0$. By complementary slackness (18), we have

$$(c(z_k) + \nabla c(z_k)(z_{k+1} - z_k))_i = \rho(\lambda_{k+1} - \lambda_k)_i. \tag{36}$$

Expanding $c(z_k)$ in a Taylor expansion around z_{k+1}, utilizing (32), and taking absolute values yields

$$\begin{aligned} |(c(z_k) + \nabla c(z_k)(z_{k+1} - z_k))_i| &\le |c_i(z_{k+1})| + \beta\|z_{k+1} - z_k\|^2 \\ &= |c_i(z_{k+1}) - c_i(z_*)| + \beta\|z_{k+1} - z_k\|^2 \le \beta\|z_{k+1} - z_*\| + \beta\bar{E}_k \le \beta\bar{E}_k. \end{aligned} \tag{37}$$

Dividing (36) by ρ and utilizing (37) and (26) gives (35).

(C3) $(\lambda_{k+1})_i = 0$ and $(\lambda_k)_i > 0$. By (17), we have

$$0 < \rho(\lambda_k)_i = -\rho(\lambda_{k+1} - \lambda_k)_i \le |(c(z_k) + \nabla c(z_k)(z_{k+1} - z_k))_i|.$$

Dividing this by ρ and again utilizing (37) and (26) gives (35).

This completes the proof of both (33) and (29).

Consider the following system of linear equations and inequalities in λ:

$$\nabla_z \mathcal{L}(z_*, \lambda) = 0, \quad \lambda \ge 0. \tag{38}$$

This system is feasible since any $\lambda \in \mathcal{M}$ is a solution. By (29) and a result of Hoffman [14], the closest solution $\hat{\lambda}_{k+1}$ of (38) to λ_{k+1} satisfies

$$\|\lambda_{k+1} - \hat{\lambda}_{k+1}\| \le \beta\bar{E}_k. \tag{39}$$

That is, Hoffman's result states that if a linear system of inequalities is feasible, then the distance from any given point to the set of feasible points is bounded by a constant times the norm of the constraint violation at the given point. By (29), the norm of the constraint violation is at most $\beta\bar{E}_k$ at λ_{k+1}, from which it follows that the distance from λ_{k+1} to the closest solution of (38) is bounded by a constant times $\bar{E}_k$. Since $c(z_*) = 0$, this solution of (38) is contained in $\mathcal{M}$ and it is the closest element of $\mathcal{M}$ to λ_{k+1}. Relations (25) and (39) combine to complete the proof of (13).

Part 2 (Containment). Collecting results, we have shown that if

$$p = p_k = (z_k, \lambda_k, \lambda_k)$$

is sufficiently close to $p_* = (z_*, \lambda_*, \lambda_*)$, then (15) has a unique solution $(z_{k+1}, \lambda_{k+1}) \in \mathcal{N}(\rho)$ where ρ is any scalar satisfying (21), where z_{k+1} and λ_{k+1} satisfy (13), and where λ_{k+1} also satisfies (28). As σ_1 or δ in Lemma 1 decreases, the constant β in (20) can be kept fixed since the set of ρ and p that satisfies the constraints of the lemma becomes smaller. That is, if (20) holds for one set of ρ and p values, then it holds for all subsets. Let $\bar{\beta}$ be the constant appearing in (13) that we estimated in Part 1 using Lemma 1. Given any positive $\epsilon < 1$, let us choose σ_1 and δ of Lemma 1 small enough that

$$\bar{\beta}(\|z_k - z_*\|^2 + \|\lambda_k - \hat{\lambda}_k\|^2 + \rho_k\|\lambda_k - \hat{\lambda}_k\|) \le \epsilon(\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\|)$$

for all $p_k \in \mathcal{B}_\delta(p_*)$ and $\rho_k \le \sigma_1$. From the analysis of Part 1, both (13) and (28), there exists, for all $p_k \in \mathcal{B}_\delta(p_*)$ and ρ_k satisfying

$$\sigma_0\|z_k - z_*\| \le \rho_k \le \sigma_1, \tag{40}$$

a unique solution $(z_{k+1}, \lambda_{k+1}) \in \mathcal{N}(\rho_k)$ to (15), and we have

$$\|z_{k+1} - z_*\| + \|\lambda_{k+1} - \hat{\lambda}_{k+1}\| \le \epsilon(\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\|), \tag{41}$$

and

$$\|\lambda_{k+1} - \lambda_*\| \le \|\lambda_k - \lambda_*\| + \beta_0(\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\|), \tag{42}$$

where β_0 denotes the specific constant β appearing in (28).

We now show in an inductive fashion that for (z_0, λ_0) sufficiently close to (z_*, λ_*), there exists a sequence (z_k, λ_k), $k = 0, 1, \ldots$, where (z_{k+1}, λ_{k+1}) is the unique solution to (15) in $\mathcal{N}(\rho_k)$ corresponding to $p = (z_k, \lambda_k, \lambda_k)$, and to ρ_k satisfying (40). In particular, let r_0 be chosen small enough that

$$r_1 := 2r_0\left(1 + \frac{\beta_0}{1-\epsilon}\right) \le \delta/2.$$

Starting from any $(z_0, \lambda_0) \in \mathcal{B}_{r_0}(z_*, \lambda_*)$, we proceed by induction and suppose that (z_0, λ_0), $(z_1, \lambda_1), \ldots, (z_j, \lambda_j)$ are all contained in $\mathcal{B}_{r_1}(z_*, \lambda_*)$. Since $r_1 \le \delta/2$, there exists a unique solution $(z_{j+1}, \lambda_{j+1}) \in \mathcal{N}(\rho_j)$ to (15) for $p = (z_j, \lambda_j, \lambda_j)$. By (41), it follows that for $0 \le k \le j+1$,

$$\begin{aligned}\|z_k - z_*\| + \|\lambda_k - \hat{\lambda}_k\| &\le \epsilon^k(\|z_0 - z_*\| + \|\lambda_0 - \hat{\lambda}_0\|) \\ &\le \epsilon^k(\|z_0 - z_*\| + \|\lambda_0 - \lambda_*\|) \le r_0 \le r_1/2.\end{aligned} \tag{43}$$

By (42) and (43), we have

$$\begin{aligned}\|\lambda_{j+1} - \lambda_*\| &\le \|\lambda_j - \lambda_*\| + \beta_0\epsilon^j(\|z_0 - z_*\| + \|\lambda_0 - \hat{\lambda}_0\|) \\ &\le \|\lambda_0 - \lambda_*\| + \beta_0(\|z_0 - z_*\| + \|\lambda_0 - \hat{\lambda}_0\|)\sum_{k=0}^{j}\epsilon^k \\ &\le \|\lambda_0 - \lambda_*\| + \frac{\beta_0}{1-\epsilon}(\|z_0 - z_*\| + \|\lambda_0 - \hat{\lambda}_0\|) \\ &\le \|\lambda_0 - \lambda_*\| + \frac{\beta_0}{1-\epsilon}(\|z_0 - z_*\| + \|\lambda_0 - \lambda_*\|) \\ &\le r_0 + \frac{\beta_0 r_0}{1-\epsilon} \le r_1/2.\end{aligned} \tag{44}$$

Combining (43) and (44) yields

$$\|z_{j+1} - z_*\| + \|\lambda_{j+1} - \lambda_*\| \le r_1/2 + r_1/2 = r_1.$$

Hence, $(z_{j+1}, \lambda_{j+1}) \in \mathcal{B}_{r_1}(z_*, \lambda_*)$ and the induction is complete.

Part 3 (Local minimizer). Finally, we show that z_{k+1} is a local minimizer for (5). Since $\lambda_* = (\mu_*, \pi_*)$ with $\pi_* > 0$, it follows that by taking r_0 sufficiently small, $\pi_{k+1} > 0$ for all k. By complementary slackness (18), we have

$$\pi_{k+1} = \pi_k + (h(z_k) + \nabla h(z_k)(z_{k+1} - z_k))/\rho.$$

As noted after (18), if $(z, \lambda) = (z_{k+1}, \lambda_{k+1})$ is a solution of (15), then $\lambda_{k+1} = (\mu_{k+1}, \pi_{k+1})$ achieves the maximum in (5) for $z = z_{k+1}$. Since the maximizing λ in (5) is a continuous function of z (see [3, Lemma 4]), we conclude that for z near z_{k+1}, the maximizing $\lambda = (\mu, \pi)$ has $\pi > 0$; hence, by complementary slackness and for z near z_{k+1}, the maximizing π is given by

$$\pi = \pi_k + (h(z_k) + \nabla h(z_k)(z - z_k))/\rho.$$

After making this substitution in (5), the cost function of the minimax problem can be decomposed into the sum of a convex function of z:

$$\max_{\mu \ge 0} \; \mu^{\mathsf{T}}(g(z_k) + \nabla g(z_k)(z - z_k)) - \frac{1}{2}\rho_k\|\mu - \mu_k\|^2,$$

and a strongly convex part

$$(z - z_k)^{\mathsf{T}}\nabla f(z_k) + \frac{1}{2}(z - z_k)^{\mathsf{T}}\nabla_z^2\mathcal{L}(z_k, \lambda_k)(z - z_k)$$
$$+ (h(z_k) + \nabla h(z_k)(z - z_k))^{\mathsf{T}}\left(\pi_k + \frac{1}{2\rho_k}(h(z_k) + \nabla h(z_k)(z - z_k))\right).$$

The first part is convex since the extremand is a linear function of z and the max of a sum is less than or equal to the sum of the maxs. The second part is strongly convex since the Hessian matrix

$$\nabla_z^2\mathcal{L}(z_k, \lambda_k) + \frac{1}{\rho_k}\nabla h(z_k)^{\mathsf{T}}\nabla h(z_k)$$

is positive definite for ρ_k and r_0 sufficiently small by Lemma 3 in the Appendix. Hence, the cost function of (5) is a strongly convex function of z in a neighborhood of z_{k+1}, and since the derivative vanishes at z_{k+1} by (16), z_{k+1} is a local minimum. This completes the proof of Theorem 1. □

3. Stability for the linearized system

The proof of Lemma 1 is based on the following result, which is a variation of Lemma 2.1 in [6].

Lemma 2. *Let X be a subset of $\mathbf{R}^n$ and let $\|\cdot\|_\rho$ denote the norm on X. Given $w_* \in X$ and $\tau > 0$, define*

$$W = \{x \in \text{cl } X : \|x - w_*\|_\rho \leq \tau\}.$$

In other words, W is the intersection of the closure of X and the ball with center w_* and radius τ. Suppose that F maps W to the subsets of $\mathbf{R}^m$, and $T : W \times P \to \mathbf{R}^m$, where P is a set. Let $p_* \in P$ with $T(w_*, p_*) \in F(w_*)$, let L be an $m \times n$ matrix, and let τ, η, ϵ, and γ denote any positive numbers for which $\epsilon\gamma < 1$, $\tau \geq \eta\gamma/(1-\epsilon\gamma)$, and the following properties hold:

(P1) $\|T(w_*, p_*) - T(w_*, p)\| \leq \eta$ *for all* $p \in P$.

(P2) $\|T(w_2, p) - T(w_1, p) - L(w_2 - w_1)\| \leq \epsilon\|w_2 - w_1\|_\rho$ *for all* $w_1, w_2 \in W$ *and* $p \in P$.

(P3) *For some set $\mathcal{N} \supset \{T(w, p) - Lw : w \in W,\ p \in P\}$, the following problem has a unique solution for each $\psi \in \mathcal{N}$:*

$$\text{Find } x \in X \text{ such that } Lx + \psi \in F(x) \tag{45}$$

and if $x(\psi)$ denotes the solution corresponding to ψ, we have

$$\|x(\psi_2) - x(\psi_1)\|_\rho \leq \gamma\|\psi_2 - \psi_1\| \tag{46}$$

for each $\psi_1, \psi_2 \in \mathcal{N}$.

Then for each $p \in P$, there exists a unique $w \in W$ such that $T(w, p) \in F(w)$. Moreover, for every $p_i \in P$, $i = 1, 2$, if w_i denotes the w associated with p_i, then we have

$$\|w_2 - w_1\|_\rho \leq \frac{\gamma}{1 - \gamma\epsilon}\|T(w_1, p_2) - T(w_1, p_1)\|. \tag{47}$$

Proof: Fix $p \in P$ and for $w \in W$, let $\Phi(w)$ denote the solution to (45) corresponding to $\psi = T(w, p) - Lw$. That is, $\Phi(w) = x[T(w, p) - Lw]$. For $w_i \in W$, $i = 1, 2$, define $\psi_i = T(w_i, p) - Lw_i$. Observe that

$$\begin{aligned} \|\Phi(w_1) - \Phi(w_2)\|_\rho &= \|x(\psi_1) - x(\psi_2)\|_\rho \leq \gamma\|\psi_1 - \psi_2\| \\ &= \gamma\|T(w_1, p) - T(w_2, p) - L(w_1 - w_2)\| \\ &\leq \gamma\epsilon\|w_1 - w_2\|_\rho \end{aligned}$$

for each $w_1, w_2 \in W$. Since $\gamma\epsilon < 1$, Φ is a contraction on W with contraction constant $\gamma\epsilon$. From the assumption $T(w_*, p_*) \in F(w_*)$, it follows that

$$w_* = x[T(w_*, p_*) - Lw_*].$$

Given $w \in W$, we have

$$\begin{aligned}\|\Phi(w) - w_*\|_\rho &= \|x[T(w, p) - Lw] - x[T(w_*, p_*) - Lw_*]\|_\rho \\ &\le \gamma \|T(w, p) - T(w_*, p_*) - L(w - w_*)\| \\ &\le \gamma(\|T(w, p) - T(w_*, p) - L(w - w_*)\| \\ &\quad + \|T(w_*, p) - T(w_*, p_*)\|) \\ &\le \gamma(\epsilon\|w - w_*\|_\rho + \eta) \le \gamma(\epsilon\tau + \eta) \le \tau\end{aligned}$$

since $\|w - w_*\|_\rho \le \tau$ for all $w \in W$ and $\tau \ge \gamma\eta/(1 - \gamma\epsilon)$. Thus Φ maps W into itself. By the Banach contraction mapping principle, there exists a unique $w \in W$ such that $w = \Phi(w)$. Since $w = \Phi(w)$ is equivalent to $T(w, p) \in F(w)$ for $w \in W$ and $p \in P$, we conclude that for each $p \in P$, there is a unique $w \in W$ such that $T(w, p) \in F(w)$. For $p = p_i \in P$, $i = 1, 2$, let $w = w_i$ denote the associated solutions to $T(w, p) \in F(w)$. We have

$$\begin{aligned}\|w_1 - w_2\|_\rho &= \|x[T(w_1, p_1) - Lw_1] - x[T(w_2, p_2) - Lw_2]\|_\rho \\ &\le \gamma\|T(w_1, p_1) - T(w_2, p_2) - L(w_1 - w_2)\| \\ &\le \gamma\|T(w_1, p_1) - T(w_1, p_2)\| \\ &\quad + \gamma\|T(w_1, p_2) - T(w_2, p_2) - L(w_1 - w_2)\| \\ &\le \gamma\|T(w_1, p_1) - T(w_1, p_2)\| + \gamma\epsilon\|w_1 - w_2\|_\rho.\end{aligned}$$

Rearranging this inequality, the proof is complete. □

Proof of Lemma 1: In order to apply Lemma 2 to T defined in (14), we identify w or x with the pair (z, λ), we identify p with the triple $(\underline{z}, \underline{\lambda}_1, \underline{\lambda}_2)$, and we choose

$$F(w) = F(z, \lambda) = \begin{pmatrix} 0 \\ N(\lambda) \end{pmatrix},$$

and

$$X = \{(z, \lambda) \in \mathbf{R}^n \times \mathbf{R}^m : \lambda = (\mu, \pi),\ \mu \ge 0,\ \pi > 0\}.$$

The set P, chosen later, is a neighborhood of $(z_*, \lambda_*, \lambda_*)$. In presenting the linearization L of Lemma 2, we partition both the constraint function c and the multiplier λ into their components (g, h) and (μ, π) respectively. The linearization L of $T(\cdot, p_*)$ around w_* is given by

$$L\begin{pmatrix} z \\ \mu \\ \pi \end{pmatrix} = \begin{pmatrix} Qz + A^{\mathsf T}\mu + B^{\mathsf T}\pi \\ Az - \rho\mu \\ Bz - \rho\pi \end{pmatrix},$$

where

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

In order to apply Lemma 2 to the function T in (14), we need to establish the Lipschitz property (46). This leads us to consider the problem: Find $x \in X$ such that $L(x) + \psi \in F(x)$. Since L has three components, we partition $\psi = (\varphi, r, s)$, and the linearized problem takes the form: Find $(z, \mu, \pi) \in X$ such that

$$Qz + A^{\mathsf{T}}\mu + B^{\mathsf{T}}\pi + \varphi = 0, \tag{48}$$

$$\mu \geq 0, \ \ Az - \rho\mu + r \in N(\mu), \tag{49}$$

$$Bz - \rho\pi + s = 0, \tag{50}$$

where in the last equation (50), we exploit the fact that $\pi > 0$ for all $(z, \mu, \pi) \in X$.

In order to analyze the linearization (48)–(50), we introduce the following auxiliary problem:

$$\min_{z} \max_{\mu \geq 0} \ z^{\mathsf{T}}\varphi + \frac{1}{2}z^{\mathsf{T}}Qz + \frac{1}{2\rho}\|Bz + s\|^2 + \mu^{\mathsf{T}}(Az + r) - \frac{\rho}{2}\|\mu\|^2. \tag{51}$$

By (11) and Lemma 3 in the Appendix, the matrix $Q + B^{\mathsf{T}}B/\rho$ is positive definite with smallest eigenvalue at least $\alpha/2$ for ρ sufficiently small, where α appears in (11). Hence, the extremand in (51) is strongly convex in z and strongly concave in μ. By [8, Proposition 2.2, p. 173], the max and the min can be interchanged. For fixed μ, the min in (51) is attained by the solution z of the following linear equation:

$$\left(Q + \frac{1}{\rho}B^{\mathsf{T}}B\right)z + \varphi + A^{\mathsf{T}}\mu + B^{\mathsf{T}}s/\rho = 0. \tag{52}$$

After substituting this z in (51), we obtain an *equivalent* strongly concave maximization problem in the variable μ and the parameters φ, r, and s appear linearly in the cost function. Since strongly concave maximization problems are Lipschitz continuous functions of linear parameters in the cost (for example, see [3, Lemma 4]), the maximizing μ is a Lipschitz continuous function of the parameter ψ, and by (52), the minimizing z is also a Lipschitz continuous function of ψ.

Since (48)–(50) are the first-order conditions for a solution of (51), and since the first-order conditions are necessary and sufficient for optimality in this convex/concave setting, we conclude that (48)–(50) have a unique solution $(z(\psi), \lambda(\psi))$ depending Lipschitz continuously on the parameters φ, r, and s. We now apply [10, Theorem 2.1] in order to determine more precisely how the Lipschitz constant of $(z(\psi), \lambda(\psi))$ depends on ρ. Defining the set

$$c(\psi) = \{i : (Az(\psi) - \rho\mu(\psi) + r)_i = 0\}, \tag{53}$$

where $\lambda(\psi) = (\mu(\psi), \pi(\psi))$, it follows from [10, Theorem 2.1], that if γ_1 and γ_2 satisfy

$$\|z(\psi_2) - z(\psi_1)\| \leq \gamma_1\|\psi_1 - \psi_2\|, \quad \|\lambda(\psi_2) - \lambda(\psi_1)\| \leq \gamma_2\|\psi_2 - \psi_1\|,$$

whenever $c(\psi_1) = c(\psi_2)$, then these same Lipschitz constants work for all ψ_1 and ψ_2.

After substituting for μ in (52), using the relation $(Az(\psi) - \rho\mu(\psi) + r)_i = 0$ for $i \in c(\psi)$, we see that $z = z(\psi)$ satisfies

$$\left(Q + \frac{1}{\rho}C^{\mathsf{T}}C\right)z + \varphi + C^{\mathsf{T}}t/\rho = 0, \tag{54}$$

where C and t are gotten by augmenting B and s by the rows of A and the components of r associated with $i \in c(\psi)$. Let UR denote an orthogonal decomposition of C where R is right triangular (that is, $R_{ij} = 0$ if $i > j$) with linearly independent rows and U has orthonormal columns. After substituting $C = UR$ in (54), we obtain the equivalent system

$$\begin{pmatrix} Q & R^{\mathsf{T}} \\ R & -\rho I \end{pmatrix}\begin{pmatrix} z \\ \chi \end{pmatrix} = \begin{pmatrix} -\varphi \\ -U^{\mathsf{T}}t \end{pmatrix}.$$

The second equation $\chi = (Rz + U^{\mathsf{T}}t)/\rho$ in this system is the definition of χ and the first equation in this system is (54). Since the coefficient matrix is nonsingular for ρ sufficiently small (see [1, Lemma 1.27]), both $z(\psi)$ and $\chi(\psi)$ are Lipschitz continuous functions of ψ, where the Lipschitz constant is independent of ρ for ρ sufficiently small:

$$\|z(\psi_2) - z(\psi_1)\| + \|\chi(\psi_2) - \chi(\psi_1)\| \le \beta\|\psi_1 - \psi_2\| \tag{55}$$

Let V have orthonormal columns chosen so that the matrix $(U \mid V)$ is orthogonal. The vector $\pi(\psi)$ satisfies (50) and the components $\mu_0(\psi)$ of $\mu(\psi)$ associated with $i \in c(\psi)$ satisfy an analogous relation in (53). Hence, we have

$$\begin{pmatrix} \mu_0(\psi) \\ \pi(\psi) \end{pmatrix} = (Cz(\psi) + t)/\rho.$$

Multiplying by $(U \mid V)^{\mathsf{T}}$ yields:

$$\begin{aligned}(U \mid V)^{\mathsf{T}}\begin{pmatrix} \mu_0(\psi) \\ \pi(\psi) \end{pmatrix} &= (U \mid V)^{\mathsf{T}}(Cz(\psi) + t)/\rho \\ &= \frac{1}{\rho}\begin{pmatrix} U^{\mathsf{T}}(Cz(\psi) + t) \\ V^{\mathsf{T}}t \end{pmatrix} = \begin{pmatrix} \chi(\psi) \\ V^{\mathsf{T}}t/\rho \end{pmatrix}\end{aligned}$$

Multiplying again by $(U \mid V)$ gives

$$\begin{pmatrix} \mu_0(\psi) \\ \pi(\psi) \end{pmatrix} = U\chi(\psi) + VV^{\mathsf{T}}t/\rho. \tag{56}$$

Since $\chi(\psi)$ is a Lipschitz continuous function of ψ, it follows from (56) that $\mu_0(\psi)$ and $\pi(\psi)$ are Lipschitz continuous functions of ψ, while the remaining components of $\mu(\psi)$

vanish. Therefore, when $c(\psi_1) = c(\psi_2)$, (55) and (56) give us the estimates

$$\|z(\psi_2) - z(\psi_1)\| \le \beta\|\psi_2 - \psi_1\|, \tag{57}$$

$$\|\lambda(\psi_2) - \lambda(\psi_1)\| \le \beta\|\psi_2 - \psi_1\| + \|q_2 - q_1\|/\rho, \tag{58}$$

where $q = (r, s)$ and β is independent of ρ for ρ sufficiently small. By [10, Theorem 2.1], this estimate is valid for arbitrary choices of the parameters.

Given a fixed positive scalar σ_1, we assume that ρ is always $\le \sigma_1$. Hence, after multiplying (58) by ρ and adding to (57), we conclude that

$$\|w(\psi_1) - w(\psi_2)\|_\rho \le \gamma\|\psi_1 - \psi_2\|, \quad w(\psi) = (z(\psi), \lambda(\psi)), \tag{59}$$

for some constant γ independent of ρ, where

$$\|w\|_\rho = \|(z, \lambda)\|_\rho = \|z\| + \rho\|\lambda\|.$$

For the choice $\psi = \psi_* = (\varphi_*, r_*, s_*) = T(w_*, p_*) - L(w_*)$ where

$$\varphi_* = -Qz_* - A^{\mathsf T}\mu_* - B^{\mathsf T}\pi_*, \quad r_* = -Az_* + \rho\mu_*, \quad s_* = -Bz_* + \rho\pi_*,$$

(48)–(50) have the solution $z = z_*$ and $\lambda = \lambda_* = (\mu_*, \pi_*)$. Defining the parameter

$$\Delta = \frac{1}{2\gamma} \min_i \, (\pi_*)_i, \tag{60}$$

it follows from (59) that for all $\psi \in \mathcal{B}_{\rho\Delta}(\psi_*)$ and for all j,

$$\begin{aligned} |(\pi(\psi) - \pi_*)_j| &\le \|\pi(\psi) - \pi_*\| \\ &= \|\pi(\psi) - \pi(\psi_*)\| \le (\gamma/\rho)\|\psi - \psi_*\| \le \min_i \, (\pi_*)_i/2. \end{aligned}$$

Hence, $\pi(\psi) > 0$ for all $\psi \in \mathcal{B}_{\rho\Delta}(\psi_*)$, from which it follows that $(z(\psi), \lambda(\psi)) \in X$ for all $\psi \in \mathcal{B}_{\rho\Delta}(\psi_*)$. Combining this with (59), we conclude that (45) has a unique solution and (46) holds for all $\psi \in \mathcal{B}_{\rho\Delta}(\psi_*)$.

Given an arbitrary scalar $\sigma_1 > 0$, and positive scalars σ_0 and δ, chosen shortly, we define

$$P = \{(\underline{z}, \underline{\lambda}_1, \underline{\lambda}_2) \in \mathcal{B}_\delta(p_*) : \sigma_0\|\underline{z} - z_*\| \le \rho\}, \tag{61}$$

where $p_* = (z_*, \lambda_*, \lambda_*)$. By choosing σ_0 sufficiently large and δ sufficiently small, we will satisfy the condition $\epsilon\gamma < 1$ of Lemma 2, and by choosing δ smaller if necessary, the remaining conditions of Lemma 2 will be satisfied.

(P1) Observe that $T(w_*, p_*) = 0$ where $w_* = (z_*, \lambda_*)$. Defining $p = (\underline{z}, \underline{\lambda}_1, \underline{\lambda}_2)$, we have

$$T(w_*, p_*) - T(w_*, p) = \begin{pmatrix} \nabla_z\mathcal{L}(\underline{z}, \lambda_*) + \nabla_z^2\mathcal{L}(\underline{z}, \underline{\lambda}_1)(z_* - \underline{z}) \\ c(\underline{z}) + \nabla c(\underline{z})(z_* - \underline{z}) - \rho(\lambda_* - \underline{\lambda}_2) \end{pmatrix}.$$

Expanding in a Taylor series around p_* gives

$$\begin{aligned}&\|T(w_*, p_*) - T(w_*, p)\| \\ &\quad \le \beta(\|\underline{z} - z_*\|^2 + \|\underline{\lambda}_1 - \lambda_*\| \|\underline{z} - z_*\| + \rho\|\underline{\lambda}_2 - \lambda_*\|) \end{aligned} \tag{62}$$

for all $p \in P$. Since the right side of (62) is bounded by $\beta\delta$, the constant η in (P1) can be made arbitrarily small by taking δ small.

(P2) Let ϵ be any positive number small enough that $\epsilon\gamma < 1$ where γ appears in (59). Observe that

$$\begin{aligned}&T(w_1, p) - T(w_2, p) - L(w_1 - w_2) \\ &\quad = \begin{pmatrix} \left(\nabla_z^2 \mathcal{L}(\underline{z}, \underline{\lambda}_1) - \nabla_z^2 \mathcal{L}(z_*, \lambda_*)\right)(z_2 - z_1) + (\nabla c(\underline{z}) - \nabla c(z_*))^{\mathsf{T}}(\lambda_2 - \lambda_1) \\ (\nabla c(\underline{z}) - \nabla c(z_*))(z_2 - z_1) \end{pmatrix},\end{aligned}$$

where $w_1 = (z_1, \lambda_1)$ and $w_2 = (z_2, \lambda_2)$. By the assumed Lipschitz continuity of the derivatives, and by (61), we have, for all $p \in P$ and for any choice of w_1 and w_2,

$$\begin{aligned}&\|T(w_1, p) - T(w_2, p) - L(w_1 - w_2)\| \\ &\quad \le \beta\|\underline{z} - z_*\| \|w_1 - w_2\| + \beta\|\underline{\lambda}_1 - \lambda_*\| \|z_1 - z_2\| \\ &\quad \le \frac{\beta\rho}{\sigma_0}\|w_1 - w_2\| + \beta\delta\|z_1 - z_2\| \\ &\quad \le \frac{\beta}{\sigma_0}(\rho\|z_1 - z_2\| + \rho\|\lambda_1 - \lambda_2\|) + \beta\delta\|z_1 - z_2\| \\ &\quad \le \frac{\beta}{\sigma_0}(\sigma_1\|w_1 - w_2\|_\rho + \|w_1 - w_2\|_\rho) + \beta\delta\|w_1 - w_2\|_\rho \\ &\quad = \beta\left(\frac{(1 + \sigma_1)}{\sigma_0} + \delta\right)\|w_1 - w_2\|_\rho. \end{aligned} \tag{63}$$

Choose σ_0 large enough and δ small enough that the factor multiplying $\|w_1 - w_2\|_\rho$ in (63) is $\le \epsilon$. This establishes (P2) and $\epsilon\gamma < 1$.

(P3) Choosing $\tau = \rho$, the set W of Lemma 2 is

$$W = \{w = (z, \lambda) \in \mathbf{R}^n \times \mathbf{R}^m : \lambda = (\mu, \pi),\ \mu \ge 0,\ \pi > 0,\ \|w - w_*\|_\rho \le \rho\}.$$

By (62) and (63), we have for all $w \in W$ and $p \in P$,

$$\begin{aligned}&\|T(w, p) - L(w) - (T(w_*, p_*) - L(w_*))\| \\ &\quad \le \|(T(w, p) - T(w_*, p)) - L(w - w_*)\| + \|T(w_*, p) - T(w_*, p_*)\| \\ &\quad \le \epsilon\|w - w_*\|_\rho + \beta(\|\underline{z} - z_*\|^2 + \|\underline{\lambda}_1 - \lambda_*\| \|\underline{z} - z_*\| + \rho\|\underline{\lambda}_2 - \lambda_*\|) \\ &\quad \le \rho(\epsilon + \beta\delta) \end{aligned} \tag{64}$$

since $\|w - w_*\|_\rho \le \tau = \rho$, $\|\underline{z} - z_*\| \le \rho/\sigma_0$, and $(\underline{z}, \underline{\lambda}_1, \underline{\lambda}_2) \in \mathcal{B}_\delta(p_*)$. Choose ϵ and δ smaller if necessary so that

$$\epsilon + \beta\delta \le \Delta,$$

where Δ is defined in (60). Hence, by (64), we have

$$\|T(w, p) - L(w) - (T(w_*, p_*) - L(w_*))\| \le \Delta\rho$$

for all $w \in W$ and $p \in P$. Since $\psi_* = T(w_*, p_*) - L(w_*)$, it follows that

$$T(w, p) - L(w) \in \mathcal{B}_{\rho\Delta}(\psi_*)$$

for all $w \in W$ and $p \in P$. This completes the proof of (P3) since we already showed that (45) has a unique solution satisfying (46) for all $\psi \in \mathcal{B}_{\rho\Delta}(\psi_*)$.

Finally, let us consider the condition

$$\tau \ge \gamma\eta/(1 - \epsilon\gamma) \tag{65}$$

of Lemma 2, where $\eta = \sup\{\|T(w_*, p_*) - T(w_*, p)\| : p \in P\}$. Recalling that $\tau = \rho$, and utilizing (62), we see that (65) is satisfied if

$$\rho \ge \beta(\|\underline{z} - z_*\|^2 + \|\underline{\lambda}_1 - \lambda_*\|\|\underline{z} - z_*\| + \rho\|\underline{\lambda}_2 - \lambda_*\|) \tag{66}$$

for each $(\underline{z}, \underline{\lambda}_1, \underline{\lambda}_2) \in P$; here the factor $\gamma/(1 - \epsilon\gamma)$ of (65) is absorbed into β. Assuming δ is small enough that $\beta\|\underline{\lambda}_2 - \lambda_*\| < 1$, we rearrange (66) to obtain the equivalent relation

$$\rho \ge \frac{\beta(\|\underline{z} - z_*\| + \|\underline{\lambda}_1 - \lambda_*\|)\|\underline{z} - z_*\|}{1 - \beta\|\underline{\lambda}_2 - \lambda_*\|}. \tag{67}$$

By the definition of P, $\rho \ge \sigma_0\|\underline{z} - z_*\|$ for all $p = (\underline{z}, \underline{\lambda}_1, \underline{\lambda}_2) \in P$. Hence, if

$$\sigma_0 \ge \frac{\beta(\|\underline{z} - z_*\| + \|\underline{\lambda}_1 - \lambda_*\|)}{1 - \beta\|\underline{\lambda}_2 - \lambda_*\|}, \tag{68}$$

(67) will be satisfied. Choosing δ small enough that (68) is satisfied, it follows that (67) holds, which implies in turn (65). Since all the assumptions of Lemma 2 are satisfied, Lemma 1 follows almost directly. The neighborhood $\mathcal{N}(\rho)$ of Lemma 1 coincides with W of Lemma 2, while the ball $\mathcal{B}_\delta$ of Lemma 1 is the same ball appearing in the definition of P in (61). The constant β of Lemma 1 is the expression $\gamma/(1 - \gamma\epsilon)$ of (47). □

Appendix: A matrix bound

Lemma 3. *Given matrices Q_* and B_* where Q_* is symmetric, suppose that*

$$w^{\mathsf{T}} Q_* w \ge \alpha\|w\|^2 \quad \textit{whenever } B_* w = 0, \quad w \in \mathbf{R}^n. \tag{69}$$

Then given any $\delta > 0$, *there exists* $\sigma > 0$ *and neighborhoods* $\mathcal{B}$ *of* B_* *and* $\mathcal{Q}$ *of* Q_* *such that*

$$v^{\mathsf{T}}\left(Q + \frac{1}{\rho}B^{\mathsf{T}}B\right)v \geq (\alpha - \delta)\|v\|^2$$

for all $v \in \mathbf{R}^n$, $0 < \rho \leq \sigma$, $B \in \mathcal{B}$, *and* $Q \in \mathcal{Q}$.

Proof: If w lies in the null space of B_*, then

$$w^{\mathsf{T}}(Q_* + B_*^{\mathsf{T}}B_*/\rho)w \geq \alpha\|w\|^2$$

by (69). There exists a scalar $\tau > 0$ such that $\|B_*u\| \geq \tau\|u\|$ for all u in the row space of B_*. Hence, for u in the row space of B_*, we have

$$u^{\mathsf{T}}(Q_* + B_*^{\mathsf{T}}B_*/\rho)u = u^{\mathsf{T}}Q_*u + \|B_*u\|^2/\rho \geq (\tau^2/\rho - \|Q_*\|)\|u\|^2.$$

An arbitrary vector in $v \in \mathbf{R}^n$ has the orthogonal decomposition $v = u + w$ where u is in the row space of B_* and w is in the null space of B_*. Since $B_*w = 0$, it follows that

$$\begin{aligned} v^{\mathsf{T}}(Q_* + B_*^{\mathsf{T}}B_*/\rho)v &= (u + w)^{\mathsf{T}}(Q_* + B_*^{\mathsf{T}}B_*/\rho)(u + w) \\ &= w^{\mathsf{T}}Q_*w + u^{\mathsf{T}}(Q_* + B_*^{\mathsf{T}}B_*/\rho)u + 2u^{\mathsf{T}}Q_*w \\ &\geq \alpha\|w\|^2 + \left(\frac{\tau^2}{\rho} - \|Q_*\|\right)\|u\|^2 - 2\|u\|\|w\|\|Q_*\|. \end{aligned} \tag{70}$$

Utilizing the inequality

$$ab \leq \epsilon a^2 + b^2/4\epsilon,$$

with $a = \|w\|$ and $b = 2\|u\|\|Q_*\|$ gives

$$2\|u\|\|w\|\|Q_*\| \leq \epsilon\|w\|^2 + \|Q_*\|^2\|u\|^2/\epsilon.$$

Inserting this in (70), we have

$$v^{\mathsf{T}}(Q_* + B_*^{\mathsf{T}}B_*/\rho)v \geq (\alpha - \epsilon)\|w\|^2 + \left(\frac{\tau^2}{\rho} - \|Q_*\| - \frac{\|Q_*\|^2}{\epsilon}\right)\|u\|^2.$$

Let us choose σ small enough that

$$\frac{\tau^2}{\sigma} - \|Q_*\| - \frac{\|Q_*\|^2}{\epsilon} \geq \alpha - \epsilon.$$

Since $\|w\|^2 = \|u\|^2 + \|v\|^2$, it follows that

$$v^{\mathsf{T}}(Q_* + B_*^{\mathsf{T}}B_*/\rho)v \geq (\alpha - \epsilon)\|v\|^2$$

for all v and for all $0 < \rho \leq \sigma$. Since the expression $Q + B^{\mathsf{T}}B/\sigma$ is a continuous function of B and Q, there exists neighborhoods $\mathcal{Q}$ of Q_* and $\mathcal{B}$ of B_* such that

$$v^{\mathsf{T}}(Q + B^{\mathsf{T}}B/\sigma)v \geq (\alpha - 2\epsilon)\|v\|^2$$

for all v and for all $Q \in \mathcal{Q}$ and $B \in \mathcal{B}$. When $0 < \rho \leq \sigma$, we have

$$v^{\mathsf{T}}(Q + B^{\mathsf{T}}B/\rho)v \geq v^{\mathsf{T}}(Q + B^{\mathsf{T}}B/\sigma)v \geq (\alpha - 2\epsilon)\|v\|^2.$$

Taking $\delta = 2\epsilon$, the proof is complete. □

References

1. D.P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Academic Press: New York, 1982.
2. F.H. Clarke, "Generalized gradients and applications," Trans. of the Amer. Math. Soc., vol. 205, pp. 247–262, 1975.
3. A.L. Dontchev and W.W. Hager, "Lipschitzian stability in nonlinear control and optimization," SIAM J. Control Optim., vol. 31, pp. 569–603, 1993.
4. A.L. Dontchev, W.W. Hager, A.B. Poore, and B. Yang, "Optimality, stability and convergence in nonlinear control," Appl. Math. Optim., vol. 31, pp. 297–326, 1995.
5. A.L. Dontchev and W.W. Hager, "The Euler approximation in state constrained optimal control," Department of Mathematics, University of Florida, Gainesville, FL 32611, November, 1997.
6. A.L. Dontchev and W.W. Hager, "Lipschitzian stability for state constrained nonlinear optimal control," SIAM J. Control Optim., vol. 35, pp. 698–718, 1998.
7. A.L. Dontchev and R.T. Rockafellar, "Characterizations of strong regularity for variational inequalities over polyhedral convex sets," SIAM J. Optim., vol. 6, pp. 1087–1105, 1996.
8. I. Ekeland and R. Temam, Convex Analysis and Variational Problems, North-Holland: Amsterdam, 1976.
9. A. Fischer, "Modified Wilson method for nonlinear programs with nonunique multipliers," Technische Universität Dresden, Germany, February, 1997.
10. W.W. Hager, "Lipschitz continuity for constrained processes," SIAM J. Control Optim., vol. 17, pp. 321–338, 1979.
11. W.W. Hager, "Approximations to the multiplier method," SIAM J. Numer. Anal., vol. 22, pp. 16–46, 1985.
12. W.W. Hager, "Convergence of Wright's stabilized SQP algorithm," Mathematics Department, University of Florida, Gainesville, FL 32611, January, 1997.
13. W.W. Hager and M.S. Gowda, "Stability in the presence of degeneracy and error estimation," Mathematics Department, University of Florida, Gainesville, FL 32611, November 22, 1997 (to appear in Math. Programming).
14. A.J. Hoffman, "On approximate solutions of systems of linear inequalities," J. Res. Nat. Bur. Standards, vol. 49, pp. 263–265, 1952.
15. O.L. Mangasarian and S. Fromovitz, "The Fritz-John necessary optimality conditions in the presence of equality and inequality constraints," J. of Math. Anal. and Appl., vol. 17, pp. 37–47, 1967.
16. S.M. Robinson, "Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear-programming algorithms," Math. Programming, vol. 7, pp. 1–16, 1974.
17. S.M. Robinson, "Strongly regular generalized equations," Math. Oper. Res., vol. 5, pp. 43–62, 1980.
18. R.T. Rockafellar, "The multiplier method of Hestenes and Powell applied to convex programming," J. Optim. Theory Appl., vol. 12, pp. 555–562, 1973.
19. S.J. Wright, "Superlinear convergence of a stabilized SQP method to a degenerate solution," Comput. Optim. Appl., vol. 11, pp. 253–275, 1998.