# On the Formulation and Theory of the Newton Interior-Point Method for Nonlinear Programming[1]

A. S. El-Bakry[2], R. A. Tapia[3], T. Tsuchiya[4] and Y. Zhang[5]

**Abstract**

In this work we first study in detail the formulation of the primal-dual interior-point method for linear programming. We show that, contrary to popular belief, it cannot be viewed as the damped Newton's method applied to the Karush-Kuhn-Tucker conditions for the logarithmic barrier function problem. Next we extend the formulation to general nonlinear programming, and then validate this extension by demonstrating that this algorithm can be implemented so that it is locally and Q-quadratically convergent under only the standard Newton's method assumptions. We also establish a global convergence theory for this algorithm and include promising numerical experimentation.

**Keywords:** Interior-point methods, primal-dual methods, nonlinear programming, superlinear and quadratic convergence, global convergence.

## 1 Introduction

Motivated by the impressive computational performance of primal-dual interior-point methods for linear programming (see for example Lustig, Marsten, and Shanno (Ref. 1)), it is natural that researchers have directed their attention to the, generally more difficult, area of nonlinear

[1]The authors would like to thank Sandra Santos for painstakingly proofreading an earlier version of this paper.

[2]Department of Mathematics, Faculty of Science, Alexandria University, Alexandria, Egypt and the Center for Research on Parallel Computations, Rice University, Houston, Texas, 77521-1892(elbakry@rice.edu). This author was supported in part by NSF Coop. Agr. No. CCR-8809615 and AFOSR 89-0363 and the REDI Foundation.

[3]Department of Computational and Applied Mathematics and the Center for Research on Parallel Computations, Rice University, Houston, Texas, 77521-1892(rat@rice.edu). This author was supported in part by NSF Coop. Agr. No. CCR-8809615, AFOSR 89-0363, DOE DEFG05-86ER25017 and ARO 9DAAL03-90-G-0093.

[4]The Institute of Statistical Mathematics, 4–6–7 Minami-Azabu, Minato-Ku, Tokyo, 106, Japan.

[5]Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21228, and visiting member of the Center for Research on Parallel Computations, Rice University, Houston, Texas, 77521-1892(yzhang@math.umbc.edu). This author was supported in part by NSF DMS-9102761 and DOE DE-FG02-93ER25171.

programming. Recently there has been considerable activity in the area of interior-point methods for quadratic and convex programming. We shall not attempt to list these research efforts, and restrict our attention to interior-point methods for nonconvex programming. In the area of barrier methods we mention M. Wright (Ref. 2) and Nash and Sofer (Ref. 3). S. Wright (Ref. 4) considers the monotone nonlinear complementarity problem and Monteiro, Pang, and Wang (Ref. 5) consider the nonmonotone nonlinear complementarity problem. S. Wright (Ref. 6) considered the linearly constrained nonlinear programming problem. Lasdon, Yu, and Plummer (Ref. 7) considered various interior-point method formulations for the general nonlinear programming problem. An algorithm and corresponding theory was given by Yamashita (Ref. 8). Other work in the area of interior-point methods for nonlinear programming include McCormick (Ref. 9), Anstreicher and Vial (Ref. 10), Kojima, Megiddo, and Noma (Ref. 11), and Monteiro and Wright (Ref. 12).

The primary objective of this paper is to carry over from linear programming a viable formulation of an interior-point method for the general nonlinear programming problem. In order to accomplish this objective, we first study in extensive detail the formulation of the highly successful Kojima-Mizuno-Yoshise (Ref. 13) primal-dual interior-point method for linear programming. It has been our basic perception that the fundamental ingredient in this formulation is the perturbed Karush-Kuhn-Tucker conditions and the relationship between these conditions and logarithmic barrier function method has not been clearly delineated. Hence Sections 2-4 are devoted to this concern. Of particular interest in this context is Proposition 2.3 which shows that Newton's method applied to the Karush-Kuhn-Tucker conditions for the logarithmic barrier function formulation of the primal linear program and Newton's method applied to the perturbed Karush-Kuhn-Tucker conditions (i.e. the Kojima-Mizuno-Yoshise primal-dual method) never coincide.

In Section 4 we state what we consider to be a basic formulation of an interior-point method for the general nonlinear programming problem. The viability of this formulation is reinforced by the local theory developed in Section 5. Here we demonstrate that local, superlinear, and quadratic convergence can all be obtained for the interior-point method, under exactly the conditions needed for the standard Newton's method theory. The global convergence theory is the subject of Section 6. In Section 7 we present some preliminary numerical experimentation using the 2-norm of the residual as our merit function. Finally in Section 8 we give some concluding remarks.

The choice of merit function for interior-point methods is not a focus of the current research. Such activity is of importance and merits further investigation. Our globalization theory conveniently and effectively uses the 2-norm of the residual as merit function. At the very least it can be viewed as a demonstration of the viability of such theory for general interior-point methods.

## 2 Interpretation of the LP Formulation

Consider the primal linear program in the standard form

$$\begin{array}{ll} \textit{minimize} & c^T x \\ \textit{subject to} & Ax = b \\ & x \geq 0 \end{array} \tag{1}$$

where $c, x \in \mathbf{R}^n$, $b \in \mathbf{R}^m$, and $A \in \mathbf{R}^{m \times n}$. The dual linear program can be written

$$\begin{array}{ll} \textit{maximize} & b^T y \\ \textit{subject to} & A^T y + z = c \\ & z \geq 0 \end{array} \tag{2}$$

and $z \in \mathbf{R}^n$ is called the vector of dual slack variables.

**Basic Assumption:** The matrix $A$ has full rank.

As is done in this area, we use $X$ to denote the diagonal matrix with diagonal $x$ and employ analogous notation for other quantities. Also $e$ is a vector of all ones whose dimension will vary with the context.

A point $x \in \mathbf{R}^n$ is said to be strictly feasible for problem (1) if it is both feasible and positive. A point $z \in \mathbf{R}^n$ is said to be feasible for problem (2) if there exists $y \in \mathbf{R}^m$ such that $(y, z)$ is feasible for problem (2). Moreover, $z$ (or $(y, z)$) is said to be strictly feasible (for problem (2)) if it is feasible and $z$ is positive. A pair $(x, z)$ is said to be on the *central path* (at $\mu > 0$) if $x_i z_i = \mu$ for all $i$, and $x$ is feasible for problem (1), and $z$ is feasible for problem (2). We also say that $x$ is on the central path (at $\mu > 0$) if $(x, \mu X^{-1} e)$ is on the central path, i.e., if $\mu X^{-1} e$ is feasible for problem (2).

The first-order or Karush-Kuhn-Tucker (KKT) optimality conditions for problem (2) are

$$F(x, y, z) \equiv \begin{pmatrix} Ax - b \\ A^T y + z - c \\ XZe \end{pmatrix} = 0, \quad (x, z) \geq 0. \tag{3}$$

By the perturbed KKT conditions for problem (1) we mean

$$F_\mu(x, y, z) \equiv \begin{pmatrix} Ax - b \\ A^T y + z - c \\ XZe - \mu e \end{pmatrix} = 0, \quad (x, z) > 0, \quad \mu > 0. \tag{4}$$

Observe that the perturbation is made only to the complementarity equation. Fiacco and McCormick (Ref. 14) were probably the first to consider the perturbed KKT conditions. They did so in the context of the general inequality constrained nonlinear programming problem. They made several key observations including the fact that the sufficiency conditions for the unconstrained minimization of the logarithmic barrier function were implied locally by the perturbed KKT conditions and the standard second-order sufficiency conditions.

In 1987 Kojima, Mizuno, and Yoshise (Ref. 13) proposed the now celebrated primal-dual interior-point method for linear programming. In essence, their algorithm is damped Newton applied to the perturbed KKT conditions (4). These authors state that their algorithm is based on Megiddo's (Ref. 15) work concerning the classical logarithmic barrier function method. This pioneering work of Kojima, Mizuno, and Yoshise has motivated considerable research activity in the general area of primal-dual interior-point methods for linear programming, quadratic programming, convex programming, linear complementarity problems, and some activity in general nonlinear programming. However, the relationship between the perturbed KKT conditions and the logarithmic barrier function problem seems not to have been well articulated and is often misstated. Therefore, we will rigorously pursue a study of this relationship.

Our intention is to demonstrate the following. While the perturbed KKT conditions are in an obvious sense equivalent to the KKT conditions for the logarithmic barrier function problem, they are not the KKT conditions for this problem or for any other problem. Furthermore, the primal-dual Newton interior-point method cannot be viewed as Newton's method applied to the KKT conditions for the logarithmic barrier function problem; indeed these latter iterates and the former iterates never coincide. Towards this end we begin by considering the logarithmic barrier function problem associated with problem (1)

$$\begin{array}{ll} minimize & c^T x - \mu \sum_{i=1}^{n} log(x_i) \\ subject\ to & Ax = b \\ & x > 0 \end{array} \tag{5}$$

for a fixed $\mu > 0$. The KKT conditions for problem (5) are

$$\hat{F}_\mu(x,y) \equiv \begin{pmatrix} A^T y + \mu X^{-1} e - c \\ Ax - b \end{pmatrix} = 0, \quad x > 0 \tag{6}$$

**Proposition 2.1** *The perturbed KKT conditions for problem (1) given by (4), and the KKT conditions for the logarithmic barrier function problem (5) given by (6) are equivalent in the sense that they have the same solutions, i.e.,* $\hat{F}_\mu(x,y) = 0$ *if and only if* $F_\mu(x,y,\mu X^{-1}e) = 0$.

**Proof:** The proof is straightforward. □

In spite of the equivalence described in Proposition 2.1, we have the following anomaly.

**Proposition 2.2** *The perturbed KKT conditions for problem (1), i.e.,* $F_\mu(x,y,z) = 0$, *or any permutation of these equations, are not the KKT conditions for the logarithmic barrier function problem (5) or any other (smooth) unconstrained or equality constrained optimization problem.*

**Proof:** If $F_\mu(x,y,z) = 0$ were the KKT conditions for some equality constrained optimization problem we would have that there exists a Lagrangian function $L$ such that

$$\nabla L(x,y,z) = F_\mu(x,y,z).$$

It would then follow that

$$\nabla^2 L(x,y,z) = F'_\mu(x,y,z).$$

However $\nabla^2 L(x,y,z)$ is a Hessian matrix and is therefore symmetric. But direct calculations show that $F'_\mu(x,y,z)$ or any permutations of its rows is not symmetric. This argument also excludes unconstrained optimization problems. $\square$

We tacitly assumed that $L(x,y,z)$ in the proof of Proposition 2.2 was of class $C^2$.

Observe that the perturbed KKT conditions (4) are obtained from (6), the KKT conditions for the logarithmic barrier function problem, by introducing the auxiliary variables

$$z = \mu X^{-1} e$$

and then expressing these nonlinear defining relations in the form

$$XZe = \mu e.$$

Considerably more will be said about this nonlinear transformation in Section 3. We now demonstrate exactly how nonlinear this transformation is by showing that the equivalence depicted in Proposition 2.1 in no way carries over to a Newton algorithmic equivalence. Certainly, the possibility of such an equivalence is not precluded by Proposition 2.2 alone.

**Proposition 2.3** *Consider a triple $(x,y,z)$ such that $x$ is strictly feasible for problem (1) and $(y,z)$ is strictly feasible for problem (2). Let $(\Delta x, \Delta y, \Delta z)$ denote the Newton correction at $(x,y,z)$ obtained from the nonlinear system $F_\mu(x,y,z) = 0$ given by (4). Also let $(\Delta x', \Delta y')$ denote the Newton correction at $(x,y)$ obtained from the nonlinear system $\hat{F}_\mu(x,y) = 0$ given by (6). Then the following are equivalent:*

1. $(\Delta x, \Delta y) = (\Delta x', \Delta y')$
2. $\Delta x = 0$
3. $\Delta x' = 0$
4. *$x$ is on the central path at $\mu$*

**Proof:** The two Newton systems that we are concerned with are

$$\begin{aligned} A^T\Delta y' - \mu X^{-2}\Delta x' &= -A^T y - \mu X^{-1}e + c \\ A\Delta x' &= 0 \end{aligned} \tag{7}$$

and

$$\begin{aligned} A^T\Delta y + \Delta z &= 0 \\ A\Delta x &= 0 \\ Z\Delta x + X\Delta z &= -XZe + \mu e \end{aligned} \tag{8}$$

These two linear systems have unique solutions under the assumptions that $(x, z) > 0$ and the matrix $A$ has full rank. We briefly outline a proof for (7). A proof for (8) is only slightly more difficult. Consider the homogeneous system

$$\begin{aligned} A^T\Delta y' - \mu X^{-2}\Delta x' = 0 \\ A\Delta x' = 0 \end{aligned} \tag{9}$$

If we multiply the first equation of (9) by $AX^2$ and use the second equation we obtain

$$(AX^2A^T)\Delta y' = 0. \tag{10}$$

Moreover $AX^2A^T$ is invertible under our assumptions. Hence $\Delta y' = 0$ and therefore from (9) $\Delta x' = 0$. This implies that our system has a unique solution.

Proof of $1 \Rightarrow 2$

Solving the last equation of (8) for $\Delta z$, substituting in the first, and observing that by feasibility $z = c - A^T y$ leads to

$$A^T\Delta y - X^{-1}Z\Delta x = -A^T y - \mu X^{-1}e + c.$$

Comparing the last equation with the first in (7) gives

$$XZ\Delta x = \mu\Delta x'. \tag{11}$$

From the first two equations in (8) we see that $\Delta x^T\Delta z = 0$, i.e.,

$$\Delta x_1\Delta z_1 + \ldots + \Delta x_n\Delta z_n = 0. \tag{12}$$

Define a subset $I$ of $\{1, \ldots, n\}$ as follows. The index $i \in I$ if and only if $\Delta x_i \neq 0$. Now by way of contradiction suppose that $I$ is not empty. From the last equation in (8) and (11) we have that

$$z_i\Delta x_i + x_i\Delta z_i = 0 \quad \text{for} \quad i \in I.$$

Since $z_i > 0$ and $x_i > 0$, the last equation implies that $\Delta x_i$ and $\Delta z_i$ are both not zero and are of opposite sign. However, this contradicts (12). This is the contradiction that we were searching for and we may now conclude that I is empty. Hence $\Delta x = 0$ and we have shown that $1 \Rightarrow 2$.

Proof of 2 $\Rightarrow$ 3
Suppose that $\Delta x = 0$. Then from the first and third equation in (8) we see that

$$A^T \Delta y = z - \mu X^{-1} e.$$

Hence $(0, \Delta y)$ also solves (7).

Proof of 3 $\Rightarrow$ 4
If $\Delta x' = 0$, then from the first equation in (7)

$$A^T (y + \Delta y') + \mu X^{-1} e - c = 0.$$

Therefore $\mu X^{-1} e$ is strictly feasible for problem (2). This says that $x$ is on the central path at $\mu$.

Proof of 4 $\Rightarrow$ 1
Suppose that $x$ is on the central path. This means that $\mu X^{-1} e$ is feasible for problem (2), i.e., there exists $\hat{y}$ such that $(\hat{y}, \mu X^{-1} e)$ is feasible for problem (2). It follows that $(0, \hat{y} - y)$ solves (7). Also $(0, \hat{y} - y, \mu X^{-1} e - z)$ solves (8). Consequently $(\Delta x', \Delta y') = (\Delta x, \Delta y)$ and we have established that 4 $\Rightarrow$ 1, and finally the proposition. $\square$

**Remark** Proposition 2.3 is extremely restrictive. It is incorrect to interpret it as saying that the two Newton iterates agree only if the current $x$ is on the central path. It says that these iterates agree if and only if there is no movement in $x$. This characterizes the redundant situation when $x$ is on the central path at $\mu$ and we are trying to find an $x$ which is on the central path at $\mu$. If $x$ is on the central path at $\mu$ and we are trying to find a point on the central path at $\hat{\mu} \neq \mu$, then the two Newton iterates will not generate $(\Delta x, \Delta y) = (\Delta x', \Delta y')$. Simply stated, the two Newton iterates never coincide.

# 3 Interpretation of the Perturbed KKT Conditions

There is a philosophical parallel between the modification of the penalty function method that leads to the multiplier method and the modification of the KKT conditions for the logarithmic barrier function problem that leads to the perturbed KKT conditions. The similarity is that both modifications introduce an auxiliary variable to serve as approximation to the multiplier vector and use this as a vehicle for removing inherent ill-conditioning from the formulation. However, the roles that the two auxiliary variables play in the removal of inherent ill-conditioning are quite different. We believe that this parallel adds prospective to the role of the perturbed KKT conditions and therefore pursue it in some detail. The following comments are an attempt to shed understanding on the perturbed KKT conditions and are not intended to be viewed as mathematical theory.

For the sake of simplicity our constrained problems will have only one constraint. And for the sake of illustration the multiplier associated with this constraint will be nonzero at the solution. The amount of smoothness required is not an issue and all functions will be as smooth as the context requires.

Consider the equality constrained optimization problem.

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & h(x) = 0 \end{array} \tag{13}$$

where $f, h : \mathbf{R}^n \to \mathbf{R}$. The KKT conditions for problem (13) are

$$\begin{aligned} \nabla f(x) + \lambda \nabla h(x) &= 0, \\ h(x) &= 0. \end{aligned} \tag{14}$$

The $\ell_2$-penalty function associated with problem (13) is

$$P(x;\rho) = f(x) + \frac{\rho}{2} h(x)^T h(x). \tag{15}$$

The gradient of $P$ is given by

$$\nabla P(x;\rho) = \nabla f(x) + \rho h(x) \nabla h(x) \tag{16}$$

and the Hessian of $P$ is given by

$$\nabla^2 P(x;\rho) = \nabla^2 f(x) + \rho h(x) \nabla^2 h(x) + \rho \nabla h(x) \nabla h(x)^T.$$

The *penalty function method* consists of the generation of the sequence $\{x_k\}$ defined by

$$x_k = \arg\min P(x;\rho_k).$$

Suppose that $x_k \to x^*$, a solution of (13), and let $\lambda^*$ be the associated multiplier. Then we must have $\rho_k h(x_k) \to \lambda^*$. Since $h(x_k) \to 0$, and we are assuming that $\lambda^* \neq 0$, necessarily $\rho_k \to +\infty$. However as $\rho_k \to +\infty$ the conditioning of the Hessian matrix $\nabla^2 P(x_k;\rho_k)$ becomes arbitrarily bad. The problem here is that we are asking too much from the penalty parameter $\rho_k$. We are asking it to contribute to good global behavior by penalizing constraint violation and we are asking it to contribute to good local behavior by forcing $\rho_k h(x_k)$ to approximate the multiplier. Hestenes (Ref. 16) in 1969 proposed a way of circumventing the conditioning deficiency. He introduced an auxiliary variable $\lambda$ and replaced $\rho h(x)$ in (16) with $\lambda + \rho h(x)$. This modification effectively converts the penalty function into the augmented Lagrangian. The role of the auxiliary variable $\lambda$ estimating the multiplier was relegated to that of parameter in that $\lambda_k$ was held fixed during a minimization phase of the augmented Lagrangian for the determination of $x_k$ and then updated according to the formula $\lambda_{k+1} = \lambda_k + \rho_k h(x_k)$. In this way the role of $\rho_k h(x_k)$ is no longer one of estimating the multiplier, but one of estimating the correction to the multiplier. Hence it is most appropriate for $\rho_k h(x_k) \to 0$ and the requirement that $\rho_k \to +\infty$ is no longer necessary. The multiplier method has enjoyed considerable success in the computational sciences marketplace.

Now, consider the inequality constrained optimization problem

$$\begin{array}{ll} minimize & f(x) \\ subject\ to & g(x) \geq 0 \end{array} \tag{17}$$

where $f, g : \mathbf{R}^n \to \mathbf{R}$. The KKT conditions for this problem are

$$\begin{aligned} \nabla f(x) - z\nabla g(x) &= 0 \\ zg(x) &= 0 \\ g(x) &\geq 0 \\ z &\geq 0. \end{aligned} \tag{18}$$

The logarithmic barrier function associated with problem (17) is

$$B(x;\mu) = f(x) - \mu\ \log(g(x)), \quad \mu > 0. \tag{19}$$

The gradient of $B$ is given by

$$\nabla B(x;\mu) = \nabla f(x) - \frac{\mu}{g(x)}\nabla g(x);$$

and the Hessian of $B$ is given by

$$\nabla^2 B(x;\mu) = \nabla^2 f(x) - \frac{\mu}{g(x)}\nabla^2 g(x) + \frac{\mu}{g(x)^2}\nabla g(x)\nabla g(x)^T.$$

The logarithmic barrier function method consists of generating a sequence of iterates $\{x_k\}$ as solutions of the essentially unconstrained problem

$$\begin{array}{ll} minimize & B(x;\mu_k) \\ subject\ to & g(x) > 0. \end{array} \tag{20}$$

Suppose that the constraint $g(x)$ is binding at a solution $x^*$ of problem (17). As before we see that convergence of $\{x_k\}$ to $x^*$ requires that $\frac{\mu_k}{g(x_k)} \to z^*$, where $z^*$ is the multiplier associated with the solution $x^*$. Since $\frac{\mu_k}{g(x_k)} \to z^*$ and $g(x_k) \to 0$ we see that $\frac{\mu_k}{g(x_k)^2} \to +\infty$ and the Hessian of the logarithmic barrier function becomes arbitrarily badly conditioned. As in the case of the penalty function method we are asking the penalty parameter sequence (barrier parameter sequence in this case) to do too much and the price once again is inherent ill-conditioning. Now introduce the auxiliary variable $z = \frac{\mu}{g(x)}$ and write this defining relationship in the benign form $zg(x) = \mu$, so that differentiation will not create ill-conditioning.

In this fashion the KKT conditions for the logarithmic barrier function problem (20); namely

$$\begin{aligned} \nabla f(x) - \tfrac{\mu}{g(x)}\nabla g(x) &= 0 \\ g(x) &> 0 \end{aligned}$$

are transformed into the perturbed KKT conditions

$$\begin{aligned} \nabla f(x) - z\nabla g(x) &= 0 \\ zg(x) &= \mu \\ g(x) &> 0 \end{aligned}$$

as proposed and discussed in Fiacco and McCormick (Ref. 14).

We now summarize. In the penalty function method the quantity $\rho h(x)$ must approximate the multiplier, necessitating $\rho_k \to +\infty$. Hence the derivative of $\rho h(x)$ becomes arbitrarily large leading to arbitrarily bad conditioning of the Hessian matrix. On the other hand in the logarithmic barrier function method the quantity $\mu/g(x)$ must approximate the multiplier. Hence $\mu$ cannot go to zero too fast and the derivative of $\mu/g(x)$ becomes arbitrarily large leading to arbitrarily bad conditioning of the Hessian matrix. In the former case the difficulty arises from the fact that $\rho \to +\infty$. The introduction of the auxiliary variable $\lambda$ in the multiplier method allows one to remove this requirement; hence the removal of ill-conditioning. In the latter case the difficulty arises from the differentiation of the functional form $\mu/g(x)$. The introduction of the auxiliary variable $z$ allows one to change the functional form so that differentiation no longer leads to ill-conditioning. Hence, while there is certainly a philosophical similarity between the two approaches, there is no doubt that the latter is more satisfying and mathematically more elegant. While this transformation seems rather straightforward, we stress that it leads to significant changes, i.e. the removal of ill-conditioning and the effect of Proposition 2.3. The main point of the current discussion is to focus on the similarity between the multiplier methods as a vehicle for removing inherent ill-conditioning from the penalty function method and the perturbed KKT conditions as a vehicle for removing inherent ill-conditioning from the logarithmic barrier function problem. The extent to which ill-conditioning is reflected in computation is not a discussion issue here.

It is perhaps of interest to point out that the auxiliary variable $z$ estimating the multiplier can be introduced in a logical fashion from a logarithmic barrier function formulation. Towards this end consider the slack variable form of problem (17)

$$\begin{aligned} &minimize && f(x) \\ &subject\ to && g(x) - s = 0 \\ & && s \geq 0. \end{aligned} \tag{21}$$

The KKT conditions for this problem are

$$\begin{aligned}
\nabla f(x) - z\nabla g(x) &= 0\\
z - w &= 0\\
g(x) - s &= 0\\
ws &= 0\\
(w, s) &\geq 0.
\end{aligned} \tag{22}$$

The system (22) is equivalent, and Newton algorithmically equivalent, to the system

$$\begin{aligned}
\nabla f(x) - z\nabla g(x) &= 0\\
g(x) - s &= 0\\
zs &= 0\\
(s, z) &\geq 0.
\end{aligned} \tag{23}$$

The logarithmic barrier function problem for (21) is

$$\begin{array}{ll}
\textit{minimize} & f(x) - \mu\ \log(s)\\
\textit{subject to} & g(x) - s = 0\\
 & (s > 0).
\end{array} \tag{24}$$

The KKT conditions for (24) are

$$\begin{aligned}
\nabla f(x) - z\nabla g(x) &= 0\\
z - \tfrac{\mu}{s} &= 0\\
g(x) - s &= 0\\
s &> 0\\
z &\geq 0
\end{aligned} \tag{25}$$

By writing $z - \mu/s = 0$ as $sz = \mu$ in (25) we arrive at the perturbed version of the KKT conditions (23). Once more we stress that such a transformation gives an equivalent problem, removes inherent ill-conditioning, but does not preserve Newton algorithmic equivalence (see Proposition 2.3). What we have witnessed here is the following. The pure logarithmic barrier function method deals with an unconstrained problem. Hence there are no multipliers in the formulation. However, if we first add nonnegativity slack variables, then the logarithmic barrier function problem is an equality constrained problem and therefore the corresponding first-order conditions involve multipliers.

We now briefly motivate the perturbed KKT conditions in a manner that has nothing to do with the logarithmic barrier function. Consider the complementarity equation for problem (1)

$$XZe = 0. \tag{26}$$

In any Newton's method formulation we deal with linearized complementarity

$$Z\Delta x + X\Delta z = -XZe. \tag{27}$$

Linearized complementarity leads to several remarkable algorithmic properties. This was observed by Tapia (Ref. 17) in 1980 for the general nonlinear programming problem and was developed and expounded by El-Bakry, Tapia, and Zhang (Ref. 18) for the application of the primal-dual interior-point methods to linear programming. In spite of its local strengths, globally, linearized complementarity has a serious flaw. It forces iterates to stick to the boundary of the feasible region once they approach that boundary. That is, if a component $[x_k]_i$ of a current iterate becomes zero and $[z_k]_i > 0$, then from the linearized complementarity equation (27) we see that $[x_l]_i = 0$ for all $l > k$, i.e., this component will remain zero in all future iterations. The analogous situation is true for the $z$ variable. Such an undesirable attribute clearly precludes the global convergence of the algorithm. An obvious correction is to modify the Newton formulation so that zero variables can become nonzero in subsequent iterations. This can be accomplished by replacing the complementarity equation $XZe = 0$ with perturbed complementarity $XZe = \mu e$ ($\mu > 0$). Of course this is exactly the introduction of the notion of adherence to the central path. It is known that such adherence tends to keep the iterates away from the boundary and promotes the global convergence of the Newton interior-point method. It is this central path interpretation that we feel best motivates the perturbed KKT conditions.

# 4 Nonlinear Programming Formulation

In this section we formulate the primal-dual Newton interior-point method for the general nonlinear programming problem. Our approach will be to consider damped Newton applied to the perturbed KKT conditions. In order to fully imitate the formulation used in the linear programming case we will transform inequalities into equalities by adjoining nonnegative slack variables.

Consider the general nonlinear programming problem

$$\begin{array}{ll} minimize & f(x) \\ subject\ to & h(x) = 0 \\ & g(x) \geq 0 \end{array} \tag{28}$$

where $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^m$ ($m < n$), and $g : \mathbf{R}^n \to \mathbf{R}^p$. The Lagrangian associated with problem (28) is

$$L(x, y, z) = f(x) + y^T h(x) - z^T g(x).$$

If $x$ is feasible for problem (28), then we let $\mathcal{B}(x)$ denote the set of indices of binding inequality constraints at $x$, i.e.,

$$\mathcal{B}(x) = \{i : g_i(x) = 0, i = 1, \ldots, p\}.$$

The KKT conditions for problem (28) are

$$
\begin{aligned}
\nabla_x L(x,y,z) &= 0 \\
h(x) &= 0 \\
g(x) &\geq 0 \\
Zg(x) &= 0 \\
z &\geq 0 \ ,
\end{aligned}
\tag{29}
$$

where $\nabla_x L(x,y,z) = \nabla f(x) + \nabla h(x) y - \nabla g(x) z$.

The standard Newton's method assumptions for problem (28) are

**A4.1:** (Existence) There exists $(x^*, y^*, z^*)$, solution to problem (28) and associated multipliers, satisfying the KKT conditions (29).

**A4.2:** (Smoothness) The Hessian matrices $\nabla^2 f(x), \nabla^2 h_i(x), \nabla^2 g_i(x)$ for all $i$ exist and are locally Lipschitz continuous at $x^*$.

**A4.3:** (Regularity) The set $\{\nabla h_1(x^*), \ldots, \nabla h_m(x^*)\} \cup \{\nabla g_i(x^*) : i \in \mathcal{B}(x^*)\}$ is linearly independent.

**A4.4:** (Second-order Sufficiency) For all $\eta \neq 0$ satisfying $\nabla h_i(x^*)^T \eta = 0, \ i = 1, \ldots, m$ and $\nabla g_i(x^*)^T \eta = 0, \ i \in \mathcal{B}(x^*)$, we have $\eta^T {\nabla_x}^2 L(x^*) \eta > 0$.

**A4.5:** (Strict Complementarity) For all $i$, $z_i^* + g_i(x^*) > 0$.

The KKT conditions (29) can be written in slack variable form as

$$
F(x,y,s,z) \equiv \begin{pmatrix} \nabla_x L(x,y,z) \\ h(x) \\ g(x) - s \\ ZSe \end{pmatrix} = 0, \ \ (s,z) \geq 0. \tag{30}
$$

The following proposition is fundamental to our work

**Proposition 4.1** *Let conditions A4.1 and A4.2 hold. Also let $s^* = g(x^*)$. The following statements are equivalent:*

1. *Conditions A4.3-A4.5 also hold.*
2. *The Jacobian matrix $F'(x^*, y^*, s^*, z^*)$ of $F(x,y,s,z)$ in (30) is nonsingular.*

**Proof:** Such an equivalence is reasonably well-known for the equality constrained optimization problem. Hence we base our proof on that equivalence. To begin with observe that

$$F'(x^*, y^*, s^*, z^*) = \begin{pmatrix} \nabla_x^2 L_* & \nabla h(x^*) & -\nabla g(x^*) & 0 \\ \nabla h(x^*)^T & 0 & 0 & 0 \\ \nabla g(x^*)^T & 0 & 0 & -I \\ 0 & 0 & S^* & Z^* \end{pmatrix}, \tag{31}$$

where $\nabla_x^2 L_* = \nabla_x^2 L(y^*, x^*, z^*)$. Consider the equality constrained optimization problem

$$\begin{array}{ll} minimize & f(x) \\ subject\ to & h(x) = 0 \\ & g_i(x) = 0, \quad i \in \mathcal{B}(x^*) . \end{array}$$

Observe that the regularity condition A4.3 is regularity for this problem and the second-order sufficiency condition A4.4 is second-order sufficiency for this problem. Hence from the theory of equality constrained optimization we see that A4.3 and A.4.4 are equivalent to the nonsingularity of the matrix

$$\hat{F}'(x^*, y^*, s^*, z^*) = \begin{pmatrix} \nabla_x^2 L_* & \nabla h(x^*) & -\nabla \hat{g}(x^*) \\ \nabla h(x^*)^T & 0 & 0 \\ \nabla \hat{g}(x^*)^T & 0 & 0 \end{pmatrix}, \tag{32}$$

where $\nabla \hat{g}(x^*)$ is the matrix whose columns are $\{\nabla g_i(x^*) : i \in \mathcal{B}(x^*)\}$. It is not difficult to see that the nonsingularity of (31) is equivalent to strict complementarity A4.5 and the nonsingularity of $\hat{F}'(x^*, y^*, s^*, z^*)$ . □

We loose a small amount of flexibility by adding slack variables to the KKT conditions (29) and then working with the resulting system (30), instead of adding slack variables directly to the optimization problem (28) and then working with the resulting KKT conditions. This small observation is quite subtle; but will play a role in the formulation of our interior-point method. Hence we now pursue it in some detail.

Consider the following equivalent slack variable form of problem (28)

$$\begin{array}{ll} minimize & f(x) \\ subject\ to & h(x) = 0 \\ & g(x) - s = 0 \\ & s \geq 0 . \end{array} \tag{33}$$

The KKT conditions for problem (33) are

$$
\begin{aligned}
\nabla f(x) + \nabla h(x) y - \nabla g(x) w &= 0 \\
w - z &= 0 \\
h(x) &= 0 \\
g(x) - s &= 0 \\
ZSe &= 0 \\
(s, z) &\geq 0.
\end{aligned}
\tag{34}
$$

The equation $w - z = 0$ in (34) says that at the solution the multipliers associated with the equality constraints $g(x) - s = 0$ are the same as the multipliers corresponding to the inequality constraints $s \geq 0$. Moreover, due to the linearity of this equation, the Newton corrections $\Delta w$ and $\Delta z$ will also be the same. However the damped Newton step $w + \alpha_w \Delta w$ and the damped Newton step $z + \alpha_z \Delta z$ will be the same if and only if $\alpha_w = \alpha_z$ (assuming $\Delta w$ and $\Delta z$ are not both zero). We have learned from numerical experimentation that there is value in taking different steplengths for the $w$ and $z$ variables. Hence our interior-point method will be based on (34). In particular we base our algorithm on the perturbed KKT conditions

$$
F_\mu(x, y, s, w, z) = \begin{pmatrix} \nabla f(x) + \nabla h(x) y - \nabla g(x) w \\ w - z \\ h(x) \\ g(x) - s \\ ZSe - \mu e \end{pmatrix} = 0, \quad (s, w, z) \geq 0 \; . \tag{35}
$$

Proposition 4.1 readily extends to $F'_\mu(x, y, s, w, z)$.

We now describe our primal-dual Newton interior-point method for the general nonlinear optimization problem (28). At the $k^{\underline{th}}$ iteration, let $v_k = (x_k, y_k, s_k, w_k, z_k)$. We obtain our perturbed Newton correction corresponding to the parameter $\mu_k$, $\Delta v_k = (\Delta x_k, \Delta y_k, \Delta s_k, \Delta w_k, \Delta z_k)$, as the solution of the perturbed Newton linear system

$$
F'_\mu(v_k) \Delta v = -F_\mu(v_k). \tag{36}
$$

We allow the flexibility of choosing different steplengths for the various components of $v_k$. If our choice of steplengths are $\alpha_x, \alpha_y, \alpha_s, \alpha_w$, and $\alpha_z$, we construct the expanded vector of steplengths

$$
\alpha_k = (\alpha_x, \ldots, \alpha_x, \alpha_y, \ldots, \alpha_y, \alpha_s, \ldots, \alpha_s, \alpha_w, \ldots, \alpha_w, \alpha_z, \ldots, \alpha_z),
$$

where the frequencies of occurances are $n, m, p, p$, and $p$ respectively. Now we let

$$
\Lambda_k = diag(\alpha_k), \tag{37}
$$

and the subsequent iterate $v_{k+1}$ can be written as

$$v_{k+1} = v_k + \Lambda_k \Delta v.$$

Now we are ready to state our generic primal-dual Newton interior-point method for the general nonlinear optimization problem (28). For global convergence consideration a merit function $\phi(v)$, that measures the progress towards the solution $v^* = (x^*, y^*, s^*, s^*, z^*)$, should be used.

---

**Algorithm 1 (Interior-Point Algorithm)**

*Let $v_0 = (x_0, y_0, s_0, w_0, z_0)$ be an initial point satisfying $(s_0, w_0, z_0) > 0$. For $k = 0, 1, 2, \ldots$, do*

**(1)** *Test for convergence.*

**(2)** *Choose $\mu_k > 0$.*

**(3)** *Solve the linear system (36) for $\Delta v = (\Delta x, \Delta y, \Delta s, \Delta w, \Delta z)$.*

**(4)** *Compute the quantities*

$$\hat{\alpha}_s = \frac{-1}{\min((S_k)^{-1}\Delta s_k, -1)}$$

$$\hat{\alpha}_w = \frac{-1}{\min((W_k)^{-1}\Delta w_k, -1)}$$

$$\hat{\alpha}_z = \frac{-1}{\min((Z_k)^{-1}\Delta z_k, -1)}$$

**(5)** *Choose $\tau_k \in (0, 1]$ and $\alpha_p \in (0, 1]$ satisfying*

$$\phi(v_k + \Lambda_k \Delta v) \leq \phi(v_k) + \beta \alpha_p \nabla \phi(v_k)^T \Delta v_k, \tag{38}$$

*for some fixed $\beta \in (0, 1)$, where $\Lambda_k$ is described in (37) with the steplength choices*

$$\alpha_x = \alpha_p$$

$$\alpha_y = \alpha_p$$

$$\alpha_s = \min(1, \tau_k \hat{\alpha}_s)$$

$$\alpha_w = \min(1, \tau_k \hat{\alpha}_w)$$

$$\alpha_z = \min(1, \tau_k \hat{\alpha}_z).$$

**(6)** *Set $v_{k+1} = v_k + \Lambda_k \Delta v_k$ and $k \leftarrow k + 1$. Go to Step (1).*

---

If one prefers equal steplengths for the various component functions, then there is no value in carrying $w$ as a separate variable and it should be set equal to $z$. Moreover, in this case the obvious choice for the steplength for the $s$ and $z$ components is

$$\min(1, \tau_k \hat{\alpha}_s, \tau_k \hat{\alpha}_z). \tag{39}$$

It is a straightforward matter to employ backtracking on (39) in order to satisfy the sufficient decrease condition (38). Our local analysis will be given with the steplength choice (39). A reasonable modification of this approach would be to choose $\alpha_k$ via backtracking and then choose $\alpha_p$, the steplength for the $(x, y)-$variables, such that $\alpha_p > \alpha_k$ and the sufficient decrease condition (38) is still maintained.

# 5 Local Convergence Properties

In this section we will demonstrate that our perturbed and damped interior-point Newton's method can be implemented so that the highly desirable properties of the standard Newton's method are retained. We find this demonstration particularly satisfying since it adds credibility to our choice of formulation. The major issue here concerning fast convergence is the same as it was in the linear programming application. There it was dealt with successfully by Zhang, Tapia, and Dennis (Ref. 19), and Zhang and Tapia (Ref. 20). This issue is – Is it possible to choose the algorithmic parameters $\tau_k$ (percentage of movement to the boundary) and $\mu_k$ (perturbation) in such a way that the perturbed and damped step approaches the Newton step sufficiently fast so that quadratic convergence will be retained ?. We stress the point that the choice $\alpha_p = 1$ and $\tau_k = 1$ do not necessarily imply that the steplength $\alpha_k$ is 1.

We begin by giving a formal definition of the perturbed damped Newton's method and then deriving some facts that will be useful concerning the convergence rate of the perturbed damped Newton's method. Towards this end consider the general nonlinear equation problem

$$F(x) = 0, \tag{40}$$

where $F : \mathbf{R}^n \to \mathbf{R}^n$. Recall that the standard Newton's method assumptions for problem (40) are

**A5.1:** There exists $x^* \in \mathbf{R}^n$ such that $F(x^*) = 0$.

**A5.2:** The Jacobian matrix $F'(x^*)$ is nonsingular.

**A5.3:** The Jacobian operator $F'$ is locally Lipschitz continuous at $x^*$.

By the perturbed damped Newton's method for problem (40) we mean the construction of the iteration sequence

$$x_{k+1} = x_k - \alpha_k F'(x_k)^{-1}[F(x_k) - \mu_k \hat{p}], \;\; k = 0, 1, 2, \ldots \tag{41}$$

where $0 < \alpha_k \leq 1$, $\mu_k > 0$, and $\hat{p}$ is a fixed vector in $\mathbf{R}^n$.

**Proposition 5.1** *Consider a sequence $\{x_k\}$ generated by the perturbed damped Newton's method (41) for problem (40). Let $x_k \to x^*$ such that $F(x^*) = 0$ and the standard assumptions A5.1-A5.3 hold at $x^*$.*

1. *If $\alpha_k \to 1$ and $\mu_k = o(\|F(x_k)\|)$, then the sequence $\{x_k\}$ converges to $x^*$ Q-superlinearly.*

2. *If $\alpha_k = 1 + O(\|F(x_k)\|)$ and $\mu_k = O(\|F(x_k)\|^2)$, then the sequence $\{x_k\}$ converges to $x^*$ Q-quadratically.*

**Proof:** Standard Newton's method analysis arguments (see Dennis and Schnabel (Ref. 21) for example) can be used to show that

$$\|x_{k+1} - x^*\| = (1 - \alpha_k)\|x_k - x^*\| + \mu_k\|F'(x_k)^{-1}\hat{p}\| + O(\|x_k - x^*\|^2), \tag{42}$$

and

$$\|F(x_k)\| = O(\|x_k - x^*\|) \tag{43}$$

for all $x_k$ sufficiently near $x^*$. The proof now follows by considering (42) and (43). □

We are now ready to establish convergence rate results for our perturbed damped interior-point Newton's method for problem (40), i.e. Algorithm 1. First we introduce some notation and make several observations. We let $w = z$ and choose the steplength $\alpha_k$ given by (39). Our algorithm is the perturbed damped Newton's method applied to the nonlinear system $F(x, y, s, z) = 0$ given in (34). Observe that the conditions A4.1-A4.5 imply the conditions A5.1-A5.3 according to Proposition 4.1. In the following presentation it will be convenient to write

$$\mu_k = \sigma_k \min(S_k Z_k e)$$

and state our conditions in terms of $\sigma_k$.

**Theorem 5.1 (Convergence Rate)** *Consider a sequence $\{v_k\}$ generated by Algorithm 1. Assume that $\{v_k\}$ converges to a solution $v^*$ such that the standard assumptions A4.1-A4.5 for problem (28) hold at $v^*$.*

1. *If $\tau_k \to 1$ and $\sigma_k \to 0$, then the sequence $\{v_k\}$ converges to $v^*$ Q-superlinearly.*

2. *If $\tau_k = 1 + O(\|F(v_k)\|)$ and $\sigma_k = O(\|F(v_k)\|)$, then the sequence $\{v_k\}$ converges to $v^*$ Q-quadratically.*

**Proof:** The proof of the theorem will follow directly from Proposition 5.1 once we establish that $\alpha_k$ satisfies a relationship of the form

$$\alpha_k = \min(1, \tau_k + O(\sigma_k) + O(\|F(v_k)\|)). \tag{44}$$

We now turn our attention to this task. Since $\Delta v = -F'(v_k)^{-1}(F(v_k) - \mu_k \hat{e})$, where the vector $\hat{e} = (0, \ldots, 0, 1, \ldots, 1)$ with $p$ ones, we see that

$$\|\Delta s_k\| = O(\|F(v_k)\|) + O(\mu_k), \tag{45}$$

and

$$\|\Delta z_k\| = O(\|F(v_k)\|) + O(\mu_k). \tag{46}$$

Hence both $\Delta s_k$ and $\Delta z_k$ converge to zero.

From linearized perturbed complementarity we have

$$S_k{}^{-1}\Delta s + Z_k{}^{-1}\Delta z = -e + \mu_k S_k{}^{-1} Z_k{}^{-1} e. \tag{47}$$

It follows from strict complementarity, (45), (46), and (47) that if $i$ is an index such that $s_i{}^* = 0$, then

$$\frac{[\Delta s_k]_i}{[s_k]_i} = -1 + O(\|F(v_k)\|) + O(\sigma_k),$$

while if it is an index such that $[s^*]_i > 0$, then

$$\frac{[\Delta s_k]_i}{[s_k]_i} \to 0.$$

Similar relationships hold for the $z$-variables. Hence

$$\min(S_k{}^{-1}\Delta s, Z_k{}^{-1}\Delta z) = -1 + O(\|F(v_k)\|) + O(\sigma_k). \tag{48}$$

So

$$\alpha_k = \min(1, \tau_k / (1 + O(\|F(v_k)\|) + O(\sigma_k))). \tag{49}$$

However, if $\alpha_k$ satisfies a relationship of the form (49), then it satisfies a relationship of the form (44). $\square$

**Theorem 5.2 (Local Convergence)** *Consider problem (28) and a solution $v^*$ such that the standard assumptions A4.1-A4.5 hold at $v^*$. Given $\hat{\tau} \in (0,1)$ there exists a neighborhood $D$ of $v^*$ and a constant $\hat{\sigma} > 0$ such that for any $v_0 \in D$ and any choice of algorithmic parameters $\tau_k \in [\hat{\tau}, 1]$ and $\sigma_k \in (0, \hat{\sigma}]$, Algorithm 1 is well defined and the iteration sequence converges Q-linearly to $v^*$.*

**Proof:** We first observe that the estimates constructed in the proof of Proposition 5.1 and Theorem 5.1 above do not depend on the fact that we assumed convergence of the iteration sequence. Clearly they depend strongly on the standard assumptions. By using (42), (43), and (44) we can derive

$$\|v_{k+1} - v^*\| \leq (1 - \tau_k + O(\sigma_k) + O(\|v_k - v^*\|))\|v_k - v^*\|. \tag{50}$$

In (50) we used the fact that

$$\mu_k = \sigma_k O(\|F(v_k)\|) = \sigma_k O(\|v_k - v^*\|).$$

The proof now follows from (50). □

# 6 Global Convergence Theory

In this section we establish a global convergence theory for a primal-dual Newton interior-point algorithm. The algorithm that we consider here has the same basic structure as Algorithm 1 with a particular choice for the merit function $\phi$. The main result is Theorem 6.1 which states that any limit point of the sequence generated by our algorithm is a KKT point of problem (28).

We start by recalling that the slack-variable form of the KKT conditions of problem (28) is

$$F(x, y, s, z) \equiv \begin{pmatrix} \nabla_x L(x, y, s, z) \\ h(x) \\ g(x) - s \\ ZSe \end{pmatrix} = 0, \quad (s, z) \geq 0,$$

which can be written as

$$F(x, y, s, z) \equiv \begin{pmatrix} G(x, y, s, z) \\ ZSe \end{pmatrix} = 0, \quad (s, z) \geq 0,$$

where

$$G(x, y, s, z) \equiv \begin{pmatrix} \nabla_x L(x, y, s, z) \\ h(x) \\ g(x) - s \end{pmatrix}. \tag{51}$$

As before we will use the following notation

$$v = (x, y, s, z).$$

At a current point $v = (x, y, s, z)$ and for a chosen steplength $\alpha$, the subsequent iterate is calculated as

$$v(\alpha) = (x(\alpha), y(\alpha), s(\alpha), z(\alpha)) = (x, y, s, z) + \alpha(\Delta x, \Delta y, \Delta s, \Delta z), \tag{52}$$

where $(\Delta x, \Delta y, \Delta s, \Delta z)$ is the solution of the system

$$F'(v)\Delta v = -F(v) + \mu \hat{e}. \tag{53}$$

To specify the selection of $\alpha$, we first introduce some quantities and functions that we will make use of later. For a given starting point $v_0 = (x_0, y_0, z_0, s_0)$ with $(s_0, z_0) > 0$, let

$$\tau_1 = \min(Z_0 S_0 e)/[(z_0)^T s_0/p], \quad \tau_2 = (z_0)^T s_0/\|G(v_0)\|_2.$$

Define

$$f^I(\alpha) = \min(Z(\alpha)s(\alpha)) - \gamma\tau_1 z(\alpha)^T s(\alpha)/p, \tag{54}$$

and

$$f^{II}(\alpha) = z(\alpha)^T s(\alpha) - \gamma\tau_2\|G(v(\alpha))\|_2, \tag{55}$$

where $\gamma \in (0,1)$ is a constant. We note that the functions $f^i(\alpha)$, $i = I, II$, depend on the iteration count $k$, though for simplicity we choose not to explicitly write out this dependency. It is also worth noting that

**(i)** for $v = v_0$ and $\gamma = 1$, $f^i(0) = 0$ for $i = I, II$;

**(ii)** $f^I(\alpha)$ is a piecewise quadratic and $f^{II}(\alpha)$ is generally nonlinear.

It is known that if $\alpha_k$ are chosen such that $f^I(\alpha) \geq 0$ for all $\alpha \in [0, \alpha_k]$ at every iteration, then $(z_k, s_k) > 0$ and

$$\min(Z_k S_k e)/[(z_k)^T s_k/p] \geq \gamma_k \tau_1,$$

where $\gamma_k \in (0,1)$. This is a familiar centrality condition for interior-point methods.

Based on these observations, in choosing the steplength $\alpha_k$ at every iteration, we will require $\alpha_k$ to satisfy $f^i(\alpha_k) \geq 0$, $i = I, II$, and $f^I(\alpha) \geq 0$ for all $\alpha \in [0, \alpha_k]$.

For $i = I, II$, define

$$\alpha^i = \max_{\alpha \in [0,1]} \{\alpha : f^i(\alpha') \geq 0 \text{ for all } \alpha' \leq \alpha\}, \tag{56}$$

i.e., $\alpha^i$ are either one or the smallest positive root for the functions $f^i(\alpha)$ in $(0, 1]$ (it will be shown later that $\alpha^i > 0$). Since $f^I(\alpha)$ is a piecewise quadratic, $\alpha^I$ is easy to find.

Our globalized algorithm is a perturbed and damped Newton method with a back-tracking linesearch. The merit function used for linesearch is the squared $\ell_2$ norm of the residual, i.e.,

$$\phi(v) = \|F(v)\|_2^2. \tag{57}$$

We use the notation $\phi_k$ to denote the value of the function $\phi(v)$ evaluated at $v_k$. Similar notation will be used for other quantities depending on $v_k$. Moreover, we use $\phi_k(\alpha)$ to denote $\phi(v_k + \alpha\Delta v_k)$. Clearly, $\phi_k = \phi_k(0) = \phi(v_k)$.

It is not difficult to obtain a condition under which the perturbed Newton step

$$\Delta v = -F'_\mu(v_k)^{-1}F_\mu(v_k),$$

gives descent for the merit function $\phi(v)$. The derivative of $\phi_k(\alpha)$ at $\alpha = 0$ is

$$\begin{aligned}(\nabla\phi)^T\Delta v = &\ 2(F'(v)^T F(v))^T[F'(v)^{-1}(-F(v)+\mu\hat{e})] \\ = &\ 2F(v)^T(-F(v)+\mu\hat{e}) \\ = &\ 2(-\|F(v)\|_2^2 + \mu F(v)^T\hat{e}),\end{aligned}$$

hence

$$(\nabla\phi)^T\Delta v < 0 \text{ if and only if } \mu < \|F(v)\|_2^2/s^Tz.$$

Now we describe the globalized primal-dual Newton interior-point algorithm.

---

**Algorithm 2 (Global Algorithm)**
*Choose* $v_0 = (x_0, y_0, s_0, z_0)$ *such that* $(s_0, z_0) > 0$, $\rho \in (0,1)$ *and* $\beta \in (0, 1/2]$. *Set* $k = 0$, $\gamma_{k-1} = 1$, *and compute* $\phi_0 = \phi(v_0)$. *For* $k = 0, 1, 2, \dots,$ *do*

**(1)** *Test for convergence: if* $\phi_k \le \epsilon_{exit}$, *stop.*

**(2)** *Choose* $\sigma_k \in (0,1)$ *and for* $v = v_k$ *compute the perturbed Newton direction* $\Delta v_k$ *from (53) with*

$$\mu_k = \sigma_k \frac{(s_k)^T z_k}{p}.$$

**(3)** *Steplength selection:*

**(3.1)** *Choose* $1/2 \le \gamma_k \le \gamma_{k-1}$, *compute* $\alpha^i$, $i = I, II$, *from (56) and let*

$$\bar{\alpha}_k = \min(\alpha^I, \alpha^{II}). \tag{58}$$

**(3.2)** *Let* $\alpha_k = \rho^t\bar{\alpha}_k$, *where* $t$ *is the smallest nonnegative integer such that* $\alpha_k$ *satisfies*

$$\phi_k(\alpha_k) \le \phi_k(0) + \alpha_k\beta\phi'_k(0). \tag{59}$$

**(4)** *Let* $v_{k+1} = v_k + \alpha_k\Delta v_k$ *and* $k \leftarrow k+1$. *Go to Step (1).*

---

The question as to whether the perturbed Newton direction is a descent direction for the merit function $\phi$ (for the choice of $\mu$ given in Algorithm 2) is answered in the affirmative in the following proposition.

**Proposition 6.1** *The direction $\Delta v_k$ generated by Algorithm 2 is a descent direction for the merit function $\phi(v)$ at $v_k$. Moreover if condition (59) is satisfied, then*

$$\phi_k(\alpha_k) \leq [1 - 2\alpha_k \beta(1 - \sigma_k)]\phi_k(0).$$

**Proof:** We will suppress the subscript $k$ in the proof. Note that

$$\nabla \phi^T \Delta v = -2(\phi - \mu z^T s).$$

Since $\mu = \sigma z^T s/p$ and

$$(z^T s)^2/p = (\|ZSe\|_1/\sqrt{p})^2 \leq \|ZSe\|_2^2 \leq \|G\|_2^2 + \|ZSe\|_2^2 = \phi, \tag{60}$$

it follows that

$$\nabla \phi^T \Delta v \leq -2(1 - \sigma)\phi < 0.$$

So the perturbed Newton direction indeed gives descent. Moreover condition (59) can be written

$$\phi(\alpha) \leq [1 - 2\alpha\beta(1 - \sigma)]\phi(0).$$

This proves the proposition. □

This proposition asserts also that the sequence $\{\phi_k\}$ is monotone and non-increasing, therefore,

$$\phi_k \leq \phi_0 \text{ for all } k.$$

Moreover, we have global Q-linear convergence of the values of the merit function $\phi$ to zero if $\{\alpha_k\}$ is bounded away from zero, and $\sigma_k$ is bounded away from from one. It is also worth noting that the above inequality is equivalent to

$$\frac{\|F(v_{k+1})\|_2}{\|F(v_k)\|_2} \leq [1 - 2\alpha_k \beta(1 - \sigma_k)]^{1/2}.$$

One problem that may preclude global convergence is that the sequence $\{\|Z_k S_k e\|\}$ converges to zero, but $\{\phi(v_k)\}$ does not. The following proposition shows that Step (3.1) in Algorithm 2 plays a key role in preventing such behavior from occurring.

**Proposition 6.2** *Let $\{v_k\}$ be generated by Algorithm 2. Then*

$$\ell\phi(v_k) \leq [(z_k)^T s_k]^2 \leq p\phi(v_k),$$

*where $\ell = [\min(1, 0.5\tau_2)/2]^2$.*

**Proof:** We will again suppress the subscript $k$ in the proof. The second inequality follows from (60). So we only need to prove the first one.

Since $\alpha_k \leq \bar{\alpha}_k$, we have $f^i(\alpha_k) \geq 0$, $i = I, II$. From (55) and the choice $\gamma_k \geq 1/2$,

$$z^T s \geq \frac{1}{2}(\|ZSe\|_2 + 0.5\tau_2\|G\|_2) \geq \frac{1}{2}\min(1, 0.5\tau_2)\|F\|_2.$$

This completes the proof. $\square$

Given $\epsilon \geq 0$, let us define the set

$$\Omega(\epsilon) \equiv \left\{ v : \ \epsilon \leq \phi(v) \leq \phi_0, \ \frac{\min(ZSe)}{z^T s/p} \geq \frac{\tau_1}{2}, \ \frac{z^T s}{\|G(v)\|_2} \geq \frac{\tau_2}{2} \right\}. \tag{61}$$

This set will play a pivotal role in establishing our global convergence theory. For this set, the following observations are in order.

**(a)** $\Omega(\epsilon)$ is a closed set.

**(b)** From the construction of the algorithm, in particular, $\gamma_k \geq 1/2$,

$$\{v_k\} \subset \Omega(0). \tag{62}$$

**(c)** In $\Omega(\epsilon)$ where $\epsilon > 0$, $z^T s$ is bounded above and bounded away from zero.

**(d)** In $\Omega(\epsilon)$ where $\epsilon > 0$, all components of $ZSe$ are bounded above and bounded away from zero.

We will establish global convergence of the algorithm under the following assumptions.

**A6.1:** In the set $\Omega(0)$, the functions $f(x), h(x)$, and $g(x)$ are twice continuously differentiable and the derivative of $G(v)$, given by equation (51), is Lipschitz continuous with constant $L$. Moreover, the columns of $\nabla h(x)$ are linearly independent.

**A6.2:** The iteration sequence $\{x_k\}$ is bounded. (This can be ensured by enforcing box constraints $-M \leq x \leq M$ for sufficiently large $M > 0$).

**A6.3:** The matrix $\nabla_x^2 L(x, y, z) + \nabla g(x) S^{-1} Z \nabla^T g(x)$ is invertible for $v$ in any compact subset of $\Omega(0)$ where $s$ is bounded away from zero.

**A6.4:** Let $I_s^0$ be the index set $\{i : 1 \leq i \leq p, \ \liminf [s_k]_i = 0\}$. Then the set of gradients $\{\nabla h_1(x_k), \ldots, \nabla h_m(x_k), \nabla g_i(x_k), i \in I_s^0\}$ is linearly independent for $k$ sufficiently large.

We note that if we have $g(x_k) - s_k \to 0$ in the algorithm, then Assumption 6.4 is equivalent to the linear independence of the gradients for active constraints, which is a standard regularity assumption in constrained optimization.

**Proposition 6.3** *Assume that condition A6.1 holds. Then for $v \in \Omega(\epsilon)$ and $x$ in a compact set, there exists a positive constant $M_2$ such that,*

$$\|y\| \le M_2(1 + \|z\|).$$

**Proof:** We have $\nabla_x L(x, y, s, z) = \nabla f(x) + \nabla h(x)y - \nabla g(x)z$. Then by Assumption 6.1,

$$y = [\nabla h(x)^T \nabla h(x)]^{-1} \nabla h(x)^T (\nabla_x L(x, y, s, z) - \nabla f(x) + \nabla g(x)z).$$

This implies the proposition. □

In the remaining part of this section, we concentrate our effort on proving the following fact: given any $\epsilon > 0$, as long as the iteration sequence $v_k$ generated by the algorithm satisfies

$$v_k \in \Omega(\epsilon), \quad \epsilon > 0, \tag{63}$$

then the steps $\Delta v_k$ and the steplength $\alpha_k$ are uniformly bounded above and away from zero, respectively, in the algorithm. This fact implies the convergence of the algorithm.

**Lemma 6.1** *If $\{v_k\} \subset \Omega(\epsilon)$, then the iteration sequence $\{v_k\}$ is bounded above and in addition $\{(z_k, s_k)\}$ is component-wise bounded away from zero.*

**Proof:** From Assumption 6.2, $\{x_k\}$ is bounded. By Proposition 6.3, it suffices to prove that $\{(z_k, s_k)\}$ is bounded above and component-wise away from zero.

The boundedness of $\{x_k\}$ in $\Omega(\epsilon)$ implies that $\{\|g(x_k)\|\}$ is bounded above, say, by $M > 0$. Therefore, if follows from the definition of (61) and the fact that $\{\|F(v_k)\|_2\}$ is monotonically decreasing that

$$\|s_k\| \le \|g(x_k) - s_k\| + \|g(x_k)\| \le \sqrt{\phi_0} + M.$$

This proves that $\{s_k\}$ is bounded above.

Since in $\Omega(\epsilon)$, $[z_k]_i[s_k]_i$, $i = 1, 2, \ldots, p$, are all bounded away from zero. Hence all components of $\{z_k\}$ are bounded away from zero because $\{s_k\}$ is bounded above. Moreover, $\{s_k\}$ will be bounded away from zero if $\{z_k\}$ is bounded above. This will be proved next by contradiction.

Suppose that, if necessary considering a subsequence, $[z_k]_i \to \infty$ for $i$ in some index set. Then the boundedness of $[z_k]_i[s_k]_i$ implies that $\liminf [s_k]_i = 0$ and the corresponding index set is $I_s^0$. Since $\|\nabla f(x_k) + \nabla h(x_k)y - \nabla g(x_k)z\|$ is bounded in $\Omega(\epsilon)$, so is $\|\nabla h(x_k)y - \nabla g(x_k)z\|$ because $\|\nabla f(x_k)\|$ is bounded. Since $\|z_k\| \to \infty$,

$$\frac{\|\nabla h(x_k)y - \nabla g(x_k)z\|}{\|(y_k, z_k)\|} \to 0.$$

Let $w^*$ be any limit point of $\{(y_k, z_k)/\|(y_k, z_k)\|\}$. Clearly, $\|w^*\| = 1$, and the components of $w^*$ corresponding to those indices for which $\{[z_i]_k\} < +\infty$, i.e., $i \notin I_s^0$, are zero. Let $\hat{w}^*$ be the vector consisting of the components of $w^*$ but excluding those corresponding to $i \notin I_s^0$. So $\|\hat{w}^*\| = \|w^*\| = 1$. The above relation implies that at least for a subsequence of $\{x_k\}$,

$$[\nabla h(x_k), \nabla g_i(x_k), i \in I_s^0]\hat{w}^* \to 0.$$

This, however, contradicts Assumption A6.4. So $\{z_k\}$ is bounded above and $\{s_k\}$ is bounded away from zero. □

**Lemma 6.2** *If $\{v_k\} \subset \Omega(\epsilon)$, then $\{[F'(v_k)]^{-1}\}$ is bounded.*

**Proof:** For simplicity, we will suppress the arguments and subscripts in this proof. Rearranging the order of rows and columns of $F'(v)$, we have the following matrix.

$$F'(v) = \begin{pmatrix} A & B \\ -B^T & C \end{pmatrix} \equiv \begin{pmatrix} Z & S & 0 & 0 \\ -I & 0 & 0 & \nabla g^T \\ 0 & 0 & 0 & \nabla h^T \\ 0 & -\nabla g & -\nabla h & \nabla_x^2 L \end{pmatrix},$$

where

$$A \equiv \begin{pmatrix} Z & S \\ -I & 0 \end{pmatrix}, \; B \equiv \begin{pmatrix} 0 & 0 \\ 0 & \nabla g^T \end{pmatrix}, \; C \equiv \begin{pmatrix} 0 & \nabla h^T \\ \nabla h & \nabla_x^2 L \end{pmatrix}.$$

From Lemma 6.1

$$A^{-1} = \begin{pmatrix} 0 & -I \\ S^{-1} & S^{-1}Z \end{pmatrix}$$

exists in $\Omega(\epsilon)$ and is uniformly bounded. Furthermore, by Assumptions A6.1, A6.3, and Lemma 6.1 the matrix

$$H \equiv (B^T A^{-1} B + C) = \begin{pmatrix} 0 & \nabla h^T \\ \nabla h & \nabla_x^2 L + \nabla g S^{-1} Z \nabla g^T \end{pmatrix}$$

is invertible and $\|H^{-1}\|$ is uniformly bounded in $\Omega(\epsilon)$.

A straightforward calculation shows that

$$\begin{pmatrix} A & B \\ -B^T & C \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} - A^{-1}BH^{-1}B^TA^{-1} & -A^{-1}BH^{-1} \\ H^{-1}B^TA^{-1} & H^{-1} \end{pmatrix},$$

which is bounded since every matrix involved is bounded. This implies that $(F'(v))^{-1}$ is uniformly bounded in $\Omega(\epsilon)$ and proves the lemma. □

The following corollary follows directly from Lemma 6.2.

**Corollary 6.1** *If $\{v_k\} \subset \Omega(\epsilon)$, then the sequence of search steps $\{\Delta v_k\}$ generated by Algorithm 2 is bounded.*

Now we prove that $\{\bar{\alpha}_k\}$ given by Step (3.1) of Algorithm 2 is bounded away from zero.

**Lemma 6.3** *If $\{v_k\} \subset \Omega(\epsilon)$ and $\{\sigma_k\}$ is bounded away from zero, then $\{\bar{\alpha}_k\}$ is bounded away from zero.*

**Proof:** Let us suppress the subscript $k$. Since $\bar{\alpha} = \min(\alpha^I, \alpha^{II})$, where

$$\alpha^i = \max_{\alpha \in [0,1]} \{\alpha : f^i(\alpha') \geq 0 \text{ for all } \alpha' \leq \alpha\}, \; i = I, II.$$

it suffices to show that $\{\alpha^i\}$, $i = I, II$, are bounded away from zero.

From the definition of $\alpha^I$ and $f^I(\alpha)$, $\alpha^I$ is the largest number in $[0,1]$ such that

$$z_i(\alpha)s_i(\alpha) - \gamma\tau_1 z(\alpha)^T s(\alpha)/p \geq 0, \;\; \alpha \in [0, \alpha^I], \;\; i = 1, 2, \ldots, p.$$

Let

$$\eta_i = |\Delta z_i \Delta s_i - \gamma\tau_1 \Delta z^T \Delta s|.$$

From the boundedness of $\Delta v$ (see Corollary 6.1), we have for some positive constant $M_3$,

$$\eta_i \leq M_3.$$

A straightforward calculation shows that for $\alpha \in [0,1]$

$$\begin{aligned}
& z_i(\alpha)s_i(\alpha) - \gamma\tau_1 z(\alpha)^T s(\alpha)/p \\
= \;& (1-\alpha)(z_i s_i - \gamma\tau_1 \tfrac{s^T z}{p}) + (1-\gamma\tau_1)\mu\alpha + (\Delta z_i \Delta s_i - \tfrac{\gamma\tau_1}{p}\Delta z^T \Delta s)\alpha^2 \\
\geq \;& (1-\gamma\tau_1)\mu\alpha - |\Delta z_i \Delta s_i - \tfrac{\gamma\tau_1}{p}\Delta z^T \Delta s|\alpha^2 \\
= \;& (1-\gamma\tau_1)\mu\alpha - \eta_i \alpha^2 \\
\geq \;& (1-\gamma\tau_1)\mu\alpha - M_3\alpha^2.
\end{aligned}$$

From the definition of $\alpha^I$ (see (56)), clearly,

$$\alpha^I \geq \frac{(1-\gamma\tau_1)\mu}{M_3}.$$

Observe that $\mu = \sigma s^T z/p$ is bounded below in $\Omega(\epsilon)$ for $\sigma$ bounded away from zero. Hence $\alpha^I$ is bounded away from zero in $\Omega(\epsilon)$.

Now we show that $\{\alpha_k^{II}\}$ generated by Step 2 of Algorithm 2 is bounded away from zero. By the mean-value theorem for vector-valued functions,

$$\begin{aligned}
G(v + \alpha\Delta v) &= G(v) + \alpha \left(\int_0^1 G'(v + t\alpha\Delta v)dt\right)\Delta v \\
&= G(v) + \alpha G'(v)\Delta v + \alpha \left(\int_0^1 (G'(v + t\alpha\Delta v) - G'(v))dt\right)\Delta v \\
&= G(v)(1-\alpha) + \alpha \left(\int_0^1 (G'(v + t\alpha\Delta v) - G'(v))dt\right)\Delta v.
\end{aligned}$$

Invoking Lipschitz continuity for the derivative of $G(v)$ (Assumption A6.1), we obtain

$$\|G(v+\alpha\Delta v)\| \le \|G(v)\|(1-\alpha) + L\|\Delta v\|^2\alpha^2.$$

Using the above inequality, we have

$$\begin{array}{rcl} f^{II}(\alpha) & = & z(\alpha)^T s(\alpha) - \gamma\tau_2\|G(v+\alpha\Delta v)\| \\ & \ge & z^T s(1-\alpha) + z^T s\sigma\alpha + (\Delta z)^T\Delta s\alpha^2 \\ & & -\gamma\tau_2(\|G(v)\|(1-\alpha) + L\|\Delta v\|^2\alpha^2) \\ & = & (z^T s - \gamma\tau_2\|G(v)\|)(1-\alpha) \\ & & +z^T s\sigma\alpha + [(\Delta z)^T\Delta s - \gamma\tau_2 L\|\Delta v\|^2]\alpha^2 \\ & \ge & \alpha[z^T s\sigma - |(\Delta z)^T\Delta s - \gamma\tau_2 L\|\Delta v\|^2|\alpha]. \end{array}$$

Since $\|\Delta v\|$ is uniformly bounded, there exists a constant $M_4 > 0$ such that

$$|(\Delta z)^T\Delta s - \gamma\tau_2 L\|\Delta v\|^2| \le M_4.$$

Hence

$$f^{II}(\alpha) \ge \alpha(z^T s\sigma - M_4\alpha).$$

This implies that

$$\alpha^{II} \ge \frac{z^T s\sigma}{M_4}.$$

Since $\{s^T z\}$ and $\{\sigma_k\}$ are bounded away from zero in $\Omega(\epsilon)$, then $\{\alpha_k^{II}\}$ is bounded away from zero. This completes the proof. □

**Theorem 6.1** *Let $\{v_k\}$ be generated by Algorithm 2 with $\epsilon_{exit} = 0$, and $\{\sigma_k\} \subset (0,1)$ bounded away from zero and one. Under Assumptions A6.1-A6.4, $\{F(v_k)\}$ converges to zero and for any limit point $v^* = (x^*, y^*, z^*, s^*)$ of $\{v_k\}$, $x^*$ is a KKT point of problem (28).*

**Proof:** Note that $\{\|F(v_k)\|\}$ is monotone decreasing, ; hence convergent. By contradiction, suppose that $\{\|F(v_k)\|\}$ does not converge to zero. Then $\{v_k\} \subset \Omega(\epsilon)$ for some $\epsilon > 0$. If for infinitely many iterations, $\alpha_k = \bar{\alpha}_k$, then it follows from the inequality

$$\frac{\phi(v_{k+1})}{\phi(v_k)} \le 1 - 2\alpha_k\beta(1-\sigma_k)$$

and Lemma 6.3 that the corresponding subsequence of $\{\phi_k\}$ converges to zero $Q$-linearly. This gives a contradiction. Now assume that $\alpha_k < \bar{\alpha}_k$ for $k$ sufficiently large. Since $\{\bar{\alpha}_k\}$ is bounded away from zero, then the back-tracking linesearch used in Algorithm 2 produces

$$\frac{\nabla\phi(v_k)^T\Delta v_k}{\|\Delta v_k\|} = \frac{-2(\phi(v_k) - \mu_k(z_k)^T s_k)}{\|\Delta v_k\|} \to 0,$$

see Ortega and Rheinboldt (Ref. 22) and Byrd and Nocedal (Ref. 23). Since $\{\|\Delta v_k\|\}$ is bounded according to Corollary 6.1,

$$\phi(v_k) - \mu_k(z_k)^T s_k \to 0.$$

However, it follows from (60) that

$$\phi(v_k) - \mu_k(z_k)^T s_k \geq (1 - \sigma_k)\phi(v_k).$$

Therefore, it must hold that $\phi(v_k) \to 0$ because $\{\sigma_k\}$ is bounded away from one. This again leads to a contradiction. So $\{\|F(v_k)\|\}$ must converge to zero.

Since the KKT conditions for problem (28), $F(x, y, z, s) = 0$ and $(z, s) \geq 0$, are satisfied by $v^*$, clearly $x^*$ is a KKT point. □

# 7 Computational Experience

In this section we report our preliminary numerical experience with Algorithm 2. The numerical experiments were done on a Sun 4/490 WorkStation running SunOS Operating System Release 4.1.3 with 64 Megabytes of memory. The programs were written in *MATLAB* and run under version 4.1.

We implemented Algorithm 2 with a slight simplification, i.e., we did not enforce condition (55) in our linesearch in order to avoid possible complication caused by the nonlinear function $f^{II}(\alpha)$ in condition (55).

We chose the algorithmic parameters for Algorithm 2 as follows. In Step (2), we choose $\sigma_k = \min(\eta_1, \eta_2 s_k^T z_k)$, where $\eta_1 = 0.2$ and $\eta_2 = 100$. Moreover, we used $\beta = 10^{-4}$ in condition (59) of Step (3.2), and set the back-tracking factor $\rho$ to 0.5.

In our implementation, we used a finite-difference approximation to the Hessian of the Lagrangian function. The numerical experiments were performed on a subset of the Hock and Schittkowski's test problems (Ref. 24 and 25). For most problems, we used the standard starting points listed in (Ref. 24 and 25). However, for some problems, the standard starting point are too close to the solution and we instead selected more challenging starting points.

The results of our numerical experience are summarized in Table 1. The first column gives the problem number as given in (Ref. 24 and 25). The $n$, $m$, and $p$ columns give the dimension (number of variables, not including slack variables), the number of equality constraints and the number of inequality constraints, respectively. The *Iterations* column gives the number of iteration required by Algorithm 2 to obtain a point that satisfies the stopping criterion

$$\frac{\|F(v_k)\|_2}{1 + \|v_k\|_2} \leq \epsilon_{exit} = 10^{-8}.$$

We summarize the results of our numerical experimentation in the following comments

1. The implemented algorithm solved all the problems tested to the given tolerance, except for problems 13 and 23. For problem 23 we had to take different step sizes with respect to the $s$-variables and $z$-variables in order to converge. For problem 13, where regularity does not hold, we only obtained a small decrease in the merit function. After 100 iterations the norm of the residual was $3.21 \times 10^{-2}$ and $\|g(x) - s\|_2$ was of order $10^{-8}$.

2. The quadratic rate of convergence is observed in problems where second order sufficiency is satisfied.

3. In the absence of strict complementarity, the algorithm was globally convergent but the local convergence was slow. This observation is compatible with our convergence theory. Strict complementarity is needed only for fast local convergence.

# 8 Concluding Remarks

Some understanding of the relationship between the logarithmic barrier function formulation and the perturbed Karush-Kuhn-Tucker conditions was presented in Sections 2-3. In summary; the logarithmic barrier function method has an inherent flaw of ill-conditioning. This conditioning deficiency can be circumvented by introducing an auxiliary variable and writing the defining relationship for this auxiliary variable in a particularly nice manner which can be viewed as perturbed complementarity. The resulting system is the perturbed KKT conditions. This approach of deriving the perturbed KKT conditions from the KKT conditions of the logarithmic barrier function problem involves auxiliary variables and a nonlinear transformation and is akin to Hestenes' derivation of the multiplier method from the penalty function method. Hence attributing algorithmic strengths resulting from the use of the perturbed KKT conditions to the KKT conditions for the logarithmic barrier function is inappropriate and analogous to crediting the penalty function method for the algorithmic strengths of the multiplier method. In Section 4 we presented a formulation of a generic line-search primal-dual interior-point method for the general nonlinear programming problem. The viability of the formulation was demonstrated in Sections 5 and 6. In Section 5, we established the standard Newton's method local convergence and convergence rate results for our interior-point formulation. In Section 6, we devised a globalization strategy using the $\ell_2$-norm-residual merit function and established a global convergence theory for this strategy. Finally, our preliminary numerical results obtained from the globalized algorithm appear to be promising.

## References


1. LUSTIG, J., MARSTEN, R. E., and SHANNO, D. F., *On implementing Mehrotra's predictor-corrector interior point method for linear programming,* SIAM Journal on Optimization, Vol. 2, No. 3, pp. 435–449, 1992.

2. WRIGHT, M. H., *Interior methods for constrained optimization,* Numer. Analysis Manuscript 91-10, AT&T Bell Laboratories, Murray Hill, New Jersey, 1991.

3. NASH, S. G. and SOFER, A., *A barrier method for large scale constrained optimization,* Technical Report 91-10, Department of Operations Research and Applied Statistics, George Mason University, Fairfax, Virginia 22030, 1991.

4. WRIGHT, S. J., *A superlinear infeasible-interior-point algorithm for monotone nonlinear complementarity problems,* Technical Report MCS-P344-1292, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, 1992.

5. MONTEIRO, R. C., PANG, J., and WANG, T., *A positive algorithm for the nonlinear complementarity problem,* Technical report, Department of Systems and Industrial Engineering, University of Arizona, Tucson, Arizona 85721, 1992.

6. WRIGHT, S.J., *An interior point algorithm for linearly constrained optimization,* SIAM J. on Optimization, Vol. 2, pp. 450–473, 1992.

7. LASDON, L., YU. G., and PLUMMER, J., *An interior-point algorithm for solving general nonlinear programming problems,* Talk presented at the SIAM Conference on Optimization in Chicago, May 1992.

8. YAMASHITA, H., *A globally convergent primal-dual interior point method for constrained optimization,* Technical report, Mathematical Systems Institute, Inc.. Japan, 1992.

9. McCORMICK, G. P., *The superlinear convergence of a nonlinear primal-dual algorithm,* Technical Report T-550/91, School of Engineering and Applied Science, George Washington University, Washington, D.C. 20052, 1991.

10. ANSTREICHER, K. M. and VIAL, J., *On the convergence of an infeasible primal-dual interior-point method for convex programming,* to appear in Optimizations Methods and Software, 1993.

11. KOJIMA, M., MEGIDDO, N., and NOMA, N., *Homotopy continuation methods for nonlinear complementarity problems,* Mathematics of Operations Research, Vol. 16, pp. 754–774, 1991.

12. MONTEIRO, R. C., and WRIGHT, S. J., *A globally and superlinearly convergent potential reduction interior point method for convex programming,* Manuscript 1992. .

13. Kojima, M., Mizuno, S., and Yoshise, A., *A primal-dual interior point method for linear programming,* Progress in mathematical programming interior-point and related methods, edited by N. Megiddo, Springer-Verlag, New York, 1989.

14. FIACCO, A. V. and McCORMICK, G. P., *Nonlinear Programming, Sequential Unconstrained Minimization Techniques,* John Wiley and Sons, New York, 1968.

15. MEGIDDO, N., *Pathways to the optimal set in linear programming,* Progress in mathematical programming interior-point and related methods, edited by N. Megiddo, Springer-Verlag, New York, 1989.

16. HESTENES, M. R., *Multiplier and gradient methods,* Journal of Optimization Theory and Applications, Vol. 4, pp. 303–329, 1969.

17. TAPIA, R. A., *On the role of slack variables in quasi-Newton methods for constrained optimization,* Numerical Optimization of dynamic systems, edited by Dixon, L. C. W. and Szegö, G. P., North-Holland, 1980.

18. EL-BAKRY, A. S., TAPIA, R. A., and ZHANG, Y., em A study of indicators for identifying zero variables in interior-point methods, SIAM Review, Vol. 36, pp. 45–72, 1994.

19. ZHANG, Y., TAPIA, R. A., and DENNIS, J. E., Jr., *On the superlinear and quadratic convergence of primal-dual interior-point linear programming algorithms,* SIAM Journal on Optimization, Vol. 2, pp. 304–324, 1992.

20. ZHANG, Y. and TAPIA, R. A., *Superlinear and quadratic convergence of primal-dual interior-point algorithms for linear programming revisited,* Journal of Optimization and its Applications, Vol. 73, pp. 229–242, 1992.

21. DENNIS, J. E., Jr. and SCHNABEL, R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Prentice-Hall Inc., Englewood cliffs, New Jersey, 1983.

22. ORTEGA, J. M. and RHEINBOLDT, W. C., *Iterative solution of nonlinear equations in several variables,* Academic Press, New York and London, 1970.

23. BYRD, R. H. and NOCEDAL, J., *A tool for the analysis of Quasi-Newton methods with application to unconstrained minimization,* SIAM Journal on Numerical Analysis, Vol. 26, pp. 727–739, June 1989.

24. HOCK, W and SCHITTKOWSKI, K., *Test Examples for Nonlinear Programming Codes,* Springer-Verlag, New York, 1981.

25. SCHITTKOWSKI, K., *More Test Examples for Nonlinear Programming Codes,* Springer-Verlag, New York, 1987.

Table 1: Numerical experience

| Problem | $n$ | $m$ | $p$ | Iterations | | Problem | $n$ | $m$ | $p$ | Iterations |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 0 | 1 | 70 | | 55 | 6 | 6 | 8 | 12 |
| 2 | 2 | 0 | 1 | 9 | | 55 | 2 | 0 | 3 | 62 |
| 3 | 2 | 0 | 1 | 6 | | 60 | 3 | 1 | 6 | 9 |
| 4 | 2 | 0 | 2 | 6 | | 62 | 3 | 1 | 6 | 9 |
| 5 | 2 | 0 | 4 | 7 | | 63 | 3 | 2 | 3 | 8 |
| 10 | 2 | 0 | 1 | 10 | | 64 | 3 | 0 | 4 | 24 |
| 11 | 2 | 0 | 1 | 9 | | 65 | 3 | 0 | 7 | 20 |
| 12 | 2 | 0 | 1 | 10 | | 66 | 3 | 0 | 8 | 10 |
| 13 | 2 | 0 | 3 | >100 | | 71 | 4 | 1 | 9 | 18 |
| 14 | 2 | 1 | 1 | 7 | | 72 | 4 | 0 | 10 | 13 |
| 15 | 2 | 0 | 3 | 15 | | 73 | 4 | 1 | 6 | 17 |
| 16 | 2 | 0 | 5 | 19 | | 74 | 4 | 3 | 10 | 19 |
| 17 | 2 | 0 | 5 | 34 | | 75 | 4 | 3 | 10 | 16 |
| 18 | 2 | 0 | 6 | 18 | | 76 | 4 | 0 | 7 | 8 |
| 19 | 2 | 0 | 6 | 15 | | 80 | 5 | 3 | 10 | 6 |
| 20 | 2 | 0 | 5 | 13 | | 81 | 9 | 13 | 13 | 13 |
| 21 | 2 | 0 | 5 | 13 | | 83 | 5 | 0 | 16 | 23 |
| 22 | 2 | 0 | 2 | 7 | | 84 | 5 | 0 | 16 | 17 |
| 23 | 2 | 0 | 9 | 21 | | 86 | 5 | 0 | 15 | 18 |
| 24 | 2 | 0 | 5 | 8 | | 93 | 6 | 0 | 8 | 10 |
| 25 | 3 | 0 | 6 | 9 | | 100 | 7 | 0 | 4 | 10 |
| 26 | 3 | 1 | 0 | 22 | | 104 | 8 | 0 | 22 | 12 |
| 29 | 3 | 0 | 1 | 13 | | 106 | 8 | 0 | 22 | 37 |
| 30 | 3 | 0 | 7 | 13 | | 226 | 2 | 0 | 4 | 7 |
| 31 | 3 | 0 | 7 | 9 | | 227 | 2 | 0 | 2 | 7 |
| 32 | 3 | 1 | 4 | 15 | | 231 | 2 | 0 | 2 | 57 |
| 33 | 3 | 0 | 6 | 10 | | 233 | 2 | 0 | 1 | 56 |
| 34 | 3 | 0 | 8 | 9 | | 250 | 3 | 0 | 8 | 8 |
| 35 | 3 | 0 | 4 | 7 | | 251 | 3 | 0 | 7 | 9 |
| 36 | 3 | 0 | 7 | 9 | | 263 | 4 | 2 | 2 | 19 |
| 37 | 3 | 0 | 8 | 8 | | 325 | 2 | 1 | 2 | 7 |
| 38 | 4 | 0 | 8 | 11 | | 339 | 3 | 0 | 4 | 8 |
| 41 | 4 | 1 | 8 | 12 | | 340 | 3 | 0 | 2 | 8 |
| 43 | 4 | 0 | 3 | 12 | | 341 | 3 | 0 | 4 | 9 |
| 44 | 4 | 0 | 10 | 9 | | 342 | 3 | 0 | 4 | 14 |
| 45 | 5 | 0 | 10 | 9 | | 353 | 4 | 1 | 6 | 10 |
| 53 | 5 | 3 | 10 | 6 | | 354 | 4 | 0 | 5 | 11 |