The Predictor Corrector Interior-Point Method as a Composite Newton Method

R. Tapia
Yin Zhang
M. Saltzman
A. Weiser

CRPC-TR90183
December, 1990

Center for Research on Parallel Computation
Rice University
P.O. Box 1892
Houston, TX 77251-1892

Revised August, 1991.

The Predictor-Corrector Interior-Point Method as a Composite Newton Method*

R. Tapia[†] Y. Zhang[‡] M. Saltzman[§] A. Weiser[¶]

July, 1990
(Revised August 1991)

*This research was supported in part by NSF Coop. Agr. No. CCR-8809615 and was initiated early in 1990 while the second and third authors were at Rice University as visiting members of the Center for Research in Parallel Computation.

†Department of Mathematical Sciences and Center for Research on Parallel Computation, Rice University, Houston, Texas 77251-1892. This author was supported in part by AFOSR 89-0363; DOE DEFG05-86ER25017; and ARO 9DAAL03-90-G-0093.

‡Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21228. This author was supported in part by NSF Grant DMS-9102761 and DOE Grant DE-FG05-91ER25100.

§Department of Mathematical Sciences, Clemson University, Clemson, South Carolina 29634.

¶Center for Research on Parallel Computation, Rice University, Houston, Texas 77251-1892.

Abstract

The simplified Newton method reduces the work required by Newton's method per iteration by reusing the initial Jacobian matrix. However, fast convergence is sacrificed. The level-m composite Newton method attempts to balance the trade-off between work and fast convergence by composing one Newton step with m simplified Newton steps. In this work we demonstrate that Mehrotra's predictor-corrector interior-point method is the level-1 composite Newton variant of the Newton interior-point method. The level-1 composite Newton method is well known to give cubic convergence. Hence we find it mathematically enlightening that the interior-point aspect of the predictor-corrector method, i.e. choosing the steplength so that the iterates remain nonnegative, precludes the standard proof of cubic convergence, but does support the proof of quadratic convergence. We next demonstrate that by choosing steplength one in a neighborhood of the solution (therefore allowing negative iterates) cubic convergence can be retained for nondegenerate problems. Preliminary numerical experiments with this locally modified algorithm are most impressive and are an important and appropriate part of a companion paper that numerically studies the local behavior of the predictor-corrector algorithm.

1 Introduction

In Subsection 1.1 we review the composite Newton method, in 1.2 we recall the Newton interior-point method, in 1.3 we present Mehrotra's predictor-corrector interior-point method, and in 1.4 we present our composite Newton interior-point method. Section 2 contains equivalence results between Newton predictor-corrector methods and the level-1 composite Newton method. Since the level-1 composite Newton method is known to be cubically convergent, in Section 3 we study the cubic convergence aspect of the Mehrotra predictor-corrector interior-point method via our equivalence result. It is interesting to learn that the interior-point feature of the method, i.e., the step is damped so that iterates remain positive, precludes the standard proof of cubic convergence of the method. However, for nondegenerate problems it is possible to retain quadratic convergence. Recall that recently Zhang, Tapia and Dennis (1990) demonstrated that the Newton interior-point method can attain quadratic convergence for nondegenerate problems. We then prove that by choosing steplength one in a neighborhood of the solution, cubic convergence can be attained by the predictor-corrector interior-point method for nondegenerate problems. It should be emphasized that choosing the step to the boundary of the positive orthant is not sufficient to preserve the cubic convergence. Iterates must be allowed to become negative when steplength one calls for such action. This sequence of events is satisfying since it adds credibility to the nonlinear programming adage that enforcing feasibility with respect to inequality constraints may compromise fast convergence. In this sense it is somewhat noteworthy that Zhang, Tapia and Dennis were able to establish quadratic convergence of the Newton interior-point method, since it is a method which does not necessarily give steplength one, even if the step is all the way to the boundary of the positive orthant.

Numerical experimentation with the cubically convergent modification is most impressive and has been relegated to the companion paper, El-Bakry, Tapia, Zhang [2], which numerically studies the local behavior of the predictor-corrector algorithm. Clearly an optimal implementation of the composite Newton interior-point method would allow m (the number of simplified Newton steps) to vary at each Newton step. This issue is not the subject of the current work, but probably merits further study. Finally, in Section 4 we give some concluding remarks.

1.1 The Composite Newton Method

Consider the nonlinear equation

$$F(x) = 0 \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$. By the *damped Newton method* for problem (1.1) we mean the iterative process

$$\begin{array}{lll} \text{solve} & F'(x_k)(\Delta x) = -F(x_k) & \text{for } \Delta x \\ \text{set} & x_{k+1} = x_k + \alpha_k \Delta x \ , & k = 0, 1, \dots \ . \end{array} \tag{1.2}$$

The flexibility of being able to choose α_k less than one is important from global convergence considerations. When the choice of steplength is $\alpha_k = 1$ we drop the qualifier *damped.*

Under standard assumptions Newton's method is known to give Q-quadratic convergence. Not counting the work required to evaluate the function F or its Jacobian, the algebra required per iteration is $O(n^3)$, since the dominant task is the factorizing of the $n \times n$ Jacobian matrix $F'(x_k)$. For large n this can be a very serious concern.

A particularly obvious technique for reducing the amount of algebra needed at each iteration is given by the *damped simplified Newton method*

$$\begin{array}{lll} \text{solve} & F'(x_0)(\Delta x) = -F(x_k) & \text{for } \Delta x \\ \text{set} & x_{k+1} = x_k + \alpha_k \Delta x \ , & k = 0, 1, \dots \ . \end{array} \tag{1.3}$$

The simplified Newton method requires an initial factorization of $F'(x_0)$ and then a solve at each iteration; hence it requires only $O(n^2)$ algebra per iteration. However, it gives only Q-linear convergence and it is not at all clear in what cases it should be preferred to Newton's method, since the slow convergence might force a prohibitive number of iterations.

In an effort to cover the middle ground between the extremes of Newton and simplified Newton it is very natural to consider the variant of Newton's method which takes m simplified Newton steps between every two Newton steps. By the *damped (level-m) composite Newton method* we mean the iterative procedure

$$\begin{array}{lll} \text{solve} & F'(x_k)(\Delta x_i) = -F(x_k + \Delta x_0 + \cdots + \Delta x_{i-1}) & \text{for } \Delta x_i\,, i = 0, \ldots, m \\ \text{set} & x_{k+1} = x_k + \alpha_k(\Delta x_0 + \Delta x_1 + \cdots + \Delta x_m)\,, & k = 0, 1, \ldots\,. \end{array} \tag{1.4}$$

Of course it is possible to introduce a different steplength control $\alpha_{k,i}$ for each correction Δx_i, $i = 0, \ldots, m$; however we have no need to consider such flexibility.

It is reasonably well known that, under the standard Newton's method assumptions, the level-m composite Newton method has a Q-convergence rate of $m + 2$. A proof can be found in Chapter 10 of Ortega and Rheinboldt (1970). The damped level-1 composite Newton method where one Newton step is composed with one simplified Newton step is of particular interest to us. It can be written

$$\begin{array}{lll} \text{solve} & F'(x_k)(\Delta x_N) = -F(x_k) & \text{for } \Delta x_N \\ \text{solve} & F'(x_k)(\Delta x_S) = -F(x_k + \Delta x_N) & \text{for } \Delta x_S \\ \text{set} & x_{k+1} = x_k + \alpha_k(\Delta x_N + \Delta x_S)\,, & k = 0, 1, \ldots,\,. \end{array} \tag{1.5}$$

Ortega and Rheinboldt (1970) credit the cubic convergence of the level-1 composite Newton method to Traub (1964). However, the notion of composing Newton steps with simplified Newton steps is much older and a part of the folklore of Newton's method. It is generally felt by practitioners that the formulation of composite Newton steps is of value when n is large and the function F can be evaluated cheaply; this is clearly the situation for the Newton interior-point method for linear programming described in Subsection 1.2.

Observe that each level-m composite Newton iterate can be viewed as a major iterate and is the result of $m + 1$ inner iterations. The average amount of algebra per inner iteration is $O((n^3 + mn^2)/\ (m + 1))$ and is $O(n^2)$ for large m. The average convergence rate for the inner iterates is the $(m + 1)$-st root of $m + 2$ and behaves like 1 for large m. It is no surprise then that for large m the level-m composite

Newton method behaves like the simplified Newton method. It follows that an optimal implementation of composite Newton would not only vary m at each Newton step but would keep m relatively small.

1.2 The Newton Interior-Point Method

Consider a linear program in the standard form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \tag{1.6}$$

where $c, x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ $(m < n)$ and A has full rank m.

The first-order optimality conditions for the linear program (1.6) can be written

$$F(x, y, \lambda) \equiv \begin{pmatrix} Ax - b \\ A^T\lambda + y - c \\ XYe \end{pmatrix} = 0\,, \quad (x, y) \geq 0 \tag{1.7}$$

where y and λ are dual variables, $X = \text{diag}(x)$, $Y = \text{diag}(y)$, and $e^T = (1, \ldots, 1) \in \mathbb{R}^n$.

The point (x, y, λ) is said to be feasible for problem (1.7) if $Ax = b$, $A^T\lambda + y - c = 0$, and $(x, y) \geq 0$. A feasible point (x, y, λ) is strictly feasible if $(x, y) > 0$. We tacitly assume that strictly feasible points exist.

It is now well understood how the primal-dual logarithmic barrier function interior-point method introduced by Megiddo (1989) and its numerous subsequent modifications can be stated in the framework of a damped and perturbed Newton's method applied to problem (1.7). For more detail, see Zhang, Tapia and Dennis (1990) for example. In presenting this algorithmic framework we will write $z = (x, y, \lambda)$, $\Delta z = (\Delta x, \Delta y, \Delta\lambda)$, $\Delta X = \text{diag}(\Delta x)$ and $\Delta Y = \text{diag}(\Delta y)$. We also let $\min(u)$ denote the smallest component of the vector u and $\hat{e}$ denote the vector $(0,\ldots,0,1,\ldots,1)^T$ where the number of zeros is $n+m$ and the number of ones is n. Finally, we will write the choice of the barrier parameter as $\mu(z)$ where z is the current iterate. When it is appropriate we write μ_k for $\mu(z_k)$.

Algorithm 1 (Newton Interior-Point Method)

Given $z_0 = (x_0, y_0, \lambda_0)$ with $(x_0, y_0) > 0$, for $k = 0, 1, \ldots$, do

(1) Solve $F'(z_k)(\Delta z) = -F(z_k) + \mu(z_k)\hat{e}$ for Δz (1.8)

(2) Choose $\tau_k \in (0,1)$ and set

$$\alpha_k = \min\left(1, \frac{-\tau_k}{\min(X_k^{-1}\Delta x_k)}, \frac{-\tau_k}{\min(Y_k^{-1}\Delta y_k)}\right) \tag{1.9}$$

(3) Set $z_{k+1} = z_k + \alpha_k \Delta z$

Actually in most implementations the formula (1.9) for α_k is further broken down and one steplength is used to update the x-variable and another is used to update the y-variable and the λ-variable. While this distinction is of value in practice; it is not an issue in the present work and consequently will be ignored.

Recently, under mild assumptions, Zhang, Tapia and Dennis (1990) demonstrated that for nondegenerate and degenerate problems Q-superlinear convergence could be attained by Algorithm 1 by merely letting $\sigma_k \to 0$ and $\tau_k \to 1$, where σ_k is defined by $\mu_k = \sigma_k x_k^T y_k/n$. Moreover, for nondegenerate problems Q-quadratic convergence could be attained by letting $\sigma_k = O(x_k^T y_k)$ and $\tau_k = 1 + O(x_k^T y_k)$.

For the current iterate z let $\Delta z_N = -F'(z)^{-1}F(z)$ denote the Newton step and let $\Delta z_\mu = \mu(z)F'(z)^{-1}\hat{e}$ denote the centering step. The Newton step can very likely point toward the boundary of the positive orthant, necessitating a very small choice for the steplength α. The major role of the centering step Δz_μ is to remedy this situation. Hence it seems quite reasonable that the choice for the barrier function parameter μ should also be a function of the Newton step Δz_N. The problem with implementing this idea is that the Newton step Δz_N must be calculated before the centering step Δz_μ is calculated. This entails an additional solve. However, we believe that this idea deserves further study. We are interested in a barrier parameter choice function of the form $\mu(z, \Delta z)$ and the following adaptive form of Algorithm 1, in which the barrier parameter μ is not only a function of the current iterate but also a function of the Newton step.

Adaptive Newton Interior-Point Method

The algorithm is the same as Algorithm 1, except that Step (1) is replaced with

$$\begin{aligned} (1) \qquad & \text{Solve } F'(z_k)(\Delta z_N) = -F(z_k) \text{ for } \Delta z_N \\ & \text{Solve } F'(z_k)(\Delta z_\mu) = \mu(z_k, \Delta z_N)\hat{e} \text{ for } \Delta z_\mu \\ & \text{Set } \Delta z = \Delta z_N + \Delta z_\mu \end{aligned} \tag{1.10}$$

1.3 The Predictor-Corrector Interior-Point Method

Mizuno, Todd, and Ye (1989) suggested and studied an algorithm which they labeled a predictor-corrector algorithm. In their algorithm the predictor step is a damped Newton step for problem (1.7), producing a new strictly feasible iterate. The subsequent corrector step is a centered Newton step. In this corrector step, the choice of μ, the barrier parameter, is based on the predictor step. Both the predictor and the corrector steps require essentially the same amount of work, namely, the evaluation and factorization of the Jacobian matrix.

Mehrotra (1989) later presented the following variant of Algorithm 1, which he also referred to as a predictor-corrector method. A common feature in these two predictor-corrector approaches is that the value of the barrier parameter in the corrector step depends on the predictor step. However, unlike Mizuno, Todd and Ye's corrector step, Mehrotra's corrector step does not evaluate a fresh Jacobian matrix. Instead, it reuses the Jacobian matrix used by the predictor step.

Algorithm 2 (Predictor-Corrector Interior-Point Method)

Given $z_0 = (x_0, y_0, \lambda_0)$ with $(x_0, y_0) > 0$, for $k = 0, 1, \ldots$ do

(1) Solve $F'(z_k)(\Delta z_p) = -F(z_k)$ for Δz_p

(2) Solve
$$F'(z_k)(\Delta z_c) = -\begin{pmatrix} 0 \\ 0 \\ \Delta X_p \Delta y_p \end{pmatrix} + \mu(z_k, \Delta z_p)\hat{e} \tag{1.11}$$
for Δz_c

(3) Set $\Delta z = \Delta z_p + \Delta z_c$

(4) Choose $\tau_k \in (0,1)$ and set
$$\alpha_k = \min\left(1\,, \frac{-\tau_k}{\min(X_k^{-1}\Delta x_k)}\,, \frac{-\tau_k}{\min(Y_k^{-1}\Delta y_k)}\right)$$

(5) Set $z_{k+1} = z_k + \alpha_k \Delta z$

While in the present section we are not concerned with the specific choice of the initial iterate z_0 or the various algorithmic parameters, we emphasize that Mehrotra suggested choices that allowed him to obtain very impressive numerical results.

1.4 The Composite Newton Interior-Point Method

In this subsection we present our composite Newton interior-point method for problem (1.7). As in the predictor-corrector methods of Mizuno, Todd, and Ye (1989) and Mehrotra (1989) we assume that the barrier parameter choice function is of the form $\mu(z, \Delta z)$. Recall that $\hat{e} = (0, \ldots, 0, 1, \ldots, 1)^T$.

Algorithm 3 (Level-m Composite Newton Interior-Point Method)

Given $z_0 = (x_0, y_0, \lambda_0)$ with $(x_0, y_0) > 0$ for $k = 0, 1, \ldots,$ do

(1) Solve $F'(z_k)(\Delta z_0) = -F(z_k)$ for Δz_0

(2) For $i = 1, \ldots, m$ do

$$\text{Solve } F'(z_k)(\Delta z_i) = -F(z_k + \sum_{j=0}^{i-1} \Delta z_j) + \mu_k(z_k, \sum_{j=0}^{i-1} \Delta z_j)\hat{e} \tag{1.12}$$

for Δz_i

(3) Set $\Delta z = \sum_{i=0}^{m} \Delta z_i$

(4) Choose $\tau_k \in (0, 1)$ and set

$$\alpha_k = \min\left(1, \frac{-\tau_k}{\min(X_k^{-1}\Delta x_k)}, \frac{-\tau_k}{\min(Y_k^{-1}\Delta y_k)}\right)$$

(5) Set $z_{k+1} = z_k + \alpha_k \Delta z$.

As was the case with the Newton interior-point method we will also be interested in the adaptive form of our algorithm.

1.4.1 Adaptive Composite Newton Interior-Point Method

The algorithm is the same as Algorithm 3, except that Steps (1) and (2) are replaced with

(1) Solve $F'(z_k)(\Delta z_C) = \hat{e}$ for the centering step Δz_C

(2) For $i = 0, \ldots, m$ do

Solve $F'(z_k)(\Delta z_i) = -F(z_k + \sum_{j=0}^{i-1} \Delta z_j)$ for Δz_i

Replace Δz_i with $\Delta z_i + \mu(z_k, \sum_{j=0}^{i} \Delta z_j)\Delta z_C$

Observe that the adaptive form of the Newton interior-point method and the adaptive form of the composite Newton interior-point method require one additional solve.

2 Predictor-Corrector as Composite Newton

We say that two algorithms are equivalent if given a current iterate they produce the same subsequent iterate for the same choice of common algorithmic parameters.

Theorem 2.1 *The predictor-corrector interior-point method (Algorithm 2) is equivalent to the level-1 composite Newton interior-point method (Algorithm 3).*

Proof. Let $z = (x, y, \lambda)$ be the current iterate and let $\Delta z_p = (\Delta x_p, \Delta y_p, \Delta\lambda_p)$ be the predictor step for problem (1.7), i.e., Δz_p is obtained from Step (1) of Algorithm 2. By comparing Algorithm 2 with Algorithm 3 ($m = 1$), we see that our proof will be complete once we show that

$$F(z + \Delta z_p) = \begin{pmatrix} 0 \\ 0 \\ \Delta X_p \Delta y_p \end{pmatrix} . \tag{2.1}$$

Writing (2.1) in further detail gives

$$A(x + \Delta x_p) - b = 0 \tag{2.2}$$

$$A^T(\lambda + \Delta\lambda_p) + (y + \Delta y_p) - c = 0 \tag{2.3}$$

$$[x + \Delta x_p]_i[y + \Delta y_p]_i = -[\Delta x_p]_i[\Delta y_p]_i \, , \quad i = 1, \dots, n \, . \tag{2.4}$$

By expanding (2.4) we see that (2.1) holds if and only if

$$[x]_i[y]_i + [y]_i[\Delta x_p]_i + [x]_i[\Delta y_p]_i = 0 \, , \quad i = 1, \dots, n \, . \tag{2.5}$$

However, (2.2), (2.3) and (2.5) are exactly the defining relations for the Newton step. □

While in Mehrotra (1989) no explanation for the predictor-corrector method is given, in a more recent paper, Mehrotra (1990), Mehrotra offers an interpretation of a related, but somewhat different, algorithm. Following the lead of Monteiro, Adler, and Resende (1988) he constructs a standard homotopy in a parameter, say δ, between problem (1.7) and a problem which had the current iteration as its solution. The primal-dual trajectory path parametrized by δ gives the solution of problem (1.7) for $\delta = 0$ and the current iterate for $\delta = 1$. He then views the iterate obtained from the predictor-corrector method as a point on a quadratic path which approximates the primal-dual trajectory path.

The equivalence represented by Theorem 2.1 was conjectured while listening to Mehrotra discuss his predictor-corrector method in Asilomar, California earlier this year. After proving Theorem 2.1 and while preparing this paper we received the recent paper of Lustig, Marsten, and Shanno (1990). In this paper the authors describe a comprehensive implementation of the Mehrotra predictor-corrector method and present impressive numerical results.

Lustig, Marsten and Shanno (1990) present Mehrotra's predictor-corrector method in the following manner. Rather than applying Newton's method to (1.7) to generate correction terms to the current iterate, substitute the new iterate into (1.8) directly, yielding

$$A(x + \Delta x) = b \tag{2.6a}$$

$$A^T(\lambda + \Delta\lambda) - (y + \Delta y) = c \tag{2.6b}$$

$$[x + \Delta x]_i[y + \Delta y]_i = \mu\ , \quad i = 1, \ldots, n\ . \tag{2.6c}$$

Simple algebra reduces (2.6) to the equivalent system

$$A\Delta x = b - Ax \tag{2.7a}$$

$$A^T\Delta\lambda - \Delta y = c - A^T\lambda + y \tag{2.7b}$$

$$[x]_i[\Delta y]_i + [y]_i[\Delta x]_i = \mu - [x]_i[y]_i - [\Delta x]_i[\Delta y]_i\ , \quad i = 1, \ldots, n\ . \tag{2.7c}$$

Observe that (2.7) defines the step $(\Delta x, \Delta y, \Delta\lambda)$ implicitly, i.e., in a nonlinear manner. In order to determine a step approximately satisfying (2.7) Mehrotra suggests first solving (1.7) for the Newton predictor step $(\Delta x_p, \Delta y_p, \Delta\lambda_p)$ and then using Δx_p and Δy_p on the right-hand side of (2.7). The new step is then obtained by solving (2.7) with this modified right-hand side.

It should be clear that presentation (2.7), with Δx and Δy replaced by Δx_p and Δy_p, reflects the level-1 composite Newton method written in the form

$$F'(z_k)\Delta z = -[F(z_k) + F(z_k - F'(z_k)^{-1}F(z_k))] \; ; \qquad (2.8)$$

while Mehrotra's presentation reflects the slightly different form (1.5).

Lustig, Marsten and Shanno (1990) attempt an explanation of the predictor-corrector notion in terms of trajectories parametrized by the parameter μ. Their explanation contains some ambiguity in that it is not clear to what trajectories they are referring. Moreover, any explanation based on issues derived from μ cannot give a complete picture, since the predictor-corrector notion still makes sense even when the problem formulation is free of μ, i.e. $\mu = 0$ in all cases. However, implicit in these authors' comments is the understanding that the corrector step can be viewed as a perturbed simplified Newton step.

The presentation given above of the Mehrotra predictor-corrector method implies a general Newton predictor-corrector method. We now abstract this implication.

Consider the nonlinear equation problem

$$F(z) = (f_1(z), \ldots, f_n(z))^T = 0 \, .$$

Suppose we have found g_i such that

$$f_i(z + \Delta z) = f_i(z) + f_i'(z)(\Delta z) + g_i(z, \Delta z) \, , \quad i = 1, \ldots, n \, . \qquad (2.9)$$

Then using the notation $G = (g_1, \ldots, g_n)^T$ we see from (2.9) that

$$F(z + \Delta z) = 0 \qquad (2.10)$$

can be written equivalently as

$$F'(z)(\Delta z) = -F(z) - G(z, \Delta z) . \tag{2.11}$$

The general Newton predictor-corrector method therefore consists of first computing the predictor step Δz_p as the Newton step, i.e., as the solution of $F'(z)(\Delta z) = -F(z)$, and then computing the composite step Δz_c as the solution of

$$F'(z)(\Delta z) = -F(z) - G(z, \Delta z_p) . \tag{2.12}$$

In the case of the Mehrotra predictor-corrector method for problem (1.7) we see from (2.6) and (2.7) (or directly) that

$$G(z, \Delta z) = \begin{pmatrix} 0 \\ 0 \\ \Delta X \Delta y \end{pmatrix} . \tag{2.13}$$

Now, observe that if Δz_N, the Newton step for F at z exists, then necessarily

$$F(z + \Delta z_N) = G(z, \Delta z_N) . \tag{2.14}$$

This fact follows immediately from (2.9) and the fact that $f_i'(z)(\Delta z_N) = -f_i(z)$. By substituting (2.14) into (2.12) we see that when the predictor step is the Newton step ($\Delta z_p = \Delta z_N$) we can write (2.12) as

$$F'(z)(\Delta z) = -F(z) - F(z - F'(z)^{-1}F(z)) . \tag{2.15}$$

Hence, our abstraction of the predictor-corrector notion is also no more than level-1 composite Newton.

We end this section with the following observation. From the development given of the Mehrotra predictor-corrector method in terms of (2.6) and (2.7) we see that the method was probably derived as a method which first identified the exact higher-order term $G(z, \Delta z)$ given by (2.13), and then used this representation of the higher-order information in the computation of the corrector step, i.e., used $G(z, \Delta z_p)$, where Δz_p is the Newton predictor step, in this computation. We therefore find it noteworthy that, in this sense, the level-1 composite Newton method uses exact higher-order information in all applications, even though the representation of the higher-order term $G(z, \Delta z)$ is unknown. It does this by working with $F(z+\Delta z)$ instead of $G(z, \Delta z)$, see (2.14).

3 Cubic Convergence

As before we consider problem (1.7) and use the notation $z = (x, y, \lambda)$. Also, recall that $\hat{e} = (0, \ldots, 0, 1, \ldots, 1)^T$ where the number of zeros is $n + m$ and the number of ones is n. The pure Newton method can be written

$$N(z) = z - F'(z)^{-1}F(z) \tag{3.1}$$

and the predictor-corrector interior-point method can be written

$$\hat{N}(z) = z - \alpha F'(z)^{-1}[F(z) + F(N(z)) - \mu\hat{e}] \,. \tag{3.2}$$

Therefore

$$\begin{aligned}
\hat{N}(z) - z_* &= z - z_* - F'(z)^{-1}[F(z) + F(N(z))] \\
&\quad +(1-\alpha)F'(z)^{-1}[F(z) + F(N(z))] + \alpha\mu F'(z)^{-1}\hat{e} \\
&= F'(z)^{-1}[F'(z)(N(z)) - F(N(z)) - F'(z)(z_*)] \\
&\quad +(1-\alpha)F'(z)[F(z) + F(N(z))] + \alpha\mu F'(z)^{-1}\hat{e} \\
&= -F'(z)^{-1}\{[F(N(z)) - F(z_*) - F'(z_*)(N(z) - z_*)] \\
&\quad +[F'(z_*) - F'(z)](N(z) - z_*)\} \\
&\quad +(1-\alpha)F'(z)[F(z) + F(N(z))] + \alpha\mu F'(z)^{-1}\hat{e}.
\end{aligned} \tag{3.3}$$

Now, locally, i.e. in a neighborhood of the solution z_*, we know from standard Newton's method analysis that

$$\|N(z) - z_*\| = O(\|z - z_*\|^2) .$$

See, for example, Dennis and Schnabel (1983) or Ortega and Rheinboldt (1970). Hence, we can rewrite the four groups of terms on the right-hand side in (3.3) and obtain

$$\|\hat{N}(z) - z_*\| = O(\|z - z_*\|^4) + O(\|z - z_*\|^3) + |1 - \alpha| O(\|z - z_*\|) + \mu O(1) ;$$

which simplifies to

$$\|\hat{N}(z) - z_*\| = O(\|z - z_*\|^3) + |1 - \alpha| O(\|z - z_*\|) + \mu O(1) . \tag{3.4}$$

In deriving (3.4) we used the fact that $\|F(z)\| = O(\|z - z_*\|)$ and $\|F(N(z))\| = O(\|z - z_*\|^2)$.

The term $\mu O(1)$ can be made $O(\|z - z_*\|^3)$ by the choice of μ. Everything now hinges on the term $|1 - \alpha| O(\|z - z_*\|)$. We must therefore take a very close look at the quantity $1 - \alpha$. Clearly, for cubic convergence, we need $|1 - \alpha|$ to be $O(\|z - z_*\|^2)$.

Assuming strict complementarity, z_* is a nondegenerate vertex solution, and z_k is feasible. Zhang, Tapia and Dennis (1990) obtained the useful expression

$$1 - \alpha_k = \frac{1 - \tau_k - \sigma_k \theta_k}{1 - \sigma_k \theta_k} + O(x_k^T y_k) \tag{3.5}$$

for the Newton interior-point method. See (3.7) of Zhang, Tapia and Dennis (1990). In (3.5), τ_k and σ_k are as in Algorithm 1, $\theta_k \in (\frac{1}{n}, 1]$ and $O(x_k^T y_k)$ is not necessarily zero and is exactly first order. Observe that $O(x_k^T y_k) = O(\|z - z_*\|)$, since for feasible z_k we have $x_k^T y_k = \|F(z_k)\|_1$.

For the present purpose of studying $1 - \alpha_k$, the Newton predictor-corrector interior point method and the Newton interior-point method are philosophically the same, i.e., both can be viewed as perturbed Newton. In the former case the perturbation to the right-hand side of the defining relation is $\mu\hat{e} - F(z - F'(z)^{-1}F(z))$, while in the latter case the perturbation is merely $\mu\hat{e}$. Observe that these two perturbation terms differ by a term which is order $O(\|z - z_*\|^2)$ or equivalently $O((x_k^T y_k)^2)$. Hence (3.5) is also valid for the Newton predictor-corrector interior-point method. It can now be seen from (3.5) that independent of the choices for τ_k and σ_k, the term $|1 - \alpha_k|$ is at best $O(\|z - z_*\|)$ and the Newton predictor-corrector interior-point method, even for nondegenerate problems, cannot be shown to be cubically convergent by the standard approach. However, by choosing $\alpha_k = 1$ near the solution and $\mu_k = O((x_k^T y_k)^3)$ we see from (3.4) that it is possible to obtain cubic convergence. We formally state these observations as the following theorem.

Theorem 3.1 *Let $\{x_k, y_k, \lambda_k\}$ be produced by Mehrotra's predictor-corrector interior-point method with z_0 strictly feasible. Assume*

(i) strict complementarity,

(ii) x_* is a nondegenerate vertex, and

(iii) $\{(x_k, y_k, \lambda_k)\}$ converges to (x_*, y_*, λ_*).

If the choices of σ_k and τ_k satisfy

$$0 \leq \sigma_k \leq min(\sigma, c_1(x_k^T y_k)) \tag{3.6}$$

and

$$0 < 1 - \tau_k \leq min(1 - \tau, c_2 x_k^T y_k) \tag{3.7}$$

where $\sigma \in [0, 1), \tau \in (0, 1)$ and $c_1, c_2 > 0$, then the convergence is Q-quadratic, i.e. there exist $\gamma_2 > 0$ such that for k large

$$\|(x_{k+1}, y_{k+1}, \lambda_{k+1}) - (x_*, y_*, \lambda_*)\| \leq \gamma_2 \|(x_k, y_k, \lambda_k) - (x_*, y_*, \lambda_*)\|^2 .$$

On the other hand, if instead of (3.6) we have

$$0 \leq \sigma_k \leq min(\sigma, c_1(x_k^T y_k)^2) \tag{3.8}$$

and instead of (3.7) we have that for large k

$$\alpha_k = 1 \ , \tag{3.9}$$

then the convergence is Q-cubic, i.e. there exist $\gamma_3 > 0$ *such that for k large*

$$\|(x_{k+1}, y_{k+1}, \lambda_{k+1}) - (x_*, y_*, \lambda_*)\| \leq \gamma_3 \|(x_k, y_k, \lambda_k) - (x_*, y_*, \lambda_*)\|^3 \ .$$

Proof. The proof follows from combining the discussion given above with details given in Zhang, Tapia, and Dennis (1990) for the proof of Theorem 4.1. □

4 Concluding Remarks

In this paper we have abstracted the Mehrotra Newton predictor-corrector philosophy and demonstrated that it is equivalent to the level-1 composite Newton philosophy.

We were intrigued by the discovery that, while the level-1 composite Newton method is known to be cubically convergent, this standard convergence rate proof applied to the predictor-corrector interior-point method gives at best quadratic convergence. The limitation of the standard proof results from the constrictive steplength choice forced on the method by the interior point philosophy, i.e., requiring the iterates to remain feasible with respect to the nonnegativity constraints. We demonstrated that if one drops the interior-point aspect of the predictor-corrector method locally, i.e., in a neighborhood of the solution steplength one is selected, and also chooses the barrier parameter to be of the order of the duality gap cubed, then cubic convergence can be attained for nondegenerate problems.

The research presented in Zhang, Tapia, and Dennis (1990), in Zhang, Tapia and Potra (1990), and the present research leads us to conjecture that we should implement Newton interior-point methods and their variants in a manner which near the solution sets the barrier parameter to zero and takes steplength one, i.e., as old-fashioned Newton. Our preliminary numerical experiments employing this idea were impressive and motivated the more general study described in the companion paper El-Bakry, Tapia and Zhang [2]. The reader is referred to that paper for numerical results.

References

[1] J.E. DENNIS Jr. and R.B. SCHNABEL. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, NJ, 1983. Russian edition, Mir Publishing Office, Moscow, 1988, O. Burdakov, translator.

[2] A. EL-BAKRY, R.A. TAPIA, and Y. ZHANG. Numerical comparisons of local convergence strategies for interior-point methods in linear programming. Technical Report TR91-18, Department of Mathematical Sciences, Rice University, Houston, Texas 77151-1892, September 1991.

[3] I.J. LUSTIG, R.E. MARSTEN, and D.F. SHANNO. On implementing Mehrotra's predictor-corrector interior point method for linear programming. Technical Report SOR 90-03, Department of Civil Engineering and Operations Research, Princeton University, Princeton, New Jersey 08544, April 1990.

[4] N. MEGIDDO. Pathways to the optimal set in linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming, Interior Point and Related Methods*, pages 131–158. Springer-Verlag, New York, 1989.

[5] S. MEHROTRA. On finding a vertex solution using interior-point methods. Technical Report 89-22, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, 1989. Revised January 1990.

[6] S. MEHROTRA. On the implementation of a (primal-dual) interior point method. Technical Report 90-03, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, March 1990.

[7] R.D.C. MONTEIRO, I. ADLER, and M.G.C. RESENDE. A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension. ESRC Report 88-8, Department of Industrial Engineering and Operations Research, University of California, Berkeley, California 94720, March 1988.

[8] S.MIZUNO, M.J. TODD and Y. YE. Anticipated behavior of the path following algorithms for linear programming. Technical Report No. 878, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853, December 1989.

[9] J.M. ORTEGA and W.C. RHEINBOLDT. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York, NY, 1970.

[10] J. TRAUB. *Iterative Methods for the Solution of Equations.* Prentice Hall, Englewood Cliffs, New Jersey, 1964.

[11] Y. ZHANG, R.A. TAPIA, and J.E. DENNIS. On the superlinear and quadratic convergence of primal-dual interior-point linear programming algorithms. Technical Report TR90-6, Department of Mathematical Sciences, Rice University, Houston, Texas 77251-1892, January 1990. To appear in *SIAM J. Optimization.*

[12] Y. ZHANG, R.A. TAPIA, and F. POTRA. On the superlinear convergence of interior-point algorithms for a general class of problems. Technical Report TR90-9, Department of Mathematical Sciences, Rice University, Houston, Texas 77251-1892, March 1990. To appear in *SIAM J. Optimization.*