PREPRINT MCS-P556-0196, MATHEMATICS AND COMPUTER SCIENCE DIVISION, ARGONNE NATIONAL LABORATORY

Superlinear Convergence of an Interior-Point Method for Monotone Variational Inequalities

Daniel Ralph* and Stephen Wright†

March 5, 1996

Abstract

We describe an infeasible-interior-point algorithm for monotone variational inequality problems and prove that it converges globally and superlinearly under standard conditions plus a constant rank constraint qualification. The latter condition represents a generalization of the two types of assumptions made in existing superlinear analyses; namely, linearity of the constraints and linear independence of the active constraint gradients.

1 Introduction

We consider the monotone variational inequality over a closed convex set $\mathcal{C} \subset \mathbb{R}^N$:

$$\text{Find } z \in \mathcal{C} \text{ such that } \quad (z' - z)^T \Phi(z) \geq 0, \quad \text{for all } z' \in \mathcal{C}. \tag{1}$$

The mapping $\Phi : \mathbb{R}^N \to \mathbb{R}^N$ is assumed to be continuously differentiable (C^1) and monotone; the latter property means that

$$(z' - z)^T (\Phi(z') - \Phi(z)) \geq 0 \text{ for all } z', z \in \mathbb{R}^N.$$

We assume that $\mathcal{C}$ is defined as an intersection of finitely many algebraic inequalities; that is,

$$\mathcal{C} \;=\; \{z \in \mathbb{R}^N \,|\, g(z) \leq 0\}, \tag{2}$$

where $g : \mathbb{R}^N \to \mathbb{R}^P$ is a C^2 function for which each component function g_i, $i = 1, 2, \ldots, P$, is convex.

*Department of Mathematics, The University of Melbourne, Parkville, Victoria 3052, Australia. The work of this author was supported by the Australian Research Council.

†Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, U.S.A. This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Computational and Technology Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

The mixed nonlinear complementarity (NCP) formulation of this problem is: Find the vector triple (z, λ, y) such that

$$\begin{bmatrix} 0 \\ y \end{bmatrix} = \begin{bmatrix} f(z,\lambda) \\ -g(z) \end{bmatrix}, \qquad (\lambda, y) \geq 0, \qquad \lambda^T y = 0, \tag{3}$$

where $f : \mathbb{R}^N \to \mathbb{R}^N$ is the C^1 function defined by

$$f(z,\lambda) = \Phi(z) + Dg(z)^T \lambda. \tag{4}$$

Note that f is monotone with respect to $z \in \mathbb{R}^N$ for all vectors $\lambda \in \mathbb{R}^P$ with nonnegative components (that is, $\lambda \in \mathbb{R}^P_+$). The mapping

$$(z,\lambda) \to \begin{bmatrix} f(z,\lambda) \\ -g(z) \end{bmatrix} \tag{5}$$

is monotone because monotonicity of Φ and of each function Dg_i means that its Jacobian matrix

$$\begin{bmatrix} D_z f & Dg^T \\ -Dg & D_\lambda h \end{bmatrix} = \begin{bmatrix} D\Phi(z) + \sum_{i=1}^P \lambda_i D^2 g_i(z) & Dg(z)^T \\ -Dg(z) & 0 \end{bmatrix} \tag{6}$$

is positive semidefinite for all $(z, \lambda) \in \mathbb{R}^N \times \mathbb{R}^P_+$.

It is well known [2] that, under suitable conditions on g such as the famous Slater constraint qualification, z solves (1) if and only if there exists a multiplier λ such that (z, λ) solves (3).

We solve (1) by a method based on the interior-point algorithm of Wright and Ralph [10]. Besides being easier to adapt to the case of mixed NCP (3), it is also considerably simpler than the algorithm in [10], in fact, closer in spirit to the method of Wright [8] for monotone linear complementarity problems. We show that under certain assumptions the method converges globally and superlinearly to the solution set of (3), even in some situations in which the solution does not satisfy a strong uniqueness and nondegeneracy condition.

Superlinear convergence for interior-point methods was discussed first by Zhang, Tapia, and Dennis [15]; see also Zhang and Tapia [14] and Ye, Güler, Tapia, and Zhang [13]. Infeasible-interior-point methods for the latter class were described by Wright in [9], with improvements in [7, 8]. For *nonlinear* monotone complementarity problems, Wright and Ralph [10] describe a superlinearly convergent method that requires invertibility of the principal submatrix of the Jacobian corresponding to basic rows and columns. This condition actually guarantees uniqueness of the solution point (z^*, λ^*), that is, uniqueness of the multiplier λ^* in (3). Similar assumptions almost always are made in the asymptotic analysis of nonlinear programming algorithms. The main point of this paper is to show that superlinear convergence also occurs under weaker assumptions that allow the multiplier λ to be nonunique. In fact, the algorithm here is the only one we know of for nonlinear programs with nonlinear constraints and nonunique multipliers for which convergence is superlinear.

Loosening of degeneracy assumptions has practical importance for large-scale problems, where degeneracy or near-degeneracy at solution points is typical. In this paper, we assume that the active constraint gradients satisfy a *constant rank constraint qualification* at the solution. This condition can be thought of as an interpolation between the two most commonly made assumptions, namely, linear independence of the active constraint gradients and linearity of the constraint function $g(z)$.

Possibly the best known application of (1) is the convex programming problem defined by

$$\min_z \phi(z) \qquad \text{subject to } z \in \mathcal{C}, \tag{7}$$

where $\phi : \mathbb{R}^N \to \mathbb{R}$ is C^2 and convex. Let $\Phi = D\phi$. It is easy to show that the NCP formulation (3),(4) is equivalent to the standard Karush-Kuhn-Tucker (KKT) conditions for (7). If a constraint qualification holds, the solutions of (1) and (7) coincide.

The paper is developed as follows. In the remainder of this section, we summarize the notation and terminology to be used in the paper. (Because of the technical nature of our analysis, it is useful to have this material gathered in one place.) In Section 2, we describe the algorithm for solving (3), but omit some of the details because of the similarity to Wright and Ralph [10]. In Section 3, we prove the global convergence result for this algorithm and state the local superlinear convergence result. The analysis in this section is quite similar to that of [10], but it differs in some of the details. The rest of the paper is devoted to outlining and proving the superlinear convergence theorem. In Section 4 we state and discuss the assumptions that are used in this theorem. Section 5 shows that the steps generated by the algorithm during its final stages satisfy the estimate required by the proof of the superlinear convergence theorem. We divide Section 5 into subsections and provide ample motivating discussion so that readers can see the thrust of our argument without our going into the details. Section 6 describes conditions under which one of our key assumptions—existence of a limit point—is satisfied, and also proves some auxiliary results that follow from the assumptions of Section 4.

Notation and Terminology

Unless otherwise specified, $\|\cdot\|$ denotes the Euclidean norm of a vector, while

$$\mathbb{R}^P_+ = \{y \in \mathbb{R}^P \mid y \geq 0\}, \qquad \mathbb{R}^P_{++} = \{y \in \mathbb{R}^P \mid y > 0\}.$$

For any two vectors c and d, we frequently use (c, d) as shorthand for $(c^T, d^T)^T$. The vector $(1, 1, \ldots, 1)$ is denoted by e, while z_+ is obtained by replacing all negative components in the vector z by zero. The closed unit ball is denoted by $\mathbb{B}$. Derivatives are indicated by D, or D_z for a partial derivative with respect to z.

Iteration indices (usually k) appear as superscripts on vectors and matrices and as subscripts on scalars. Subscripts are used to indicate components of vectors and matrices.

If ψ is a function mapping $\mathbb{R}_+$ to $\mathbb{R}_+$, we write $\psi(\tau) = O(\tau)$ if there are constants $\bar{\tau} > 0$ and $C > 0$ such that $\psi(\tau) \leq C\tau$ for $\tau \in (0, \bar{\tau}]$.

The kernel or null space of a matrix $H \in \mathbb{R}^{p \times q}$ is

$$\ker H \;=\; \{d \in \mathbb{R}^q \mid Hd = 0\},$$

while the range space is denoted by

$$\operatorname{ran} H \;=\; \{Hd \,|\, d \in \mathbb{R}^q\}.$$

Given $\emptyset \neq \mathcal{I} \subset \{1, 2, \ldots, p\}$ and $\emptyset \neq \mathcal{J} \subset \{1, 2, \ldots, q\}$, we define three submatrices of H as follows:

$$H_{\mathcal{I}\mathcal{J}} = [H_{ij}]_{i\in\mathcal{I}, j\in\mathcal{J}}, \qquad H_{\cdot\mathcal{J}} = [H_{ij}]_{i=1,\ldots,p; j\in\mathcal{J}}, \qquad H_{\mathcal{I}\cdot} = [H_{ij}]_{i\in\mathcal{I}; j=1,\ldots,q}.$$

If $w \in \mathbb{R}^p$ and $\mathcal{I} \subset \{1, 2, \ldots, p\}$, then $w_\mathcal{I}$ denotes the subvector $[w_i]_{i\in\mathcal{I}}$. In dealing with the function $g : \mathbb{R}^N \to \mathbb{R}^P$ in (2), we use $Dg_\mathcal{I}(z)$ to denote the $|\mathcal{I}| \times N$ Jacobian of $g_\mathcal{I}$ with respect to z.

Often the arguments are omitted from the functions and Jacobians $f(z, \lambda)$, $g(z)$, $D_z f(z, \lambda)$, and so on. In such cases, the arguments should be assumed to be z, λ, and y, or any applicable combination thereof.

We use $\mathcal{S}$ to denote the solution set for (3) and $\mathcal{S}_Z$ to denote its projection onto its first $N + P$ components; that is,

$$\mathcal{S} = \{(z, \lambda, y) \,|\, (z, \lambda, y) \text{ solves } (3)\}, \qquad \mathcal{S}_Z = \{(z, \lambda) \,|\, (z, \lambda, -g(z)) \in \mathcal{S}\}. \tag{8}$$

We can partition $\{1, 2, \ldots, P\}$ into two index sets $\mathcal{B}$ and $\mathcal{N}$ such that

$$\lambda^*_\mathcal{N} = 0, \qquad y^*_\mathcal{B} = 0, \qquad \text{all } (z^*, \lambda^*, y^*) \in \mathcal{S}. \tag{9}$$

The solution (z^*, λ^*, y^*) is *strictly complementary* if $\lambda^* + y^* > 0$; that is, $\lambda^*_\mathcal{B} > 0$ and $y^*_\mathcal{N} > 0$. We also use this term when referring to just the z and λ components of the solution. That is, we say (z^*, λ^*) is *strictly complementary* if $\lambda^*_\mathcal{B} > 0$ and $g_i(z^*) < 0$ for $i \in \mathcal{N}$.

The distance of a vector $w \in \mathbb{R}^p$ to a set $\mathcal{T} \subset \mathbb{R}^p$ is

$$\operatorname{dist}_\mathcal{T}(w) = \inf\{\|w - w^*\| \,|\, w^* \in \mathcal{T}\}.$$

Given $\mathcal{H} \subset \mathbb{R}^{p\times q}$, we say $\mathcal{H}$ has *constant column rank* (CCR) if for each sequence $\{H^k\} \subset \mathcal{H}$ converging to some $H \in \mathbb{R}^{p\times q}$ and each $\emptyset \neq \mathcal{J} \subset \{1, 2, \ldots, q\}$, we have

$$\operatorname{rank} H^k_{\cdot\mathcal{J}} \quad\to\quad \operatorname{rank} H_{\cdot\mathcal{J}}.$$

Given the current point (z, λ, y) and a search direction $(\Delta z, \Delta\lambda, \Delta y)$, we define the *complementarity measure* μ as

$$\mu = \lambda^T y/P,$$

and the intermediate quantities $(z(\alpha), \lambda(\alpha), y(\alpha))$ and $\mu(\alpha)$ by

$$(z(\alpha), \lambda(\alpha), y(\alpha)) = (z, \lambda, y) + \alpha(\Delta z, \Delta\lambda, \Delta y), \qquad \mu(\alpha) = \lambda(\alpha)^T y(\alpha)/P.$$

2 An Algorithm for Mixed NCP

We now outline an infeasible-interior-point algorithm for mixed NCP that synthesizes two earlier methods: the algorithm described by Wright and Ralph [10] for monotone NCP and the algorithm of Wright [8] for linear complementarity problems. Neither of these formulations applies explicitly to the mixed problem. In the case of linear problems, a mixed framework is unnecessary in any case, since there are strong equivalence relationships between mixed problems and nonmixed problems.

Our description is terse because much of the motivation can be found in the papers cited above.

Given a starting point (z^0, λ^0, y^0) with $(\lambda^0, y^0) > (0,0)$, the algorithm generates a sequence of iterates (z^k, λ^k, y^k) that satisfies this same positivity condition. For each vector triple (z, λ, y) for which $(\lambda, y) > 0$, we define the residuals r_f and r_g by

$$\begin{bmatrix} r_f(z,\lambda) \\ r_g(z,y) \end{bmatrix} = \begin{bmatrix} -f(z,\lambda) \\ y + g(z) \end{bmatrix}. \tag{10}$$

Another useful quantity is the vector e, defined by $e = (1, 1, \ldots, 1)^T$. As is usual in descriptions of interior-point methods, we turn positive vectors into diagonal matrices by capitalizing their names; that is,

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m), \qquad Y = \mathrm{diag}(y_1, y_2, \ldots, y_m).$$

When $(z, \lambda, y) = (z^k, \lambda^k, y^k)$, we sometimes attach a subscript or superscript k to the quantities μ, r, Λ, Y to make the dependence on (z^k, λ^k, y^k) explicit.

The algorithm can be thought of as a modified Newton algorithm applied to the following system of constrained nonlinear equations.

$$\begin{bmatrix} -f(z,\lambda) \\ y + g(z) \\ -\Lambda Y e \end{bmatrix} = \begin{bmatrix} r_f(z,\lambda) \\ r_g(z,y) \\ -\Lambda Y e \end{bmatrix} = 0, \qquad (\lambda, y) \geq 0. \tag{11}$$

The "modifications" are needed to keep λ^k and y^k from prematurely approaching the boundary of the feasible region defined by the conditions $y \geq 0$ and $\lambda \geq 0$. Line searches are used and, on some iterates, the search direction is skewed toward the interior of the positive orthant, so that longer steps can be taken without violating positivity. Near the solution, the algorithm reverts to pure Newton steps, allowing the rapid local convergence properties of this method to take effect.

The major computational operation in the algorithm is the repeated solution of $2P + N$-dimensional linear systems of the form

$$\begin{bmatrix} D_z f & Dg^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta\lambda \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_f(z,\lambda) \\ r_g(z,y) \\ -\Lambda Y e + \tilde{\sigma}\mu_k e \end{bmatrix}, \tag{12}$$

where the *centering parameter* $\tilde{\sigma}$ lies in the range $[0, \frac{1}{2}]$. These equations are simply the Newton equations for the nonlinear system mentioned earlier, except for the $\tilde{\sigma}$ term. The algorithms searches along the direction obtained from (12).

In the algorithm of Wright and Ralph [10] (which applies to nonmixed NCP), the search for α takes place along a curved arc rather than a straight line. The curvature on this arc ensures that the residual term decreases linearly with α. It is not clear how to extend this strategy to the mixed case, so the algorithm in this paper uses a simpler straight-line search. The global and local convergence properties are essentially the same as in [10].

At each iteration, the algorithm computes a fast step—a pure Newton step for which $\tilde{\sigma} = 0$ in (12). If the fast step fails to give a sufficiently large decrease in μ, we revert to a safe step by assigning a positive value to $\tilde{\sigma}$. This modification allows a longer step to be taken, so that a certain minimal amount of progress toward the solution can be made. In choosing the step length α, we require not only that all iterates (z^k, λ^k, y^k) remain strictly positive, but also that they satisfy

$$\lambda_i^k y_i^k \geq \gamma_k \mu_k, \qquad i = 1, 2, \ldots, P, \tag{13}$$

for positive values of γ_k bounded away from zero. This condition ensures that the pairwise products $\lambda_i y_i$ stay roughly in balance as they approach zero, so that no single one of them vanishes much faster than the others. On fast steps, we expand this region by decreasing γ slightly, to allow steps of length near 1 to be taken.

The algorithm is parametrized by a variety of positive scalar constants, which we specify now for easy reference. Their roles are explained as they arise in subsequent discussions:

$$\begin{gathered} \chi \in (0,1), \qquad \bar{\sigma} \in (0, \tfrac{1}{2}), \qquad \bar{\alpha} \in (0,1], \qquad \kappa \in (0,1), \qquad \hat{\tau} \in (0,1), \\ \beta_{\min} > 0 \ \text{ such that } \|r_f^0\| \leq \beta_{\min}\mu_0 \text{ and } \|r_g^0\| \leq \beta_{\min}\mu_0, \qquad \beta_{\max} = \beta_{\min} e^{3/2}, \\ 0 < \gamma_{\min} < \gamma_{\max} \leq \tfrac{1}{2}, \qquad \bar{\gamma} \in (0, \tfrac{1}{2}), \qquad \rho \in (0, \min((\tfrac{1}{2}\bar{\gamma})^{1/\hat{\tau}}, 1-\kappa)). \end{gathered} \tag{14}$$

The starting point (z^0, λ^0, y^0) is assumed to satisfy

$$\lambda_i^0 y_i^0 \geq \gamma_{\max}\mu_0. \tag{15}$$

The main algorithm can now be specified.

$t_0 \leftarrow 0$; $\gamma_0 \leftarrow \gamma_{\max}$; $\beta_0 \leftarrow \beta_{\min}$;
for $k = 0, 1, 2, \ldots,$
  if $\mu_k = 0$,
    terminate with solution (z^k, λ^k, y^k);

  $(z^{k+1}, \lambda^{k+1}, y^{k+1}) \leftarrow$ **fast**$(z^k, \lambda^k, y^k, t_k, \gamma_k, \beta_k)$;
  if $\mu_{k+1} \leq \rho\mu_k$
    $\gamma_{k+1} \leftarrow \gamma_{\min} + \bar{\gamma}^{t_k}(\gamma_{\max} - \gamma_{\min})$; $\beta_{k+1} \leftarrow (1 + \bar{\gamma}^{t_k+1})\beta_k$;
    $t_{k+1} \leftarrow t_k + 1$;

else

$(z^{k+1}, \lambda^{k+1}, y^{k+1}) \leftarrow \textbf{safe}(z^k, \lambda^k, y^k, t_k, \gamma_k, \beta_k)$;
$\gamma_{k+1} \leftarrow \gamma_k$; $\beta_{k+1} \leftarrow \beta_k$;
$t_{k+1} \leftarrow t_k$;

end for.

The fast step is taken only if it decreases the complementarity gap μ by at least a factor of ρ. The counter t_k keeps track of the number of successful fast steps prior to iteration k. As we see in the definitions of the subroutines **fast** and **safe** below, the value of t_k indirectly governs the distance α_k that we move along the current search direction.

The coefficient matrix in (12) is the same for both fast and safe steps, so only one matrix factorization is required per iteration.

The safe-step procedure is defined as follows.

safe$(z, \lambda, y, t, \gamma, \beta)$:
choose $\tilde{\sigma} \in [\bar{\sigma}, \frac{1}{2}]$, $\alpha^0 \in [\bar{\alpha}, 1]$;
solve (12) to find $(\Delta z, \Delta\lambda, \Delta y)$;
choose α to be the first element in the sequence $\alpha^0, \chi\alpha^0, \chi^2\alpha^0, \ldots$,
such that the following conditions are satisfied:

$$\begin{aligned} \lambda_i(\alpha) y_i(\alpha) &\geq \gamma\, \mu(\alpha), && \text{(16a)} \\ \|r_f(z(\alpha), \lambda(\alpha))\| &\leq \beta\mu(\alpha); && \text{(16b)} \\ \|r_g(z(\alpha), y(\alpha))\| &\leq \beta\mu(\alpha); && \text{(16c)} \\ \mu(\alpha) &\leq [1 - \alpha\kappa(1-\tilde{\sigma})]\, \mu && \text{(16d)} \end{aligned}$$

return $(z(\alpha), \lambda(\alpha), y(\alpha))$.

A nonzero centering term is used, allowing us to move a nontrivial distance along the search direction while staying in the set defined by

$$\{(z, \lambda, y) \,|\, \lambda_i y_i \geq \gamma\mu\}. \tag{17}$$

The second and third acceptance conditions (16b), (16c) ensure that the infeasibility remains bounded by a multiple of the complementarity. The infeasibility is "squeezed" to zero at least as rapidly as the complementarity measure. Similar conditions are enforced in infeasible-interior-point algorithms for linear complementarity and linear programming; see, for example, Wright [8]. The fourth condition (16d) is a "sufficient decrease" condition of the kind often found in algorithms for nonlinear optimization. Its purpose is to ensure that the decrease in objective function (in this case, μ) achieves at least a fraction κ of the decrease promised by the linearized model (12).

Fast-step calculations are a little more complicated. Since they use no centering ($\tilde{\sigma} = 0$), it may not be possible to satisfy the acceptance criteria (16) regardless of how small we choose

α. Hence, these criteria must be relaxed but not abandoned. The amount of relaxation is large enough to allow near-unit steps to be taken near the solution, but small anough to keep the iterates inside a neighborhood of the central path. These opposing considerations are balanced by making the amount of relaxation geometric in the fast step counter t.

fast$(z, \lambda, y, t, \gamma, \beta)$:
solve (12) with $\tilde{\sigma} = 0$ to find $(\Delta z, \Delta\lambda, \Delta y)$;
set $\tilde{\gamma} = \gamma_{\min} + \bar{\gamma}^{t+1}(\gamma_{\max} - \gamma_{\min})$; set $\tilde{\beta} = (1 + \bar{\gamma}^{t+1})\beta$;
define

$$\alpha^0 = 1 - \frac{\mu^{\hat{\tau}}}{\bar{\gamma}^t}; \tag{18}$$

if $\alpha^0 \leq 0$ **return**(z, λ, y);
choose α to be the first element in the sequence $\alpha^0, \chi\alpha^0, \chi^2\alpha^0, \ldots$, such that the following conditions are satisfied:

$$\begin{aligned} \lambda_i(\alpha) y_i(\alpha) &\geq \tilde{\gamma}\, \mu(\alpha), & \text{(19a)} \\ \|r_f(z(\alpha), \lambda(\alpha))\| &\leq \tilde{\beta}\mu(\alpha); & \text{(19b)} \\ \|r_g(z(\alpha), y(\alpha))\| &\leq \tilde{\beta}\mu(\alpha); & \text{(19c)} \end{aligned}$$

return $(z(\alpha), \lambda(\alpha), y(\alpha))$.

Note that a sufficient decrease condition is not needed in (19); the acceptance test $\mu_{k+1} \leq \rho\mu_k$ in the main algorithm performs this check.

Before embarking on the convergence analysis, we note that the following conditions are satisfied by every iterate (z^k, λ^k, y^k):

$$\begin{aligned} \lambda_i^k y_i^k &\geq \gamma_k \mu_k \geq \gamma_{\min}\mu_k, \qquad i = 1, 2, \ldots, P, & \text{(20a)} \\ \max(\|r_f^k\|, \|r_g^k\|) &\leq \beta_k \mu_k. & \text{(20b)} \end{aligned}$$

Note too that β_k is bounded. In fact,

$$\beta_{\min} \leq \beta_k = \beta_{\min} \prod_{j=1}^{t_k} (1 + \bar{\gamma}^j) \leq \beta_{\min} \prod_{j=1}^{\infty} (1 + (\tfrac{1}{2})^j) \leq \beta_{\min} e^{3/2} = \beta_{\max}, \tag{21}$$

where e in this case is Euler's constant and not the vector of 1s.

3 Convergence

In this section we first prove global convergence and then discuss superlinear local convergence.

3.1 Global Convergence

We prove here a global convergence result: either the sequence of iterates terminates finitely at a solution, or all limit points are solutions of (3). To prove this result, we use a simple technique due to Polak [6, Chapter 1].

We start by formalizing our assumptions on Φ and g.

Assumption 1 $\Phi : \mathbb{R}^N \to \mathbb{R}^N$ *is* C^1 *and monotone; and each component function* g_i *of* $g : \mathbb{R}^N \to \mathbb{R}^P$ *is* C^2 *and convex.*

It follows immediately from this definition and (6) that

$$\begin{bmatrix} D_z f & Dg^T \\ -Dg & 0 \end{bmatrix}$$

is postive semidefinite for each $(z, \lambda) \in \mathbb{R}^N \times \mathbb{R}^P_+$.

Recall that $\mathcal{S}$ is the solution set for (3). All iterates of the algorithm are confined to the set Ω, defined by

$$\begin{aligned} \Omega = \{(z, \lambda, y) \mid (\lambda, y) \geq 0, & \\ \|r_f(z, \lambda)\| \leq \beta_{\max}\mu, \ \ \|r_g(z, y)\| \leq \beta_{\max}\mu, \ \ \lambda_i y_i \geq \gamma_{\min}\mu, \ \ i = 1, 2, \ldots, P\}. \end{aligned} \tag{22}$$

We also define

$$\Omega_{++} = \Omega \cap (\mathbb{R}^N \times \mathbb{R}^P_{++} \times \mathbb{R}^P_{++})$$

and note that

$$\Omega = \Omega_{++} \cup \mathcal{S}, \qquad \Omega_{++} \cap \mathcal{S} = \emptyset.$$

In this definition, $\mathbb{R}^P_{++}$ is the strictly positive orthant in $\mathbb{R}^P$ and $\mu = \lambda^T y / P$ as before.

The result that $(z^k, \lambda^k, y^k) \in \Omega$ for all k follows from (14) and (21).

By monotonicity, we know that the submatrix $D_z f$ in the Jacobian is positive semidefinite. To ensure that the Newton-like equations (12) have a unique solution, we impose a slightly stronger condition.

Assumption 2 *The two sided projection of the matrix*

$$D_z f(z, \lambda) = D\Phi(z) + \sum_{i=1}^{P} \lambda_i D^2 g_i(z)$$

into $\ker Dg(z)$ *is positive definite for all* $z \in \mathbb{R}^N$ *and* $\lambda \in \mathbb{R}^P_{++}$.

To verify that (12) has a unique solution, eliminate Δy and $\Delta\lambda$ from (12), and note that the coefficient matrix in the reduced system defined by

$$\left[D_z f + (Dg)^T \Lambda Y^{-1}(Dg)\right] \Delta z = -f(z, \lambda) - (Dg)^T \Lambda Y^{-1}(g(z) + \mu_k \tilde{\sigma} \Lambda^{-1} e)$$

is positive definite.

Assumptions 1 and 2 imply that the algorithm takes a nontrivial step α_k along the computed search direction—and therefore makes a nontrivial amount of progress—at every iteration. The first result indicates that this claim is true in the case of safe steps.

Lemma 1 *Suppose that Assumptions 1 and 2 hold. Let $(\hat{z}, \hat{\lambda}, \hat{y}) \in \Omega \backslash \mathcal{S}$. Then there are scalars $\hat{\delta} > 0$ and $\hat{\alpha} \in (0, 1]$ such that if the algorithm takes a safe step from any point (z, λ, y) satisfying*

$$(z, \lambda, y) \in \hat{\mathcal{B}} \stackrel{\triangle}{=} (\hat{z}, \hat{\lambda}, \hat{y}) + \hat{\delta}\mathrm{I\!B}, \tag{23}$$

the calculated step length α will satisfy $\alpha \geq \hat{\alpha}$.

Proof. We define $\hat{\delta}$ by

$$\hat{\delta} = \tfrac{1}{2} \min_{i=1,2,\ldots,P} \left(\min(\hat{\lambda}_i, \hat{y}_i) \right) > 0.$$

For $(z, \lambda, y) \in \hat{\mathcal{B}}$, we then have

$$\lambda_i y_i \geq (\hat{\lambda}_i - \hat{\delta})(\hat{y}_i - \hat{\delta}) \geq \hat{\delta}^2, \qquad \mu = \lambda^T y / P \geq \hat{\delta}^2. \tag{24}$$

Note from (20a) that, if the safe step routine is called at the point (z, λ, y), then

$$\lambda_i y_i \geq \gamma \mu, \qquad i = 1, 2, \ldots, P,$$

for the value of γ that is passed to the routine **safe**.

Since $\lambda > 0$ and $y > 0$ for all $(z, \lambda, y) \in \hat{\mathcal{B}}$, the coefficient matrix in (12) is nonsingular and continuous in an open set containing $\hat{\mathcal{B}}$. The right-hand side in (12) is also continuous with respect to (z, λ, y) and $\tilde{\sigma}$. Hence, there is a constant $C_6 > 0$ such that

$$\|(\Delta z, \Delta \lambda, \Delta y)\| \leq C_6 \tag{25}$$

for all $(z, \lambda, y) \in \hat{\mathcal{B}}$, $\tilde{\sigma} \in [\bar{\sigma}, \frac{1}{2}]$.

Define $\hat{\alpha}^{(1)} = \hat{\delta}/(2C_6)$. We then have for all $\alpha \in [0, \hat{\alpha}^{(1)}]$ that

$$\lambda_i + \alpha \Delta \lambda_i \geq \hat{\lambda}_i - \hat{\delta} - \frac{\hat{\delta}}{2C_6} |\Delta \lambda_i| \geq 2\hat{\delta} - \hat{\delta} - \tfrac{1}{2}\hat{\delta} > 0,$$

and similarly for $y_i + \alpha \Delta y_i$.

Now define

$$\hat{\alpha}^{(2)} = \min \left(\hat{\alpha}^{(1)}, \frac{\bar{\sigma}(1 - \gamma_{\max})\hat{\delta}^2}{2C_6^2} \right).$$

We now show that the first acceptance criterion (16a) is satisfied for all $\alpha \in [0, \hat{\alpha}^{(2)}]$. From the last block row in (12), we have

$$\begin{aligned} \lambda_i(\alpha) y_i(\alpha) &= \lambda_i y_i - \alpha \lambda_i y_i + \alpha \tilde{\sigma} \mu + \alpha^2 \Delta \lambda_i \Delta y_i \\ &\geq \gamma \mu (1 - \alpha) + \alpha \tilde{\sigma} \mu - \alpha^2 C_6^2, \end{aligned}$$

since $\lambda_i y_i \geq \gamma$. Using (12) again, we also have

$$\begin{aligned} \lambda(\alpha)^T y(\alpha) &= \lambda^T y - \alpha(1 - \tilde{\sigma}) \lambda^T y + \alpha^2 \Delta \lambda^T \Delta y \\ &\leq \lambda^T y - \alpha(1 - \tilde{\sigma}) \lambda^T y + \alpha^2 C_6^2. \end{aligned} \tag{26}$$

By combining these two estimates, we find that (16a) is satisfied if

$$\gamma\mu(1-\alpha)+\alpha\tilde{\sigma}\mu-\alpha^2 C_6^2 \geq \gamma\mu(1-\alpha)+\alpha\gamma\tilde{\sigma}\mu+\alpha^2\gamma C_6^2/P,$$

which, in turn, is satisfied if

$$\tilde{\sigma}\alpha(1-\gamma)\mu \geq 2\alpha^2 C_6^2.$$

Since $\mu \geq \hat{\delta}^2$, $\gamma \in (\gamma_{\min}, \gamma_{\max}]$, and $\tilde{\sigma} \geq \bar{\sigma}$, this last condition holds for all $\alpha \in [0, \hat{\alpha}^{(2)}]$, so the condition (16a) is satisfied for α in this range.

We now prove that the fourth condition (16d) holds for all $\alpha \in [0, \hat{\alpha}^{(3)}]$, where

$$\hat{\alpha}^{(3)} = \min\left(\hat{\alpha}^{(2)}, P(1-\kappa)\frac{\hat{\delta}^2}{2C_6^2}\right).$$

For α in this range, we have from $\mu \geq \hat{\delta}^2$, in (24), and $\tilde{\sigma} \leq \frac{1}{2}$ that

$$\alpha^2 C_6^2 \leq \alpha P(1-\kappa)\frac{\hat{\delta}^2}{2} \leq \alpha(1-\kappa)(1-\tilde{\sigma})\lambda^T y.$$

Hence, from (26), we have

$$\lambda(\alpha)^T y(\alpha) \leq \lambda^T y - \alpha(1-\tilde{\sigma})\lambda^T y + \alpha(1-\kappa)(1-\tilde{\sigma})\lambda^T y \leq [1-\alpha\kappa(1-\tilde{\sigma})]\lambda^T y,$$

as required.

We turn next to the second condition (16b). From Taylor's theorem and (12), we have

$$\begin{aligned} f(z(\alpha),\lambda(\alpha)) &= f(z,\lambda)+\alpha\begin{bmatrix} D_z f & D_\lambda f \end{bmatrix}\begin{bmatrix} \Delta z \\ \Delta\lambda \end{bmatrix} \\ &\quad +\alpha\int_0^1 [Df(z+\theta\alpha\Delta z, \lambda+\theta\alpha\Delta\lambda)-Df(z,\lambda)]\begin{bmatrix} \Delta z \\ \Delta\lambda \end{bmatrix} d\theta \\ &= (1-\alpha)f(z,\lambda)+\alpha\Delta r_f, \end{aligned} \tag{27}$$

where we have defined

$$\Delta r_f = \int_0^1 [Df(z+\theta\alpha\Delta z, \lambda+\theta\alpha\Delta\lambda)-Df(z,\lambda)]\begin{bmatrix} \Delta z \\ \Delta\lambda \end{bmatrix} d\theta. \tag{28}$$

By taking norms, we obtain

$$\|\Delta r_f\| \leq \max_{\theta\in(0,1)} \|Df(z+\theta\alpha\Delta z, \lambda+\theta\alpha\Delta\lambda)-Df(z,\lambda)\| \, \|(\Delta z, \Delta\lambda)\|. \tag{29}$$

Therefore, by continuity of Df (Assumption 1) and the bound (25), there is a scalar $\hat{\alpha}^{(4)} \in (0, \hat{\alpha}^{(3)}]$ such that

$$\alpha \in [0, \hat{\alpha}^{(4)}] \quad \Rightarrow \quad \|\Delta r_f\| \leq \tfrac{1}{2}\bar{\sigma}\beta_{\min}\hat{\delta}^2, \tag{30}$$

for all $(z, \lambda, y) \in \hat{\mathcal{B}}$ from which a safe step is calculated. By reducing $\hat{\alpha}^{(4)}$, if necessary, we can also assert that

$$\alpha \in [0, \hat{\alpha}^{(4)}] \quad \Rightarrow \quad \alpha C_6^2 \leq \tfrac{1}{2}\bar{\sigma} P \hat{\delta}^2. \tag{31}$$

By taking norms in (27) and using (20b), we find that

$$\begin{aligned} \|f(z(\alpha), \lambda(\alpha))\| &\leq (1-\alpha)\|f(z,\lambda)\| + \alpha\|\Delta r_f\| \\ &\leq (1-\alpha)\beta\mu + \alpha\|\Delta r_f\|. \end{aligned} \tag{32}$$

Meanwhile, we have by a slight change to (26) (bounding below instead of above) that

$$\lambda(\alpha)^T y(\alpha) \geq \lambda^T y(1 - \alpha + \alpha\tilde{\sigma}) - \alpha^2 C_6^2.$$

Trivial rearrangement of this expression gives

$$(1-\alpha)\mu \leq \mu(\alpha) - \alpha\tilde{\sigma}\mu + \alpha^2 C_6^2/P.$$

By substituting into (32), we obtain

$$\begin{aligned} \|f(z(\alpha), \lambda(\alpha))\| &\leq \beta\mu(\alpha) - \beta\alpha\tilde{\sigma}\mu + \beta\alpha^2 C_6^2/P + \alpha\|\Delta r_f\| \\ &= \beta\mu(\alpha) - \alpha\left[\beta\tilde{\sigma}\mu - \beta\alpha C_6^2/P - \|\Delta r_f\|\right]. \end{aligned} \tag{33}$$

Since $\tilde{\sigma} \geq \bar{\sigma}$ and $\beta \geq \beta_{\min}$, we have from (24), (30), and (31) that

$$\|\Delta r_f\| \leq \tfrac{1}{2}\tilde{\sigma}\beta\mu, \qquad \beta\alpha C_6^2/P \leq \tfrac{1}{2}\tilde{\sigma}\beta\mu,$$

for all $\alpha \in [0, \hat{\alpha}^{(4)}]$. Hence, the bracketed term in (33) is nonnegative, and we have

$$\|r_f(z(\alpha), \lambda(\alpha))\| = \|f(z(\alpha), \lambda(\alpha))\| \leq \beta\mu(\alpha),$$

for all $\alpha \in [0, \hat{\alpha}^{(4)}]$, as required.

By an almost identical argument, we can show that the third condition (16c) holds for $\alpha \in [0, \hat{\alpha}^{(4)}]$, though we may have to choose $\hat{\alpha}^{(4)}$ smaller (but still positive).

We have shown that the criteria (16) are satisfied for all $\alpha \in [0, \hat{\alpha}^{(4)}]$. Hence, the step length selected by **safe** will be at least as long as the first value of α below $\hat{\alpha}^{(4)}$ that is tried by the Armijo backtracking strategy. We deduce that

$$\alpha \geq \hat{\alpha} \stackrel{\triangle}{=} \min(\bar{\alpha}, \chi\hat{\alpha}^{(4)}),$$

and our proof is complete. ∎

The global convergence result and its proof are similar to Theorem 3.3 of Wright and Ralph [10].

Theorem 1 *Suppose that Assumptions 1 and 2 hold. Then either*

(A) $(z^k, \lambda^k, y^k) \in \mathcal{S}$ *for some* $k < \infty$*, or*

(B) all limit points of $\{(z^k, \lambda^k, y^k)\}$ *belong to* $\mathcal{S}$*.*

Proof. Suppose for contradiction that the sequence $\{(z^k, \lambda^k, y^k)\}$ is infinite, with a limit point $(\hat{z}, \hat{\lambda}, \hat{y})$ that does not belong to $\mathcal{S}$. Since the sequence is contained entirely in the closed set Ω, we must have $(\hat{z}, \hat{\lambda}, \hat{y}) \in \Omega \backslash \mathcal{S}$. We must have $(\hat{\lambda}, \hat{y}) > 0$, since otherwise it would follow from the definition of Ω that $\hat{\mu} = \hat{\lambda}^T \hat{y}/P = 0$ and hence $(\hat{z}, \hat{\lambda}, \hat{y}) \in \mathcal{S}$. Hence, $\hat{\mu} > 0$.

Let $\mathcal{K}$ be an infinite subsequence such that

$$\{(z^k, \lambda^k, y^k)\}_{k \in \mathcal{K}} \to (\hat{z}, \hat{\lambda}, \hat{y}).$$

Since $\{\mu_k\}$ is monotone decreasing, we have $\mu_k \geq \hat{\mu}$ for all k. If a safe step is taken from the k-th iterate, for some $k \in \mathcal{K}$, we have from (16d) and Lemma 1 that the $(k+1)$-th iterate must satisfy

$$\mu_{k+1} \leq [1 - \alpha_k \kappa (1 - \sigma_k)]\, \mu_k \leq \mu_k - \frac{\hat{\alpha}\kappa}{2} \hat{\mu}. \tag{34}$$

If a fast step is taken, we have from the acceptance test in the main algorithm that

$$\mu_{k+1} \leq \rho \mu_k = \mu_k - (1 - \rho)\mu_k \leq \mu_k - \frac{(1-\rho)}{2} \hat{\mu}, \tag{35}$$

The estimates (34) and (35) show that, whatever kind of step is taken, the reduction in μ from iterate k is at least a small constant. Therefore, since $\{\mu_k\}$ is monotone decreasing and K is infinite, we have $\mu_k \downarrow -\infty$. This is a contradiction, since μ_k is bounded below by zero, so the proof is complete. ∎

3.2 Superlinear Local Convergence

By making various assumptions about the functions Φ and g and about the solution set $\mathcal{S}$ (see the next section), we can show that the algorithm converges superlinearly. The sequence of duality measures $\{\mu_k\}$ converges with Q-order at least $1 + \hat{\tau}$, where $\hat{\tau} \in (0, 1)$ is the parameter used to choose the initial step length for the fast step in (18).

We state our main result here. The remainder of the paper lays the groundwork for its proof, which is given at the end.

Theorem 2 *Suppose that Assumptions 1, 2, 3, 4, 5, 6, and 7 are satisfied and that the sequence* $\{(z^k, \lambda^k, y^k)\}$ *is infinite, with a limit point* (z^*, λ^*, y^*) *(in the solution set* $\mathcal{S}$*). Then the algorithm eventually always takes fast steps, and*

(i) the sequence $\{\mu_k\}$ *converges superlinearly to zero with Q-order at least* $1 + \hat{\tau}$*, and*

(ii) the sequence $\{(z^k, \lambda^k, y^k)\}$ *converges superlinearly to* (z^*, λ^*, y^*) *with R-order at least* $1 + \hat{\tau}$*.*

4 Assumptions for Superlinear Convergence

We have already shown in Section 3.1 that Assumptions 1 and 2 are enough to guarantee global convergence of the kind described in Theorem 1. In the remainder of the paper, we focus on case (B) of this theorem, in which the iterate sequence has a limit point in the solution set $\mathcal{S}$. In this section, we state and describe the assumptions that will be used in the proof of Theorem 2.

Assumption 3 is the Slater constraint qualification.

Assumption 3 *There is a vector $\bar{z} \in \mathcal{C}$ such that $g(\bar{z}) < 0$.*

Assumption 4 concerns strict complementarity for at least one member of the solution set.

Assumption 4 *There is a strictly complementary solution (z^*, λ^*, y^*), that is, (z^*, λ^*, y^*) satisfies (3) with $\lambda^* + y^* > 0$.*

The next assumption concerns smoothness of Φ and Dg around the vector z^* defined by Assumption 4.

Assumption 5 *The matrix-valued functions $D\Phi$ and $D^2 g_i$, $i = 1, 2, \ldots, P$ are Lipschitz continuous in a neighborhood of z^*.*

We show in Lemma 3 below that the z^* component of the solution is unique. This fact, together with Assumption 5 and the observation that $D_z f(z, \lambda)$ is linear in λ, ensures that $D_z f(z, \lambda)$ and $Dg(z)$ are Lipschitz continuous in a neighborhood of $\mathcal{S}_Z$.

For the next assumption, we recall the definition of the index sets $\mathcal{B}$ and $\mathcal{N}$ from Section 1. All strictly complementary solutions (z^*, λ^*, y^*) have $\lambda^*_{\mathcal{B}} > 0$, $\lambda^*_{\mathcal{N}} = 0$, $y^*_{\mathcal{B}} = 0$, and $y^*_{\mathcal{N}} > 0$. This assumption concerns invertibility of the projection of $D_z f(z^*, \lambda)$ onto the null space of the active constraints, which are the components $g_i(z)$ for $i \in \mathcal{B}$.

Assumption 6 *Let $\mathcal{S}_Z$ and $\mathcal{B}$ be defined as in Section 1, and z^* be as defined in Assumption 4. Let Λ^* be the set of $\lambda \in \mathbb{R}^P$ such that $(z^*, \lambda) \in \mathcal{S}_Z$. Then for each extreme point λ^e of Λ^*, the two-sided projection of $D_z f(z^*, \lambda^e)$ onto $\ker(Dg^*_{\mathcal{B}})$ is invertible; that is, for any basis Z of $\ker(Dg^*_{\mathcal{B}})$, the matrix $Z^T D_z f(z^*, \lambda^e) Z$ is invertible.*

This assumption looks similar to Assumption 2, but it applies to a different set of points (z, λ) and also refers to a different subspace—that of the *active* constraint Jacobian, not of the entire constraint Jacobian.

Assumption 6 appears to be weaker than the more usual condition, in the context of nonlinear programming, that $D_z f(z^*, \lambda)$ is positive definite on $\ker(Dg^*_{\mathcal{B}})$ for each $\lambda \in \Lambda^*$. It is an easy exercise, however, to show that these two conditions are equivalent, though checking the former is certainly more convenient in that it requires consideration of only finitely many matrices.

Lemma 2 *Suppose that Assumptions 1, 2, 3, 4, and 6 are satisfied. Then the set of multipliers Λ^* defined in Assumption 6 is polyhedral, convex, and compact, hence is equal to the convex hull of its extreme points.*

Proof. Clearly Λ^* is a polyhedral, convex set. Boundedness follows from Gauvin [1] if we can show that the Mangasarian-Fromovitz constraint qualification holds at z^*. Given the Slater point $(\bar{z}, \bar{\lambda})$ from Assumption 3, we have for $i \in \mathcal{B}$ that

$$g_i(\bar{z}) \ \geq \ g_i(z^*) + Dg_i(z^*)(\bar{z} - z^*) \ = \ Dg_i(z^*)(z - z^*),$$

so that $Dg_i(z^*)(\bar{z} - z^*) < 0$ for all $i \in \mathcal{B}$, as required. ∎

We return to our earlier claim that the z^* solution component is uniquely determined.

Lemma 3 *If Assumptions 1, 3, 4, and 6 hold, then*

$$\mathcal{S}_Z \quad = \quad \{z^*\} \times \Lambda^*,$$

where $\Lambda^* \subset \mathbb{R}^P_+$ is the set of multipliers referred to in Assumption 6.

Proof. Convexity of $\mathcal{S}_Z$ follows from Proposition 3.1 of Harker and Pang [2], since the NCP (3) is a equivalent to a monotone variational inequality over a closed convex set. The invertibility condition, Assumption 6, implies that for (z, λ) in $\mathcal{S}_Z$ near (z^*, λ^*), we must have $z = z^*$.

Suppose $\mathcal{S}_Z$ contains (z, λ), where z is remote from z^*. By convexity, we also have

$$(1 - \alpha)(z^*, \lambda^*) + \alpha(z, \lambda) \in \mathcal{S}_Z$$

for all $\alpha \in [0, 1]$. Since $(1 - \alpha)z^* + \alpha z \to z^*$ as $\alpha \downarrow 0$, it follows from local uniqueness that $z = z^*$. ∎

Note that Lemmas 2 and 3 together imply that $\mathcal{S}_Z$ is compact.

Finally, we state the constant rank assumption. See Pang and Ralph [5] for some discussion on this and related conditions.

Assumption 7 *The constant rank constraint qualification (CRCQ) holds for the system $g(z) \leq 0$ at z^*: For some neighborhood U of z^*, the set of matrices $\{Dg_{\mathcal{B}}(z) \,|\, z \in U\}$ has constant column rank.*

Clearly the CRCQ holds if g is affine. It also holds if $(Dg^*_{\mathcal{B}})^T$ has full column rank (that is, if the linear independence constraint qualification holds).

5 Proof of the Superlinear Convergence Result

In this section, we prove the main result, Theorem 2. Most of the effort is spent in estimating the size of fast steps $(\Delta z, \Delta\lambda, \Delta y)$ that are calculated from points $(z, \lambda, y) \in \Omega$ close to the limit point $(\bar{z}, \bar{\lambda}, \bar{y})$. The ultimate result, Corollary 1, shows that the estimate

$$\|(\Delta z, \Delta\lambda, \Delta y)\| \leq C_0 \mu \tag{36}$$

holds for all steps of this type. In Subsection 5.6, this estimate is used together with Lipschitz continuity to complete the proof of Theorem 2.

The task of proving the estimate (36) turns out to be highly technical, so we have organized our argument into subsections and provided considerable motivating discussion. Readers should be able to follow the outline of our argument without delving into the details. The difficulty is due entirely to our wish to use weaker conditions than the usual nondegeneracy conditions. When the latter hold, the condition (36) follows from a simple application of the implicit function theorem.

Most results in this section follow from the same set of assumptions, which we define here to avoid repetition:

> *Standing Assumptions:* These are the assumptions of Theorem 2; namely, Assumptions 1, 2, 3, 4, 5, 6, and 7, together with an assumption that the sequence has a limit point but does not terminate finitely.

Assumption 7 is needed only from Subsection 5.4 onwards, but we include it among the standing assumptions for simplicity.

In Subsection 5.1, we define a partition of the vector $(\Delta z, \Delta\lambda, \Delta y)$ into two components (t, u, v) and (t', u', v'). Subsection 5.2 gives a relatively easy part of the proof: showing that the components $\Delta\lambda_{\mathcal{N}}$ and $\Delta y_{\mathcal{B}}$ are $O(\mu_k)$. Subsections 5.3 and 5.4 show that (t', u', v') and $(u_{\mathcal{B}}, v_{\mathcal{N}})$, respectively, are also $O(\mu)$. All these results, taken together, establish $\|(\Delta\lambda, \Delta y)\| = O(\mu)$. We summarize this result in Subsection 5.5 and deduce that the remaining step component $\|\Delta z\|$ is also $O(\mu)$.

Throughout the section, we assume that the sequence (z^k, λ^k, y^k) has a limit point that we denote by $(\hat{z}, \hat{\lambda}, \hat{y})$. Of course, we know from Theorem 1 that $(\hat{z}, \hat{\lambda}, \hat{y}) \in \mathcal{S}$. When Assumption 4 and the result of Lemma 3 hold, all solutions have the vector z^* as their z component. In this case we have $\hat{z} = z^*$, so we sometimes write the limit point as $(z^*, \hat{\lambda}, y^*)$, where $y^* = -g(z^*)$.

Another quantity that appears repeatedly in the remaining analysis of this section is the restricted neighborhood $\Omega(\delta)$ of the limit point defined by

$$\Omega(\delta) \stackrel{\triangle}{=} \{(z, \lambda, y) \in \Omega \,|\, \|(z, \lambda, y) - (z^*, \hat{\lambda}, y^*)\| \leq \delta\}. \tag{37}$$

5.1 Computation of Fast Steps

Recall that each fast step is obtained by solving (12) with $\tilde{\sigma} = 0$; that is,

$$\begin{bmatrix} D_z f & (Dg)^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta\lambda \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_f \\ r_g \\ -\Lambda Y e \end{bmatrix}. \tag{38}$$

For convenience, we restate the following notational definitions from Section 2:

$$r_f = -f(z,\lambda), \qquad r_g = y + g(z), \qquad \mu = \lambda^T y / P.$$

We are particularly interested in the fast step calculation when the current iterate (z, λ, y) is close to the limit point $(z^*, \hat{\lambda}, y^*)$. To establish bounds on the step $(\Delta z, \Delta\lambda, \Delta y)$ in this situation, we split it into two pieces. The splitting is defined implicitly in terms of the following minimization problem:

$$(z^*, \bar{\pi}) \in \operatorname*{argmin}_{(z^*,\pi)\in \mathcal{S}_Z} \left\| \begin{bmatrix} f(z^*,\pi) - [f(z,\lambda) + D_z f(z,\lambda)(z^* - z) + Dg(z)^T(\pi - \lambda)] \\ g(z^*) - [g(z) + Dg(z)(z^* - z)] \end{bmatrix} \right\|. \tag{39}$$

Existence of the vector $(z^*, \bar{\pi})$ follows from compactness of $\mathcal{S}_Z$. We use $(z^*, \bar{\pi})$ to define the vectors η_f, η_g, ϵ_f, ϵ_g as follows:

$$\begin{aligned} \eta_f &= D_z f(z,\lambda)(z^* - z) + Dg(z)^T(\bar{\pi} - \lambda), && \text{(40a)} \\ \eta_g &= y - Dg(z)(z^* - z) + g(z^*) && \text{(40b)} \\ \epsilon_f &= -f(z,\lambda) - D_z f(z,\lambda)(z^* - z) - Dg(z)^T(\bar{\pi} - \lambda), && \text{(40c)} \\ \epsilon_g &= g(z) - g(z^*) + Dg(z)(z^* - z). && \text{(40d)} \end{aligned}$$

The right-hand side of (38) can now be partitioned as

$$\begin{bmatrix} r_f \\ r_g \\ -\Lambda Y e \end{bmatrix} = \begin{bmatrix} \eta_f \\ \eta_g \\ -Y\Lambda e \end{bmatrix} + \begin{bmatrix} \epsilon_f \\ \epsilon_g \\ 0 \end{bmatrix},$$

and the splitting $(\Delta z, \Delta\lambda, \Delta y) = (t, u, v) + (t', u', v')$ of the right-hand side follows accordingly:

$$\begin{bmatrix} D_z f & (Dg)^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} t \\ u \\ v \end{bmatrix} = \begin{bmatrix} \eta_f \\ \eta_g \\ -\Lambda Y e \end{bmatrix}, \tag{41}$$

$$\begin{bmatrix} D_z f & (Dg)^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} t' \\ u' \\ v' \end{bmatrix} = \begin{bmatrix} \epsilon_f \\ \epsilon_g \\ 0 \end{bmatrix}. \tag{42}$$

Because of Assumption 2, the systems (38), (41), and (42) all have unique solutions.

5.2 Bounds for $\Delta\lambda_{\mathcal{N}}$ and $\Delta y_{\mathcal{B}}$

It is relatively easy to obtain size estimates for about $\Delta\lambda_{\mathcal{N}}$ and $\Delta y_{\mathcal{B}}$, which together make up half the components of $(\Delta\lambda, \Delta y)$. We start by deriving some upper and lower bounds on the components of λ and y for (z, λ, y) in a neighborhood of the form (37), which will prove useful throughout the remainder of this section.

Lemma 4 *Suppose that the standing assumptions hold. Then there is a constant C_4 such that the following bounds hold for all $(z, \lambda, y) \in \Omega(1)$:*

$$\lambda_i \le C_4\mu \;\; (i \in \mathcal{N}), \qquad y_i \le C_4\mu \;\; (i \in \mathcal{B}), \tag{43a}$$

$$\lambda_i \ge \gamma_{\min}/C_4 \;\; (i \in \mathcal{B}), \qquad y_i \ge \gamma_{\min}/C_4 \;\; (i \in \mathcal{N}), \tag{43b}$$

$$y_i \ge \gamma_{\min}\mu/C_4 \;\; (i \in \mathcal{B}), \qquad \lambda_i \ge \gamma_{\min}\mu/C_4 \;\; (i \in \mathcal{N}). \tag{43c}$$

Proof. Let (z^*, λ^*, y^*) denote the strictly complementary solution from Assumption 4. By monotonicity of the mapping (5), (10), and the fact that $g(z^*) = -y^*$, we have

$$0 \le \left[\begin{array}{c} f(z,\lambda) - f(z^*,\lambda^*) \\ -g(z) + g(z^*) \end{array}\right]^T \left[\begin{array}{c} z - z^* \\ \lambda - \lambda^* \end{array}\right] = \left[\begin{array}{c} -r_f \\ y - r_g - y^* \end{array}\right]^T \left[\begin{array}{c} z - z^* \\ \lambda - \lambda^* \end{array}\right].$$

By rearranging this expression, we have from $(\lambda^*)^T y^* = 0$, (20b), and (21) that

$$\begin{aligned} (\lambda^*)^T y + (y^*)^T \lambda &\le \lambda^T y + \|r_f\|\|z - z^*\| + \|r_g\|\|\lambda - \lambda^*\| \\ &\le P\mu + \beta_{\max}\mu \left(\|z\| + \|z^*\| + \|\lambda\| + \|\lambda^*\|\right). \end{aligned}$$

Since $(z, \lambda, y) \in \Omega(1)$, we have

$$\|(z,\lambda)\| \le \|(z^*, \hat{\lambda})\| + \|(z^* - z, \hat{\lambda} - \lambda)\| \le \|(z^*, \hat{\lambda})\| + 1, \tag{44}$$

so we can bound the term in parentheses by a constant, giving

$$(\lambda^*)^T y + (y^*)^T \lambda \le \bar{C}_4\mu,$$

for some positive constant $\bar{C}_4$. Since $\lambda^*_{\mathcal{N}} = 0$ and $y^*_{\mathcal{B}} = 0$, this inequality implies that

$$\sum_{i\in\mathcal{B}} \lambda_i^* y_i + \sum_{i\in\mathcal{N}} y_i^* \lambda_i \le \bar{C}_4\mu.$$

Since $(\lambda^*_{\mathcal{B}}, y^*_{\mathcal{N}}) > 0$ and $(\lambda, y) > 0$, each term in the summations is positive, so we have

$$\lambda_i \le \frac{1}{y_i^*}\bar{C}_4\mu, \quad i \in \mathcal{N}; \qquad y_i \le \frac{1}{\lambda_i^*}\bar{C}_4\mu, \quad i \in \mathcal{B}.$$

From these bounds, we can define C_4 is an obvious way to satisfy (43a).

For any $i \in \mathcal{B}$, we have from (22) and (43a) that

$$\lambda_i \ge \frac{\gamma_{\min}\mu}{y_i} \ge \frac{\gamma_{\min}\mu}{C_4\mu} = \frac{\gamma_{\min}}{C_4},$$

giving the first part of (43b). The second part is proved similarly.

For $i \in \mathcal{B}$, we have from (22) and our choice of $(z, \lambda, y) \in \Omega(1)$ that

$$y_i \geq \frac{\gamma_{\min}\mu}{\lambda_i} \geq \frac{\gamma_{\min}\mu}{\hat{\lambda}_i + 1}.$$

A similar lower bound can be proved for λ_i, $i \in \mathcal{N}$. Hence (43c) holds, for a suitable redefinition of C_4. ∎

Lemma 5 *Suppose that the standing assumptions are satisfied. Then there are constants $\delta_1 \in (0,1]$ and $C_9 > 0$ such that for all $(z, \lambda, y) \in \Omega(\delta_1)$, the solution of the linear system*

$$\begin{bmatrix} D_z f & (Dg)^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \widehat{\Delta z} \\ \widehat{\Delta \lambda} \\ \widehat{\Delta y} \end{bmatrix} = \begin{bmatrix} r_f \\ r_g \\ 0 \end{bmatrix} \tag{45}$$

satisfies

$$\|\widehat{\Delta z}\| \leq C_9 \left(\mu + \|\widehat{\Delta \lambda}_{\mathcal{B}}\| \right).$$

Proof. Because $\delta_1 \leq 1$, the estimates (43) apply for points $(z, \lambda, y) \in \Omega(\delta_1)$. Note too that these points also satisfy $\mu = O(\delta_1)$, since

$$P\mu = \lambda^T y = \lambda_{\mathcal{B}}^T y_{\mathcal{B}} + \lambda_{\mathcal{N}}^T y_{\mathcal{N}} \leq (\|\hat{\lambda}_{\mathcal{B}}\| + \delta_1)\delta_1 + \delta_1(\|y_{\mathcal{N}}\| + \delta_1) = O(\delta_1).$$

By eliminating $\widehat{\Delta y}$ and $\widehat{\Delta \lambda}_{\mathcal{N}}$ from the system (45), we obtain

$$\begin{bmatrix} (D_z f) + (Dg_{\mathcal{N}})^T \Lambda_{\mathcal{N}} (Y_{\mathcal{N}})^{-1} Dg_{\mathcal{N}} & (Dg_{\mathcal{B}})^T \\ -Dg_{\mathcal{B}} & (\Lambda_{\mathcal{B}})^{-1} Y_{\mathcal{B}} \end{bmatrix} \begin{bmatrix} \widehat{\Delta z} \\ \widehat{\Delta \lambda}_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} r_f - (Dg_{\mathcal{N}})^T \Lambda_{\mathcal{N}} (Y_{\mathcal{N}})^{-1} (r_g)_{\mathcal{N}} \\ (r_g)_{\mathcal{B}} \end{bmatrix}. \tag{46}$$

From Lemma 4, we have $\|\Lambda_{\mathcal{N}}(Y_{\mathcal{N}})^{-1}\| = O(\mu)$ and $\|\Lambda_{\mathcal{B}}^{-1} Y_{\mathcal{B}}\| = O(\mu)$. Because of Lipschitz continuity (Assumption 5) and $(z, \lambda, y) \in \Omega(\delta_1)$, we have

$$\begin{aligned} Dg(z) - Dg(z^*) &= O(\|z - z^*\|) = O(\delta_1) \\ D_z f(z, \lambda) - D_z f(z^*, \hat{\lambda}) &= O(\|z - z^*\|) + O(\|\lambda - \hat{\lambda}\|) = O(\delta_1). \end{aligned}$$

By perturbing the coefficient matrix in (46) and substituting these estimates, along with $\mu = O(\delta_1)$, we obtain

$$\begin{aligned} &\begin{bmatrix} D_z f(z^*, \hat{\lambda}) & Dg_{\mathcal{B}}(z^*)^T \\ -Dg_{\mathcal{B}}(z^*) & 0 \end{bmatrix} \begin{bmatrix} \widehat{\Delta z} \\ \widehat{\Delta \lambda}_{\mathcal{B}} \end{bmatrix} \\ &= \begin{bmatrix} r_f - (Dg_{\mathcal{N}})^T \Lambda_{\mathcal{N}} (Y_{\mathcal{N}})^{-1} (r_g)_{\mathcal{N}} \\ (r_g)_{\mathcal{B}} \end{bmatrix} + O(\delta_1) \begin{bmatrix} \widehat{\Delta z} \\ \widehat{\Delta \lambda}_{\mathcal{B}} \end{bmatrix}. \end{aligned} \tag{47}$$

By partitioning $\widehat{\Delta z}$ into its components in $\ker Dg_{\mathcal{B}}(z^*)$ and $\operatorname{ran} Dg_{\mathcal{B}}(z^*)^T$, we have from Assumption 6 that $\widehat{\Delta z}$ is bounded in norm by the right-hand side of (47). Hence, since $\|r_f\|$ and $\|r_g\|$ are both $O(\mu)$, and $Dg_{\mathcal{N}}$ is bounded on bounded sets, we can write

$$\|\widehat{\Delta z}\| \leq \bar{C}_9 \left(\mu + \delta_1 (\|\widehat{\Delta z}\| + \|\widehat{\Delta\lambda}_{\mathcal{B}}\|) \right),$$

for some constant $\bar{C}_9$. By choosing δ_1 small enough that

$$\bar{C}_9 \delta_1 \leq \tfrac{1}{2},$$

we can combine terms in $\|\widehat{\Delta z}\|$ on the left-hand side and divide to obtain

$$\|\widehat{\Delta z}\| \leq 2\bar{C}_9 \left(\mu + \delta_1 \|\widehat{\Delta\lambda}_{\mathcal{B}}\| \right) \leq 2\bar{C}_9 \mu + \|\widehat{\Delta\lambda}_{\mathcal{B}}\|,$$

proving the result. ∎

In subsequent results, we often will refer to the positive definite diagonal matrix D^k defined by

$$D = \Lambda^{-1/2} Y^{1/2}. \tag{48}$$

We can obtain bounds on $\|D\|$ and $\|D^{-1}\|$ for points $(z, \lambda, y) \in \Omega(1)$ by applying Lemma 4. For $\|D^{-1}\|$, we have

$$\|D^{-1}\| = \max_{i=1,\ldots,n} \frac{\lambda_i^{1/2}}{y_i^{1/2}} \leq \frac{(\hat{\lambda}_i + 1)^{1/2}}{(\gamma_{\min} \min(\mu, 1)/C_4)^{1/2}} \leq C_7 \mu^{-1/2}, \tag{49}$$

for some constant C_7. Similar logic shows that

$$\|D\| \leq C_7 \mu^{-1/2}, \tag{50}$$

after a possible redefinition of C_7.

The next result is a bound on the scaled vectors $D\Delta\lambda$ and $D^{-1}\Delta y$.

Lemma 6 *Suppose that the standing assumptions hold. Then for the constant δ_1 defined in Lemma 5, there is a constant $C_3 > 0$ such that the solution $(\Delta z, \Delta\lambda, \Delta y)$ of (38) satisfies*

$$\|D\Delta\lambda\| \leq C_3 \mu^{1/2}, \qquad \|D^{-1}\Delta y\| \leq C_3 \mu^{1/2}, \tag{51}$$

for all $(z, \lambda, y) \in \Omega(\delta_1)$.

Proof. We break the solution into two pieces and prove that the required bounds hold for each part. We write

$$(\Delta z, \Delta\lambda, \Delta y) = (\overline{\Delta z}, \overline{\Delta\lambda}, \overline{\Delta y}) + (\widehat{\Delta z}, \widehat{\Delta\lambda}, \widehat{\Delta y}),$$

where

$$\begin{bmatrix} D_z f & Dg^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \overline{\Delta z} \\ \overline{\Delta \lambda} \\ \overline{\Delta y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -Y\Lambda e \end{bmatrix}, \qquad \begin{bmatrix} D_z f & Dg^T & 0 \\ -Dg & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \widehat{\Delta z} \\ \widehat{\Delta \lambda} \\ \widehat{\Delta y} \end{bmatrix} = \begin{bmatrix} r_f \\ r_g \\ 0 \end{bmatrix}. \tag{52}$$

For the first component, we multiply the last block row by the diagonal matrix $(Y\Lambda)^{-1/2}$ to obtain

$$D\overline{\Delta\lambda} + D^{-1}\overline{\Delta y} = -(Y\Lambda)^{1/2}e. \tag{53}$$

From (52), we also have

$$\overline{\Delta\lambda}^T\overline{\Delta y} = -\overline{\Delta\lambda}^T(Dg)\overline{\Delta z} = \overline{\Delta z}^T(D_z f)\overline{\Delta z} \geq 0,$$

so by taking inner products in (53), we obtain

$$\|D\overline{\Delta\lambda}\|^2 + \|D^{-1}\overline{\Delta y}\|^2 \leq \|D\overline{\Delta\lambda}\|^2 + \overline{\Delta z}^T(D_z f)\overline{\Delta z} + \|D^{-1}\overline{\Delta y}\|^2 = \|(Y\Lambda)^{1/2}e\|^2 = P\mu.$$

Hence, we have

$$\|D\overline{\Delta\lambda}\| \leq P^{1/2}\mu^{1/2}, \qquad \|D^{-1}\overline{\Delta y}\| \leq P^{1/2}\mu^{1/2}. \tag{54}$$

For the second component of the solution, we obtain from the last block row in (52) that

$$D\widehat{\Delta\lambda} = -D^{-1}\widehat{\Delta y} \;\; \Rightarrow \;\; \|D\widehat{\Delta\lambda}\| = \|D^{-1}\widehat{\Delta y}\|, \tag{55}$$

and so we seek a bound for $\|D\widehat{\Delta\lambda}\|$. Using (52) again, we obtain

$$-Dg\widehat{\Delta z} - \widehat{\Delta y} = r_g \;\; \Rightarrow \;\; -Dg\widehat{\Delta z} + D^2\widehat{\Delta\lambda} = r_g.$$

By taking inner products with $\widehat{\Delta\lambda}$, we obtain

$$\|D\widehat{\Delta\lambda}\|^2 = \widehat{\Delta\lambda}^T r_g + \widehat{\Delta\lambda}^T(Dg)\widehat{\Delta z}.$$

From the first block row in (52), we have by positive semidefiniteness of $D_z f(z,\lambda)$ that

$$\widehat{\Delta\lambda}^T(Dg)\widehat{\Delta z} = (r_f - (D_z f)\widehat{\Delta z})^T\widehat{\Delta z} \leq r_f^T\widehat{\Delta z}.$$

By combining the last two expressions, we obtain

$$\|D\widehat{\Delta\lambda}\|^2 \leq \widehat{\Delta\lambda}^T r_g + r_f^T\widehat{\Delta z} \leq \|\widehat{\Delta\lambda}\|\|r_g\| + \|r_f\|\|\widehat{\Delta z}\|. \tag{56}$$

Because of (20b) and Lemma 5, we have

$$\|r_f\| \leq \beta_{\max}\mu, \qquad \|r_g\| \leq \beta_{\max}\mu, \qquad \|\widehat{\Delta z}\| \leq C_9(\mu + \|\widehat{\Delta\lambda}\|).$$

It follows from (49) that

$$\|\widehat{\Delta\lambda}\| \le \|D^{-1}\|\|D\widehat{\Delta\lambda}\| \le C_7\mu^{-1/2}\|D\widehat{\Delta\lambda}\|.$$

By substituting all these estimates into the right-hand side of (56), we find that there is a constant C_{10} such that

$$\|D\widehat{\Delta\lambda}\|^2 \le C_{10}\left(\mu^{1/2}\|D\widehat{\Delta\lambda}\| + \mu^2\right).$$

It follows immediately from this expression and (55) that

$$\|D\widehat{\Delta\lambda}\| \le \bar{C}_3\mu^{1/2}, \qquad \|D^{-1}\widehat{\Delta y}\| \le \bar{C}_3\mu^{1/2},$$

for some constant $\bar{C}_3$. The result of the lemma is obtained by combining this estimate with (54). ■

Bounds on half the components of $(\Delta\lambda, \Delta y)$ follow easily.

Theorem 3 *Suppose that the standing assumptions hold. Then for the constant δ_1 defined in Lemma 5, there is a positive constant C_5 such that the solution $(\Delta z, \Delta\lambda, \Delta y)$ of (38) satisfies*

$$\|\Delta\lambda_{\mathcal{N}}\| \le C_5\mu, \qquad \|\Delta y_{\mathcal{B}}\| \le C_5\mu, \tag{57}$$

for all $(z, \lambda, y) \in \Omega(\delta_1)$.

Proof. From the definition (48) and the bounds (51), we have that

$$\left|\left(\frac{y_i}{\lambda_i}\right)^{1/2}\Delta\lambda_i\right| \le \|D\Delta\lambda\| \le C_3\mu^{1/2}.$$

Hence from (43a) and (43b), we have for $i \in \mathcal{N}$ that

$$|\Delta\lambda_i| \le \left(\frac{\lambda_i}{y_i}\right)^{1/2} C_3\mu^{1/2} \le \frac{C_4\mu^{1/2}}{\gamma_{\min}^{1/2}} C_3\mu^{1/2},$$

which proves that $\|\Delta\lambda_{\mathcal{N}}\| \le C_5\mu_k$ for an obvious definition of C_5. The bound on $\|\Delta y_{\mathcal{B}}\|$ is derived in the same way. ■

5.3 A Bound for (t', u', v')

In this subsection we find bounds for the components (t', u', v') defined by (42). The difficult part of the analysis appears in the following two lemmas, in which we estimate the size of (ϵ_f, ϵ_g) in (40c),(40d).

Under our standing assumptions, we can define the following set:

$$\mathcal{S}_Z^\infty \;=\; \{(z,\lambda) \in \mathcal{S}_Z \,|\, \lambda_i \ge \gamma_{\min}/C_4,\ i \in \mathcal{B};\ \ g_i(z) \le -\gamma_{\min}/C_4,\ i \in \mathcal{N}\}, \tag{58}$$

where C_4 is defined in Lemma 4. Because of (43b), all limit points of the sequence $\{(z^k, \lambda^k)\}$ lie in $\mathcal{S}_Z^\infty$; in particular, $(z^*, \hat{\lambda}) \in \mathcal{S}_Z^\infty$. Obviously, $(z, \lambda, -g(z))$ is a strictly complementary solution of (3) whenever $(z, \lambda) \in \mathcal{S}_Z^\infty$.

Our first result, like the results in the preceding subsections, considers points $(z, \lambda, y) \in \Omega$ near the solution set $\mathcal{S}$ and shows that $\mathrm{dist}_{\mathcal{S}_Z}(z, \lambda)$ can be bounded in terms of the amount by which $(z, \lambda, -g(z))$ violates feasibility and complementarity.

Lemma 7 *Suppose that the standing assumptions hold. Then there exist constants $\tilde{L}$ and $\delta_2 \in (0, \delta_1]$ such that the following bound holds for all points $(z, \lambda) \in \mathbb{R}^N \times \mathbb{R}^P_+$ with $\mathrm{dist}_{\mathcal{S}_Z^\infty}(z, \lambda) \le \delta_2$:*

$$\mathrm{dist}_{\mathcal{S}_Z}(z, \lambda) \;\le\; \tilde{L} \| \left(f(z, \lambda), g(z)_+, \lambda^T g(z) \right) \|. \tag{59}$$

Proof. By Lemmas 2 and 3, we know that $\mathcal{S}_Z$ is compact. Since $\mathcal{S}_Z^\infty \subset \mathcal{S}_Z$ and $\mathcal{S}_Z^\infty$ is closed, $\mathcal{S}_Z^\infty$ too is compact.

We prove the result by contradiction. If the claim is false, we can choose a sequence $\{(\xi_z^k, \xi_\lambda^k)\} \subset \mathbb{R}^N \times \mathbb{R}^P_+$ with the properties

$$\mathrm{dist}_{\mathcal{S}_Z^\infty}(\xi_z^k, \xi_\lambda^k) \downarrow 0, \tag{60}$$

and

$$\|(\xi_z^k, \xi_\lambda^k) - (z^*, \bar{\lambda}^k)\| \;\ge\; k \| \left(f(\xi_z^k, \xi_\lambda^k), g(\xi_z^k)_+, (\xi_\lambda^k)^T g(\xi_z^k) \right) \|, \tag{61}$$

where $(z^*, \bar{\lambda}^k)$ is the nearest point in $\mathcal{S}_Z$ to ξ^k for each k. (Note that $(z^*, \bar{\lambda}^k)$ exists, by compactness of $\mathcal{S}_Z$, and that the z^* component is uniquely defined.) By compactness of $\mathcal{S}_Z^\infty$ and (60), we can take subsequences if necessary and assume that both $\{(\xi_z^k, \xi_\lambda^k)\}$ and $\{(z^*, \bar{\lambda}^k)\}$ converge to $(z^*, \bar{\lambda}) \in \mathcal{S}_Z^\infty$. By defining $\tau_k = \|(\xi_z^k, \xi_\lambda^k) - (z^*, \bar{\lambda}^k)\|$ and taking a further subsequence, we can assume that there is a vector $(d_z, d_\lambda) \in (\mathbb{R}^N \times \mathbb{R}^P_+) \setminus \{0\}$ such that

$$\frac{(\xi_z^k, \xi_\lambda^k) - (z^*, \bar{\lambda}^k)}{\tau_k} \;\to\; (d_z, d_\lambda).$$

(In fact, (d_z, d_λ) is a unit vector.) Since $\bar{\lambda}_{\mathcal{N}}^k = 0$ and $\bar{\lambda}_{\mathcal{B}}^k > 0$ for all k sufficiently large, the solution $(z^*, \bar{\lambda}^k)$ is strictly complementary for all k sufficiently large.

The following analysis is devoted to showing that $(d_z, d_\lambda) = 0$, a contradiction that proves the result. First, we show that (d_z, d_λ) is in the normal cone to $\mathcal{S}_Z$ at $(z^*, \bar{\lambda})$, namely,

$$\begin{pmatrix} d_z \\ d_\lambda \end{pmatrix}^T \begin{pmatrix} z^* - z^* \\ \lambda - \bar{\lambda} \end{pmatrix} \;\le\; 0 \quad \text{for all} \quad (z^*, \lambda) \in \mathcal{S}_Z. \tag{62}$$

Second, we show that (d_z, d_λ) is in the tangent cone to $\mathcal{S}_Z$ at $(z^*, \bar{\lambda})$, indeed that

$$(z^*, \bar{\lambda}) + \tau (d_z, d_\lambda) \in \mathcal{S}_Z \quad \text{for small } \tau > 0. \tag{63}$$

Together, these two results imply that $\|(d_z, d_\lambda)\| = 0$, as required.

To show (62), we note that, since $(z^*, \bar\lambda^k)$ is the projection of (ξ_z^k, ξ_λ^k) onto $\mathcal{S}_Z$, we have

$$\left(\begin{array}{c} \xi_z^k - z^* \\ \xi_\lambda^k - \bar\lambda^k \end{array} \right)^T \left(\begin{array}{c} z^* - z^* \\ \lambda - \bar\lambda^k \end{array} \right) \le 0, \qquad \text{for all } (z^*, \lambda) \in \mathcal{S}_Z.$$

We obtain (62) by dividing this expression by τ_k and taking limits.

The proof of (63) is longer. By the smoothness properties of f, and the fact that $f(z^*, \bar\lambda^k) = 0$, we have

$$\frac{f(\xi_z^k, \xi_\lambda^k)}{\tau_k} = \frac{f(\xi_z^k, \xi_\lambda^k) - f(z^*, \bar\lambda^k)}{\tau_k} \to D_z f(z^*, \bar\lambda) d_z + Dg(z^*)^T d_\lambda. \tag{64}$$

Taking $i \in \mathcal{B}$, we have $g_i(z^*) = 0$ and so

$$\frac{g_i(\xi_z^k)_+}{\tau_k} = \left[\frac{g_i(\xi_z^k) - g_i(z^*)}{\tau_k} \right]_+ \to [Dg_i(z^*) d_z]_+, \quad \text{for all } i \in \mathcal{B}. \tag{65}$$

For the nonbasic components, we have

$$g_i(z^*) < 0 \;\; \Rightarrow \;\; g_i(\xi_z^k)_+ = 0, \quad \text{for all } i \in \mathcal{N}, \tag{66}$$

and all k sufficiently large. Also, we have

$$\frac{(\xi_\lambda^k)^T g(\xi_z^k)}{\tau_k} = \frac{(\xi_\lambda^k)^T g(\xi_z^k) - (\bar\lambda^k)^T g(z^*)}{\tau_k} \to g(z^*)^T d_\lambda + \bar\lambda^T Dg(z^*) d_z. \tag{67}$$

By combining (64), (65), (66), and (67) and the property (61), we obtain

$$\begin{aligned} 0 &= \lim_k \frac{\|(f(\xi_z^k, \xi_\lambda^k), g(\xi_z^k)_+, (\xi_\lambda^k)^T g(\xi_z^k))\|}{\tau_k} \\ &= \|(D_z f(z^*, \bar\lambda) d_z + Dg(z^*)^T d_\lambda, [Dg_{\mathcal{B}}(z^*) d_z]_+, g(z^*)^T d_\lambda + \bar\lambda^T Dg(z^*) d_z)\|. \end{aligned}$$

It follows immediately that

$$D_z f(z^*, \bar\lambda) d_z + Dg(z^*) d_\lambda = 0, \tag{68a}$$
$$Dg_{\mathcal{B}}(z^*) d_z \le 0, \tag{68b}$$
$$g(z^*)^T d_\lambda + \bar\lambda^T Dg(z^*) d_z = 0. \tag{68c}$$

Since $g_{\mathcal{B}}(z^*) = 0$ and $\bar\lambda_{\mathcal{N}} = 0$, we can rewrite (68c) as

$$\sum_{i \in \mathcal{N}} g_i(z^*) (d_\lambda)_i + \sum_{i \in \mathcal{B}} \bar\lambda_i Dg_i(z^*) d_z = 0. \tag{69}$$

Since $(\bar\lambda^k)_{\mathcal{N}} = 0$ and $\xi_\lambda^k \ge 0$ for all k, we have $(d_\lambda)_{\mathcal{N}} \ge 0$. Therefore all product terms in both summations in (69) are nonpositive, so we can use $g_{\mathcal{N}}(z^*) < 0$ and $\bar\lambda_{\mathcal{B}} > 0$ to deduce that

$$(d_\lambda)_{\mathcal{N}} = 0, \qquad Dg_{\mathcal{B}}(z^*) d_z = 0. \tag{70}$$

By multiplying (68a) by d_z^T and using (70), we obtain

$$d_z^T Df_z(z^*, \bar{\lambda}) d_z = -d_z^T Dg(z^*)^T d_\lambda = 0. \tag{71}$$

Assumption 6, together with $d_z \in \ker Dg_{\mathcal{B}}(z^*)$ (from (70)) and (71), implies that $d_z = 0$. Hence, (68a) reduces to

$$Dg(z^*)^T d_\lambda = 0. \tag{72}$$

Finally, we are in a position to verify that (63) is satisfied. Because of $d_z = 0$ and $(d_\lambda)_{\mathcal{N}} = 0$, we have

$$\begin{array}{rcll} g_{\mathcal{B}}(z^* + \tau d_z) = g_{\mathcal{B}}(z^*) & = & 0, & \\ g_{\mathcal{N}}(z^* + \tau d_z) = g_{\mathcal{N}}(z^*) & < & 0, & \\ \bar{\lambda}_{\mathcal{N}} + \tau (d_\lambda)_{\mathcal{N}} & = & 0, & \\ \bar{\lambda}_{\mathcal{B}} + \tau (d_\lambda)_{\mathcal{B}} & > & 0, & \text{for } \tau > 0 \text{ sufficiently small.} \end{array}$$

From (72) and the fact that f is linear in λ, we have

$$f(z^* + \tau d_z, \bar{\lambda} + \tau d_\lambda) = f(z^*, \bar{\lambda} + \tau d_\lambda) = f(z^*, \bar{\lambda}) + \tau Dg(z^*)^T d_\lambda = 0.$$

Together, these formulae indicate that (63) holds, so we are done. ∎

Lemma 8 *Suppose that our standing assumptions are satisfied. Then there exist constants $\hat{L} > 0$, $\check{L} > 0$, and $\delta_3 \in (0, \delta_2]$ (where δ_2 is defined in Lemma 7) such that for each $(z, \lambda, y) \in \Omega(\delta_3)$ we have*

$$\left\| \begin{bmatrix} f(z, \lambda) + D_z f(z, \lambda)(z^* - z) + Dg(z)^T (\bar{\pi} - \lambda) \\ g(z) - g(z^*) + Dg(z)(z^* - z) \end{bmatrix} \right\| \le \hat{L} \mu^2, \tag{73}$$

and

$$\left\| \begin{bmatrix} D_z f(z, \lambda)(z^* - z) + Dg(z)^T (\bar{\pi} - \lambda) \\ y - Dg(z)(z^* - z) + g(z^*) \end{bmatrix} \right\| \le \check{L} \mu. \tag{74}$$

where, as in (39), $\bar{\pi}$ is chosen from the optimal Lagrange multiplier set Λ^* to minimize the left-hand side of (73).

Proof. We start by proving (73). As in (39), we denote the minimand of the left-hand side in (73) by $(z^*, \bar{\pi})$, whose existence follows from compactness of $\mathcal{S}_Z$. We show first that $\|(z^*, \bar{\pi}) - (z, \lambda)\| = O(\mu)$ and then prove the result by a Lipschitz continuity argument.

By considering $(z, \lambda, y) \in \Omega(\delta_2)$, we have from (10), (22), and the fact that $y \ge 0$ that $\|f(z, \lambda)\| \le \beta_{\max} \mu$ and

$$\|g(z)_+\| = \|[r_g - y]_+\| \le \|r_g\| \le \beta_{\max} \mu. \tag{75}$$

Since for all $(z,\lambda,y) \in \Omega(\delta_2)$, we have $\|(z,\lambda)\| \le \hat{C}_1$ for some constant $\hat{C}_1$, it follows that

$$(z,\lambda,y) \in \Omega(\delta_2) \quad \Rightarrow \quad |\lambda^T g(z)| = |\lambda^T(r_g - y)| \le \|\lambda\|\|r_g\| + |\lambda^T y| \le (\hat{C}_1\beta_{\max} + P)\mu.$$

We have shown that the right-hand side in (59) is $O(\mu)$ and therefore, by the result of Lemma 7, we have

$$\|(z,\lambda) - (z^*,\bar{\pi})\| \le \bar{C}_1\mu \tag{76}$$

for some constant $\bar{C}_1$ and all $(z,\lambda,y) \in \Omega(\delta_2)$.

By the Lipschitz continuity assumption (see Assumption 5 and the comments that follow) we can choose $\delta_3 \in (0,\delta_2]$ such that $D_z f(z,\lambda)$ and $Dg(z)$ are Lipschitz continuous for $\mathrm{dist}_{\mathcal{S}_Z^\infty}(z,\lambda) \le \delta_3$. Therefore, the matrix function

$$\left[\begin{array}{cc} D_z f(z,\lambda) & Dg(z)^T \\ Dg(z) & 0 \end{array}\right]$$

is also Lipschitz continuous as a function of (z,λ) in this neighborhood. Since $(z^*,\bar{\pi}) \in \mathcal{S}_Z$, we have $f(z^*,\bar{\pi}) = 0$ and we have that

$$\left\| \left[\begin{array}{c} f(z,\lambda) + D_z f(z,\lambda)(z^*-z) + Dg(z)^T(\bar{\pi}-\lambda) \\ g(z) - g(z^*) + Dg(z)(z^*-z) \end{array}\right] \right\| \le \bar{L}\|(z^*,\bar{\pi}) - (z,\lambda)\|^2, \tag{77}$$

for some constant $\bar{L} > 0$ and all (z,λ) with $\mathrm{dist}_{\mathcal{S}_Z^\infty}(z,\lambda) \le \delta_3$. We obtain the result (73) by combining (76) with (77) and defining $\hat{L} = \bar{L}\bar{C}_1^2$.

For (74), we have that

$$\begin{aligned} &\left[\begin{array}{c} -D_z f(z,\lambda)(z^*-z) - Dg(z)^T(\bar{\pi}-\lambda) \\ y - Dg(z)(z^*-z) + g(z^*) \end{array}\right] \\ &\quad = \left[\begin{array}{c} r_f \\ r_g \end{array}\right] - \left[\begin{array}{c} f(z,\lambda) + D_z f(z,\lambda)(z^*-z) + Dg(z)^T(\bar{\pi}-\lambda) \\ g(z) - g(z^*) + Dg(z)(z^*-z) \end{array}\right], \end{aligned}$$

and therefore

$$\left\| \left[\begin{array}{c} D_z f(z,\lambda)(z^*-z) + Dg(z)^T(\bar{\pi}-\lambda) \\ y - Dg(z)(z^*-z) + g(z^*) \end{array}\right] \right\| \le \left\| \left[\begin{array}{c} r_f \\ r_g \end{array}\right] \right\| + \hat{L}\mu^2,$$

where the last term is a consequence of (73). Since $\|(r_f,r_g)\| = O(\mu)$ by (22), we have the result. ∎

We use Lemma 8 to estimate the quantities η_f, η_g, ϵ_f, and ϵ_g defined by (40). For $(z,\lambda,y) \in \Omega(\delta_3)$, we have from (39), (40c), (40d), and (73) that

$$\|\epsilon_f\| \le \hat{L}\mu^2, \qquad \|\epsilon_g\| \le \hat{L}\mu^2. \tag{78}$$

Similarly, we have from (39), (40a), (40b), and (74) that

$$\|\eta_f\| \le \breve{L}\mu, \qquad \|\eta_g\| \le \breve{L}\mu. \tag{79}$$

Lemma 9 *Suppose that the standing assumptions hold and δ_3 is given by Lemma 8. Then there is a constant C_{11} such that the solution (t', u', v') of (42) satisfies $\|(t', u', v')\| \leq C_{11}\mu$ for all $(z, \lambda, y) \in \Omega(\delta_3)$.*

Proof. Note that (t', u', v') satisfies the equations (45) if we replace $(r_f, r_g, 0)$ on the right-hand side by $(\epsilon_f, \epsilon_g, 0)$. The main difference between the two systems is the size of the right-hand sides: $O(\mu)$ in (45), $O(\mu^2)$ here from (78). By using the same technique of proof as in Lemma 5, we can show that

$$\|t'\| \leq C_9(\mu^2 + \|u'_{\mathcal{B}}\|), \tag{80}$$

for some constant C_9. This estimate, together with the techniques of the second part of the proof of Lemma 6, implies that

$$\|Du'\| \leq C_3\mu^{3/2}, \qquad \|D^{-1}v'\| \leq C_3\mu^{3/2}, \tag{81}$$

where D is defined as in (48). The estimates $\|D\| \leq C_7\mu^{-1/2}$ and $\|D^{-1}\| \leq C_7\mu^{-1/2}$ obtained from (49) and (50) can now be combined with (80) and (81) to complete the proof. ∎

5.4 Bounds for $u_{\mathcal{B}}$ and $v_{\mathcal{N}}$

In this subsection we address the most difficult part of the proof: showing that the components $u_{\mathcal{B}}$ and $v_{\mathcal{N}}$ from (41) are $O(\mu_k)$. As in the case of affine f, the key to our result is to show that $(u_{\mathcal{B}}, v_{\mathcal{N}})$ is the solution of a certain quadratic program (Theorem 5 below). Unlike the affine case, however, the coefficient matrix in this quadratic program does not remain constant. Instead, this matrix satisfies a constant column rank condition (Theorem 4), and this condition is enough to yield the desired bound (Lemmas 10 and 11).

We start by proving a novel variant of a lemma from Monteiro and Wright [4, Lemma 2.2]. The definition of constant column rank appears at the end of Section 1.

Lemma 10 *If $\mathcal{H}$ is a bounded set in $\mathbb{R}^{p\times q}$ with constant column rank and $\|\cdot\|$ is any norm on $\mathbb{R}^q$, there exists a nonnegative constant $L = L(\mathcal{H})$ with the property that for each $H \in \mathcal{H}$ and $h \in \operatorname{ran} H$, there is a solution $w \in \mathbb{R}^q$ of the equation $Hw = h$ for which*

$$\|w\| \leq L\|h\|.$$

Proof. The case of $h = 0$ is trivial, so we assume throughout the proof that $h \neq 0$.

To obtain a contradiction, assume there exist $\{H^k\} \subset \mathcal{H}$ and $\{h^k\} \subset \mathbb{R}^q \setminus \{0\}$ such that, for each k, $h^k \in \operatorname{ran} H^k$ and

$$\operatorname{dist}_{(H^k)^{-1}h^k}(0) \; > \; k\|h^k\|. \tag{82}$$

We may assume without loss of generality (by taking subsequences and dividing by $\|h^k\|$ if necessary) that $H^k \to H \in \mathbb{R}^{p\times q}$ and $h^k \to h \in \mathbb{R}^q \setminus \{0\}$.

Let $\mathcal{J}$ be a maximal set of column indices of H such that $H_{\cdot\mathcal{J}}$ has linearly independent columns. By the assumption of constant column rank, we find that for large enough k, $\mathcal{J}$ is also a maximal set of column indices of H^k for which $H^k_{\cdot\mathcal{J}}$ has linearly independent columns. Since $h^k \in \operatorname{ran} H^k$, it follows that, for large k, there is a (unique) solution $w^k_{\mathcal{J}}$, of the system

$$H^k_{\cdot\mathcal{J}} w^k_{\mathcal{J}} = h^k.$$

Now choose a subset $\mathcal{I}$ of the rows of H such that the submatrix $H_{\mathcal{I}\mathcal{J}}$ is invertible, and let

$$w_{\mathcal{J}} = H^{-1}_{\mathcal{I}\mathcal{J}} h_{\mathcal{I}}.$$

It follows that $w^k_{\mathcal{J}} \to w_{\mathcal{J}}$.

For each k we augment $w^k_{\mathcal{J}}$ to form $w^k \in (H^k)^{-1} h^k$, by setting $w^k_j = 0$ for $j \notin \mathcal{J}$. Similarly, we can augment $w_{\mathcal{J}}$ above by setting $w_j = 0$ for $j \notin \mathcal{J}$, to form $w \in H^{-1} h$. Of course $w^k \to w$, and since $h^k \to h \neq 0$, we have

$$\frac{\|w^k\|}{\|h^k\|} \to \frac{\|w\|}{\|h\|} < \infty,$$

contradicting (82). ∎

On the one hand, Lemma 10 extends Hoffman's lemma [3] by allowing H to vary within a set $\mathcal{H}$ rather than remain constant. On the other hand, Hoffman's lemma is more general in that it applies to linear systems of inequalities as well as equalities. We believe, however, that the above result and proof can be adapted to linear systems that include inequalities.

In the following result, we partition the matrix $H \in \mathcal{H} \subset \mathbb{R}^{p\times q}$ as

$$H = \begin{bmatrix} \tilde{H} & \hat{H} \end{bmatrix},$$

where $\tilde{H} \in \mathbb{R}^{p\times\tilde{q}}$ and $\hat{H} \in \mathbb{R}^{p\times\hat{q}}$, with $\tilde{q} + \hat{q} = q$. We use $\tilde{w}$ and $\hat{w}$ to denote vectors in $\mathbb{R}^{\tilde{q}}$ and $\mathbb{R}^{\hat{q}}$, respectively. Below, as usual, $\|\cdot\|$ is the 2-norm.

Lemma 11 *Let $\mathcal{H}$ be a bounded subset of $\mathbb{R}^{p\times q}$ with constant column rank. Then there exists a nonnegative constant $L = L(\mathcal{H})$ with the property that for any $\hat{q}\times\hat{q}$ diagonal matrix $S > 0$, matrix $H = \begin{bmatrix} \tilde{H} & \hat{H} \end{bmatrix} \in \mathcal{H}$ and vector $h \in \operatorname{ran} H$, the (unique) $\hat{w}$ component of the solution of the following problem*

$$\min_{(\tilde{w},\hat{w})} \tfrac{1}{2}\|S\hat{w}\|^2, \qquad \text{subject to } \tilde{H}\tilde{w} + \hat{H}\hat{w} = h \tag{83}$$

satisfies

$$\|\hat{w}\|_\infty \leq L\|h\|_\infty.$$

Proof. We adapt the proof of Monteiro and Wright [4, Lemma 7].

Assume for a contradiction that there exist sequences of positive diagonal matrices $\{S^k\}$, matrices $\{H^k\} \subset \mathcal{H}$, and vectors $\{h^k\}$ such that $h^k \in \operatorname{ran} H^k$ for each k, and

$$\lim_{k\to\infty} \frac{\|\hat{w}^k\|_\infty}{\|h^k\|_\infty} = \infty,$$

where $(\tilde{w}^k, \hat{w}^k)$ is a solution of (83), unique in the $\hat{w}^k$ component, with $S = S^k$, $H = H^k$ and $h = h^k$. By taking a subsequence if necessary, we can define a constant $L_1 > 0$ and a nonempty index set $\mathcal{J} \subset \{1, 2, \ldots, \hat{q}\}$ such that

$$\begin{aligned} \frac{|\hat{w}_j^k|}{\|h^k\|_\infty} &\le L_1, \quad \forall j \notin \mathcal{J}; &\text{(84a)}\\ \lim \frac{|\hat{w}_j^k|}{\|h^k\|_\infty} &= \infty, \quad \forall j \in \mathcal{J}. &\text{(84b)} \end{aligned}$$

Consider the following linear system

$$\begin{aligned} \tilde{H}^k \tilde{w} + \hat{H}^k \hat{w} &= h^k, \\ \hat{w}_j &= \hat{w}_j^k, \quad \forall j \notin \mathcal{J}, \end{aligned} \tag{85}$$

and note that $(\tilde{w}^k, \hat{w}^k)$ is a solution of this system.

Consider the coefficient matrix in (85), which is $\begin{bmatrix} \tilde{H} & \hat{H} \end{bmatrix}$ followed by the row vectors $\begin{bmatrix} 0 & (e^j)^T \end{bmatrix}$, $j \notin \mathcal{J}$, where e^j is the vector in $\mathbb{R}^{\hat{q}}$ composed of 0s except for a 1 in its jth entry. The rank of this matrix is the sum of the cardinality of the set $\{1, 2, \ldots, \hat{q}\} \setminus \mathcal{J}$ and the rank of $\begin{bmatrix} \tilde{H} & \hat{H}_{\cdot\mathcal{J}} \end{bmatrix}$. Hence, the family of coefficient matrices of (85) has constant column rank. By Lemma 10, the system (85) has a solution $(\tilde{x}^k, \hat{x}^k)$ such that

$$\|\hat{x}^k\|_\infty \le \|(\tilde{x}^k, \hat{x}^k)\|_\infty \le L_2 \left\{ \|h^k\|_\infty + \max_{j \notin \mathcal{J}} |\hat{w}_j^k| \right\},$$

where L_2 is a constant depending only on $\mathcal{H}$ and $\mathcal{J}$. Therefore from (84a), we have

$$\|\hat{x}^k\|_\infty \le L_3 \|h^k\|_\infty,$$

where $L_3 = L_2(1 + L_1)$. From (84b) there exists $K \ge 0$ such that for all $k \ge K$ we have

$$|\hat{w}_j^k| > L_3 \|h^k\|_\infty, \quad \forall j \in \mathcal{J},$$

and therefore

$$|\hat{w}_j^k| > \|\hat{x}^k\|_\infty, \quad \forall j \in \mathcal{J}, \ \forall k \ge K.$$

From this relation and the fact that $\hat{w}^k$ satisfies the second equation of (85), we obtain

$$\begin{aligned} \|S^k \hat{x}^k\|^2 &= \sum_{j\in\mathcal{J}} (S_{jj}^k \hat{x}_j^k)^2 + \sum_{j\notin\mathcal{J}} (S_{jj}^k \hat{x}_j^k)^2 \\ &< \sum_{j\in\mathcal{J}} (S_{jj}^k \hat{w}_j^k)^2 + \sum_{j\notin\mathcal{J}} (S_{jj}^k \hat{w}_j^k)^2 \\ &= \|S^k \hat{w}^k\|^2, \qquad \text{for all } k \ge K. \end{aligned} \tag{86}$$

This relation, together with the fact that $\hat{x}^k$ satisfies the first equation of (85), contradicts the assertion that $\hat{w}^k$ is an optimal solution of (83) with $S = S^k$, $H = H^k$, and $h = h^k$. ∎

In Theorem 4 below, we identify the matrix set $\mathcal{H}$ in Lemmas 10 and 11 with the set

$$\left\{ \begin{bmatrix} D_z f(z,\lambda) & Dg_{\mathcal{B}}(z)^T & 0 \\ -Dg(z) & 0 & -I_{\cdot\mathcal{N}} \end{bmatrix} \; : \; \mathrm{dist}_{\mathcal{S}_Z^\infty}(z,\lambda) \le \epsilon \right\}, \tag{87}$$

for some $\epsilon > 0$. To apply this result, we need to show that this set has constant column rank, as we do in the next technical lemma and Theorem 4.

Lemma 12 *Let $\emptyset \neq \mathcal{J} \subset \mathcal{B}$ and $\emptyset \neq \mathcal{K} \subset \mathcal{N}$. Let I denote the identity in $\mathbb{R}^{P\times P}$. If the two-sided projection of $D_z f(z,\lambda)$ onto $\ker(Dg_{\mathcal{B}})$ is positive definite, then for $t \in \mathbb{R}^n$ and $\pi_{\mathcal{J}} \in \mathbb{R}^{|\mathcal{J}|}$, we have*

$$(t, \pi_{\mathcal{J}}) \in \ker \begin{bmatrix} D_z f(z,\lambda) & Dg_{\mathcal{J}}(z^k)^T \\ -Dg_{\mathcal{B}}(z^k) & 0 \end{bmatrix} \tag{88}$$

if and only if $t = 0$ and $\pi_{\mathcal{J}} \in \ker(Dg_{\mathcal{J}})^T$. In addition, we have

$$\dim\ker \begin{bmatrix} D_z f & (Dg_{\mathcal{J}})^T & 0 \\ -Dg & 0 & -I_{\cdot\mathcal{K}} \end{bmatrix} = \dim\ker(Dg_{\mathcal{J}})^T. \tag{89}$$

Proof. The reverse implication in the first statement is obvious. To prove the forward implication, assume first that (88) holds. We then have

$$(D_z f)t \in \mathrm{ran}\,(Dg_{\mathcal{J}})^T \subset \mathrm{ran}\,(Dg_{\mathcal{B}})^T. \tag{90}$$

Let Z be a basis of $\ker(Dg_{\mathcal{B}})$, so that $Z^T \mathrm{ran}\,(Dg_{\mathcal{B}})^T = 0$. Because $Dg_{\mathcal{B}} t = 0$, we have $t = Z\tilde{t}$ for some $\tilde{t}$. From (90), we have $Z^T(D_z f)t = 0$, and so $Z^T(D_z f)Z\tilde{t} = 0$. Because of our nonsingularity assumption on the projection of $D_z f(z,\lambda)$, we have $\tilde{t} = 0$ and therefore $t = 0$. Hence, by substituting in (90), we obtain $\pi_{\mathcal{J}} \in \ker(Dg_{\mathcal{J}})^T$, so the proof of the first part is complete.

We now prove (89). Let the vector $(t, \pi_{\mathcal{J}}, s_{\mathcal{K}})$ have the property that

$$(t, \pi_{\mathcal{J}}, s_{\mathcal{K}}) \in \ker \begin{bmatrix} D_z f & (Dg_{\mathcal{J}})^T & 0 \\ -Dg & 0 & -I_{\cdot\mathcal{K}} \end{bmatrix}.$$

By partitioning appropriately, we have

$$\begin{aligned} (D_z f)t + (Dg_{\mathcal{J}})^T \pi_{\mathcal{J}} &= 0 && \text{(91a)} \\ -Dg_{\mathcal{B}} t &= 0, && \text{(91b)} \\ -(Dg_{\mathcal{N}})t - I_{\mathcal{N}\mathcal{K}} s_{\mathcal{K}} &= 0. && \text{(91c)} \end{aligned}$$

Now we can apply the first part of the theorem to (91a) and (91b) to find that the system (91) can be written equivalently as

$$\begin{aligned} t &= 0, \\ (Dg_{\mathcal{J}})^T \pi_{\mathcal{J}} &= 0, \\ -I_{\mathcal{N}\mathcal{K}} s_{\mathcal{K}} &= 0. \end{aligned}$$

Since $\mathcal{K} \subset \mathcal{N}$, the last of these equations implies that $s_{\mathcal{K}} = 0$. Therefore the solutions of (91) are the vectors of the form $(t, \pi_{\mathcal{J}}, s_{\mathcal{K}}) = (0, \pi_{\mathcal{J}}, 0)$, for all $\pi_{\mathcal{J}} \in \ker(Dg_{\mathcal{J}})^T$, and the proof is complete. ∎

Under certain assumptions (including Assumption 7), it follows from (89) that the set (87) has constant column rank for some $\epsilon > 0$. We state the result formally.

Theorem 4 *Suppose that the standing assumptions are satisfied. Then there is a constant $\epsilon > 0$ such that the bounded set (87) has constant column rank.*

Proof. Because of Assumption 6 and continuity of $D_z f(z, \lambda)$ and $Dg(z)$ with respect to z, we can choose $\epsilon > 0$ so that

- $D_z f(z, \lambda)$ and $Dg(z)$ are bounded on the bounded set $\mathcal{S}_Z + \epsilon \mathbb{B}$, and
- the two-sided projection of $D_z f(z, \lambda)$ onto $\ker Dg_{\mathcal{B}}(z)$ is invertible.

Hence, Lemma 12 applies.

Suppose for contradiction that (87) does not have constant column rank for any $\epsilon > 0$. Then there is a sequence $\{(z^k, \lambda^k)\}$ converging to some $(z^*, \lambda^\infty) \in \mathcal{S}_Z^\infty$ (hence, $D_z f(z^k, \lambda^k) \to D_z f(z^*, \lambda^\infty)$), and some index sets $\mathcal{J} \subset \mathcal{B}$, $\mathcal{K} \subset \mathcal{N}$ such that

$$\dim\ker \begin{bmatrix} D_z f(z^k, \lambda^k) & Dg_{\mathcal{J}}(z^k)^T & 0 \\ -Dg(z^k) & 0 & -I_{\cdot\mathcal{K}} \end{bmatrix} < \dim\ker \begin{bmatrix} D_z f(z^*, \lambda^\infty) & (Dg^*_{\mathcal{J}})^T & 0 \\ -Dg^* & 0 & -I_{\cdot\mathcal{K}} \end{bmatrix}.$$

Hence, from (89), we must have

$$\dim\ker(Dg^k_{\mathcal{J}})^T < \dim\ker(Dg^*_{\mathcal{J}})^T$$

for all k. This inequality contradicts Assumption 7, so no such sequence exists, and the proof is complete. ∎

Finally, we state the quadratic program for which $(t, u_{\mathcal{B}}, v_{\mathcal{N}})$ is a solution, and we use the results above to estimate the size of these components. See (40) and (41) for the definitions of η_f, η_g and t, u, v respectively.

Theorem 5 *Suppose that the standing assumptions hold, and let $(z, \lambda, y) \in \Omega(\delta_4)$, where $\delta_4 = \min(\delta_3, \epsilon)$, and δ_3 and ϵ are defined in Lemma 8 and Theorem 4, respectively. Then the solution (t, u, v) of (41) is also the solution of the following convex quadratic program:*

$$
\begin{aligned}
& \min_{(\bar{t}, \bar{u}_{\mathcal{B}}, \bar{v}_{\mathcal{N}})} \tfrac{1}{2}\|D_{\mathcal{BB}}\bar{u}_{\mathcal{B}}\|^2 + \tfrac{1}{2}\|(D_{\mathcal{NN}})^{-1}\bar{v}_{\mathcal{N}}\|^2, \\
& \text{subject to } \begin{bmatrix} D_z f(z,\lambda) & Dg_{\mathcal{B}}(z)^T & 0 \\ -Dg(z) & 0 & -(I_{\cdot\mathcal{N}}) \end{bmatrix} \begin{bmatrix} \bar{t} \\ \bar{u}_{\mathcal{B}} \\ \bar{v}_{\mathcal{N}} \end{bmatrix} = \begin{bmatrix} \eta_f - Dg_{\mathcal{N}}(z)^T u_{\mathcal{N}} \\ \eta_g + I_{\cdot\mathcal{B}} v_{\mathcal{B}} \end{bmatrix}.
\end{aligned} \tag{92}
$$

Moreover, there is a constant C_{12} such that

$$
\|(u_{\mathcal{B}}, v_{\mathcal{N}})\| \le C_{12}\|(\eta_f, \eta_g, u_{\mathcal{N}}, v_{\mathcal{B}})\|. \tag{93}
$$

Proof. Note first that the matrices D, D^{-1} (see (48)) are well defined because of the restriction $(z, \lambda, y) \in \Omega(\delta_4)$.

It is immediate from (41) that $(t, u_{\mathcal{B}}, v_{\mathcal{N}})$ is feasible for (92). To prove optimality, we need to show that the remaining KKT conditions hold; that is,

$$
\begin{bmatrix} 0 \\ D^2_{\mathcal{BB}} u_{\mathcal{B}} \\ D^{-2}_{\mathcal{NN}} v_{\mathcal{N}} \end{bmatrix} \in \operatorname{ran}\left(\begin{bmatrix} (D_z f)^T & -Dg^T \\ Dg_{\mathcal{B}} & 0 \\ 0 & -I_{\mathcal{N}\cdot} \end{bmatrix} \right).
$$

By using arguments similar to those of Ye and Anstreicher [12, Section 3], we can show that

$$
\operatorname{ran}\left(\begin{bmatrix} (D_z f)^T & -Dg^T \\ Dg_{\mathcal{B}} & 0 \\ 0 & -I_{\mathcal{N}\cdot} \end{bmatrix} \right) = \operatorname{ran}\left(\begin{bmatrix} -D_z f & -Dg^T \\ Dg_{\mathcal{B}} & 0 \\ 0 & -I_{\mathcal{N}\cdot} \end{bmatrix} \right).
$$

Hence, it suffices to show that

$$
\begin{bmatrix} 0 \\ D^2_{\mathcal{BB}} u_{\mathcal{B}} \\ D^{-2}_{\mathcal{NN}} v_{\mathcal{N}} \end{bmatrix} = \begin{bmatrix} -D_z f & -Dg^T \\ Dg_{\mathcal{B}} & 0 \\ 0 & -I_{\mathcal{N}\cdot} \end{bmatrix} \begin{bmatrix} z + t - z^* \\ \lambda + u - \bar{\pi} \end{bmatrix}, \tag{94}
$$

where $\bar{\pi}$ is defined in (39). To verify this claim, note first that by (40a) and (41), we have

$$
D_z f(z,\lambda)t + Dg(z)^T u = \eta_f = D_z f(z,\lambda)(z^* - z) + Dg(z)^T(\bar{\pi} - \lambda),
$$

and therefore

$$
0 = -D_z f(z,\lambda)(z + t - z^*) - Dg(z)^T(\lambda + u - \bar{\pi}).
$$

For the second part of (94), we have from (40b) and (41) that

$$
\begin{aligned}
-(Dg_{\mathcal{B}})t &= v_{\mathcal{B}} + (\eta_g)_{\mathcal{B}} = v_{\mathcal{B}} + y_{\mathcal{B}} - (Dg_{\mathcal{B}})(z^* - z), \\
D^2 u &= \Lambda^{-1}(Yu) = \Lambda^{-1}(-\Lambda Y e - \Lambda v) = -y - v,
\end{aligned}
$$

and therefore

$$D^2_{\mathcal{BB}} u_{\mathcal{B}} = (Dg_{\mathcal{B}})(z + t - z^*).$$

Finally, we use (41) together with $\bar{\pi}_{\mathcal{N}} = 0$ to write

$$D^{-2}_{\mathcal{NN}} v_{\mathcal{N}} = Y^{-1}_{\mathcal{NN}} \Lambda_{\mathcal{NN}} v_{\mathcal{N}} = -\lambda_{\mathcal{N}} - u_{\mathcal{N}} = -I_{\mathcal{N}\cdot}(\lambda + u - \bar{\pi}).$$

We now prove (93). For $(z, \lambda, y) \in \Omega(\delta_4)$, we have

$$\mathrm{dist}_{\mathcal{S}^\infty_Z}(z, \lambda) \le \|(z, \lambda) - (z^*, \hat{\lambda})\| \le \delta_4 \le \epsilon.$$

It therefore follows from Theorem 4 that the coefficient matrix in (92) lies in the set (87), which has constant column rank. Our claim is proved by applying Lemma 11 to the quadratic program (92). ■

5.5 The Fast Step Estimate

We are now in a position to tie together the results of Subsections 5.2, 5.3, and 5.4 and therefore obtain an estimate for the length of the fast step.

Corollary 1 *Suppose that the standing assumptions hold. Then for the positive constant δ_4 defined in Theorem 5 and all $(z, \lambda, y) \in \Omega(\delta_4)$, the fast step $(\Delta z, \Delta\lambda, \Delta y)$ calculated by setting $\tilde{\sigma} = 0$ in (12) satisfies*

$$\|(\Delta z, \Delta\lambda, \Delta y)\| \le C_0 \mu, \tag{95}$$

for some constant C_0.

Proof. From Theorem 3, we have $\|(\Delta\lambda_{\mathcal{N}}, \Delta y_{\mathcal{B}})\| = O(\mu)$ whenever $(z, \lambda, y) \in \Omega(\delta_4) \subset \Omega(\delta_1)$. We seek similar bounds on the remaining components, which are $(\Delta\lambda_{\mathcal{B}}, \Delta y_{\mathcal{N}})$ and Δz.

From Lemma 9, we have for $(z, \lambda, y) \in \Omega(\delta_3)$ that $\|(t', u', v')\| \le C_{11}\mu$. Therefore,

$$\|(u_{\mathcal{N}}, v_{\mathcal{B}})\| \le \|(\Delta\lambda_{\mathcal{N}}, \Delta y_{\mathcal{B}})\| + \|(u'_{\mathcal{N}}, v'_{\mathcal{B}})\| = O(\mu).$$

Since η_f and η_g are bounded by $\check{L}\mu$ over the set $\Omega(\delta_3)$ (Lemma 8 and (79)), the right-hand side of (93) is $O(\mu)$. Hence, the second part of Theorem 5 yields $\|(u_{\mathcal{B}}, v_{\mathcal{N}})\| = O(\mu)$. Hence,

$$\|(\Delta\lambda_{\mathcal{B}}, \Delta y_{\mathcal{N}})\| \le \|(u_{\mathcal{B}}, v_{\mathcal{N}})\| + \|(u'_{\mathcal{B}}, v'_{\mathcal{N}})\| = O(\mu). \tag{96}$$

Finally, we show that the desired estimate holds for Δz as well. The proof is almost the same as the proof of Lemma 5, so we skip the details. Starting with (12), we perform block elimination to obtain a system with the same coefficient matrix as in (46), but a different right-hand side; namely,

$$\begin{aligned} &\begin{bmatrix} r_f - (Dg_{\mathcal{N}})^T \Lambda_{\mathcal{N}} Y^{-1}_{\mathcal{N}}((r_g)_{\mathcal{N}} - y_{\mathcal{N}}) \\ (r_g)_{\mathcal{B}} - y_{\mathcal{B}} \end{bmatrix} \\ &= \begin{bmatrix} r_f - (Dg_{\mathcal{N}})^T \Lambda_{\mathcal{N}} Y^{-1}_{\mathcal{N}}(r_g)_{\mathcal{N}} \\ (r_g)_{\mathcal{B}} \end{bmatrix} + \begin{bmatrix} (Dg_{\mathcal{N}})^T \lambda_{\mathcal{N}} \\ -y_{\mathcal{B}} \end{bmatrix}. \end{aligned} \tag{97}$$

The first vector on the right is exactly the right hand side of (46), hence its norm is $O(\mu)$ as shown in the proof of Lemma 5. The second vector on the right of the above equation is also $O(\mu)$ from Lemma 4. Thus the vector on the left hand side of (97) is $O(\mu)$ for $(z, \lambda, y) \in \Omega(\delta_4)$. Hence, as in (47), we have that

$$\begin{aligned} \begin{bmatrix} D_z f(z^*, \hat{\lambda}) & (Dg_{\mathcal{B}}^*)^T \\ -Dg_{\mathcal{B}}^* & 0 \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta\lambda_{\mathcal{B}} \end{bmatrix} &= O(\mu) + O(\mu + \|z - z^*\| + \|\lambda - \hat{\lambda}\|) \begin{bmatrix} \Delta z \\ \Delta\lambda_{\mathcal{B}} \end{bmatrix} \\ &= O(\mu) + O(\mu + \delta_4) \begin{bmatrix} \Delta z \\ \Delta\lambda_{\mathcal{B}} \end{bmatrix}. \end{aligned} \tag{98}$$

By using the same argument as in Lemma 5, we have that $\|\Delta z\| = O(\mu) + O(\|\Delta\lambda_{\mathcal{B}}\|)$. (A careful analysis shows that it is not even necessary to decrease δ_4 to obtain this estimate.) Because $\|\Delta\lambda_{\mathcal{B}}\| = O(\mu)$ by (96), we have $\|\Delta z\| = O(\mu)$, as required. ■

5.6 Proof of Theorem 2

At long last, we are in a position to prove Theorem 2. We look at a subsequence that approaches the limit point $(z^*, \hat{\lambda}, y^*)$, and we show that once this subsequence enters a sufficiently small neighborhood of this point, with a sufficiently large iteration count, the following things happen:

- When the fast step is tried, the initial choice (18) for α satisfies the conditions (19), and the new iterate satisfies $\mu_{k+1} \le \rho\mu_k$ and is accepted by the main algorithm.
- The new iterate and all subsequent iterates cannot escape a (slightly larger) neighborhood of $(z^*, \hat{\lambda}, y^*)$, and fast steps are taken at all these iterates too.
- The entire sequence converges superlinearly to the limit point $(z^*, \hat{\lambda}, y^*)$.

Proof. (Theorem 2) To prove the assertion that the initial choice of fast step length (18) is eventually always accepted, we collect a few relevant facts.

First, note from the choice of constant δ_3 in the proof of Lemma 8 and the fact that $\delta_4 \in (0, \delta_3]$ that $Df(z, \lambda)$ and $Dg(z)$ are Lipschitz continuous on an open neighborhood of $\Omega(\delta_4)$. We denote the relevant Lipschitz constant by L.

Second, note that the sequence $\{\mu_k^{\hat{\tau}}/\bar{\gamma}^{t_k}\}$ decreases monotonically to zero. On safe steps, we have $\mu_{k+1} < \mu_k$ while t_k (and therefore the denominator) remain unchanged. On fast steps, we have from the relationship between ρ, $\bar{\gamma}$, and $\hat{\tau}$ in (14) that

$$\frac{\mu_{k+1}^{\hat{\tau}}}{\bar{\gamma}^{t_{k+1}}} \le \frac{\rho^{\hat{\tau}}\mu_k^{\hat{\tau}}}{\bar{\gamma}\bar{\gamma}^{t_k}} \le \frac{\bar{\gamma}\mu_k^{\hat{\tau}}}{2\bar{\gamma}\bar{\gamma}^{t_k}} = \frac{1}{2}\frac{\mu_k^{\hat{\tau}}}{\bar{\gamma}^{t_k}}. \tag{99}$$

If there are infinitely many fast steps, the sequence is driven to zero because the factor 1/2 in (99) occurs infinitely often. If there are only finitely many fast steps, the denominator $\bar{\gamma}^{t_k}$

eventually settles down to a constant, and the sequence is driven to zero by the fact that $\mu_k \downarrow 0$.

We now proceed with the main part of the proof. Let $\{k_j\}_{j=0}^{\infty}$ be the sequence of indices such that

$$\lim_{j\to\infty} (z^{k_j}, \lambda^{k_j}, y^{k_j}) = (z^*, \hat{\lambda}, y^*). \tag{100}$$

Now choose the index J sufficiently large that the following conditions are satisfied:

$$\begin{aligned}
(z^{k_J}, \lambda^{k_J}, y^{k_J}) &\in \Omega(\delta_4/4), && (101a)\\
\mu_{k_J} &\leq \frac{(1-\rho)\delta_4}{4C_0}, && (101b)\\
\mu_{k_J}^{1-\hat{\tau}} &\leq \frac{(1-\bar{\gamma})(\gamma_{\max}-\gamma_{\min})}{2C_0^2}, && (101c)\\
\mu_{k_J}^{1-\hat{\tau}} &\leq \frac{\bar{\gamma}\beta_{\min}}{(L/2+\beta_{\max})C_0^2}, && (101d)\\
\mu_{k_J}^{\hat{\tau}}/\bar{\gamma}^{t_{k_J}} &\leq \rho/2, && (101e)\\
\mu_{k_J} &\leq \frac{\rho}{2C_0^2}. && (101f)
\end{aligned}$$

Let us first show that the value $\alpha = 1 - \mu_{k_J}^{\hat{\tau}}/\bar{\gamma}^{t_{k_J}}$ from (18) satisfies the condition (19a); that is,

$$\lambda_i(\alpha)y_i(\alpha) \geq (\gamma_{\min} + \bar{\gamma}^{t+1}(\gamma_{\max}-\gamma_{\min}))\mu(\alpha). \tag{102}$$

(We omit the subscript k_J here and later for clarity.) For the left-hand side of (102), we have

$$\begin{aligned}
\lambda_i(\alpha)y_i(\alpha) &= (\lambda_i + \alpha\Delta\lambda_i)(y_i + \alpha\Delta y_i)\\
&= \lambda_i y_i(1-\alpha) + \alpha^2\Delta\lambda_i\Delta y_i\\
&\geq (\gamma_{\min} + \bar{\gamma}^t(\gamma_{\max}-\gamma_{\min}))(1-\alpha)\mu - C_0^2\mu^2,
\end{aligned}$$

where we used the relationships (38), (95), and $\lambda_i y_i \geq \gamma\mu$ with $\gamma = \gamma_{\min} + \bar{\gamma}^t(\gamma_{\max}-\gamma_{\min})$. For the right-hand side of (102), we have by the same logic that

$$\begin{aligned}
\mu(\alpha) &= (\lambda + \alpha\Delta\lambda)^T(y+\alpha\Delta y)/P\\
&\leq (1-\alpha)\mu + \alpha^2\|\Delta y\|\|\Delta\lambda\|/P\\
&\leq (1-\alpha)\mu + C_0^2\mu^2. && (103)
\end{aligned}$$

Hence, for the condition (102) to hold, it suffices that

$$\begin{aligned}
&[\gamma_{\min} + \bar{\gamma}^t(\gamma_{\max}-\gamma_{\min})](1-\alpha)\mu - C_0^2\mu^2\\
&\quad\geq [\gamma_{\min} + \bar{\gamma}^{t+1}(\gamma_{\max}-\gamma_{\min})](1-\alpha)\mu + C_0^2\mu^2.
\end{aligned}$$

This inequality is equivalent to

$$(\bar{\gamma}^t - \bar{\gamma}^{t+1})(\gamma_{\max}-\gamma_{\min})\mu(1-\alpha) \geq 2C_0^2\mu^2. \tag{104}$$

By substituting $1 - \alpha = \mu^{\hat{\tau}}/\bar{\gamma}^t$ from (18) and rearranging, we find that (104) is in turn equivalent to (101c). Hence condition (19a) is satisfied.

We need the Lipschitz continuity assumption for the second condition (19b). Because of (10) and the definition of $\tilde{\beta}$ in the *fast* routine, we can rewrite this condition as

$$\|f(z(\alpha), \lambda(\alpha))\| \le (1 + \bar{\gamma}^{t+1})\beta\mu(\alpha), \tag{105}$$

where the current point (z, λ) has $\|f(z, \lambda)\| \le \beta\mu$. Taylor's theorem can be used to expand $f(z(\alpha), \lambda(\alpha))$, exactly as in (27). The difference here is that Lipschitz continuity can be used to obtain a tighter estimate of Δr_f. Note that the arguments of Df in (29) lie within the domain of Lipschitz continuity, since by (101a), (101b), and (95), we have

$$\begin{aligned} &\|(z + \theta\alpha\Delta z, \lambda + \theta\alpha\Delta\lambda) - (z^*, \hat{\lambda})\| \\ &\quad \le \quad \|(z - z^*, \lambda - \hat{\lambda})\| + \|(\Delta z, \Delta\lambda)\| \le \delta_4/4 + C_0\mu_{k_J} \le \delta_4/2. \end{aligned}$$

Therefore we have from (28) and (95) that

$$\|\Delta r_f\| \le \frac{1}{2}L\|(\Delta z, \Delta\lambda)\|^2 \le \frac{1}{2}LC_0^2\mu^2,$$

As in (27), it follows that

$$\|f(z(\alpha), \lambda(\alpha))\| \le (1 - \alpha)\beta\mu + \frac{1}{2}LC_0^2\mu^2.$$

Meanwhile, a trivial change to the estimate (103) yields

$$\mu(\alpha) \ge (1 - \alpha)\mu - C_0^2\mu^2.$$

From these last two inequalities, we see that condition (105) is satisfied if

$$(1 - \alpha)\beta\mu + \frac{1}{2}LC_0^2\mu^2 \le (1 + \bar{\gamma}^{t+1})\beta[(1 - \alpha)\mu - C_0^2\mu^2].$$

Because $(1 + \bar{\gamma}^{t+1})\beta \le \beta_{\max}$, from (21), this last condition in turn is satisfied if

$$\frac{1}{2}LC_0^2\mu^2 \le \bar{\gamma}^{t+1}\beta(1 - \alpha)\mu - \beta_{\max}C_0^2\mu^2.$$

By substituting from (18) and using the bound $\beta_{\min} \le \beta$, we find that this last condition is implied by (101d), so we conclude that (105) is also satisfied. By similar logic, we can show that the same conditions (101) also guarantee that the remaining condition (19c) holds.

Finally, we verify that $\mu_{k_J+1} \le \rho\mu_{k_J}$, so that the fast step is accepted by the main algorithm. Because of (103), this condition is satisfied if

$$(1 - \alpha)\mu + C_0^2\mu^2 \le \rho\mu,$$

which, by substitution of (18), is equivalent to

$$\mu^{\hat{\tau}}/\bar{\gamma}^t + C_0^2\mu \le \rho.$$

Conditions (101e) and (101f) together guarantee that this conditions holds, so we are done.

At this point, we have shown that a fast step is taken from $(z^{k_J}, \lambda^{k_J}, y^{k_J})$. The new iterate does not move away too far from the limit point, if at all, because

$$\begin{aligned}
\|(z^{k_J+1}, \lambda^{k_J+1}, y^{k_J+1}) - (z^*, \hat{\lambda}, y^*)\| &\le \|(z^{k_J}, \lambda^{k_J}, y^{k_J}) - (z^*, \hat{\lambda}, y^*)\| + \|(\Delta z^{k_J}, \Delta\lambda^{k_J}, \Delta y^{k_J})\| \\
&\le \delta_4/4 + C_0\mu_{k_J} \\
&\le \delta_4/2,
\end{aligned}$$

where the last inequality is a consequence of (101b) and (95). Hence, $(z^{k_J+1}, \lambda^{k_J+1}, y^{k_J+1}) \in \Omega(\frac{1}{2}\delta_4)$, and so the estimate (95) applies again at iteration k_J+1. The remaining conditions (101b)–(101f) continue to apply at the new iterate, and the same logic as above can be used to show that a fast step is again taken from this iterate. Because of these two consecutive fast steps, we have

$$\mu_{k_J+2} \le \rho\mu_{k_J+1} \le \rho^2\mu_{k_J}. \tag{106}$$

We can continue in this vein, inductively, to show that only fast steps are taken from this point onwards, and that the iterates never leave the neighborhood $\Omega(\frac{1}{2}\hat{\delta})$. The last statement follows from (95) and (106), since we have for all $s \ge 0$ that

$$\begin{aligned}
\mathrm{dist}_{\mathcal{S}_Z^\infty}(z^{k_J+s}, \lambda^{k_J+s}) &\le \hat{\delta}/4 + C_0(\mu_{k_J} + \mu_{k_J+1} + \cdots + \mu_{k_J+s-1}) \\
&\le \hat{\delta}/4 + C_0\mu_{k_J}(1 + \rho + \rho^2 + \cdots) \\
&\le \hat{\delta}/4 + \frac{C_0}{1-\rho}\mu_{k_J} \\
&\le \hat{\delta}/2.
\end{aligned}$$

We now examine the rate of convergence of $\{\mu_k\}$. From (18) and (103), we have for all $k \ge k_J$ that

$$\mu_{k+1} \le \mu_k\left(\frac{\mu_k^{\hat{\tau}}}{\bar{\gamma}^{t_k}}\right) + C_0^2\mu_k^2.$$

Hence for some $K \ge k_J$, the first term on the right-hand side dominates the second, and we have

$$\mu_{k+1} \le \mu_k^{\hat{\tau}+1}/\bar{\gamma}^{t_k}, \qquad \text{for all } k \ge K.$$

The proof that $\{\mu_k\}$ converges to zero with Q-order at least $1+\hat{\tau}$ follows by standard arguments; see Wright [9, Theorem 6.3] and Wright and Zhang [11, Theorem 5.2]. Hence, part (i) of the theorem is proved.

For (ii), we show that the sequence of iterates is Cauchy. For all $K_2 > K_1$ sufficiently large, we have from (95) that

$$\left\|(z^{K_2}, \lambda^{K_2}, y^{K_2}) - (z^{K_1}, \lambda^{K_1}, y^{K_1})\right\| \le \sum_{k=K_1}^{K_2} \alpha_k\|(\Delta z^k, \Delta\lambda^k, \Delta y^k)\|$$

$$
\begin{aligned}
&\leq\ C_0 \sum_{k=K_1}^{\infty} \mu_k \\
&\leq\ C_0 \mu_{K_1} \left[1 + \rho + \rho^2 + \cdots\right] \\
&=\ C_0 \mu_{K_1} \frac{1}{1-\rho}\ \rightarrow\ 0 \qquad \text{as } K_1 \rightarrow \infty. \qquad (107)
\end{aligned}
$$

Hence the sequence is Cauchy, so it converges to a limit point, which must be the limit point $(z^*, \hat{\lambda}, y^*)$ of the subsequence (100). Its R-order follows immediately from (107) and the result of part (i). ■

6 Existence of a Limit Point

In our main result, Theorem 2, we assumed that a limit point of the sequence $\{(z^k, \lambda^k, y^k)\}$ actually exists. This condition will follow immediately if we can show that the sequence is bounded, by compactness.

We show in Lemma 13 that boundedness of the solution set $\mathcal{S}$ is a consequence of boundedness of the feasible set $\mathcal{C}$ defined in (2). Then, in Lemma 14, we show that boundedness of the iterate sequence $\{(z^k, \lambda^k, y^k)\}$ also holds under the additional assumption that $\mu_k \downarrow 0$.

Lemma 13 *Suppose that Assumptions 1 and 3 hold and that the set $\mathcal{C}$ defined by (2) is bounded. Then the solution set $\mathcal{S}$ is nonempty, bounded, closed, and therefore compact.*

Proof. By Theorem 3.1 of Harker and Pang [2], the set of vectors z^* that solves (1) is nonempty. This set is also bounded because of the restriction $z^* \in \mathcal{C}$. Boundedness of the solution components y^* follows trivially because $y^* = g(z^*)$ and g is smooth.

We prove boundedness of the optimal λ^* components by contradiction. If the claim does not hold, we can choose a sequence of solutions $(\hat{z}^k, \hat{\lambda}^k, \hat{y}^k) \in \mathcal{S}$ such that $\|\hat{\lambda}^k\|_\infty \uparrow \infty$. (The other components $\hat{z}^k$ and $\hat{y}^k$ remain bounded, by the argument of the preceding paragraph.) We can assume without loss of generality that

$$
(\hat{z}^k, \hat{y}^k) \rightarrow (\hat{z}, \hat{y}), \quad \text{with} \quad \hat{z} \in \mathcal{C},\ \hat{y} \geq 0,
$$

and

$$
\frac{\hat{\lambda}^k}{\|\hat{\lambda}^k\|_\infty} \rightarrow \hat{\lambda}, \quad \text{with} \quad \|\hat{\lambda}\|_\infty = 1, \quad \hat{\lambda} \geq 0.
$$

Moreover, since $(\hat{\lambda}^k)^T g(\hat{z}^k) = 0$ for all k, we have that

$$
\hat{\lambda}_i > 0 \ \Rightarrow\ g_i(\hat{z}) = 0. \tag{108}
$$

Because of (3) and (4), we have that

$$
\Phi(\hat{z}^k) + \sum_{i=1}^{m} Dg_i(\hat{z}^k)\hat{\lambda}_i^k = 0, \qquad \text{for all } k.
$$

Dividing by $\|\hat{\lambda}^k\|_\infty$ and taking the limit as $k \to \infty$, we have

$$0 = \sum_{i=1}^{m} Dg_i(\hat{z})\hat{\lambda}_i = \sum_{i|\hat{\lambda}_i>0} Dg_i(\hat{z})\hat{\lambda}_i. \tag{109}$$

Given the Slater point $\bar{z}$ (Assumption 3), convexity of g, and the property (108), we have that

$$\hat{\lambda}_i > 0 \;\Rightarrow\; 0 > g_i(\bar{z}) \geq g_i(\hat{z}) + Dg_i(\hat{z})^T(\bar{z} - \hat{z}) = Dg_i(\hat{z})^T(\bar{z} - \hat{z}). \tag{110}$$

But this inequality implies that

$$\sum_{i|\hat{\lambda}_i>0} (\bar{z} - \hat{z})^T Dg_i(\hat{z})\hat{\lambda}_i < 0,$$

which contradicts (109). Hence, $\{\hat{\lambda}^k\}$ cannot be unbounded, so our proof is complete.

Closedness of $\mathcal{S}$ follows immediately from the definition. ∎

Lemma 14 *Suppose that Assumptions 1 and 3 hold and that $\mathcal{C}$ is bounded and* $\lim_{k\to\infty} \mu_k = 0$. *Then the iterate sequence $\{(z^k, \lambda^k, y^k)\}$ is bounded.*

Proof. We start by showing that there is a constant $B > 0$ such that $g_i(z^k) \leq B$ for all i and k. From this observation together with Assumption 3, we deduce that $\{z^k\}$ is bounded. Boundedness of $\{y^k\}$ follows directly from boundedness of $\{z^k\}$. The final part of the proof uses an argument like that in the proof of Lemma 13.

Since $(z^k, \lambda^k, y^k) \in \Omega$ for all k, we have from (10), (22), and $y^k \geq 0$ that

$$g_i(z^k) = [r_g^k]_i - y_i^k \leq [r_g^k]_i \leq \|r_g^k\| \leq \beta_{\max}\mu_k \leq \beta_{\max}\mu_0.$$

So if we define $B = \beta_{\max}\mu_0$, we have

$$g_i(z^k) \leq B, \qquad \text{for all } k = 0, 1, 2, \ldots \text{ and } i = 1, 2, \ldots, m. \tag{111}$$

Suppose for contradiction that $\{z^k\}$ is not bounded. If $\bar{z}$ is the vector from Assumption 3, we can choose a subsequence $\mathcal{K}$ such that

$$\|z^k - \bar{z}\| \uparrow \infty, \qquad \text{for } k \in \mathcal{K}. \tag{112}$$

We now define $\epsilon = \min_{i=1,2,\ldots,m} -g_i(\bar{z})$ and note that $\epsilon > 0$ by Assumption 3. We also define an auxiliary subsequence $\{\hat{z}^k\}$ for $k \in \mathcal{K}$ by

$$\hat{z}^k = \bar{z} + \frac{\epsilon}{B + \epsilon}(z^k - \bar{z}), \tag{113}$$

where B is defined in (111). By convexity of each g_i, we have from the definitions of B and ϵ that

$$\begin{aligned} g_i(\hat{z}^k) &\leq \left(1 - \frac{\epsilon}{B+\epsilon}\right) g_i(\bar{z}) + \frac{\epsilon}{B+\epsilon} g_i(z^k) \\ &\leq \frac{B}{B+\epsilon} g_i(\bar{z}) + \frac{\epsilon}{B+\epsilon} B \\ &\leq -\frac{B\epsilon}{B+\epsilon} + \frac{B\epsilon}{B+\epsilon} \\ &= 0, \end{aligned}$$

for all $1, 2, \ldots, m$. Hence, $\hat{z}^k \in \mathcal{C}$ by the definition (2). On the other hand, we have from (112) and (113) that

$$\|\hat{z}^k - \bar{z}\| = \frac{\epsilon}{B+\epsilon} \|z^k - \bar{z}\| \uparrow \infty, \qquad \text{for } k \in \mathcal{K},$$

which contradicts boundedness of $\mathcal{C}$. Hence, $\{z^k\}$ is bounded.

Boundedness of $\{y^k\}$ follows immediately from (10), since

$$\|y^k\| = \|r_g^k - g(z^k)\| \leq \beta_{\max}\mu_0 + \|g(z^k)\|.$$

The right-hand side of this expression is bounded because $\{z^k\}$ is bounded and g is continuous.

Assume for contradiction that $\{\lambda^k\}$ is unbounded. From (4) and (10), we have that

$$\Phi(z^k) + \sum_{i=1}^{m} Dg_i(z^k)\lambda_i^k = r_f^k. \tag{114}$$

Because $\{z^k\}$ and $\{y^k\}$ are bounded, we can choose a subsequence $\mathcal{K}$ such that

$$(z^k, y^k) \to (\hat{z}, \hat{y})$$

and

$$\frac{\lambda^k}{\|\lambda^k\|_\infty} \to \hat{\lambda}, \quad \text{with} \quad \|\hat{\lambda}\|_\infty = 1, \quad \hat{\lambda} \geq 0.$$

We have from (10) that

$$g(z^k) = r_g^k - y^k \leq \beta_{\max}\mu_k - y^k.$$

Hence, using $\mu_k \downarrow 0$ and $y^k \geq 0$, and taking the limits of both sides for $k \in \mathcal{K}$, we obtain $g(\hat{z}) = -\hat{y} \leq 0$ and hence $\hat{z} \in \mathcal{C}$. Moreover, if $\hat{\lambda}_i > 0$, we must have $g_i(\hat{z}) = 0$, since otherwise we would have

$$\lim_{k\to\infty} \mu_k \geq \lim_{k\to\infty} -\|\lambda^k\|_\infty \hat{\lambda}_i g_i(\hat{z})/P \uparrow \infty.$$

The remainder of the proof now follows exactly as in Lemma 13 above. ∎

We conclude with a corollary of Lemma 12 that throws extra light on our assumptions.

Lemma 15 *Suppose that the standing assumptions are satisfied. Then for any $(z^*, \lambda) \in \mathcal{S}_Z$, we have*

$$\begin{bmatrix} D_z f(z^*, \lambda) & Dg_{\mathcal{B}}(z^*)^T \\ -Dg_{\mathcal{B}}(z^*) & 0 \end{bmatrix} \begin{bmatrix} \delta z \\ \delta \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{115}$$

if and only if $\delta z = 0$ and $\delta\lambda \in \ker Dg_{\mathcal{B}}(z^*)^T$. In particular, the Jacobian matrix in (115) is invertible if and only if $Dg_{\mathcal{B}}(z^*)$ has full row rank.

Assumption 6 is a weak version of the better-known condition that the "active" submatrix (115) of the Jacobian (6) is invertible—an assumption that is made in most local convergence analyses of nonlinear programming algorithms including Wright and Ralph [10]. Allowing nonzero vectors $\delta\lambda$ in the null space of the above Jacobian matrix amounts to allowing nonunique optimal multipliers λ; this flexibility relies on the constant rank condition, Assumption 7. The main point of the current paper is that superlinear convergence still holds when the weaker (but more complicated!) assumptions of this paper are used instead of the standard ones.

Acknowledgments

We thank Jong-Shi Pang for comments that led to improvements in the manuscript.

References

[1] J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Mathematical Programming, 12 (1977), pp. 136–138.

[2] P. T. Harker and J.-S. Pang, *Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications*, Mathematical Programming, 48 (1990), pp. 161–220.

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

[4] R. D. C. Monteiro and S. J. Wright, *Local convergence of interior-point algorithms for degenerate monotone LCP*, Computational Optimization and Applications, 3 (1994), pp. 131–155.

[5] J.-S. Pang and D. Ralph, *Piecewise smoothness, local invertibility, and parametric analysis of normal maps*, technical report, Mathematical Sciences Department, Johns Hopkins University, Baltimore, Md., 1993. To appear in Mathematics of Operations Research.

[6] E. Polak, *Computational Methods in Optimization*, Academic Press, New York, 1970.

[7] S. J. Wright, *A path-following infeasible-interior-point algorithm for linear complementarity problems*, Optimization Methods and Software, 2 (1993), pp. 79–106.

[8] ——, *A path-following interior-point algorithm for linear and quadratic optimization problems*, Preprint MCS–P401–1293, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., December 1993. To appear in Annals of Operations Research.

[9] ——, *An infeasible-interior-point algorithm for linear complementarity problems*, Mathematical Programming, 67 (1994), pp. 29–52.

[10] S. J. Wright and D. Ralph, *A superlinear infeasible-interior-point algorithm for monotone nonlinear complementarity problems*, Preprint MCS–P344–0193, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., January 1993. revised June, 1994. To appear in Mathematics of Operations Research.

[11] S. J. Wright and Y. Zhang, *A superquadratic infeasible-interior-point method for linear complementarity problems*, Preprint MCS–P418–0294, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., February 1994. To appear in Mathematical Programming.

[12] Y. Ye and K. Anstreicher, *On quadratic and $O(\sqrt{n}L)$ convergence of a predictor-corrector algorithm for LCP*, Mathematical Programming, Series A, 62 (1993), pp. 537–551.

[13] Y. Ye, O. Güler, R. A. Tapia, and Y. Zhang, *A quadratically convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming*, Mathematical Programming, Series A, 59 (1993), pp. 151–162.

[14] Y. Zhang and R. A. Tapia, *A superlinearly convergent polynomial primal-dual interior-point algorithm for linear programming*, SIAM Journal on Optimization, 3 (1993), pp. 118–133.

[15] Y. Zhang, R. A. Tapia, and J. E. Dennis, *On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms*, SIAM Journal on Optimization, 2 (1992), pp. 304–324.