On the Fundamental Role of Interior-Point Methodology in Constrained Optimization

Richard A. Tapia

CRPC-TR97730
April 1997

Center for Research on Parallel Computation
Rice University
6100 South Main Street
CRPC - MS 41
Houston, TX 77005

Submitted November 1997; Also available as Rice CAAM TR97-09

On the Fundamental Role of Interior-Point Methodology in Constrained Optimization[1]

Richard A. Tapia[2]

Abstract

Recently primal-dual interior-point methodology has proven to be an effective tool in linear programming applications and is now being extended, with great enthusiasm to general nonlinear programming applications. The primary purpose of this current study is to develop and promote the belief that since Newton's method is a tool for square nonlinear systems of equations, the fundamental role of interior-point methodology in inequality constrained optimization is to produce, in a meaningful and effective manner, a square system of nonlinear equations that represents the inequality constrained optimization problem sufficiently well that the application of Newton's method methodology to this square system is effective and successful.

Keywords: Interior-point methods, algorithmic consistency, linear and nonlinear programming.

Abbreviated Title: Interior-Point Methodology in Constrained Optimization.

AMS(MOS) subject classifications: 65K,49M,90C

1 Introduction

The following notion should be a fundamental component in evaluating algorithms. We view it as a basic and minimal property.

[1]This research was supported in part by National Science Cooperative Agreement CCR-8809615 and Department of Energy Grant DOE DEFG05-86ER25017.

[2]Department of Computational and Applied Mathematics and Center for Research on Parallel Computation, Rice University, 6100 Main Street, Houston, Texas 77005 (rat@caam.rice.edu)

Definition (1.1) (Algorithmic consistency)
An iterative method is said to be *consistent*, with respect to problem P, if whenever the sequence it generates converges, it converges to a solution of problem P.

Consider problem constraints of the form

$$\begin{aligned} h(x) &= 0 \\ g(x) &\geq 0 \end{aligned} \tag{1.1}$$

and an iterative method of the form

$$x_{k+1} = x_k + \alpha_k \Delta x_k, \quad k = 0, 1, \dots , \tag{1.2}$$

where $\alpha_k \geq \alpha > 0$. We identify two paradigms designed to allow flexibility in the construction of the step Δx_k yet, promote algorithmic consistency.

Linearization Paradigm:
The step Δx_k is required to satisfy

$$\begin{aligned} \nabla h(x_k)^T \Delta x_k + h(x_k) &= 0 \\ \nabla g(x_k)^T \Delta x_k + g(x_k) &\geq 0. \end{aligned} \tag{1.3}$$

Interior-Point Paradigm:
The step Δx_k and the steplength α_k are required to stisfy

$$\begin{aligned} \nabla h(x_k)^T \Delta x_k + h(x_k) &= 0 \\ g(x_{k+1}) &> 0. \end{aligned} \tag{1.4}$$

Observe that both paradigms allow flexibility in the construction of the step Δx_k and promote consistency in the sense that if $x_k \rightarrow x^*$, then $h(x^*) = 0$ and $g(x^*) = 0$, under rather mild assumptions.

Definition (1.2) (Interior-point method)
By an interior-point method we mean an iterative method that generates iterates that strictly satisfy the inequalities in the problem definition.

In some sense an interior-point method ignores the inequalities and moves always staying interior with respect to these inequalities. As such it can be extremely naive, ineffective, and actually inconsistent. The interior-point challenge, which will determine success or failure of the interior-point method under construction, is to incorporate information about the "ignored" inequalities into the model subproblem defining the step Δx_k in a clever and effective manner. Our present task is to interpret the primal-dual interior-point Newton approach, proposed by Kojima, Mizuno, and Yoshise [7] in 1989 for linear programming and extended to nonlinear programming by El-Bakry, Tapia, Tsuchiya, and Zhang [2] in 1992, in the context of both existing literature and the interior-point challenge described immediately above.

Let's begin with a rather bold statement, but one that we believe captures the essence of numerical computation.

1.1 The Fundamental Tool of the Computational Sciences

The fundamental tool of the computational sciences is the numerical solution of the square nonsingular linear system

$$Ax = b. \tag{1.5}$$

where $A \in I\!R^{n \times n}$ and $b \in I\!R^n$; hence the solution x will be a member of $I\!R^n$. Since computers can only perform arithmetic it follows that the only system of equations that can be solved on a computer are linear systems. Moreover, if we ask that our method of solution be unambiguous and well-defined, we are asking that the linear system have a unique solution. Hence in the linear system (1.5) the matrix A must be square and nonsingular. It follows then the basic activity of the computational sciences is formulating problems so that their solutions can be approximated by employing strategies that require only the solutions of square nonsingular linear systems.

1.2 A Fundamental Tool

Consider the square (number of equations equals the number of variables) nonlinear system of equations

$$F(x) = 0. \tag{1.6}$$

where $F : I\!R^n \rightarrow I\!R^n$. Recall that by damped Newton's method for approximating solutions of problem (1.6) we mean the iterative procedure.

Algorithm 1.1 (Damped Newton's Method)
Given x_0, for $k = 0, 1, ...,$ do

(1) *Solve $F'(x_k)\Delta x = -F(x_k)$ for Δx_k.*

(2) *Set $x_{k+1} = x_k + \alpha_k \Delta x_k$, where $0 < \alpha_k \leq 1$.*

It is our expectation that the damped Newton sequence $\{x_k\}$ will converge to a solution x^* of problem (1.6). Moreover, we know that under rather standard assumptions this will be the case for $x_\circ$ sufficiently close to a solution x^* and the convergence will be fast, i.e., quadratic, provided $\alpha_k \to 1$ sufficiently fast.

Now, we stress that the damped Newton's method, Algorithm 1.1 is consistent.

Theorem 1.1 *(Consistency for damped Newton's method) Suppose that the Newton sequence for problem (1.6) converges to a point x^*. Assume that F' is continuous at x^* and $\alpha_k \geq \alpha > 0$. Then x^* is a solution of problem (1.6).*

Proof: The proof follows directly from Algorithm 1.1. □

Newton's method allows us to reduce the solution of a square nonlinear system of equations to the solution of a sequence of square (hopefully nonsingular) square linear systems. Hence the class of problems that can be handled numerically is significantly enlarged.

1.3 A Fundamental Activity in the Computational Sciences

A fundamental activity in the computational sciences is the modeling and formulation of problems as square nonlinear systems of equations. Once this task has been accomplished the researcher has at his or her disposal the entire arsenal of tools related to Newton's method. As we shall soon see often times a particular problem has been stated as a mathematical problem and the task at hand is to reformulate it as a square nonlinear system of equations, or as a sequence of such problems.

1.4 Computational Optimization

We now restrict our attention to the area of computational optimization. It is standard how one transforms the unconstrained optimization problem and the equality constrained optimization problem into essentially equivalent square systems of nonlinear equations. It is not at all obvious how an optimization problem with inequality constraints can be transformed into or represented by a square nonlinear system of equations. Moreover, this is the subject of the present study. It is our basic premise that the new emerging interior-point methodology should be viewed as a clever and effective way of handling optimization problems with inequality constraints by only dealing with square nonlinear systems of equations.

In the next section, Section 2, we will describe the various optimization problem classes and show how Newton's method technology can be readily applied to unconstrained optimization and equality constrained optimization. In Section 3 we will present methods from the literature for handling the inequality constrained problem by considering related square systems of nonlinear equations. We include a discussion of the strengths and weaknesses of such ad hoc procedures from the literature. In Section 4 we describe the current primal-dual interior-point formulation. We argue that this formulation has the strength of other ad hoc procedures without also sharing their weaknesses. Finally, in Section 5 we make some concluding remarks.

2 Optimization Problem Classes Hierarchy

In computational optimization it is standard to consider the following four problem classes; listed in order of increasing difficulty.

2.1 Nonlinear Equations

The nonlinear equation problem is denoted by

$$F(x) = 0, \tag{2.7}$$

where $F : I\!\!R^n \to I\!\!R^n$ and is read as find $x^* \in I\!\!R^n$ such that $F(x^*) = 0$. This problem was considered in (1.6) and we presented it as the basic framework for the definition of Newton's method.

2.2 Unconstrained Optimization

The unconstrained optimization problem is denoted by

$$minimize \quad f(x) \tag{2.8}$$

where $f : I\!R^n \to I\!R$ and is read as find $x^* \in I\!R^n$ such that $f(x^*) \leq f(x)$ for all $x \in I\!R^n$. In some applications we are interested in local minimizers and in this case we only ask that $f(x^*) \leq f(x)$ for all x such that $\|x - x^*\| < \epsilon$ for some $\epsilon > 0$. A well known necessary condition for the unconstrained minimization problem (2.8) is that any solution, including local solutions, must satisfy the nonlinear equation problem

$$\nabla f(x) = 0, \tag{2.9}$$

where $\nabla f : I\!R^n \to I\!R^n$ is the gradient operator, i.e., $\nabla f(x)$ is the vector of first order partial derivatives at x. The basic Newton's method framework for the unconstrained minimization problem (2.8) is the nonlinear equations problem (2.9). Unfortunately these two problems are not equivalent, and Newton's method may find solutions of (2.9) that are not solutions of problem (2.8). These extraneous solutions would correspond to maximizers of f or saddle points of f. However, various modifications or additions can be made to the basic Newton's method in an attempt to preclude convergence to these extraneous solutions.

2.3 Equality Constrained Optimization

The equality constrained optimization problem is denoted by

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & h_i(x) = 0 \quad i = 1, \ldots, m \end{array} \tag{2.10}$$

where $f : I\!R^n \to I\!R$, and $h_i : I\!R^n \to I\!R$ and $m < n$. The problem is to find x^* which minimizes f in the class of all x satisfying the constraints.

In an effort to derive an equivalent square nonlinear system of equations for problem (2.10) we first consider the Lagrangian function $\ell : I\!R^n \times I\!R^m \to I\!R$ defined by

$$\ell(x, \lambda) = f(x) + \lambda_1 h_1(x) + \ldots + \lambda_m h_m(x). \tag{2.11}$$

It is well-known that, under rather mild assumptions, if x^* is a solution of problem (2.10), then there exist Lagrange multipliers $\lambda^* \in I\!R^m$ so that the pair (x^*, λ^*) is a solution of the square nonlinear system of equations, the so-called first-order necessary conditions

$$\nabla \ell(x, \lambda) = 0. \tag{2.12}$$

It is not difficult to see that (2.12) has the form

$$\begin{aligned} \nabla f(x) + \lambda_1 \nabla h_1(x) + ... + \lambda_m \nabla h_m(x) &= 0 \\ h_i(x) &= 0, \quad i = 1, ..., m \end{aligned} \tag{2.13}$$

Hence, we can use this square nonlinear system of equations as the framework for applying Newton's method to the equality constrained optimization problem (2.10). As in the case of problem (2.8) and problem (2.9), we know that problem (2.10) and problem (2.13) are not mathematically equivalent. However, the equivalence is sufficiently close to allow us to build effective Newton's method theory and algorithms.

It should be clear that Newton's method will be algorithmically consistent with respect to problems (2.9) and (2.12).

2.4 Nonlinear Programming

Consider the optimization problem with both equality and inequality constraints, the so-called general nonlinear program

$$\begin{aligned} &\textit{minimize} && f(x) \\ &\textit{subject to} && h_i(x) = 0 \quad i = 1, \dots, m \\ & && g_i(x) \geq 0 \quad i = 1, \dots, p \end{aligned} \tag{2.14}$$

where f, h_i, g_i are all real-valued functions defined on $I\!R^n$. In this context we define the Lagrangian function $\ell : I\!R^n \times I\!R^m \times I\!R^p \to I\!R$ by

$$\ell(x, u, \lambda) = f(x) + \lambda^T h(x) - u^T g(x), \tag{2.15}$$

where $\lambda = (\lambda_1, \dots, \lambda_m)^T$, $h(x) = (h_1(x), \dots, h_m(x))^T$, and $g(x) = (g_1(x), ..., g_p(x))^T$. Under mild assumptions, we know that any solution of the nonlinear program (2.14 must satisfy the Karush-Kuhn-Tucker (KKT) first-order necessary conditions

KKT CONDITIONS

$$\begin{aligned} \nabla_x \ell(x, u, \lambda) &= 0 \\ ug(x) &= 0 \\ h(x) &= 0 \\ g(x) &\geq 0 \\ u &\geq 0 \end{aligned} \tag{2.16}$$

In (2.16) the equalities and inequalities are read component-wise. The first things that we notice is that the KKT system (2.16) is not a square nonlinear system. We can not imitate the situation in unconstrained and equality constrained optimization; and therefore do not know how to directly apply Newton methodology to the general nonlinear program (2.14). Indeed, how this should be done is the topic of this entire study. Before we move on, we remark that the second set of equations in the KKT conditions (2.16) are very interesting and important and are called the complementarity equations, or simply complementarity. A particular inequality constraint say g_i is either binding or active at a solution x^*, i.e., $g(x^*) = 0$, or it is non-binding and inactive, i.e., $g(x^*) > 0$. In the former case our problem does not change if we treat the inequality constraint g_i as an equality constraint; while in the latter case the flavor of the problem does not change if locally (near the solution) we ignore this inequality constraint and completely discard it from the problem formulations. It is exactly the smooth transition between these two scenarios that is the role and responsibility of the complementarity equations. As such they are extremely important; and this point of view will be greatly reinforced as we travel through the current study.

3 Standard Approaches for Inequality Constraints

In this section we study three known and rather standard approaches from the optimization literature for handling optimization problems with inequality constraints. These approaches are the squared slack variables approach, the active set strategy, and the logarithmic barrier function approach. Each of these three approaches represents the nonlinear programming problem by an optimization problem with only equality constraints. In this way we are able to apply Newton methodology to the original problem. We remind the reader that in line with the basic premise of this paper, these three approaches can be viewed as vehicles for obtaining a square nonlinear system that represents the original problem to varying degrees of accuracy and to which Newton's method can be applied.

3.1 Squared Slack Variables

Consider an inequality constraint from problem (2.14), say $g_i(x) \geq 0$. The following equivalence is immediate

$$g_i(x) \geq 0 \Leftrightarrow g_i(x) = y_i^2, \tag{3.17}$$

for some $y_i \in I\!R$. Hence, problem (2.14) can be stated equivalently as the following equality constrained optimization problem

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & h(x) = 0 \\ & g(x) = y^2 \end{array} \tag{3.18}$$

where the auxiliary variable $y \in I\!R^m$. Since (3.18) is equivalent to (2.14) and has only equality constraints, Newton methodology can be applied as described in Section 2. Much numerical experience has been gained over the years with this so-called squared-slack approach. We now formally list the advantages and disadvantages of the squared-slack approach to applying Newton methodology to the general nonlinear program (2.14).

Advantages

1. The optimization problem has no inequality constraints.
2. Newton's method is algorithmically consistent with respect to the KKT conditions (2.16) except for the nonnegativity of the multipliers.

Disadvantages

1. The dimension (number of variables) of the problem is increased.
2. The degree of nonlinearity of the inequality constraints is increased and leads to a loss of convexity.
3. The introduction of extraneous multiple solutions causes the global convergence of Newton's method to deteriorate significantly.

The first disadvantage is not particularly serious, and can actually be circumvented by a decoupling procedure described in Tapia [11]. The second disadvantage is serious and of concern. However, the third disadvantage is devastating and effectively renders the squared-slack variable approach useless, as a numerical tool. To further understand this highly critical statement, we add the following explanation. If (x^*, y^*) is a solution of problem (3.18), then so is $(x^*, \hat{y^*})$ where $\hat{y^*}$ is y^* with any component replaced by its negative value. Observe that if $y^* \in I\!R^m$, then there are 2^m possible choices for $\hat{y^*}$, hence $2^m - 1$ extraneous solutions has been introduced. Let us now argue why this exponential explosion of extraneous solutions literally makes Newton's method ineffective. On the real-line, if Newton's method is started with a point that is in the region midway between two solutions, Newton's method does not behave well and the iteration sequence may not be well-defined. In higher dimensions we observe a similar phenomenon. It is not hard to see that for the squared-slack variable formulation, the effective convergence of Newton's method will be restricted to little pockets around each solution. This restrictive behavior can be demonstrated numerically, and is unfortunately more often the rule, than the exception. In spite of this poor behavior, the use of squared-slack variables and Newton's method has been a tool in scientific computation in the past twenty years or so. This is particularly true in engineering applications.

3.2 Active Set Strategy

Consider the nonlinear program (2.14). At a solution, say x^*, there exists an active set of inequality constraints $E_* = \{i : g_i(x^*) = 0\}$ and an inactive set of inequality constraints $I_* = \{i : g_i(x^*) > 0\}$. Observe that $x^*$is a solution of both problem (2.14) and the equality constrained problem

$$\begin{array}{ll} minimize & f(x) \\ subject\ to & h_i(x) = 0 \quad i = 1, \ldots, m \\ & g_i(x) = 0 \quad i \in E_* \end{array} \tag{3.19}$$

So, if we knew which constraints were active, then we could ignore the inactive inequality constraints and formulate the problem at hand as an equality constrained problem. The *Active Set Strategy* says that at each stage of an iterative process, e.g. Newton's method, "guess" the constraints that are active at the solution. Treat these as equalities and ignore the others, i.e., work with a problem of the form (3.19).

The simplex method for linear programming is the canonical example of an active set strategy. Recall that a vertex can be characterized as a point which uniquely satisfies a subset of the constraints. It is fair to say that most active set strategies in nonlinear programming are direct descendants of the simplex idea in linear programming. Moreover, they seem to have been suggested in the decades following the birth of the simplex method. In some sense an active set strategy can be viewed as a form of model switching, and only one model is the correct model.

Advantages

1. An active set strategy leads to a square nonlinear system of equations at each iteration.

2. The square system is of lower dimension, than the original problem.

Disadvantages

1. There is an exponential combinatorial explosion in the number of possible active sets as the number of inequality constraints increases.

2. As a form of model-switching we will have success or failure as a function of how clever we are in switching models and identifying the correct model (active set). In a worst-case scenario, the correct model would not be identified and the sequence generated by the algorithm could have the property that it converges but not to a solution of the problem. That is, active set Newton's method may not be algorithmically consistent.

Our two disadvantages should make a potential user leery of the active set approach, yet today it is the most popular approach for problems with inequality constraints. Moreover, these disadvantages must speak directly to the fact that the simplex method for linear programming is known not to be a polynomial time algorithm. The now famous Klee-Minty [6] counterexample forces the simplex method to consider each vertex.

3.3 Logarithmic Barrier Function Interior-Point Method

The basic idea of the logarithmic barrier function method is to replace the optimization problem

$$\begin{array}{ll} minimize & f(x) \\ subject\ to & h(x) = 0 \\ & g(x) \geq 0 \end{array} \tag{3.20}$$

with the equality constrained optimization problem

$$\begin{array}{ll} \textit{minimize} & f(x) - \mu \log(g(x)) \\ \textit{subject to} & h(x) = 0 \end{array} \tag{3.21}$$

with $\mu > 0$. Let $x(\mu)$ denote the solution of the logarithmic barrier subproblem (3.21). It is known that under mild conditions $x(\mu)$ converges to a solution of problem (3.20) as μ decreases to zero. This fact allows us to approximate a solution of problem (3.20) by solving problem (3.21) for a sequence of μ's which decreases to zero. Implicit in problem (3.21) is the requirement that $g(x) > 0$; since otherwise the logarithm is not defined. The logarithmic term serves as a barrier and gives $x(\mu)$ such that $g(x(\mu)) > 0$. Hence this method is clearly an interior-point method in the sense that it keeps iterates strictly feasible with respect to the inequality constraints. The solution path $x(\mu)$ parameterized by μ is called the (primal) *central path* and has many beautiful theoretical properties; including that it intersects the solution set of problem (3.20) at an interesting point called the analytic center.

The logarithmic barrier function method is implicit in Frisch [4] (1955). It was promoted and popularized by Fiacco and McCormick [3] in the late 1960's. The fascinating and elegant properties of the central path were explored independently by McLinden [8], Megiddo [9], and Sonnevend [10].

Advantages

1. The logarithmic barrier function subproblem has no inequality constraints.
2. Convexity is retained in the sense, that if the original problem (3.20) is a convex program, then so is the logarithmic barrier function subproblem (3.21).
3. The method has excellent global convergence behavior.
4. The method is consistent with respect to the KKT conditions (2.16).

Disadvantages

1. The method is expensive in that it requires the solution of many nonlinear equality constrained optimization problems.

2. The method possesses inherent ill-conditioning near the solution, i.e., for small μ.

It is sufficient to illustrate this latter point with the simple case of a problem with only one inequality constraint. Towards this end consider the inequality constrained problem

$$\begin{array}{ll} minimize & f(x) \\ subject\ to & g(x) \geq 0 \end{array} \tag{3.22}$$

If we let

$$\ell_\mu(x) = f(x) - \mu \log(g(x)), \tag{3.23}$$

then we see that

$$\nabla \ell_\mu(x) = \nabla f(x) - \frac{\mu}{g(x)} \nabla g(x), \tag{3.24}$$

and

$$\nabla^2 \ell_\mu(x) = \nabla^2 f(x) - \frac{\mu}{g(x)} \nabla^2 g(x) + \frac{\mu}{g(x)^2} \nabla g(x) \nabla g(x)^T. \tag{3.25}$$

Now, if $x(\mu)$ converges to a solution of (3.22) as $\mu \to 0$, then from the KKT conditions and (3.24) we must have

$$\frac{\mu}{g(x)} \to u^* \text{ as } \mu \to 0, \tag{3.26}$$

where u^* is the Lagrange multiplier associated with the inequality constraint. Now, suppose that we are in the situation that $g(x)$ is active at the solution and $u^* \neq 0$. This is to be expected unless the constraint is redundant in the sense that the unconstrained minimizer coincides with the constrained minimizer. From this we see that for μ sufficiently small the conditioning of the Hessian matrix (3.25) becomes arbitrarily bad. The extent that this ill-conditioning manifests itself in a particular algorithmic formulation of Newton's method is not of concern here. There is some disagreement of this latter issue among computational scientists.

4 The Primal-Dual Newton Interior-Point Method

We begin with some statements that motivate and set the stage for the primal-dual Newton interior-point methods that we are about to discuss.

In an active set approach (so-called model-switching approach) we obtain a square nonlinear system of equations by ignoring various inequality constraints, their multipliers, and their complementarity equations. As a consequence our subproblems are not necessarily badly conditioned. However, this is obtained at the expense of terrible worst-case global behavior. Global information in the form of the complementarity equation has been sacrificed and is not present in the local models. Complementarity is a check on the validity of the local models. On the other hand, in the logarithmic barrier function approach we obtain a square nonlinear system and stay strictly interior with respect to the inequality constraints. As a consequence we promote excellent global convergence properties at the expense of necessarily badly conditioned subproblems. The obvious question to ask is: Can we do something that will give us the best of both worlds without the worst of either world. The answer to this rhetorical question is yes. Our development is in the spirit of the development given by Kojima, Mizuno, Yoshise [7] for the primal-dual Newton interior-point method in linear programming.

Observe that the inequality $g_i(x) \geq 0$ is equivalent to the equality $g_i(x) - s = 0$ and the inequality $s \geq 0$. Hence, without loss of generality we can consider the general nonlinear program in the following standard form

$$\begin{array}{ll} minimize & f(x) \\ subject\ to & h(x) = 0 \\ & x \geq 0 \end{array} \tag{4.27}$$

In (4.27) $f : I\!R^n \to I\!R$, $h : I\!R^n \to I\!R^m$, and $x \in I\!R^n$. Here $x \geq 0$ is notation for $x_i \geq 0$, $i = 1, ..., n$. The logarithmic barrier function formulation of problem (4.27) is

$$\begin{array}{ll} minimize & f(x) - \mu \sum_{i=1}^{n} \log(x_i) \\ subject\ to & h(x) = 0 \end{array} \tag{4.28}$$

where $\mu > 0$ and we implicitly assume $x > 0$. The KKT conditions for this problem are

$$\begin{array}{rcl} \nabla_x f(x) + \nabla h(x)^T y - \mu X^{-1} e & = & 0 \\ h(x) & = & 0, \end{array} \tag{4.29}$$

with the implicit assumption that $x > 0$. In (4.29) $e = (1, ..., 1)^T$, X is the diagonal matrix with x on its diagonal, and $\nabla h(x)^T y$ is notation for $y_1 \nabla h_1(x) + ... + y_m \nabla h_m(x)$. Consider the introduction of the auxiliary variable z defined by

$$z = \mu X^{-1} e, \tag{4.30}$$

and write (4.30) in the benign form

$$XZe = \mu e, \tag{4.31}$$

where Z is the diagonal matrix with z on the diagonal. Using the transformation (4.30) we can write (4.29) in the equivalent form

$$\begin{aligned} \nabla_x f(x) + \nabla h(x)^T y - z &= 0 \\ h(x) &= 0 \\ XZe &= \mu e \end{aligned} \tag{4.32}$$

where $\mu > 0$ and implicitly we require $(x, z) \geq 0$.

For $\mu = 0$ (4.32) are merely the KKT conditions for our original problem (4.27). Hence we call them the perturbed KKT conditions.

By the *primal-dual Newton interior-point method* we mean the algorithm that takes Newton steps on the perturbed KKT conditions (4.32) (ignoring $(x, z) \geq 0$) and damps the Newton steps so that the new x and z remain strictly positive. Of course in the process μ should change and should be chosen so that it decreases to zero. The basic philosophical approach can be described as follows.

Primal-Dual Newton Interior-Point Method Philosophy

- As a basic framework for Newton's method consider the KKT conditions for the nonlinear program.
- Start in the interior of the feasibility region defined by the inequalities in the KKT system
- In computing the Newton step ignore the inequalities in the KKT system. This gives a square nonlinear system from which a Newton step can be computed.
- Perturb the complementarity equations in the KKT system so that the Newton direction obtained from the perturbed KKT conditions does not point directly into the boundary, and allows movement away from the boundary to the central path.
- Move in the perturbed direction always staying in the interior of the region defined by the inequalities.

Observe that the primal-dual Newton interior-point formulation offers considerable flexibility and advantages. To begin with eventhough the KKT conditions for the logarithmic barrier function subproblem (4.29) and the perturbed KKT conditions (4.32) are equivalent; the Newton steps for the two systems are in general quite different. Indeed, in the case of linear programming El-Bakry, Tapia, Tsuchiya, and Zhang [2] proved that the two Newton steps are never the same. Hence in nonlinear programming we should expect similar behavior, with perhaps some pathological exceptions. If we solve (4.32) to completion for a fixed μ and then decrease μ and solve (4.32) to completion, and continue on in this fashion, then we will be mimicking the behavior of the logarithmic barrier function method; however, we will accomplish this without solving necessarily badly conditioned linear systems; as would be the case if we applied Newton's method to (4.29). However, μ can be changed more frequently in the Newton iteration process when appropriate, i.e., certainly near the solution, in an effort to obtain a more efficient implementation. El-Bakry et al [2] did not consider the flexibility of holding μ fixed for an appropriate number of Newton iterations. This flexibility is a major theme in Gonzalez-Lima, Tapia, and Potra [5] in a particular linear programming application, and is also a major theme in Argaez and Tapia [1].

We have now arrived to the point that is the underlying tenet of the current study. The primal-dual Newton interior-point method consists of Newton steps on a square nonlinear system of equations, the perturbed KKT conditions (4.32), that has in no way compromised complementarity. Hence, this feature coupled with the fact that we maintain interior (i.e., strictly feasible) iterates allows us to obtain consistency for our Newton's method with respect to the KKT conditions. So, in contrast to an active set approach, in our primal-dual Newton interior-point approach, under mild assumptions, we know that if the sequence of iterates converges, then it converges to a KKT point of the original problem. Moreover, Zhang, Tapia and Dennis [12] demonstrated that quadratic convergence could be obtained from the primal-dual interior-point method for nondegenerate linear programs. Zhang and Tapia [13] were able to extend this result to degenerate linear programs. This is a surprising result since the Jacobian matrix at a solution may be highly singular and quadratic convergence for singular Newton's method is extremely rare. El-Bakry et. al [2] demonstrated quadratic convergence in the case of nonlinear programming under the standard Newton's method theory assumptions.

Why Perturb the KKT Conditions

Let us consider the complementarity equation for a particular variable say x_i. So we have

$$x_i z_i = 0. \tag{4.33}$$

Newton's method replaces complementarity (4.33) with linearized complementarity

$$z_i \Delta x + x_i \Delta z = -x_i z_i. \tag{4.34}$$

Now, if $x_i = 0$ and $z_i \neq 0$, then from (4.34) we see that $\Delta x = 0$ and consequently $x_i + \alpha \Delta x_i = 0$ for any value of α. It follows that if we used the unperturbed KKT conditions in our Newton formulation, then the iterates would stick to the boundary of the nonnegative orthant. Such behavior would preclude any form of global convergence, or algorithmic consistency as described by Definition 1.1. Moreover, if x_i became very small, then we would expect the Newton method formulation to make small changes; since it makes no change for $x_i = 0$, and lead us to an inefficient and ineffective algorithm. However, instead of complementarity(4.33) let us consider perturbed complementarity

$$x_i z_i = \mu, \quad \mu > 0. \tag{4.35}$$

Now, if $x_i = 0$, $z_i \neq 0$, then linearized perturbed complementarity would lead to

$$\Delta x = \mu z_i^{-1}, \tag{4.36}$$

and the iterate x_i can move away from a zero (or small) value. Hence, perturbing promotes and enhances the global aspects of the primal-dual Newton interior-point method. Moreover, without perturbing global convergence is precluded. It is satisfying that this perturbation can also be motivated in terms of the logarithmic barrier function method which is known to have excellent global convergence properties. The perturbation takes us towards the central path and away from the boundary.

5 Concluding Remarks

In this study we have attempted to convince the reader that the primal-dual Newton interior-point method contains many nice philosophical features for handling inequality constrained optimization problems that other approaches from the literature lack. This argument rests on the complementarity equation. We argue that it is central to the formulation of an algorithm and contains much global information. A key ingredient in our message is that all approaches that allow Newton's method methodology to be used on an inequality constrained problem, can be viewed as approaches for constructing or extracting a square nonlinear system of equations to serve as local models for the original inequality constrained problem. Moreover, the primal-dual Newton interior-point ideology performs this task in a mathematically satisfying and numerically promising manner. These particular square nonlinear systems allow for a formulation which retains all the important information about the entire nonlinear program and also allows for excellent global and local convergence behavior. It does this while working with linear systems that do not posses inherent ill-conditioning. Moreover, one does not have to keep track of active or inactive inequality constraints.

Acknowledgments

The author acknowledges members and visiting members of the Rice interior-point methods research group for discussions and interactions on issues related to this material.

References

[1] M. Argaez and R.A. Tapia. *On the global convergence of a modified augmented Lagrangian linesearch interior point Newton method for nonlinear programming.* Technical Report TR95-38, October 1995, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005.

[2] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang. On the formulation and theory of the Newton interior-point methods for nonlinear programming *Journal of Optimization Theory and Applications*, 89:507–541, 1996.

[3] A. V. Fiacco and G. P. McCormick Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley and Sons, New York, 1968.

[4] K. R. Frisch. The logarithmic potential method of convex programming. Memorandum of May 13, 1955, University Institute of Economics, Oslo, Norway, 1955.

[5] M. Gonzalez-Lima, R.A. Tapia, and F. Potra. *On effectively computing the analytic center of the solution set by primal-dual interior-point methods.* Technical Report TR92-26, 1994, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005. To appear in SIAM Journal on Optimization.

[6] V. Klee and G.J.Minty. *How good is the simplex algorithm ?*, in Inequalities, O. Shisha, ed., Academic Press, New York, 1972, pp. 159-175.

[7] M. Kojima, S. Mizuno, and A. Yoshise. *A primal-dual interior point algorithm for linear programming*, in Progress in Mathematical Programming: Interior Point and Related Methods, N. Megiddo (editor), Springer Verlag, New York, 1989, pp. 29-47.

[8] L. McLinden. *An analogue of Moreau's proximation theorem, with applications to the nonlinear complementarity problem*, Pacific Journal of Mathematics, 88 (1980), pp. 131-158.

[9] N. Megiddo. *Pathways to the optimal set in linear programming*, in Progress in Mathematical Programming: Interior Point and Related Methods, N. Megiddo (editor), Springer Verlag, New York, 1989, Ch. 8, pp. 131-158.

[10] G. Sonnevend *An "analytic center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming*, in System Modelling and Optimization : Proceedings of the 12th IFIP Conference held in Budapest, Hungary, September 1985, A. Prekopa, J. Szelezsan, and B. Strazicky (editors). Lecture Notes in Control and Information Sciences 84, Springer Verlag, Berlin, 1986, pp. 866-876.

[11] R.A. Tapia. *On the role of slack variables in quasi-Newton methods for constrained optimization*, in Numerical Optimization of Dynamic Systems, L. C. W. Dixon and D. P. Szego, eds., North-Holland, Amestrdam, 1980, pp. 235-246.

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

[13] Y. Zhang and R.A. Tapia. *Superlinear and quadratic convergence of prima-dual interior-point methods for linear programming revisited*, Journal of Optimization Theory and Applications, 73 (1992), pp. 229-242.