CERFACS **Technical Report No. TR/PA/99/36**

Stephen J. Wright* · Dominique Orban

Properties of the Log-Barrier Function on Degenerate Nonlinear Programs

May 28, 2001

Abstract. We examine the sequence of local minimizers of the log-barrier function for a nonlinear program near a solution at which second-order sufficient conditions and the Mangasarian-Fromovitz constraint qualifications are satisfied, but the active constraint gradients are not necessarily linearly independent. When a strict complementarity condition is satisfied, we show uniqueness of the local minimizer of the barrier function in the vicinity of the nonlinear program solution, and obtain a semi-explicit characterization of this point. When strict complementarity does not hold, we obtain several other interesting characterizations, in particular, an estimate of the distance between the minimizers of the barrier function and the nonlinear program in terms of the barrier parameter, and a result about the direction of approach of the sequence of minimizers of the barrier function to the nonlinear programming solution.

1. Introduction

We consider the nonlinear programming problem

$$\min f(x) \qquad \text{subject to} \qquad c(x) \geq 0, \tag{1.1}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ and $c : \mathbf{R}^n \to \mathbf{R}^m$ are smooth (twice Lipschitz continuously differentiable) functions. We assume that second-order sufficient conditions hold at a point x^*, so that x^* is a strict local solution of (1.1).

The logarithmic barrier function for (1.1) is

$$P(x;\mu) = f(x) - \mu \sum_{i=1}^{m} \log c_i(x). \tag{1.2}$$

Under conditions assumed in this paper, and discussed in detail below, this function has a local minimizer near x^* for all μ sufficiently small. Methods based on (1.2) find approximations to the minimizer of $P(\cdot;\mu_k)$, which we denote by $x(\mu_k)$, for some sequence $\{\mu_k\}$ with $\mu_k \downarrow 0$, usually by applying some variant of Newton's method. Extrapolation techniques are sometimes used to find an appropriate initial guess after each change in μ.

In this paper, we examine properties of the sequence of minimizers of $P(\cdot;\mu)$ for small μ, in the vicinity of x^*. Previous analyses have assumed that the active constraint gradients are linearly independent at the solution—the so-called linear independence constraint qualification. By contrast,

Stephen J. Wright: Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439; wright@mcs.anl.gov

Dominique Orban: CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cédex 1, France; orban@cerfacs.fr

* Research of this author supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, U.S. Department of Energy, under Contract W-31-109-Eng-38. Travel support was provided by NATO grant CRG 960688.

we make the weaker assumption that the Mangasarian-Fromovitz constraint qualification holds. This more general condition, which is equivalent to boundedness of the set of optimal Lagrange multipliers, has been used by many authors in studying the local convergence analysis of nonlinear optimization and complementarity problems and the stability of their solutions. In Section 3, we examine the case in which at least one of the optimal multipliers satisfies the strict complementarity condition. In this case, the path of minimizers of $P(\cdot;\mu)$ behaves similarly to the case of linearly independent constraints: The minimizers are locally unique, the path traced by the minimizers is smooth (as a function of μ) with a well-defined derivative, and the corresponding sequence of multiplier estimates approaches the analytic center of the multiplier set. In Section 4, we consider the case in which the strict complementarity condition does not hold. In this case, the path traced by the set of minimizers takes on a quite different character. We prove an existence result, derive an estimate of the distance between the minimizer of $P(\cdot;\mu_k)$ in terms of μ_k, and show that any path of minimizers that converges to x^* must approach this point tangentially to the strongly active constraints.

The previous literature on the log-barrier function and properties of the minimizers of $P(\cdot;\mu)$ is plentiful. The seminal book of Fiacco and McCormick [9] presents general results about the existence of minimizers of the barrier function in the vicinity of x^* and the convergence of the minimizer sequence to x^* as $\mu_k \downarrow 0$ [9, Theorem 8]. It also shows that the path of minimizers of $P(\cdot;\mu)$ is isolated and smooth when the active constraint gradients are linearly independent and strict complementarity holds [9, Sections 5.1, 5.2]. Adler and Monteiro [1] analyze the trajectories produced by minimizers of the log-barrier function in the case of linear programming. The differences in formulation and the linearity of the problem make it difficult to relate the results of Adler and Monteiro to those of this paper. However, their Theorem 3.2 corresponds to our observation that the Lagrange multiplier estimates converge to the analytic center of the optimal multiplier set, while their Theorem 5.4 corresponds to our Theorem 3.3 in describing the direction of approach of the trajectory of minimizers to x^*. The convexity and lack of curvature in their problem gives the results a significantly different flavor, however, and their proof techniques depend strongly on the constancy of the closed subspace spanned by the active constraint gradients in the vicinity of x^*, which does not occur in (1.1) under our assumptions.

Recently, McCormick and Witzgall [17] examined the behavior of the log barrier trajectory for the case in which either the primal or dual solutions are nonunique. They assume a convex programming problem (that is, f and $-c_i$, $i = 1, 2, \dots, m$ are convex functions) and the Slater constraint qualification, which is equivalent to the Mangasarian-Fromovitz constraint qualification for convex problems. Apart from their somewhat different assumptions, Theorem 2 in [17] is similar to our Theorem 3.2, in that both show convergence of the Lagrange multiplier estimates to the analytic center of the optimal multiplier set when a strict complementarity condition holds. In other respects the aspects addressed in [17] are somewhat different from those addressed here. Unlike in [17], we do not consider nonunique primal solutions. On the other hand, we do tackle existence and uniqueness issues for minimizers of $P(\cdot;\mu)$ for sufficiently small μ (which are more complex in the nonconvex case than in the convex case), present a complete characterization of the direction of approach of the minimizers of $P(\cdot;\mu)$ to the solution x^* as $\mu \downarrow 0$ (Theorem 3.3), derive an estimate of the distance of multiplier estimates to the optimal multiplier set (Lemma 2.3), and prove a relationship between μ and the distance of a minimizer of $P(\cdot;\mu)$ to the optimal multiplier set in the non-strict complementarity case (Section 4).

The geometry of the central path has been studied extensively in a number of contexts in which convexity or monotonicity properties are present. Behavior of the (primal-dual) central path in the absence of strict complementarity has been studied for monotone complementarity problems by Kojima, Mizuno, and Noma [15] and Monteiro and Tsuchiya [22]. In [15], the authors work with nonlinear and linear monotone complementarity problems, and prove existence of the central path and (in the linear case) convergence of this path to the analytic center of the solution set. Linear monotone complementarity problems are discussed in [22], where the authors prove results related to those of Section 4 of this paper. Specifically, they show that the distance of the point parametrized by μ on the central path to its limit in the solution set varies like $\mu^{1/2}$ when strict complementarity does not hold. Monteiro and Zhou [23] discuss existence of the central path for a more general class of convex problems. Earlier fundamental work on the central path was performed by McLinden [18,19] and Megiddo [20].

Although their focus is on the log-barrier function, Fiacco and McCormick [9] actually consider a more general class of barrier functions, and also derive results for the case in which equality constraints are represented by quadratic penalty terms. Nesterov and Nemirovskii [25] study the general class of self-concordant barriers of which the log barrier is a particular instance. Following the results of Murray [24] and Lootsma [16] regarding the ill conditioning of the Hessian matrix $P_{xx}(\cdot;\mu)$ along the central path, the nature of the ill conditioning in the neighborhood of the solution is examined further by M. H. Wright [31]. The latter paper proposes techniques for calculating approximate Newton steps for the function $P(\cdot;\mu)$ that do not require the solution of ill-conditioned systems. In earlier work, Gould [13] proposed a method for computing accurate Newton steps by identifying the active indices explicitly, and forming an augmented linear system that remains well conditioned even when μ is small. The effect of finite-precision arithmetic on the calculation of Newton steps is examined by M. H. Wright [32]. Both M. H. Wright [31,32] and S. J. Wright [36] use a subspace decomposition of the Hessian $P_{xx}(\cdot;\mu)$ like the one used in Section 3 below, but there is an important distinction that we note later. The paper [36] and also Villalobos, Tapia, and Zhang [29] address the issue of domain of convergence of Newton's method applied to $P(\cdot;\mu)$, which is also addressed in Theorem 3.1 below.

The Mangasarian-Fromovitz constraint qualification has been used in place of the standard assumption of linear independence of the constraint gradients in several recent works on nonlinear programming. Ralph and Wright [28] describe a path-following method for convex nonlinear programming that achieves superlinear local convergence under this condition. S. J. Wright [35,33] and Anitescu [2] study the local convergence of sequential quadratic programming methods under this assumption.

Our paper concludes with comments about two important issues: Convergence of the Newton/log-barrier method, in which Newton's method is used to find an approximate minimizer of $P(\cdot;\mu_k)$ for each μ_k, and relevance of our results to primal-dual methods, which generate iterates with both primal and dual (Lagrange multiplier) components rather than primal components alone. Detailed study of these topics is left to future work.

2. Assumptions, Notation, and Basic Results

2.1. Assumptions

In this section, we specify the optimality conditions for the nonlinear program (1.1) and outline our assumptions on the solution x^*.

Assume first that the functions f and c are twice Lipschitz continuously differentiable in the neighborhood of interest. The Lagrangian function for (1.1) is

$$\mathcal{L}(x,\lambda) = f(x) - \lambda^T c(x), \tag{2.1}$$

where λ is the vector of Lagrange multipliers. Necessary conditions for x^* to be a solution of (1.1) are that there exists a Lagrange multiplier vector λ^* such that

$$c(x^*) \geq 0, \quad \lambda^* \geq 0, \quad (\lambda^*)^T c(x^*) = 0, \quad \mathcal{L}_x(x^*, \lambda^*) = 0. \tag{2.2}$$

The active constraints are the components of c for which $c_i(x^*) = 0$. Without loss of generality we assume these to be the first q components of c, so that

$$c_i(x^*) = 0 \quad i = 1, 2, \ldots, q, \tag{2.3a}$$
$$c_i(x^*) > 0,\ \lambda_i^* = 0, \quad i = q+1, \ldots, m. \tag{2.3b}$$

We define U_R to be an orthonormal matrix of dimensions $n \times \bar{q}$ for some $\bar{q} \leq q$ whose columns span the range space of the active constraint gradients, that is,

$$\text{Range } U_R = \text{Range } \{\nabla c_i(x^*) \,|\, i = 1, 2, \ldots, q\}. \tag{2.4}$$

We let U_N denote an orthonormal matrix of dimensions $n \times (n - \bar{q})$ whose columns span the space of vectors orthogonal to $\nabla c_i(x^*)$ for all $i = 1, 2, \ldots, q$. By the fundamental theorem of algebra, we have that

$$\begin{bmatrix} U_R \, U_N \end{bmatrix} \text{ is orthogonal.} \tag{2.5}$$

We assume that the *Mangasarian-Fromovitz constraint qualification (MFCQ)* holds at x^*, which is that there is a vector p such that

$$\nabla c_i(x^*)^T p < 0, \quad i = 1, 2, \ldots, q. \tag{2.6}$$

The stronger linear independence constraint qualification (LICQ), which assumes linear independence of the vectors $\nabla c_i(x^*)$, $i = 1, 2, \ldots, q$, is used by M. H. Wright [30], Fiacco and McCormick [9], and S. J. Wright [36], for instance. Unlike LICQ, MFCQ does not imply uniqueness of λ^*. We can use (2.1) and (2.2) to express the conditions on λ^* as

$$\nabla f(x^*) = \textstyle\sum_{i=1}^q \lambda_i^* \nabla c_i(x^*), \quad \lambda_i^* \geq 0,\ i = 1, 2, \ldots, q, \tag{2.7a}$$
$$\lambda_i^* = 0,\ i = q+1, \ldots, m. \tag{2.7b}$$

We define $\mathcal{S}_\lambda$ to be the set of multipliers satisfying these conditions at x^*, that is,

$$\mathcal{S}_\lambda \stackrel{\triangle}{=} \{\lambda^* \,|\, (x^*, \lambda^*) \text{ satisfy (2.2)}\}. \tag{2.8}$$

Gauvin [11, Theorem 1] shows that the condition (2.6) is equivalent to boundedness of the set $\mathcal{S}_\lambda$. We conclude from (2.7) that $\mathcal{S}_\lambda$ is a bounded polyhedral set.

The *strict complementarity* condition is that

$$\lambda_i^* + c_i(x^*) > 0, \qquad i = 1, 2, \ldots, m, \tag{2.9}$$

for at least one $\lambda^* \in \mathcal{S}_\lambda$. This condition is assumed in Section 3. When it holds, we can define the analytic center $\bar{\lambda}^*$ of $\mathcal{S}_\lambda$ to be

$$\bar{\lambda}^* = \arg \min_{\lambda^* \in \Lambda^*} -\sum_{i=1}^q \ln \lambda_i^*, \tag{2.10}$$

where Λ^* is the set of strictly complementary multipliers, that is,

$$\nabla f(x^*) = \textstyle\sum_{i=1}^{q} \lambda_i^* \nabla c_i(x^*), \quad \lambda_i^* > 0, \; i = 1, 2, \ldots, q, \tag{2.11a}$$
$$\lambda_i^* = 0, \; i = q+1, \ldots, m. \tag{2.11b}$$

Since the problem (2.10), (2.11) has a smooth, strictly convex objective and a convex bounded feasible set, it has a unique minimizer $\bar{\lambda}^*$ whose components $1, 2, \ldots, q$ are characterized by the first-order conditions, which is that there exists a vector $\zeta \in \mathbf{R}^n$ such that

$$\frac{1}{\bar{\lambda}_i^*} = \nabla c_i(x^*)^T \zeta > 0, \quad i = 1, 2, \ldots, q. \tag{2.12}$$

(Were the MFCQ *not* satisfied, there would not exist a vector ζ satisfying $\nabla c_i(x^*)^T \zeta > 0$, $i = 1, 2, \ldots, q$, and so the problem (2.10), (2.11) would have no solution.) Note that ζ is defined by (2.12) only up to a term in the null space of the active constraint gradients. In other words, if we decompose ζ as

$$\zeta = U_R \zeta_R + U_N \zeta_N, \tag{2.13}$$

where U_R and U_N are defined as in (2.4), (2.5), the formula (2.12) defines ζ_R uniquely while leaving ζ_N completely free. However, we will see in Section 3 that we can define ζ_N in such a way that ζ has particularly interesting properties.

Finally, we assume that the following second-order sufficient conditions for optimality are satisfied:

$$y^T \mathcal{L}_{xx}(x^*, \lambda^*) y > 0, \quad \text{for all } \lambda^* \in \mathcal{S}_\lambda \tag{2.14}$$
$$\text{and all } y \neq 0 \text{ with } \nabla f(x^*)^T y = 0 \text{ and } \nabla c_i(x^*)^T y \geq 0 \text{ for all } i = 1, 2, \ldots, q.$$

The conditions (2.2), (2.6), and (2.14) together imply that there is a $\eta > 0$ such that

$$f(x) - f(x^*) \geq \eta \|x - x^*\|^2, \;\; \text{for all feasible } x \text{ sufficiently close to } x^*; \tag{2.15}$$

see for example Bonnans and Ioffe [5].

When the strict complementarity condition (2.9) is satisfied, the second-order conditions are equivalent to the following:

$$y^T \mathcal{L}_{xx}(x^*, \lambda^*) y > 0, \quad \text{for all } \lambda^* \in \mathcal{S}_\lambda \tag{2.16a}$$
$$\text{and all } y \neq 0 \text{ with } \nabla c_i(x^*)^T y = 0 \text{ for all } i = 1, 2, \ldots, q. \tag{2.16b}$$

Using the matrix U_N defined in (2.5), we can rewrite these conditions as follows:

$$U_N^T \mathcal{L}_{xx}(x^*, \lambda^*) U_N \text{ positive definite, for all } \lambda^* \in \mathcal{S}_\lambda. \tag{2.17}$$

When the condition (2.9) is *not* satisfied, we can identify indices in $\{1, 2, \ldots, q\}$ such that $\lambda_i^* = 0$ for all $\lambda^* \in \mathcal{S}_\lambda$. We assume WLOG that these indices are $\bar{q}+1, \bar{q}+2, \ldots, 1$ for some $\bar{q} < q$. The second-order sufficient conditions (2.14) can then be rewritten as follows:

$$d^T \mathcal{L}_{xx}(x^*, \lambda^*) d > 0, \quad \text{for all } \lambda^* \in \mathcal{S}_\lambda \tag{2.18a}$$
$$\text{and all } d \neq 0 \text{ with } \nabla c_i(x^*)^T d = 0 \text{ for all } i = 1, 2, \ldots, \bar{q} \tag{2.18b}$$
$$\text{and } \nabla c_i(x^*)^T d \geq 0 \text{ for all } i = \bar{q}+1, \ldots, q. \tag{2.18c}$$

2.2. Notation

We use the following notation in the rest of the paper. For related positive quantities α and β, we say $\beta = O(\alpha)$ if there is a constant M such that $\beta \le M\alpha$ for all α sufficiently small. We say that $\beta = o(\alpha)$ if $\beta/\alpha \to 0$ as $\alpha \to 0$, $\beta = \Omega(\alpha)$ if $\alpha = O(\beta)$, and $\beta = \Theta(\alpha)$ if $\beta = O(\alpha)$ and $\alpha = O(\beta)$. It follows that the expression $\beta = O(1)$ means that $\beta \le M$ for some constant M and all values of β in the domain of interest.

For any real number β we define

$$\beta_- = \max(0, -\beta), \quad \beta_+ = \max(0, \beta). \tag{2.19}$$

For a given value of μ, we define a local minimizer of $P(\cdot;\mu)$ close to x^* generically by $x(\mu)$. (The uniqueness or at least specialness of this point is made clear in subsequent discussions.)

2.3. Basic Results

Given any strictly feasible point x and any positive value of the barrier parameter μ in (1.2), we define a vector of Lagrange multiplier estimates $\lambda(x,\mu)$ by

$$\lambda(x,\mu) = \mu C(x)^{-1}e = \left[\frac{\mu}{c_1(x)}, \dots, \frac{\mu}{c_m(x)}\right]^T, \tag{2.20}$$

where $C(x) = \operatorname{diag}(c_1(x), c_2(x), \dots, c_m(x))$. In the special case where $x = x(\mu)$ (defined above), we write

$$\lambda(\mu) \triangleq \lambda(x(\mu),\mu). \tag{2.21}$$

Moreover, in the discussions of Sections refcentral and 4, we frequently use z^k to denote a local minimizer of $P(\cdot;\mu_k)$ for some $\mu_k > 0$. In this case we use the following notation for the corresponding multiplier estimate:

$$\lambda^k \triangleq \lambda(z^k,\mu_k). \tag{2.22}$$

For future reference, we write the derivatives of the barrier function (1.2) as follows:

$$P_x(x;\mu) = \nabla f(x) - \sum_{i=1}^m \frac{\mu}{c_i(x)} \nabla c_i(x), \tag{2.23a}$$

$$P_{xx}(x;\mu) = \nabla^2 f(x) + \mu \sum_{i=1}^m \left[\frac{1}{c_i^2(x)} \nabla c_i(x) \nabla c_i(x)^T - \frac{1}{c_i(x)} \nabla^2 c_i(x)\right]. \tag{2.23b}$$

By combining (2.20) with (2.23a), we obtain

$$\nabla f(x) = \sum_{i=1}^m \lambda_i(x,\mu) \nabla c_i(x) + P_x(x;\mu), \tag{2.24}$$

while for the case $x = x(\mu)$, we have from (2.20) that

$$\nabla f(x(\mu)) = \sum_{i=1}^m \lambda_i(x(\mu),\mu) \nabla c_i(x(\mu)).$$

We denote by $\mathcal{C}$ the feasible set for (1.1), that is,

$$\mathcal{C} \stackrel{\triangle}{=} \{x \,|\, c(x) \geq 0\},$$

and by strict $\mathcal{C}$ we denote the set of points at which the inequalities are satisfied strictly, that is,

$$\text{strict } \mathcal{C} \stackrel{\triangle}{=} \{x \,|\, c(x) > 0\}.$$

It is easy to show that, under the MFCQ assumption (2.6), there is a neighborhood of x^* within which strict $\mathcal{C}$ coincides with int $\mathcal{C}$.

Lemma 2.1. *Suppose that (2.6) is satisfied. Then there is a neighborhood $\mathcal{N}$ of x^* such that*

$$\text{strict } \mathcal{C} \cap \mathcal{N} = \text{int } \mathcal{C} \cap \mathcal{N} \neq \emptyset,$$

and x^* lies in the closure of strict $\mathcal{C}$*.*

Proof. Choose $\mathcal{N}$ such that (2.6) continues to hold whenever x^* is replaced by x, for all $x \in \mathcal{C} \cap \mathcal{N}$, while the constraints $q+1, \ldots, m$ remain inactive throughout $\mathcal{N}$. We prove the result by showing that strict $\mathcal{C} \cap \mathcal{N} \subset$ int $\mathcal{C} \cap \mathcal{N}$, and then the converse.

Consider some $x \in$ strict $\mathcal{C} \cap \mathcal{N}$. By continuity of c, we can choose $\delta > 0$ such that the open Euclidean ball of radius δ around x, denoted by $\mathcal{B}_\delta(x)$, satisfies $\mathcal{B}_\delta(x) \subset \mathcal{N}$ and $c(z) > 0$ for all $z \in \mathcal{B}_\delta(x)$. Hence, $z \in$ strict $\mathcal{C} \cap \mathcal{N} \subset \mathcal{C}$, and therefore $x \in$ int $\mathcal{C}$.

Now consider a point $x \in \mathcal{N} \backslash$strict $\mathcal{C}$. If $x \notin \mathcal{C}$, then clearly $x \notin$ int $\mathcal{C} \cap \mathcal{N}$, and we are done. Otherwise, we have $c_j(x) = 0$ for some $j \in \{1, 2, \ldots, q\}$. Consider now the points $x - \alpha p$ for p defined in (2.6) and α small and positive. We have by continuity of ∇c_j that

$$c_j(x - \alpha p) = c_j(x) - \alpha \nabla c_j(x)^T p + o(\alpha) < 0,$$

for all $\alpha > 0$ sufficiently small. Therefore, $x - \alpha p \notin \mathcal{C}$, so that $x \notin$ int $\mathcal{C}$.

The claim that strict $\mathcal{C} \cap \mathcal{N} \neq \emptyset$ is proved by considering points of the form $x^* + \alpha p$, for $\alpha > 0$ and p satisfying (2.6). Consideration of the same set of points demonstrates that x^* lies in the closure of strict $\mathcal{C}$. ∎

We now show boundedness of the Lagrange multiplier estimates arising from approximate minimization of $P(\cdot; \mu)$.

Lemma 2.2. *Suppose that the first-order necessary conditions (2.2) and the MFCQ condition (2.6) hold. Then there are positive constants r and ϵ such that the following property holds: For any $\beta_1 \geq 0$ there is β_2 such that for all x and μ with*

$$\|x - x^*\| \leq r, \quad x \textit{ feasible}, \quad \mu \in (0, \epsilon], \quad \|P_x(x; \mu)\| \leq \beta_1, \tag{2.25}$$

we have $\|\lambda(x, \mu)\| \leq \beta_2$.

Proof. Let ϵ be any positive number, and choose r small enough that there are positive constants γ_1 and γ_2 such that for all x with $\|x - x^*\| \leq r$ we have

$$\nabla c_i(x)^T p \leq -\gamma_1, \quad i = 1, 2, \ldots, q, \tag{2.26}$$

$$c_i(x) \geq \gamma_2, \quad i = q+1, \ldots, m, \tag{2.27}$$

where p is the vector from (2.6).

Suppose for some β_1 that there is no β_2 with the claimed property. Then we can define sequences $\{x^k\}$ and $\{\mu_k\}$ such that (2.25) holds for $x = x^k$ and $\mu = \mu_k$ for all k, yet $\|\lambda(x^k, \mu_k)\| \uparrow \infty$. Defining

$$\bar{\lambda}^k \triangleq \frac{\lambda(x^k, \mu_k)}{\|\lambda(x^k, \mu_k)\|},$$

we can assume without loss of generality (by compactness of the unit ball) that $\bar{\lambda}^k \to \bar{\lambda}$, for some vector $\bar{\lambda}$ such that

$$\bar{\lambda} \geq 0, \quad \|\bar{\lambda}\| = 1.$$

Note that because of (2.27) and $\mu_k \leq \epsilon$, we have that

$$i = q+1, \ldots, m \;\Rightarrow\; \bar{\lambda}_i^k = \frac{\mu_k}{c_i(x^k)\|\lambda(x^k, \mu_k)\|} \leq \frac{\epsilon}{\gamma_2 \|\lambda(x^k, \mu_k)\|} \to 0,$$

so that

$$\bar{\lambda}_i = 0, \quad i = q+1, \ldots, m. \tag{2.28}$$

By compactness of $\{x \,|\, \|x - x^*\| \leq r\}$, we can choose a further subsequence if necessary and define a vector $\bar{x}$ with $\|\bar{x} - x^*\| \leq r$ such that $x^k \to \bar{x}$.

From (2.20), (2.23a), our assumption that $\|\lambda(x^k, \mu_k)\| \uparrow \infty$, and (2.28), we have that

$$\begin{aligned} 0 &= \lim_{k\to\infty} \|\lambda(x^k, \mu_k)\|^{-1} P_x(x^k; \mu_k) \\ &= -\lim_{k\to\infty} \sum_{i=1}^{m} \bar{\lambda}_i^k \nabla c_i(x^k) + \lim_{k\to\infty} \|\lambda(x^k, \mu_k)\|^{-1} \nabla f(x^k) \\ &= -\sum_{i=1}^{q} \bar{\lambda}_i \nabla c_i(\bar{x}). \end{aligned}$$

Since $\bar{x}$ satisfies the property (2.26), we have by taking inner products of this last expression with p that

$$0 = \sum_{i=1}^{q} \bar{\lambda}_i \nabla c_i(\bar{x})^T p \leq -\gamma_1 \sum_{i=1}^{q} \bar{\lambda}_i,$$

which together with $\bar{\lambda} \geq 0$ implies that $\bar{\lambda} = 0$. This contradicts $\|\bar{\lambda}\| = 1$. We conclude that the sequence of multiplier estimates $\{\lambda(x^k, \mu_k)\}$ must remain bounded, so it must be possible to choose a constant β_2, as claimed. ∎

A slight extension of this result shows conditions under which the sequence of multiplier estimates converges to the optimal multiplier set $\mathcal{S}_\lambda$.

Lemma 2.3. *Suppose that the first-order necessary conditions (2.2) and the MFCQ condition (2.6) hold. Then, given any sequences $\{\mu_k\}$ and $\{x^k\}$ with $x^k \to x^*$, $\mu_k \downarrow 0$, and $P_x(x^k; \mu_k) \to 0$, the sequence of multiplier estimates $\lambda(x^k, \mu_k)$ defined by (2.20) satisfies*

$$\operatorname{dist}_{\mathcal{S}_\lambda} \lambda(x^k, \mu_k) = O\left(\left\|x^k - x^*\right\|\right) + O(\mu_k) + O\left(\left\|P_x(x^k; \mu_k)\right\|\right).$$

Proof. From Lemma 2.2, we have that the sequence $\lambda(x^k; \mu_k)$ is bounded. Therefore, we have

$$\begin{aligned} P_x(x^k; \mu_k) &= \nabla f(x^k) - \sum_{i=1}^{m} \frac{\mu_k}{c_i(x^k)} \nabla c_i(x^k) \\ &= \nabla f(x^*) - \sum_{i=1}^{q} \lambda_i(x^k, \mu_k) \nabla c_i(x^*) + O(\mu_k) + O(\|x^k - x^*\|). \end{aligned}$$

By comparing this expression with the definition (2.7), (2.8) of $\mathcal{S}_\lambda$, and applying Hoffmann's lemma [14], we obtain the desired result. ∎

3. Behavior of the Central Path near x^* Under Strict Complementarity

In this section, we examine the properties of a path of exact minimizers $x(\mu)$ of $P(x;\mu)$. We prove that $x(\mu)$ exists and is unique for all sufficiently small μ, and that the path of minimizers

$$\mathcal{P} \stackrel{\triangle}{=} \{x(\mu), \mu > 0\}$$

is smooth and approaches x^* along a direction ζ of the form (2.13) with interesting properties. We refer to $\mathcal{P}$ as the *primal central path.*

We start in Section 3.1 by defining the direction of approach ζ and providing some motivation for this quantity in terms of the primal-dual optimality conditions. In Section 3.2, we show existence of a minimizer of $P(\cdot;\mu)$ for all μ sufficiently small in a somewhat unorthodox way; namely, by proving that Newton's method applied to $P(\cdot;\mu)$ from all starting points in a certain ball converges to a point $x(\mu)$ and that this point must be a strict local minimizer of $P(\cdot;\mu)$. The main result in this section is Theorem 3.1. We then prove in Theorem 3.2 of Section 3.3 that $x(\mu)$ is unique in the sense that there cannot be any other minimizers of $P(\cdot;\mu)$ near x^*, for small μ. Finally, in Section 3.4, we show that the direction of approach of $x(\mu)$ to x^* as $\mu \downarrow 0$ is indeed ζ.

3.1. Motivation

We introduced the vector ζ in (2.12) as the Lagrange multiplier in the problem of finding the analytic center of the optimal multiplier set $\mathcal{S}_\lambda$. Decomposing ζ as (2.13), we noted that (2.12) defined the ζ_R component. If we define the ζ_N component to be the solution of

$$U_N^T \mathcal{L}_{xx}(x^*, \bar{\lambda}^*) U_N \zeta_N = -U_N^T \mathcal{L}_{xx}(x^*, \bar{\lambda}^*) U_R \zeta_R + \sum_{i=q+1}^{m} \frac{1}{c_i(x^*)} U_N^T \nabla c_i(x^*), \tag{3.1}$$

we find that ζ takes on an additional significance as the direction of approach of the primal central path to x^*. We sketch a justification of this claim here. Note from (2.23a), (2.20), (2.21), and (2.1) that the minimizer $x(\mu)$ of $P(\cdot;\mu)$ and its associated Lagrange multiplier $\lambda(\mu)$ satisfy the following system of nonlinear equations:

$$\begin{bmatrix} \mathcal{L}_x(x,\lambda) \\ C(x)\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \mu e \end{bmatrix},$$

in addition to the feasibility conditions $c(x) > 0$ and $\lambda > 0$. By differentiating this system at $(x(\mu), \lambda(\mu))$ with respect to μ, we obtain

$$\mathcal{L}_{xx}(x(\mu), \lambda(\mu))\dot{x}(\mu) - \sum_{i=1}^{m} \dot{\lambda}_i(\mu) \nabla c_i(x(\mu)) = 0, \tag{3.2a}$$

$$\lambda_i(\mu) \nabla c_i(x(\mu))^T \dot{x}(\mu) + \dot{\lambda}_i(\mu) c_i(x(\mu)) = 1, \quad i = 1, \ldots, m. \tag{3.2b}$$

Suppose for the sake of our present argument that $\lim_{\mu\downarrow 0} \lambda(\mu) = \bar{\lambda}^*$ (2.10) and that the derivatives $\dot{\lambda}(\mu)$ and $\dot{x}(\mu)$ are bounded for μ close to zero. For the active indices, we have by taking limits in (3.2b) that

$$\nabla c_i(x^*)^T \dot{x}(0_+) = 1/\bar{\lambda}_i^* > 0, \quad i = 1, \ldots, q, \tag{3.3}$$

where $\dot{x}(0_+)$ denotes $\lim_{\mu\downarrow 0} \dot{x}(\mu)$. For the inactive indices, we have that

$$\dot{\lambda}_i(0_+) = \frac{1}{c_i(x^*)}, \quad i = q+1, \dots, m. \tag{3.4}$$

Using U_N and U_R defined in (2.4), (2.5), we have by premultiplying (3.2a) by U_N^T and substituting (3.4) that

$$U_N^T \mathcal{L}_{xx}(x^*, \bar{\lambda}^*)\dot{x}(0_+) - \sum_{i=q+1}^{m} \frac{1}{c_i(x^*)} U_N^T \nabla c_i(x^*) = 0. \tag{3.5}$$

By decomposing $\dot{x}(0_+) = U_R \dot{x}_R(0_+) + U_N \dot{x}_N(0_+)$, we observe that (3.3), (3.5) can be identified with (2.12), (3.1) when we identify $\dot{x}(0_+)$ with ζ.

A more formal proof of our claim that $\dot{x}(0_+) = \zeta$ is the main topic of this section, culminating in Theorem 3.3.

3.2. Existence of Minimizers of the Barrier Function

Our main result, Theorem 3.1, shows the existence of a minimizer $x(\mu)$ of $P(\cdot;\mu)$ that lies within a distance $\Theta(\mu)$ of x^*, and characterizes the domain of convergence of Newton's method for $P_x(\cdot;\mu)$ to this minimizer. It derives a first-order estimate of the location of this minimizer, showing that it lies within a distance $O(\mu^2)$ of the point

$$\bar{x}(\mu) = x^* + \mu\zeta, \tag{3.6}$$

where $\mu > 0$ and ζ is the vector that is uniquely specified by the formulae (2.12), (2.13), and (3.1). Our second result, Theorem 3.2, shows that the minimizer $x(\mu)$ is locally unique in a certain sense.

One key to the analysis is the partitioning of $\mathbf{R}^n$ into two orthogonal subspaces, defined by the matrices U_R and U_N of (2.4) and (2.5). This decomposition was also used in the analysis of S. J. Wright [34], but differs from those used by M. H. Wright in [31,32] and S. J. Wright [36], which define these matrices with respect to the active constraint matrix evaluated at the current iterate x, rather than at the solution x^* of (1.1). By using the latter strategy, we avoid difficulties with loss of rank in the active constraint matrix at the solution, which may occur under the MFCQ assumption of this paper, but not under the LICQ assumption used in [31,32,36].

All results in this section use the following assumption.

Assumption 3.1. *At least one constraint is active at the solution x^*, and the first-order necessary conditions (2.2), the strict complementarity condition (2.9), the second-order sufficient conditions (2.16), and the MFCQ (2.6) all hold at x^*.*

Our first lemma concerns the eigenvalues of the Hessian $P_{xx}(x;\mu)$ (2.23b) in a neighborhood of the point $\bar{x}(\mu)$.

Lemma 3.1. *Suppose that Assumption 3.1 holds. There is a positive constant χ_1 such that for all $\mu \in (0, \chi_1]$ and all x with*

$$\|x - (x^* + \mu\zeta)\| = \|x - \bar{x}(\mu)\| \le \chi_1 \mu, \tag{3.7}$$

the following property holds. Denoting

$$\begin{bmatrix} H_{RR} & H_{RN} \\ H_{NR} & H_{NN} \end{bmatrix} = \begin{bmatrix} U_R^T P_{xx}(x;\mu) U_R & U_R^T P_{xx}(x;\mu) U_N \\ U_N^T P_{xx}(x;\mu) U_R & U_N^T P_{xx}(x;\mu) U_N \end{bmatrix}, \tag{3.8}$$

and

$$\begin{bmatrix} J_{11} & J_{12} \\ J_{12}^T & J_{22} \end{bmatrix} = \begin{bmatrix} H_{RR} & H_{RN} \\ H_{RN}^T & H_{NN} \end{bmatrix}^{-1}, \tag{3.9}$$

we have that $P_{xx}(x;\mu)$ *is positive definite (so that* J *is well defined) and that*

$$J_{11} = O(\mu), \quad J_{12} = O(\mu), \quad J_{22} = O(1),$$

where the constants in the $O(\cdot)$ *terms are independent of* χ_1.

Proof. Using the definition (2.23b) together with (2.3) and Taylor's theorem, we obtain

$$\begin{aligned} &P_{xx}(x;\mu) \\ &= \nabla^2 f(x^*) + O(\mu) - \sum_{i=1}^{q} \left[\nabla c_i(x^*)^T\zeta + O(\chi_1)\right]^{-1} \left[\nabla^2 c_i(x^*) + O(\mu)\right] \\ &\quad + \sum_{i=1}^{q} \mu^{-1} \left[\nabla c_i(x^*)^T\zeta + O(\chi_1)\right]^{-2} \Big[\nabla c_i(x^*)\nabla c_i(x^*)^T \\ &\qquad + v_{1i}\nabla c_i(x^*)^T + \nabla c_i(x^*)v_{2i}^T + O(\mu^2)\Big], \end{aligned}$$

where v_{1i} and v_{2i}, $i = 1, 2, \ldots, q$ are vectors that satisfy the estimates

$$v_{1i} = O(\mu), \quad v_{2i} = O(\mu), \quad i = 1, 2, \ldots, q.$$

In expanding the last term, we have used the following relation, which holds for any two vectors a and b:

$$aa^T = bb^T + (a-b)b^T + b(a-b)^T + (a-b)(a-b)^T.$$

By using (2.12), we have that

$$\begin{aligned} &P_{xx}(x;\mu) \\ &= \nabla^2 f(x^*) - \sum_{i=1}^{q} \bar\lambda_i^* \nabla^2 c_i(x^*) + \sum_{i=1}^{q} O(\chi_1)\nabla^2 c_i(x^*) \\ &\quad + \mu^{-1}\sum_{i=1}^{q} (\bar\lambda_i^*)^2 \nabla c_i(x^*)\nabla c_i(x^*)^T + \mu^{-1}\sum_{i=1}^{q} O(\chi_1)\nabla c_i(x^*)\nabla c_i(x^*)^T \\ &\quad + \mu^{-1}\sum_{i=1}^{q} \left[\hat v_{1i}\nabla c_i(x^*)^T + \nabla c_i(x^*)\hat v_{2i}^T\right] + O(\mu), \end{aligned} \tag{3.10}$$

where $\hat v_{1i}$ and $\hat v_{2i}$ are both of size $O(\mu)$ for all $i = 1, 2, \ldots, q$. We now examine the eigenstructure of this Hessian matrix. Using G to denote the active constraint gradients, that is,

$$G \stackrel{\triangle}{=} \left[\nabla c_1(x^*), \nabla c_2(x^*), \ldots, \nabla c_q(x^*)\right],$$

with rank $G = \bar q \le q$, we have a $\bar q \times q$ matrix R with full row rank such that

$$G = U_R R.$$

Focusing on the $O(\mu^{-1})$ term in (3.10), we have that

$$\begin{aligned} P_{xx}(x;\mu) &= \mu^{-1}\sum_{i=1}^{q}(\bar{\lambda}_i^*)^2\nabla c_i(x^*)\nabla c_i(x^*)^T + \mu^{-1}O(\chi_1) + O(1) \\ &= \mu^{-1}GDG^T + \mu^{-1}O(\chi_1) + O(1) \\ &= \mu^{-1}U_R(RDR^T)U_R^T + \mu^{-1}O(\chi_1) + O(1), \end{aligned}$$

where D is the diagonal matrix whose diagonal elements are $(\bar{\lambda}_i^*)^2$, $i = 1,2,\ldots,q$, all of which are positive. Therefore the $\bar{q}$ eigenvalues of RDR^T are all of size $\Theta(\mu^{-1})$, so by choosing χ_1 small enough, we can ensure that the matrix H_{RR} defined in (3.8) also has all $\bar{q}$ of its eigenvalues of size $\Theta(\mu^{-1})$. In particular, we have

$$\|H_{RR}\| = O(\mu^{-1}), \quad \left\|H_{RR}^{-1}\right\| = O(\mu). \tag{3.11}$$

We have by the definition of U_N together with (3.10) that

$$\begin{aligned} H_{NN} &\overset{\triangle}{=} U_N^T P_{xx}(x;\mu)U_N \\ &= U_N^T\left[\nabla^2 f(x^*) - \sum_{i=1}^{q}\bar{\lambda}_i^*\nabla^2 c_i(x^*)\right]U_N + O(\chi_1) + O(\mu) \\ &= U_N^T\mathcal{L}_{xx}(x^*,\bar{\lambda}^*)U_N + O(\chi_1), \end{aligned}$$

and by the second-order sufficient condition, we have that this matrix is positive-definite with all eigenvalues of size $\Theta(1)$, provided that we choose χ_1 sufficiently small.

For the cross-term, we have that

$$H_{NR} \overset{\triangle}{=} U_N^T P_{xx}(x;\mu)U_R = U_N^T\mathcal{L}_{xx}(x^*,\bar{\lambda}^*)U_R + \mu^{-1}U_N^T\hat{V}_1G^TU_R + O(\chi_1),$$

where $\hat{V}_1 = [\hat{v}_{11},\ldots,\hat{v}_{1q}] = O(\mu)$. It follows from this estimate that

$$H_{RN}^T = H_{NR} = O(1).$$

for all χ_1 sufficiently small.

By applying a standard result for inverse of a partitioned matrix (see, for example, Wright [36, Lemma 2]) to (3.9), and using the estimates developed above, we have that

$$\begin{aligned} J_{22} &= \left(H_{NN} - H_{RN}^T H_{RR}^{-1} H_{RN}\right)^{-1} = (H_{NN} + O(\mu))^{-1} \\ &= O(1), \end{aligned} \tag{3.12a}$$

$$\begin{aligned} J_{11} &= H_{RR}^{-1} + H_{RR}^{-1}H_{RN}\left(H_{NN} - H_{RN}^T H_{RR}^{-1} H_{RN}\right)^{-1} H_{RN}^T H_{RR}^{-1} \\ &= O(\mu), \end{aligned} \tag{3.12b}$$

$$\begin{aligned} J_{12} &= -H_{RR}^{-1}H_{RN}\left(H_{NN} - H_{RN}^T H_{RR}^{-1} H_{RN}\right)^{-1} \\ &= O(\mu), \end{aligned} \tag{3.12c}$$

giving the required estimates. We omit the proof of positive definiteness of $P_{xx}(x;\mu)$, which follows from positive definiteness of J_{22}^{-1} and H_{RR}. ■

Our next lemma concerns the length of a Newton step for $P(\cdot;\mu)$, taken from a point x that is close to $\bar{x}(\mu)$.

Lemma 3.2. *Suppose that Assumption 3.1 holds. There are positive constants* χ_2 *and* C_2 *such that the following property holds. For all* $\mu \in (0, \chi_2]$ *and all* $\rho \in (0, \chi_2]$ *and all* x *with*

$$\|x - (x^* + \mu\zeta)\| = \|x - \bar{x}(\mu)\| \leq \rho\mu, \tag{3.13}$$

the point x *is strictly feasible and the Newton step* s *generated from* x *satisfies*

$$\|s\| \leq C_2(\rho + \mu)\mu.$$

Proof. We show first that χ_2 can be chosen small enough to ensure strict feasibility of x satisfying (3.13). For the inactive constraints c_i, $i = q+1, \ldots, m$, there is a constant $C_{2,0}$ such that

$$\begin{aligned} c_i(x) &\geq c_i(x^*) - C_{2,0}\|x - x^*\| \\ &\geq c_i(x^*) - C_{2,0}(\mu\|\zeta\| + \chi_2\mu) \geq (1/2)c_i(x^*) > 0, \quad i = q+1, \ldots, m, \end{aligned}$$

for χ_2 sufficiently small. For the active constraints, we have from (2.12) that there are constants $C_{2,1}$ and $C_{2,2}$ such that

$$\begin{aligned} c_i(x) &\geq c_i(x^*) + \mu\nabla c_i(x^*)^T\zeta - C_{2,1}\chi_2\mu - C_{2,2}\mu^2 \\ &\geq \mu/\bar{\lambda}_i^* - (C_{2,1} + C_{2,2})\chi_2\mu \\ &\geq \mu/(2\bar{\lambda}_i^*), \quad i = 1, 2, \ldots, q, \end{aligned}$$

for χ_2 sufficiently small.

We now examine the properties of $P_x(x;\mu)$ by expanding about x^* and using the properties (3.13) and (2.3) to obtain

$$\begin{aligned} P_x(x;\mu) = {} & \nabla f(x^*) + \mu\nabla^2 f(x^*)\zeta + \mu O(\rho) \\ & - \sum_{i=1}^{q} \left[\nabla c_i(x^*)^T\zeta + O(\rho)\right]^{-1} \left[\nabla c_i(x^*) + \mu\nabla^2 c_i(x^*)\zeta + \mu O(\rho)\right] \\ & - \sum_{i=q+1}^{m} \mu\left[c_i(x^*) + O(\mu)\right]^{-1}\left[\nabla c_i(x^*) + O(\mu)\right]. \end{aligned}$$

By choosing χ_2 sufficiently small, we have for all $\rho \in (0, \chi_2]$ that $[1 + O(\rho)]^{-1} = 1 + O(\rho)$, so

$$\begin{aligned} P_x(x;\mu) &= \nabla f(x^*) + \mu\nabla^2 f(x^*)\zeta + \mu O(\rho) \\ &\quad - \sum_{i=1}^{q} \bar{\lambda}_i^*(1 + O(\rho))\left[\nabla c_i(x^*) + \mu\nabla^2 c_i(x^*)\zeta + \mu O(\rho)\right] \\ &\quad - \sum_{i=q+1}^{m} \frac{\mu}{c_i(x^*)}\nabla c_i(x^*) + O(\mu^2) \\ &= \nabla f(x^*) - \sum_{i=1}^{q} \bar{\lambda}_i^*\nabla c_i(x^*) + \mu\left[\nabla^2 f(x^*) - \sum_{i=1}^{q} \bar{\lambda}_i^*\nabla^2 c_i(x^*)\right]\zeta \\ &\quad + \sum_{i=1}^{q} O(\rho)\nabla c_i(x^*) - \sum_{i=q+1}^{m} \frac{\mu}{c_i(x^*)}\nabla c_i(x^*) + (\rho + \mu)O(\mu). \end{aligned}$$

Hence by the definition (2.1) and the first-order conditions (2.2), we have

$$P_x(x;\mu) = \mu\mathcal{L}_{xx}(x^*, \bar{\lambda}^*)\zeta - \sum_{i=q+1}^{m} \frac{\mu}{c_i(x^*)}\nabla c_i(x^*) + \sum_{i=1}^{q} O(\rho)\nabla c_i(x^*) + (\rho + \mu)O(\mu). \tag{3.14}$$

By using the definitions (2.4) and (2.5) and the decomposition (2.13), we have that

$$\begin{aligned} U_N^T P_x(x;\mu) &= \mu U_N^T \mathcal{L}_{xx}(x^*,\bar\lambda^*) U_N \zeta_N + \mu U_N^T \mathcal{L}_{xx}(x^*,\bar\lambda^*) U_R \zeta_R \\ &\quad -\mu \sum_{i=q+1}^m \frac{1}{c_i(x^*)} U_N^T \nabla c_i(x^*) + (\rho+\mu) O(\mu). \end{aligned}$$

Hence by the definition (3.1) of ζ_N, we have that

$$U_N^T P_x(x;\mu) = (\rho+\mu) O(\mu). \tag{3.15}$$

Meanwhile, it follows immediately from (3.14) that

$$U_R^T P_x(x;\mu) = (\rho+\mu) O(1). \tag{3.16}$$

By reducing χ_2 if necessary so that $\chi_2 \le \chi_1$, we can apply Lemma 3.1 to x, ρ, μ satisfying (3.13). We have from (3.8) that

$$s = -P_{xx}(x;\mu)^{-1} P_x(x;\mu) = -\begin{bmatrix} U_R \; U_N \end{bmatrix} \begin{bmatrix} H_{RR} & H_{RN} \\ H_{RN}^T & H_{NN} \end{bmatrix}^{-1} \begin{bmatrix} U_R^T P_x(x;\mu) \\ U_N^T P_x(x;\mu) \end{bmatrix},$$

so it follows from (3.15), (3.16), and Lemma 3.1 that

$$U_R^T s = -J_{11} U_R^T P_x(x;\mu) - J_{12} U_N^T P_x(x;\mu) = (\rho\mu + \mu^2) O(1), \tag{3.17a}$$

$$U_N^T s = -J_{12}^T U_R^T P_x(x;\mu) - J_{22} U_N^T P_x(x;\mu) = (\rho\mu + \mu^2) O(1). \tag{3.17b}$$

We conclude that $\|s\| = (\rho\mu + \mu^2) O(1)$, as claimed. ∎

The next lemma concerns the *first two* Newton steps for $P(\cdot;\mu)$ taken from a point x close to $\bar x(\mu)$. It derives a bound on the second step in terms of the first.

Lemma 3.3. *Suppose that Assumption 3.1 holds. There are positive constants χ_3 and C_3 such that the following property holds. For all $\mu \in (0,\chi_3]$ and all $\tilde x$ with*

$$\|\tilde x - (x^* + \mu\zeta)\| = \|\tilde x - \bar x(\mu)\| \le \chi_3 \mu, \tag{3.18}$$

the first and second Newton steps $\tilde s$ and $\tilde s^+$ (respectively) generated from the point $\tilde x$ satisfy the bounds

$$\|\tilde s^+\| \le C_3 \mu^{-1} \|\tilde s\|^2, \qquad \|\tilde s^+\| \le (1/2)\|\tilde s\|.$$

Proof. We start by choosing χ_3 small enough that

$$(1 + 2C_2)\chi_3 \le \chi_2, \tag{3.19}$$

where χ_2 and C_2 are as defined in Lemma 3.2. We have by applying Lemma 3.2 that

$$\|\tilde s\| \le C_2(\chi_3\mu + \mu^2) \le 2C_2\chi_3\mu. \tag{3.20}$$

Therefore, we have

$$\|\tilde x + \tilde s - \bar x(\mu)\| \le \|\tilde x - \bar x(\mu)\| + \|\tilde s\| \le (\chi_3 + 2C_2\chi_3)\mu \le \chi_2\mu, \tag{3.21}$$

where we used (3.19) to derive the last inequality. We have from (3.21) and Lemma 3.2 that $\tilde x + \tilde s$ is strictly feasible, so a second Newton step $\tilde s^+$ is well defined.

We now seek a bound on $\|\tilde{s}^+\|$. By Taylor's theorem, we have

$$\begin{aligned} & P_x(\tilde{x}+\tilde{s};\mu) \\ &= P_x(\tilde{x};\mu) + P_{xx}(\tilde{x};\mu)\tilde{s} + \int_0^1 \left[P_{xx}(\tilde{x}+\tau\tilde{s};\mu) - P_{xx}(\tilde{x};\mu)\right]\tilde{s}d\tau, \\ &= \int_0^1 \left[P_{xx}(\tilde{x}+\tau\tilde{s};\mu) - P_{xx}(\tilde{x};\mu)\right]\tilde{s}d\tau. \end{aligned} \tag{3.22}$$

Techniques like those leading to Wright [36, equation (46)] can be used to analyze this integrand. We obtain

$$\int_0^1 \left[P_{xx}(\tilde{x}+\tau\tilde{s};\mu) - P_{xx}(\tilde{x};\mu)\right]\tilde{s}d\tau = \sum_{i=1}^q O(\mu^{-2}\|\tilde{s}\|^2)\nabla c_i(x^*) + O(\mu^{-1}\|\tilde{s}\|^2),$$

so we conclude from (2.4), (2.5), and (3.22), that

$$U_R^T P_x(\tilde{x}+\tilde{s};\mu) = O(\mu^{-2}\|\tilde{s}\|^2), \qquad U_N^T P_x(\tilde{x}+\tilde{s};\mu) = O(\mu^{-1}\|\tilde{s}\|^2). \tag{3.23}$$

Since, by (3.21), $\tilde{x}+\tilde{s}$ lies a the neighborhood of the form (3.13), within which the bounds (3.12) on the component blocks of the inverse Hessian $P_{xx}(\tilde{x}+\tilde{s};\mu)^{-1}$ apply, we can use these bounds together with (3.23) in the same fashion as in the argument that led to the estimate (3.17) to deduce that

$$\|\tilde{x}-\bar{x}(\mu)\| \le \chi_3\mu \;\Rightarrow\; \|\tilde{s}^+\| \le C_3\mu^{-1}\|\tilde{s}\|^2, \tag{3.24}$$

for some constant C_3. This proves the first claim.

By combining (3.24) with (3.20), we obtain that

$$\|\tilde{s}^+\| \le 2C_3C_2\chi_3\|\tilde{s}\|,$$

so that by reducing χ_3 as needed to ensure that $2C_3C_2\chi_3 \le (1/2)$, we ensure that the second part of the result also holds. ∎

We now prove our main result, which shows the existence of a minimizer $x(\mu)$ of $P(\cdot;\mu)$ close to $\bar{x}(\mu)$, and moreover proves convergence of Newton's method to this point when started from a neighborhood of $\bar{x}(\mu)$.

Theorem 3.1. *Suppose that Assumption 3.1 holds. Then there is a positive constant χ_4 such that for all $\mu \in (0,\chi_4]$ and all x^0 satisfying*

$$\|x^0 - (x^* + \mu\zeta)\| = \|x^0 - \bar{x}(\mu)\| \le \chi_4\mu, \tag{3.25}$$

the sequence $\{x^i\}$ obtained by setting

$$x^{i+1} = x^i + s^i, \;\; i = 1,2,\ldots,$$

where s^i is the Newton step for $P_x(\cdot;\mu)$ from x^i, is well defined and converges to a strict local minimizer $x(\mu)$ of $P_x(\cdot;\mu)$. Moreover we have that

$$\|x(\mu) - \bar{x}(\mu)\| = O(\mu^2) \tag{3.26}$$

and that $x(\mu)$ is the only local minimizer of $P(\cdot;\mu)$ in the neighborhood (3.7) of $\bar{x}(\mu)$.

Proof. The main part of the proof shows that for a certain choice of χ_4, the sequence $\{x^i\}$ stays inside the neighborhood defined by (3.18), and that

$$\|s^{i+1}\| \le (1/2)\|s^i\|, \quad i = 0, 1, 2, \dots. \tag{3.27}$$

It follows that $\{x^i\}$ is a Cauchy sequence and hence convergent.

We choose χ_4 to satisfy the following condition:

$$\chi_4 + 4C_2\chi_4 \le \chi_3. \tag{3.28}$$

Since $\chi_3 \le \chi_3 \le \chi_2$, we can apply Lemma 3.2 with $x = x^0$ to find that the first Newton step s^0 satisfies

$$\|s^0\| \le C_2(\chi_4 + \mu)\mu \le 2C_2\chi_4\mu. \tag{3.29}$$

Since $\chi_4 \le \chi_3$, we can also apply Lemma 3.3 with $\tilde{x} = x^0$ and deduce that the second Newton step s^1 satisfies

$$\|s^1\| \le (1/2)\|s^0\| \le C_2\chi_4\mu.$$

Since from (3.28) and (3.29) we have

$$\|x^1 - \bar{x}(\mu)\| \le \|x^0 - \bar{x}(\mu)\| + \|s^0\| \le \chi_4\mu + 2C_2\chi_4\mu \le \chi_3\mu$$

we can apply Lemma 3.3 with $\tilde{x} = x^1$ to obtain that $\|s^2\| \le (1/2)\|s^1\| \le (1/4)\|s^0\|$. A simple inductive argument along these lines shows that

$$\|x^i - \bar{x}(\mu)\| \le \left[\chi_4 + (2 - 2^{i-1})2C_2\chi_4\right]\mu \le \chi_3\mu,$$

(using (3.28) for the last bound), so that Lemma 3.3 can be applied from each x^i and that (3.27) holds. Hence the sequence is Cauchy and has a limit $x(\mu)$ that satisfies (3.18) (with $\tilde{x} = x(\mu)$) and hence also (3.7) (with $x = x(\mu)$). Hence, by Lemma 3.1 we have that $P_{xx}(x(\mu);\mu)$ is positive definite in the whole neighborhood (3.7). Moreover, by using (3.27) and setting $\tilde{x} = x^i$ and $\tilde{s} = s^i$ and taking limits as $i \to \infty$ in (3.23), we have that $P_x(x(\mu);\mu) = 0$. It follows that $x(\mu)$ is a strict local minimizer of $P(\cdot;\mu)$ and is in fact the only minimizer in the neighborhood (3.7) of $\bar{x}(\mu)$.

To verify the estimate (3.26), we consider the Newton sequence that starts at $x^0 = \bar{x}(\mu)$. This starting point certainly satisfies (3.25), so the sequence converges to $x(\mu)$ and its steps satisfy (3.27). It also satisfies (3.13) with $\rho = 0$, so by Lemma 3.2 we have that $\|s^0\| \le C_2\mu^2$. Hence from (3.27), we obtain that

$$\|x(\mu) - x^*\| \le \sum_{i=0}^{\infty} \|s^i\| \le 2\|s^0\| \le 2C_2\mu^2.$$

∎

The Q-quadratic rate of convergence of the Newton sequence to $x(\mu)$ with rate constant proportional to μ^{-1} follows from (3.24) and (3.27). Ideed, from (3.24), we can choose i sufficiently large that $\|s^{i+1}\| \le (1/4)\|s^i\|$. Meanwhile we have from (3.27) that

$$\|x^i - x^*\| \ge \|s^i\| - \sum_{j=i+1}^{\infty} \|s^j\| \ge \|s^i\| - 2\|s^{i+1}\| \ge (1/2)\|s^i\|$$

while

$$\|x^{i+1} - x^*\| \le \sum_{j=i+1}^{\infty} \|s^j\| \le 2\|s^{i+1}\|$$

By using these expressions with (3.24) we obtain

$$\|x^{i+1} - x^*\| \le 2\|s^{i+1}\| \le 2C_3\mu^{-1}\|s^i\|^2 \le 8C_3\mu^{-1}\|x^i - x^*\|^2.$$

3.3. Uniqueness of Barrier Function Minimizer

We now prove that the minimizers $x(\mu)$ of $P(\cdot;\mu)$ that we discussed in the previous subsection are unique, in the sense that there can be no other local minimizers of $P(\cdot;\mu)$ in the vicinity of x^* for μ small.

Our first step is to prove that the ratio of $c_i(z^k)$ to $\|z^k - x^*\|$ is bounded below for all active indices $i = 1, 2, \ldots, q$.

Lemma 3.4. *Suppose that Assumption 3.1 holds. For all $\{\mu_k\}$ and $\{z^k\}$ with the properties that*

$$\mu_k \downarrow 0, \quad z^k \to x^*, \quad z^k \text{ a local min of } P(\cdot;\mu_k), \tag{3.30}$$

we can choose a constant $\epsilon > 0$ such that $c_i(z^k) \geq \epsilon\|z^k - x^*\|$ for all $i = 1, 2, \ldots, q$ and all k sufficiently large.

Proof. Suppose for contradiction that there is an index $i \in \{1, 2, \ldots, q\}$ and a subsequence $\mathcal{K}$ such that

$$\lim_{k \in \mathcal{K}} \frac{c_i(z^k)}{\|z^k - x^*\|} = 0. \tag{3.31}$$

We know from Lemma 2.3 and the definition $\lambda^k \stackrel{\triangle}{=} \lambda(z^k, \mu_k)$ (2.22) that $\text{dist}_{\,\mathcal{S}_\lambda} \lambda^k \to 0$, so that all limit points of $\{\lambda^k\}$ lie in $\mathcal{S}_\lambda$, by closedness. By taking a further subsequence of $\mathcal{K}$ we can assume that there is a vectors $\hat\lambda \in \mathcal{S}_\lambda$ such that

$$\lim_{k \in \mathcal{K}} \lambda^k = \hat\lambda, \tag{3.32}$$

Similarly, by taking a further subsequence, and using compactness of the unit ball, we can identify $d \in \mathbf{R}^n$ with $\|d\| = 1$ such

$$\lim_{k \in \mathcal{K}} \frac{z^k - x^*}{\|z^k - x^*\|} = d.$$

Since $c_j(z^k) \geq 0$ and $c_j(x^*) = 0$ for each $j = 1, 2, \ldots, q$, we have that

$$\nabla c_j(x^*)^T d \geq 0, \quad j = 1, 2, \ldots, q. \tag{3.33}$$

For the index i in (3.31), we have by boundedness of $\mathcal{S}_\lambda$ and (3.32) that $\lim_{k \in \mathcal{K}} \mu_k / c_i(z^k) = \hat\lambda_i < \infty$, so that

$$\lim_{k \in \mathcal{K}} \mu_k / \|z^k - x^*\| = \lim_{k \in \mathcal{K}} \hat\lambda_i c_i(z^k) / \|z^k - x^*\| = 0. \tag{3.34}$$

Consider now the set of indices defined by

$$\mathcal{Z} \stackrel{\triangle}{=} \{j = 1, 2, \ldots, q \mid \nabla c_j(x^*)^T d = 0\}, \tag{3.35}$$

and let $\mathcal{Z}^c$ denote $\{1, 2, \ldots, q\} \backslash \mathcal{Z}$. It is easy to show that i from (3.31) belongs to $\mathcal{Z}$. Also, for any index $j \notin \mathcal{Z}$ we have

$$\hat\lambda_j = \lim_{k \in \mathcal{K}} \frac{\mu_k}{c_j(z^k)} = \lim_{k \in \mathcal{K}} \frac{\mu_k}{\nabla c_i(x^*)^T d \|z^k - x^*\| + o(\|z^k - x^*\|)} = 0. \tag{3.36}$$

Therefore from the KKT condition (2.7a), we have that

$$\nabla f(x^*) = \sum_{j=1}^{q} \hat\lambda_j \nabla c_j(x^*) = \sum_{j \in \mathcal{Z}} \hat\lambda_j \nabla c_j(x^*). \tag{3.37}$$

By the strict complementarity assumption (2.9), there is a multiplier λ^* such that

$$\nabla f(x^*) = \sum_{j=1}^{q} \lambda_j^* \nabla c_j(x^*), \quad \lambda_j^* > 0, \quad \text{for all } j = 1, 2, \dots, q. \tag{3.38}$$

By taking differences in (3.37) and (3.38), we have that

$$\sum_{j \in \mathcal{Z}} (\lambda_j^* - \hat{\lambda}_j) \nabla c_j(x^*) + \sum_{j \in \mathcal{Z}^c} \lambda_j^* \nabla c_j(x^*) = 0.$$

By taking the inner product with d, and using the definition of $\mathcal{Z}$, we have

$$\sum_{j \in \mathcal{Z}^c} \lambda_j^* \nabla c_j(x^*)^T d = 0. \tag{3.39}$$

Since $\lambda_j^* > 0$ and since $\nabla c_j(z^*)^T d \geq 0$ by (3.33), we must have $\nabla c_j(z^*)^T d = 0$, so that $\mathcal{Z}^c$ must be empty. Hence, $\mathcal{Z} = \{1, 2, \dots, q\}$.

Since z^k is a local minimizer of $P(\cdot; \mu_k)$, we have from (2.23a) and (2.1) that

$$P_x(z^k; \mu_k) = \mathcal{L}_x(z^k, \lambda^k) = 0.$$

Since $\hat{\lambda} \in \mathcal{S}_\lambda$, we have that $\mathcal{L}_x(x^*, \hat{\lambda}) = 0$. By taking differences, we have that

$$\begin{aligned}
0 &= \mathcal{L}_x(z^k, \lambda^k) - \mathcal{L}_x(x^*, \hat{\lambda}) \\
&= \mathcal{L}_x(z^k, \lambda^k) - \mathcal{L}_x(z^k, \hat{\lambda}) + \mathcal{L}_x(z^k, \hat{\lambda}) - \mathcal{L}_x(x^*, \hat{\lambda}) \\
&= -\sum_{j=1}^{m} \left[\lambda_j^k - \hat{\lambda}_j\right] \nabla c_j(z^k) + \mathcal{L}_{xx}(x^*, \hat{\lambda})(z^k - x^*) + o(\|z^k - x^*\|) \\
&= \sum_{j=1}^{q} o(1) \nabla c_i(x^*) + O(\mu_k) + \mathcal{L}_{xx}(x^*, \hat{\lambda})(z^k - x^*) + o(\|z^k - x^*\|),
\end{aligned}$$

where we have used the estimate $\lambda_j^k = \mu_k / c_j(z^k) = O(\mu_k)$ for $j = q+1, \dots, m$ to derive the final equality. Taking the inner product with d, and noting that $\nabla c_j(x^*)^T d = 0$ for all $j = 1, 2, \dots, q$, we have that

$$0 = d^T \mathcal{L}_{xx}(x^*, \hat{\lambda})(z^k - x^*) + O(\mu_k) + o(\|z^k - x^*\|).$$

If we divide by $\|z^k - x^*\|$, take the limit, and use (3.34), we obtain

$$0 = d^T \mathcal{L}_{xx}(x^*, \hat{\lambda}) d.$$

However, the second-order conditions (2.16) require that $d^T \mathcal{L}_{xx}(x^*, \hat{\lambda}) d > 0$ for all $d \neq 0$ with $\nabla c_j(x^*)^T d = 0$, giving a contradiction.

We conclude that no i with the property (3.31) can exist, giving the result. ∎

We use this result to show that $\mu_k = \Theta(\|z^k - x^*\|)$ for all sequences satisfying (3.30).

Lemma 3.5. *Suppose that Assumption 3.1 holds. Then for all sequences $\{\mu_k\}$ and $\{z^k\}$ satisfying (3.30), we have that $\mu_k = \Theta(\|z^k - x^*\|)$.*

Proof. Defining $\lambda^k \triangleq \lambda(z^k, \mu_k)$ again we show first that it is not possible to choose a subsequence $\mathcal{K}$ such that

$$\lim_{k\in\mathcal{K}} \frac{\mu_k}{\|z^k - x^*\|} = 0. \tag{3.40}$$

If this were true, we would have from Lemma 3.4, using an expansion like that in (3.36) that

$$0 \le \lim_{k\in\mathcal{K}} \lambda_j^k = \lim_{k\in\mathcal{K}} \frac{\mu_k}{c_j(z^k)} \le \lim_{k\in\mathcal{K}} \frac{\mu_k}{\epsilon\|z^k - x^*\|} = 0, \quad j = 1, 2, \ldots, q. \tag{3.41}$$

Since by Lemma 2.3 we have dist ${}_{\mathcal{S}_\lambda}\lambda^k \to 0$, (3.41) implies that $0 \in \mathcal{S}_\lambda$. Hence, given any strictly complementary solution $\lambda^* \in \mathcal{S}_\lambda$, we have that $\alpha\lambda^* \in \mathcal{S}_\lambda$ for all $\alpha \ge 0$. Hence, $\mathcal{S}_\lambda$ is unbounded, contradicting Gauvin [11, Theorem 1].

We now show that it is not possible to choose a subsequence $\mathcal{K}$ such that

$$\lim_{k\in\mathcal{K}} \frac{\mu_k}{\|z^k - x^*\|} = \infty. \tag{3.42}$$

If this were possible, we would have from smoothness of c_j, $j = 1, 2, \ldots, q$ that

$$\lim_{k\in\mathcal{K}} \lambda_j^k = \lim_{k\in\mathcal{K}} \frac{\mu_k}{c_j(z^k)} \ge \lim_{k\in\mathcal{K}} \frac{\mu_k}{M\|z^k - x^*\|}, \quad j = 1, 2, \ldots, q,$$

for some positive constant M, so that $\lim_{k\in\mathcal{K}} \lambda^k = \infty$. This contradicts boundedness of $\mathcal{S}_\lambda$ and dist ${}_{\mathcal{S}_\lambda}\lambda^k \to 0$.

We conclude that neither (3.40) nor (3.42) can occur for any subsequence $\mathcal{K}$, proving the result. ∎

Theorem 3.2. *Suppose that Assumption 3.1 holds. For all $\{\mu_k\}$ and $\{z^k\}$ with the properties (3.30), we have that $\mu_k/c_i(z^k) \to \bar{\lambda}_i^*$ for all $i = 1, 2, \ldots, m$, where $\bar{\lambda}^*$ is defined by (2.12), and in fact that $z^k = x(\mu_k)$ for all k sufficiently large.*

Proof. By Lemma 3.5, we have that the sequence $\{d^k\}$ with

$$d^k \triangleq \frac{z^k - x^*}{\mu_k} \tag{3.43}$$

lies entirely inside the compact set $\{v \mid \beta_0 \le \|v\| \le \beta_1\}$, where β_0 and β_1 are positive constants.

Since dist ${}_{\mathcal{S}_\lambda}\lambda^k \to 0$ and since $\mathcal{S}_\lambda$ is compact, we can find a limit point $\hat{\lambda} \in \mathcal{S}_\lambda$ and a subsequence $\mathcal{K}$ such that $\lim_{k\in\mathcal{K}} \lambda^k = \hat{\lambda}$. We can choose a further subsequence of $\mathcal{K}$ such that $lim_{k\in\mathcal{K}} = d$ for some vector d with $\beta_0 \le \|d\| \le \beta_1$. For any $j = 1, 2, \ldots, q$, we then have

$$\begin{aligned}\hat{\lambda}_j &= \lim_{k\in\mathcal{K}} \frac{\mu_k}{c_j(z^k)} \\ &= \lim_{k\in\mathcal{K}} \frac{\mu_k}{\nabla c_j(x^*)^T(z^k - x^*) + o(\|z^k - x^*\|)} = \lim_{k\in\mathcal{K}} \frac{1}{\nabla c_j(x^*)^T d^k + o(1)} = \frac{1}{\nabla c_j(x^*)^T d}.\end{aligned}$$

Hence, $\hat{\lambda}$ satisfies the conditions (2.12) to be the analytic center of $\mathcal{S}_\lambda$, and by uniqueness of this point, we have $\hat{\lambda} = \bar{\lambda}^*$. By compactness of $\mathcal{S}_\lambda$, the limit dist ${}_{\mathcal{S}_\lambda}\lambda^k = 0$, and our arbitrary choice of limit $\hat{\lambda}$, we conclude that

$$\lim_{k\to\infty} \lambda^k = \bar{\lambda}^*.$$

For any $j = 1, 2, \ldots, q$, we have that

$$\bar{\lambda}^* = \lim_{k\to\infty} \frac{\mu_k}{c_j(z^k)} = \lim_{k\to\infty} \frac{1}{\nabla c_j(x^*)^T d^k + o(1)},$$

so that

$$\lim_{k\to\infty} \nabla c_j(x^*)^T d^k = \bar{\lambda}_j^* = \nabla c_j(x^*)^T \zeta, \tag{3.44}$$

for the vector ζ defined in (2.12), (2.13), and (3.1). It follows from these observations together with (2.12), (3.6), (3.43), (3.44), and Theorem 3.1 that

$$\begin{aligned}\nabla c_i(x^*)^T (z^k - x^*) &= \mu_k \nabla c_i(x^*)^T \zeta + o(\mu_k) \\ &= \nabla c_i(x^*)^T \left(\bar{x}(\mu_k) - x^*\right) + o(\mu_k) \\ &= \nabla c_i(x^*)^T \left(x(\mu_k) - x^*\right) + o(\mu_k), \quad i = 1, 2, \ldots, q,\end{aligned}$$

and therefore

$$\nabla c_i(x^*)^T \left(x(\mu_k) - z^k\right) = o(\mu_k), \quad i = 1, 2, \ldots, q.$$

Hence, we have for all $\alpha \in [0,1]$ that

$$\begin{aligned}& c_i\left(z^k + \alpha(x(\mu_k) - z^k)\right) \\ &= c_i(z^k) + \alpha \nabla c_i(z^k)^T (x(\mu_k) - z^k) + O(\mu_k^2) \\ &= [\mu_k/\bar{\lambda}_i^* + o(\mu_k)] + o(\mu_k) + O(\mu_k^2) \\ &= \mu_k/\bar{\lambda}_i^* + o(\mu_k), \qquad i = 1, 2, \ldots, q.\end{aligned} \tag{3.45}$$

We now consider the Hessian $P_{xx}(\cdot;\mu_k)$ evaluated at the points $z^k + \alpha(x(\mu_k) - z^k)$, $\alpha \in [0,1]$. By using analysis similar to that in the proof of Lemma 3.2, together with the observation (3.45), we can show that this matrix is positive definite for all $\alpha \in [0,1]$, for all k sufficiently large. By Taylor's theorem, we have

$$\begin{aligned}0 &= \left(z^k - x(\mu_k)\right)^T \left[P_x(z^k;\mu_k) - P_x(x(\mu_k);\mu_k)\right] \\ &= \int_0^1 \left(z^k - x(\mu_k)\right)^T P_{xx}\left(z^k + \alpha(x(\mu_k) - z^k), \mu_k\right) \left(z^k - x(\mu_k)\right) d\alpha.\end{aligned}$$

Observe that the right-hand side of this expression is positive whenever $z^k \neq x(\mu_k)$ for all k sufficiently large. We conclude that $z^k = x(\mu_k)$ for all k sufficiently large, as required. ∎

3.4. Direction of Approach of the Central Path

We now demonstrate the differentiability of the path $x(\mu)$ described in Theorems 3.1 and 3.2. The proof again uses the decomposition of the Hessian $P_{xx}(x;\mu)$ that was first derived in the proof of Lemma 3.2.

Theorem 3.3. *Suppose that Assumption 3.1 holds. Then for the minimizers $x(\mu)$ defined in Theorem 3.1, there is a threshold $\bar{\mu}$ such that $x(\mu)$ exists and is a continuously differentiable function of μ for all $\mu \in (0, \bar{\mu}]$, and we have that*

$$\dot{x}(0_+) \overset{\triangle}{=} \lim_{\mu\downarrow 0} \dot{x}(\mu) = \zeta. \tag{3.46}$$

Proof. Setting $\bar{\mu} = \chi_4$, we have from Theorem 3.1 that $x(\mu)$ exists for all $\mu \in (0, \bar{\mu}]$.

Choose $\sigma = 2$ in Theorem 3.1, and let $\bar{\mu}$ and C_0 be defined accordingly. We now have existence of $x(\mu)$ for all $\mu \in (0, \bar{\mu}]$, and the estimate (3.26) holds. Each such $x(\mu)$ solves the equation

$$P_x(x(\mu); \mu) = 0.$$

P_{xx} is nonsingular and continuous at each $\mu \in (0, \bar{\mu}]$, by (3.9), (3.12) and the discussion preceding these expressions. Hence, we can apply the implicit function theorem (see, for example, Ortega and Rheinboldt [26, p. 128]) to conclude that $x(\cdot)$ is differentiable at μ and that the derivative $\dot{x}(\mu)$ satisfies the equation

$$P_{xx}(x(\mu), \mu)\dot{x}(\mu) - r(\mu) = 0, \tag{3.47}$$

where

$$r(\mu) \triangleq \sum_{i=1}^{m} \frac{1}{c_i(x(\mu))} \nabla c_i(x(\mu)). \tag{3.48}$$

We now show (3.46) by showing that $\dot{x}(\mu) = \zeta + u(\mu)$, where $u(\mu) = O(\mu)$ and so, in particular, $u(\mu) \to 0$ as $\mu \downarrow 0$. By substituting into (3.47), we have that $u(\mu)$ satisfies the expression

$$P_{xx}(x(\mu), \mu)u(\mu) = r(\mu) - P_{xx}(x(\mu), \mu)\zeta. \tag{3.49}$$

We can substitute $x(\mu)$ for x in the analysis of $P_{xx}(\cdot; \mu)$ that leads to (3.9), (3.12), since $x(\mu)$ certainly lies in the neighborhood (3.25) within which this estimate is valid. Hence, by applying (3.9) to (3.49), we have that

$$\begin{bmatrix} U_R^T u(\mu) \\ U_N^T u(\mu) \end{bmatrix} = \begin{bmatrix} J_{11} & J_{12} \\ J_{12}^T & J_{22} \end{bmatrix} \begin{bmatrix} U_R^T \\ U_N^T \end{bmatrix} \left[r(\mu) - P_{xx}(x(\mu), \mu)\zeta \right].$$

Hence, from the estimates (3.12), we have that

$$u(\mu) = O(\mu) \left\| U_R^T [r(\mu) - P_{xx}(x(\mu), \mu)\zeta] \right\| + O(1) \left\| U_N^T [r(\mu) - P_{xx}(x(\mu), \mu)\zeta] \right\|.$$

Therefore, our estimate $u(\mu) = O(\mu)$ will follow if we can show that

$$U_R^T [r(\mu) - P_{xx}(x(\mu), \mu)\zeta] = O(1), \tag{3.50a}$$
$$U_N^T [r(\mu) - P_{xx}(x(\mu), \mu)\zeta] = O(\mu). \tag{3.50b}$$

By substituting directly from (2.23b) and (3.48), we have that

$$\begin{aligned} P_{xx}(x(\mu); \mu)\zeta - r(\mu) = &\left[\nabla^2 f(x(\mu)) + \sum_{i=1}^{m} \frac{\mu}{c_i(x(\mu))} \nabla^2 c_i(x(\mu)) \right] \zeta \\ &+ \sum_{i=1}^{q} \frac{1}{c_i(x(\mu))} \left[\frac{\mu \nabla c_i(x(\mu))^T \zeta}{c_i(x(\mu))} - 1 \right] \nabla c_i(x(\mu)) \\ &+ \sum_{i=q+1}^{m} \frac{1}{c_i(x(\mu))} \left[\frac{\mu \nabla c_i(x(\mu))^T \zeta}{c_i(x(\mu))} - 1 \right] \nabla c_i(x(\mu)). \end{aligned} \tag{3.51}$$

From (3.26) we have

$$x(\mu) - x^* = \mu\zeta + (x(\mu) - x^* - \mu\zeta) = \mu\zeta + O(\mu^2),$$

so from (2.12) we obtain

$$\begin{aligned} c_i(x(\mu)) &= \nabla c_i(x^*)^T (x(\mu) - x^*) + O(\mu^2) \\ &= \mu \nabla c_i(x^*)^T \zeta + O(\mu^2) = \frac{\mu}{\bar{\lambda}_i^*} + O(\mu^2), \quad i = 1, 2, \ldots, q. \end{aligned} \tag{3.52}$$

Hence, Lipschitz continuity of $\nabla^2 f(\cdot)$ and $\nabla^2 c_i(\cdot)$, $i = 1, 2, \ldots, m$ implies that

$$\nabla^2 f(x(\mu)) + \sum_{i=1}^m \frac{\mu}{c_i(x(\mu))} \nabla^2 c_i(x(\mu)) = \nabla^2 f(x^*) + \sum_{i=1}^m \bar{\lambda}_i^* \nabla^2 c_i(x^*) + O(\mu). \tag{3.53}$$

For the second term on the right-hand side of (3.51), we use (3.52) again to obtain

$$\frac{\mu \nabla c_i(x(\mu))^T \zeta}{c_i(x(\mu))} - 1 = \frac{\mu/\bar{\lambda}_i^* + O(\mu^2)}{\mu/\bar{\lambda}_i^* + O(\mu^2)} - 1 = O(\mu), \quad i = 1, 2, \ldots, q. \tag{3.54}$$

Hence, by using (3.52) again together with the property $\bar{\lambda}_i^* > 0$, $i = 1, 2, \ldots, q$, we have that

$$\begin{aligned} &\sum_{i=1}^q \frac{1}{c_i(x(\mu))} \left[\frac{\mu \nabla c_i(x(\mu))^T \zeta}{c_i(x(\mu))} - 1 \right] \nabla c_i(x(\mu)) \\ &= \sum_{i=1}^q \frac{O(\mu)}{c_i(x(\mu))} \nabla c_i(x(\mu)) = \sum_{i=1}^q O(1) \nabla c_i(x^*) + O(\mu). \end{aligned} \tag{3.55}$$

For the third term on the right-hand side of (3.51), we have that, since $c_i(x(\mu))$ is bounded away from zero for μ sufficiently small,

$$\sum_{i=q+1}^m \frac{1}{c_i(x(\mu))} \left[\frac{\mu \nabla c_i(x(\mu))^T \zeta}{c_i(x(\mu))} - 1 \right] \nabla c_i(x(\mu)) = - \sum_{i=q+1}^m \frac{1}{c_i(x^*)} \nabla c_i(x^*) + O(\mu). \tag{3.56}$$

By substituting (3.53), (3.55), (3.56) into (3.51), and taking the inner product with U_R^T, we have that (3.50a) is satisfied. When we take the inner product with U_N^T, the terms involving $\nabla c_i(x^*)$, $i = 1, 2, \ldots, q$ in (3.55) are eliminated, and we are left with

$$U_N^T [P_{xx}(x(\mu); \mu)\zeta - r(\mu)] = U_N^T \mathcal{L}_{xx}(x^*, \bar{\lambda}^*)\zeta - \sum_{i=q+1}^m \frac{1}{c_i(x^*)} U_N^T \nabla c_i(x^*) + O(\mu).$$

By comparing this expression with (3.1), we conclude that (3.50b) is satisfied, completing the proof. ∎

The relation (3.46) together with (2.12) shows that the primal central path reaches x^* nontangentially to the active constraints, but rather is tangential to the linear path $\{\bar{x}(\mu) \mid \mu \in (0, \overline{\mu}]\}$ at x^*. It also shows that $p = -\zeta = -\dot{x}(0_+)$ is a direction that satisfies the MFCQ condition (2.6).

The proof of (3.46) is much simpler in the case of linearly independent active constraints. When this condition holds, Fiacco and McCormick [9, Section 5.2] replace (3.47) by an "augmented" linear system whose unknowns are both $\dot{x}(\mu)$ and $\dot{\lambda}(\mu)$ and whose coefficient matrix approaches a nonsingular limit as $\mu \downarrow 0$. The result follows by setting $\mu = 0$ and calculating the solution of this system directly. M. H. Wright performs a similar analysis [30, Section 3] and observes the nontangentiality of the path to the active constraints.

4. Relaxing the Strict Complementarity Condition

In this section, we discuss the properties of the sequence of minimizers of $P(\cdot;\mu_k)$ when strict complementarity (2.9) does not hold. That is, we have for some active constraint index $i = 1, 2, \ldots, q$ that $\lambda_i^* = 0$ for all $\lambda^* \in \mathcal{S}_\lambda$. Lemmas 2.1, 2.2, and 2.3 continue to hold when (2.9) is not satisfied. However, the problem (2.10), (2.11) that defines the analytic center is not even feasible, so the path of minimizers of $P(\cdot;\mu)$ with the particular form described in Section 3 is not defined.

Our main results are as follows. Under the second-order sufficient condition (2.18) for the non-strict complementarity case, we can show existence of a local minimizer of $P(\cdot;\mu_k)$ in the vicinity of x^*, for all μ_k sufficiently small, using a simple modification of results of M. H. Wright [30]. We show that the direction of approach of the sequence of minimizers to x^* is tangential to the strongly active constraints (those for which $\lambda_i^* > 0$ for some $\lambda^* \in \mathcal{S}_\lambda$). Finally, we show that

$$\mu_k = \Theta(\|z^k - x^*\|^2),$$

where z^k is the local minimizer of $P(\cdot;\mu_k)$. This contrasts with the strictly complementary case, in which the exponent 2 does not appear. We are not able to prove local uniqueness of the minimizer (as was seen for the strictly complementary case in Theorem 3.2), nor are we able to obtain the semi-explicit characterization of the minimizer seen in Theorem 3.1.

We can obtain some insight into the case of non-strict complementarity by considering the following simple example:

$$\min \tfrac{1}{2}\left(x_1^2 + x_2^2\right) \quad \text{subject to } x_1 \geq 1,\ x_2 \geq 0.$$

The solution is $x^* = (1, 0)^T$, with both constraints active and unique optimal Lagrange multipliers $\lambda_1^* = 1$, $\lambda_2^* = 0$. It is easy to verify that the minimizer of $P(\cdot;\mu)$ in this case is

$$x(\mu) = \left(\frac{1 + \sqrt{1 + 4\mu}}{2}, \sqrt{\mu}\right)^T \approx (1 + \mu, \sqrt{\mu})^T.$$

The path of minimizers is dramatically different from the one that would be obtained by omitting the weakly active constraint $x_2 \geq 0$ from the problem, which would be

$$x(\mu) = \left(\frac{1 + \sqrt{1 + 4\mu}}{2}, 0\right)^T \approx (1 + \mu, 0)^T.$$

Note that the path becomes tangential to the strongly active constraint $x_1 \geq 1$ and that the distance from $x(\mu)$ to the solution x^* is $O(\mu^{1/2})$ rather than $O(\mu)$, as in the case of strict complementarity.

Convergence to the analytic center of the dual multiplier set (defined by a modification of (2.10) in which we sum only over the "strictly complementary" indices) cannot be expected in this case, as shown in an example in McCormick and Witzgall [17, Section 8].

As before, we suppose that the "non-strictly complementary" indices are $\bar{q} + 1, \ldots, q$, that is,

$$\lambda_i^* = 0, \quad \text{for all } \lambda^* \in \mathcal{S}_\lambda, \text{ all } i = \bar{q} + 1, \ldots, q, \tag{4.1}$$

and recall the second-order sufficient conditions (2.18).

All results in this section use the following assumption.

Assumption 4.1. *At least one constraint is active at the solution x^*, and the first-order necessary conditions (2.2), the second-order sufficient conditions (2.18), and the MFCQ (2.6) hold at x^*. The strict complementarity condition fails to hold, that is, $\bar{q} < q$ in (4.1).*

We state first a result about the curvature of the Lagrangian Hessian along directions d that are "close" to those directions defined in (2.16b) and in (2.18b), (2.18c). The proof appears in the appendix.

Lemma 4.1. *There exist positive constants ϵ_d and $\bar{\eta}$ such that for all d with $\|d\| = 1$ and*

$$|\nabla c_i(x^*)^T d| \le \epsilon_d, \quad i = 1, 2, \ldots, \bar{q}, \tag{4.2a}$$

$$(\nabla c_i(x^*)^T d)_- \le \epsilon_d, \quad i = \bar{q}+1, \ldots, q, \tag{4.2b}$$

(where $(\cdot)_-$ is defined in (2.19)), we have that

$$d^T \mathcal{L}_{xx}(x^*, \lambda^*) d \ge \bar{\eta}, \quad \textit{for all } \lambda^* \in \mathcal{S}_\lambda.$$

We now cite a result on the existence of a sequence of minimizers of the barrier function that approaches x^*. It is a consequence of Theorem 7 in M. H. Wright [30]. Under our assumptions, x^* is a strict local minimizer of the problem (1.1), and so the set $\mathcal{M}$ in the cited result is the singleton $\{x^*\}$.

Theorem 4.1. *Suppose that Assumption 4.1 holds. Let $\{\mu_k\}$ be any sequence of positive numbers such that $\mu_k \downarrow 0$. Then*

(i) there exists a neighborhood $\mathcal{N}$ of x^* such that for all k sufficiently large, $P(\cdot; \mu_k)$ has at least one unconstrained minimizer in strict $\mathcal{C} \cap \mathcal{N}$. *Moreover, every sequence of* global *minimizers $\{\bar{x}^k\}$ of $P(\cdot; \mu_k)$ in* strict $\mathcal{C} \cap \operatorname{cl} \mathcal{N}$ *converges to x^*.*
(ii) $\lim_{k\to\infty} f(\bar{x}^k) = \lim_{k\to\infty} P(\bar{x}^k; \mu_k) = f(x^*)$.

The next three results concern the behavior of any sequences $\{\mu_k\}$ and $\{z^k\}$ with the following properties:

$$\mu_k \downarrow 0, \quad z^k \to x^*, \quad z^k \text{ a local min of } P(\cdot; \mu_k). \tag{4.3}$$

The sequence of global minimizers $\{\bar{x}_k\}$ described in Theorem 4.1 is one possible choice for $\{z^k\}$. Note that for sequences satisfying (4.3), we have from (2.1), (2.20), (2.22), and (2.23a) that

$$0 = P_x(z^k; \mu_k) = \mathcal{L}_x(z^k, \lambda^k). \tag{4.4}$$

Theorem 4.2. *Suppose that Assumption 4.1 holds. Let $\{\mu_k\}$ and $\{z^k\}$ be sequences with the properties (4.3). Then we have that*

$$\mu_k = O(\|z^k - x^*\|^2). \tag{4.5}$$

Proof. From Lemma 2.3, we have that

$$\lambda_i^k \le \operatorname{dist}_{\mathcal{S}_\lambda} \lambda^k = O\left(\|z^k - x^*\| + \mu_k\right), \quad \text{for all } i = \bar{q}+1, \ldots, q.$$

By substituting from (2.20), and using the estimate $c_i(z^k) = O\left(\|z^k - x^*\|\right)$, we have that

$$\mu_k \le c_i(z^k) O\left(\|z^k - x^*\| + \mu_k\right) \le K_1 \left(\|z^k - x^*\|^2 + \mu_k \|z^k - x^*\|\right),$$

for some $K_1 > 0$ and all k sufficiently large. Therefore, we have

$$\left(1 - K_1\|z^k - x^*\|\right)\mu_k \le K_1\|z^k - x^*\|^2,$$

so by taking k large enough that $\|z^k - x^*\| \le 1/(2K_1)$, we have the result. ■

We now show that the approach of the minimizer sequence is tangential to the strongly active constraints.

Lemma 4.2. *Suppose that Assumption 4.1 holds, and let $\{\mu_k\}$ and $\{z^k\}$ be sequences that satisfy (4.3). Then defining*

$$d^k \triangleq (z^k - x^*)/\|z^k - x^*\|, \tag{4.6}$$

we have that

$$\begin{aligned} |\nabla c_i(x^*)^T d^k| &= O(\|z^k - x^*\|), \quad i = 1, 2, \ldots, \bar{q}, \\ (\nabla c_i(x^*)^T d^k)_- &\le O(\|z^k - x^*\|), \quad i = \bar{q}+1, \ldots, q. \end{aligned}$$

Proof. We start by noting that since $c_i(x^*) = 0$ for $i = 1, 2, \ldots, q$, we have from (4.3) that

$$\nabla c_i(x^*)^T(z^k - x^*) = c_i(z^k) + O(\|z^k - x^*\|^2), \quad i = 1, 2, \ldots, q.$$

and therefore, since $\{z^k\}$ is a feasible sequence, we have

$$(\nabla c_i(x^*)^T d^k)_- \le c_i(z^k)_-/\|z^k - x^*\| + O(\|z^k - x^*\|) = O(\|z^k - x^*\|), \quad i = 1, 2, \ldots, q. \tag{4.7}$$

Therefore our result follows if we can prove in addition to (4.7) that

$$(\nabla c_i(x^*)^T d^k)_+ = O(\|z^k - x^*\|), \quad i = 1, 2, \ldots, \bar{q}. \tag{4.8}$$

The remainder of the proof is directed to showing that (4.8) holds.

We write

$$f(z^k) - f(x^*) = \left[\mathcal{L}(z^k, \lambda^k) - \mathcal{L}(x^*, \lambda^k)\right] + (\lambda^k)^T[c(z^k) - c(x^*)]. \tag{4.9}$$

For the first term on the right-hand side, we have from (4.4) that

$$\mathcal{L}(x^*, \lambda^k) - \mathcal{L}(z^k, \lambda^k) = O(\|z^k - x^*\|^2). \tag{4.10}$$

In the second term on the right-hand side of (4.9), we have by the definition (2.20), (2.22) that

$$(\lambda^k)^T c(z^k) = m\mu_k, \tag{4.11}$$

while

$$\lambda_i^k c_i(x^*) = \mu_k \frac{c_i(x^*)}{c_i(z^k)} = 0, \quad i = 1, 2, \ldots, q, \tag{4.12a}$$

$$\lambda_i^k c_i(x^*) = \mu_k \frac{c_i(x^*)}{c_i(x^*) + O(\|z^k - x^*\|)} = O(\mu_k), \quad i = q+1, \ldots, m. \tag{4.12b}$$

By substituting (4.10), (4.11), and (4.12) into (4.9), we obtain

$$f(z^k) - f(x^*) = O(\|z^k - x^*\|^2 + \mu_k).$$

Therefore, using Theorem 4.2, we obtain that

$$f(z^k) - f(x^*) = O(\|z^k - x^*\|^2),$$

so that

$$\nabla f(x^*)^T (z^k - x^*) = f(z^k) - f(x^*) + O(\|z^k - x^*\|^2) = O(\|z^k - x^*\|^2). \tag{4.13}$$

Given the convexity of $\mathcal{S}_\lambda$ and our definition of $\bar{q}$, we can choose a vector λ^* such that

$$\lambda^* \in \mathcal{S}_\lambda, \quad \lambda_i^* > 0 \text{ for } i = 1, 2, \dots, \bar{q}, \quad \lambda_i^* = 0 \text{ for } i = \bar{q}+1, \dots, m. \tag{4.14}$$

From (2.2) and (4.13), we have

$$\sum_{i=1}^{\bar{q}} \lambda_i^* \nabla c_i(x^*)^T d^k = \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*)^T d^k = \nabla f(x^*)^T d^k = O(\|z^k - x^*\|).$$

Hence, by using the identity $\beta = \beta_+ - \beta_-$ together with (4.7), we have that

$$\sum_{i=1}^{\bar{q}} \lambda_i^* (\nabla c_i(x^*)^T d^k)_+ = \sum_{i=1}^{\bar{q}} \lambda_i^* (\nabla c_i(x^*)^T d^k)_- + O(\|z^k - x^*\|) = O(\|z^k - x^*\|).$$

Because of the property (4.14) and the fact that $(\nabla c_i(x^*)^T d^k)_+ \geq 0$, we conclude that (4.8) holds, completing the proof. ∎

An immediate consequence of Lemmas 4.1 and 4.2 is that there is a positive constant $\bar{\eta}$ such that for all k sufficiently large, we have

$$(d^k)^T \mathcal{L}_{xx}(x^*, \lambda^*) d^k \geq \bar{\eta}, \quad \text{for all } \lambda^* \in \mathcal{S}_\lambda. \tag{4.15}$$

We now show that there is a lower bound on μ_k in terms of $\|z^k - x^*\|^2$ to go with the upper bound in (4.5).

Theorem 4.3. *Suppose that Assumption 4.1 holds, and let $\{\mu_k\}$ and $\{z^k\}$ be sequences with the properties (4.3). Then we have that*

$$\|z^k - x^*\|^2 = O(\mu_k). \tag{4.16}$$

Proof. Using (4.11), the definition (2.1), and the inequality $\mathcal{L}(x^*, \lambda^k) - f(x^*) = -c(x^*)^T \lambda^k \leq 0$, we have

$$\begin{aligned} m\mu_k &= c(z^k)^T \lambda^k \\ &= \left[f(z^k) - f(x^*) \right] - \left[\mathcal{L}(z^k, \lambda^k) - f(x^*) \right] \\ &\geq \left[f(z^k) - f(x^*) \right] - \left[\mathcal{L}(z^k, \lambda^k) - \mathcal{L}(x^*, \lambda^k) \right]. \end{aligned} \tag{4.17}$$

From (2.15), and by rearranging (4.17), we have for k sufficiently large that

$$\eta \|z^k - x^*\|^2 \leq m\mu_k + \left[\mathcal{L}(z^k, \lambda^k) - \mathcal{L}(x^*, \lambda^k) \right].$$

Hence, it suffices for our result to prove that the second term on the right-hand side of this expression is nonpositive.

When the sequences $\{\mu_k\}$ and $\{z^k\}$ satisfy (4.3), we have from Lemma 2.3 that

$$\text{dist}_{\mathcal{S}_\lambda} \lambda^k = O(\|z^k - x^*\| + \mu_k) \to 0.$$

Moreover, by compactness of $\mathcal{S}_\lambda$, there is a $\lambda^k_* \in \mathcal{S}_\lambda$ such that $\|\lambda^k - \lambda^k_*\| = \text{dist}_{\mathcal{S}_\lambda} \lambda^k$.

By using a Taylor series expansion of $\mathcal{L}(\cdot, \lambda^k)$ around z^k, and using (4.4), (4.6), and (4.15), we have for some ξ^k lying on the line segment between x^* and z^k that

$$\begin{aligned}
&\mathcal{L}(z^k, \lambda^k) - \mathcal{L}(x^*, \lambda^k)\\
&= -\|z^k - x^*\|^2 (d^k)^T \mathcal{L}_{xx}(\xi^k, \lambda^k) d^k\\
&= -\|z^k - x^*\|^2 \left[(d^k)^T \mathcal{L}_{xx}(x^*, \lambda^k_*) d^k + O(\|z^k - x^*\|) + \text{dist}_{\mathcal{S}_\lambda} \lambda^k \right]\\
&\le -\|z^k - x^*\|^2 \bar{\eta}/2 \le 0,
\end{aligned}$$

giving the result. ∎

This result appears similar to one of Mifflin [21, Theorem 5.4], but the assumptions on $\mathcal{L}(\cdot; \lambda^*)$ in that paper are stronger; they require $\mathcal{L}(\cdot; \lambda^*)$ to satisfy a strong convexity property over some convex set containing the iterates z^k, for all k sufficiently large.

The final result follows immediately from Theorems 4.2 and 4.3.

Corollary 4.1. *Suppose that Assumption 4.1 holds. Then any sequences $\{\mu_k\}$ and $\{z^k\}$ with the properties that*

$$\mu_k \downarrow 0, \quad z^k \to x^*, \quad z^k \text{ a local min of } P(\cdot; \mu_k), \tag{4.18}$$

will satisfy

$$\mu_k = \Theta(\|z^k - x^*\|^2). \tag{4.19}$$

5. Discussion

Motivated by the success of primal-dual interior-point methods on linear programming problems, a number of researchers recently have described primal-dual methods for nonlinear programming. In these methods, the Lagrange multipliers λ generally are treated as independent variables, rather than being defined in terms of the primal variables x by a formula such as (2.20). We mention in particular the work of Forsgren and Gill [10], El Bakry et al. [3], and Gay, Overton, and Wright [12], who use line-search methods, and Conn et al. [7] and Byrd, Gilbert, and Nocedal [6], who describe trust-region methods. Methods for nonlinear convex programming are described by Ralph and Wright [27,28], among others.

Near the solution x^*, primal-dual methods gravitate toward points on the primal-dual central path, which is parametrized by μ and defined as the set of points $(x(\mu), \lambda(\mu))$ that satisfies the conditions

$$\nabla f(x) - \sum_{i=1}^m \lambda_i \nabla c_i(x) = 0, \tag{5.1a}$$

$$\lambda_i c_i(x) = \mu, \quad \text{for all } i = 1, 2, \ldots, m, \tag{5.1b}$$

$$\lambda > 0, \quad c(x) > 0. \tag{5.1c}$$

When the LICQ and second-order sufficient conditions hold, the Jacobian matrix of the nonlinear equations formed by (5.1a) and (5.1b) is nonsingular in a neighborhood of (x^*, λ^*), where λ^* is the (unique) optimal multiplier. Fiacco and McCormick use this observation to differentiate the equations (5.1a) and (5.1b) with respect to μ, and thereby prove results about the smoothness of the trajectory $(x(\mu), \lambda(\mu))$ near (x^*, λ^*).

The results of Section 3 above show that the system (5.1) continues to have a solution in the neighborhood of $\{x^*\} \times \mathcal{S}_\lambda$ when LICQ is replaced by MFCQ. We simply take $x(\mu)$ to be the vector described in Theorems 3.1, 3.2, and 3.3, and define $\lambda(\mu)$ by (2.20). Hence, we have existence and local uniqueness of a solution even though the limiting Jacobian of (5.1a), (5.1b) is singular, and we find that the primal-dual trajectory approaches the specific limit point $(x^*, \bar{\lambda}^*)$. The smoothness properties of the path under MFCQ are not obvious, however.

In the case of no strict complementarity, the (weaker) existence results of Section 4 can again be used to deduce the existence of solutions to (5.1) near $\{x^*\} \times \mathcal{S}_\lambda$, but we cannot say much else about this case other than that the convergence rate implied by Lemma 2.3 is satisfied.

Finally, we comment about the use of Newton's method to minimize $P(\cdot;\mu)$ approximately for a decreasing sequence of values of μ, a scheme known as the Newton/log-barrier approach. Extrapolation can be used to obtain a starting point for the Newton iteration after each decrease in μ. Superlinear convergence of this approach is obtained by decreasing μ_k superlinearly to zero (that is, $\lim_{k\to\infty} \mu_{k+1}/\mu_k = 0$) while taking no more than a fixed number of Newton steps at each value of μ_k. In the case of LICQ, rapid convergence of this type has been investigated by Conn, Gould, and Toint [8], Benchakroun, Dussault, and Mansouri [4], Wright and Jarre [37], and Wright [38]. We anticipate that similar results will continue to hold when LICQ is replaced by MFCQ, because the central path continues to be smooth and the convergence domain (3.25) for Newton's method is similar in both cases. A detailed investigation of this claim and an analysis of the case in which strict complementarity fails to hold are left for future study.

Acknowledgements

The authors wish to thank an anonymous referee for a careful reading an earlier draft of this paper, for many enlightening comments, for simplifications in the proofs of Lemma 4.2 and Theorem 4.3, and for pointing out several useful references.

Proof of Lemma 4.1.

Note first that by compactness of $\mathcal{S}_\lambda$ and of the unit ball, we can choose a positive constant $\bar{\eta}$ small enough that

$$d^T \mathcal{L}_{xx}(x^*, \lambda^*) d \geq 6\bar{\eta}\|d\|^2, \quad \text{for all } \lambda^* \in \mathcal{S}_\lambda, \text{ all } d \text{ satisfying (2.18b), (2.18c).}$$

By Hoffmann's lemma [14], there is a constant M such that for any d with $\|d\| = 1$ satisfying (4.2), there is a $\bar{d}$ satisfying (2.18b), (2.18c) such that

$$\|d - \bar{d}\| \leq M\epsilon_d.$$

Therefore, assuming that $\epsilon_d \leq 1/(2M)$, we have

$$\|\bar{d}\| \geq \|d\| - \|d - \bar{d}\| \geq 1 - M\epsilon_d \geq 1/2, \qquad \|\bar{d}\| \leq 3/2.$$

For all $\lambda^* \in \mathcal{S}_\lambda$, we have from these bounds that

$$\begin{aligned} d^T \mathcal{L}_{xx}(x^*, \lambda^*) d &\geq \bar{d}^T \mathcal{L}_{xx}(x^*, \lambda^*) \bar{d} - \|\mathcal{L}_{xx}(x^*, \lambda^*)\| \left[2\|\bar{d}\| \|d - \bar{d}\| + \|d - \bar{d}\|^2 \right] \\ &\geq 6\bar{\eta}\|\bar{d}\|^2 - \|\mathcal{L}_{xx}(x^*, \lambda^*)\| [2\|\bar{d}\| M \epsilon_d + M^2 \epsilon_d^2] \\ &\geq (3/2)\bar{\eta} - \|\mathcal{L}_{xx}(x^*, \lambda^*)\| (3M\epsilon_d + M^2 \epsilon_d^2). \end{aligned}$$

Now reducing ϵ_d if necessary so that

$$\|\mathcal{L}_{xx}(x^*, \lambda^*)\| (3M\epsilon_d + M^2 \epsilon_d^2) \leq \bar{\eta}/2, \text{ for all } \lambda^* \in \mathcal{S}_\lambda,$$

we obtain the result.

References

1. I. Adler and R. D. C. Monteiro. Limiting behavior of the affine scaling continuous trajectories for linear programming problems. *Mathematical Programming*, 50:29–51, 1991.

2. M. Anitescu. On the rate of convergence of sequential quadratic programming with nondifferentiable exact penalty function in the presence of constraint degeneracy. Preprint ANL/MCS-P760-0699, Argonne National Laboratory, Argonne, Illinois, USA, 1999.

3. A. S. El Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang. On the formulation and theory of newton interior point methods for nonlinear programming. *Journal of Optimization Theory and Applications*, 89(3):507–541, 1996.

4. A. Benchakroun, J.-P. Dussault, and A. Mansouri. A two parameter mixed interior-exterior penalty algorithm. *ZOR - Mathematical Methods of Operations Research*, 41:25–55, 1995.

5. J.-F. Bonnans and A. Ioffe. Second-order sufficiency and quadratic growth for nonisolated minima. *Mathematics of Operations Research*, 20(4):801–819, 1995.

6. R. H. Byrd, J.-Ch. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming Series A*, http://dx.doi.org/10.1007/s101070000189, 2000.

7. A. R. Conn, N. I. M. Gould, D. Orban, and Ph. L. Toint. A primal-dual trust-region algorithm for non-convex nonlinear programming. *Mathematical Programming B*, 87(2):215–249, 2000.

8. A. R. Conn, N. I. M. Gould, and Ph. L. Toint. A note on using alternative second-order models for the subproblems arising in barrier function methods for minimization. *Numerische Mathematik*, 68:17–33, 1994.

9. A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Wiley, New York, 1968. Reprinted by SIAM Publications, 1990.

10. A. Forsgren and Ph. E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. *SIAM Journal on Optimization*, 8(4):1132–1152, November 1998.

11. J. Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, 12:136–138, 1977.

12. D. M. Gay, M. L. Overton, and M. H. Wright. A primal-dual interior method for nonconvex nonlinear programming. Technical Report 97-4-08, Computing Sciences Research, Bell Laboratories, Murray Hill, NJ, July 1997.

13. N. I. M. Gould. On the accurate determination of search directions for simple differentiable penalty functions. *IMA Journal of Numerical Analysis*, 6:357–372, 1986.

14. A. J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49:263–265, 1952.
15. M. Kojima, S. Mizuno, and T. Noma. Limiting behavior of trajectories by a continuation method for monotone complementarity problems. *Mathematics of Operations Research*, 15(4):662–675, 1990.
16. F. A. Lootsma. Hessian matrices of penalty functions for solving constrained optimization. *Philips Res. Repts*, 24:322–331, 1969.
17. G. P. McCormick and C. Witzgall. Logarithmic SUMT limits in convex programming. *Mathematical Programming, Series A*, 90:113–145, 2001.
18. L. McLinden. An analogue of Moreau's proximation theorem, with application to the nonlinear complementarity problem. *Pacific Journal of Mathematics*, 88:101–161, 1980.
19. L. McLinden. The complementarity problem for maximal monotone multifunctions. In R. W. Cottle, F. Gianessi, and J.-L. Lions, editors, *Variational Inequalities and Complementarity Problems*, chapter 17, pages 251–270. Wiley, New York, 1980.
20. N. Megiddo. Pathways to the optimal set in linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, pages 131–158, New-York, 1989. Springer Verlag. Identical version in: *Proceedings of the 6th Mathematical Programming Symposium of Japan, Nagoya, Japan*, pages 1–35, 1986.
21. R. Mifflin. Convergence bounds for nonlinear programming algorithms. *Mathematical Programming*, 8:251–271, 1975.
22. R.C. Monteiro and T. Tshuchiya. Limiting behavior of the derivatives of certain trajectories assiciated with a monotone horizontal linear complementarity problem. *Mathematics of Operations Research*, 21:793–814, 1996.
23. R.C. Monteiro and F. Zhou. On the existence and convergence of the central path for convex programming and some duality results. *Computational Optimization and Applications*, 10:51–77, 1998.
24. W. Murray. Analytical expressions for the eigenvalues and eigenvectors of the Hessian matrices of barrier and penalty functions. *Journal of Optimization Theory and Applications*, 7:189–196, 1971.
25. Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia, USA, 1994.
26. J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*. Academic Press, New York and London, 1970.
27. D. Ralph and S. J. Wright. Superlinear convergence of an interior-point method for monotone variational inequalities. In M. C. Ferris and J.-S. Pang, editors, *Complementarity and Variational Problems: State of the Art*, pages 345–385. SIAM Publications, Philadelphia, Penn., 1997.
28. D. Ralph and S. J. Wright. Superlinear convergence of an interior-point method despite dependent constraints. *Mathematics of Operations Research*, 25(2):179–194, May 2000.
29. M. C. Villalobos, R. A. Tapia, and Y. Zhang. The sphere of convergence of newton's method on two equivalent systems from nonlinear programming. Technical Report CRPC-TR9915, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005, September 1999.
30. M. H. Wright. Interior methods for constrained optimization. In *Acta Numerica 1992*, pages 341–407. Cambridge University Press, Cambridge, 1992.
31. M. H. Wright. Some properties of the Hessian of the logarithmic barrier function. *Mathematical Programming*, 67:265–295, 1994.
32. M. H. Wright. Ill-conditioning and computational error in interior methods for nonlinear programming. *SIAM Journal on Optimization*, 9:84–111, 1998.

33. S. J. Wright. Modifying SQP for degenerate problems. Preprint ANL/MCS-P699-1097, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., 1997. Revised June 2000.

34. S. J. Wright. Effects of finite-precision arithmetic on interior-point methods for nonlinear programming. Preprint ANL/MCS-P705-0198, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., 1998. To appear in SIAM Journal on Optimization, 2001.

35. S. J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Computational Optimization and Applications*, 11:253–275, 1998.

36. S. J. Wright. On the convergence of the Newton/log-barrier method. *Mathematical Programming, Series A*, 90:71–100, 2001.

37. S. J. Wright and F. Jarre. The role of linear objective functions in barrier methods. *Mathematical Programming, Series A*, 84:357–373, 1998.

38. S.J. Wright. On the convergence of the Newton/log-barrier method. *Mathematical Programming*, 90:71–100, 2001.