A note on using alternative second-order models for the subproblems arising in barrier function methods for minimization

by A. R. Conn[1], Nick Gould[2], and Ph. L. Toint[3]

Report 93/16 (revised)

October 13, 1993

This paper is dedicated to Professor J. Stoer on the occasion of his sixtieth birthday.

Abstract. Inequality constrained minimization problems are often solved by considering a sequence of parameterized barrier functions. Each barrier function is approximately minimized and the relevant parameters subsequently adjusted. It is common for the estimated solution to one barrier function problem to be used as a starting estimate for the next. However, this has unfortunate repercussions for the standard Newton-like methods applied to the barrier subproblem. In this note, we consider a class of alternative Newton methods which attempt to avoid such difficulties. Such schemes have already proved of use in the Harwell Subroutine Library quadratic programming codes `VE14` and `VE19`.

[1] IBM T.J. Watson Research Center, P.O.Box 218, Yorktown Heights, NY 10598, USA
Email : arconn@watson.ibm.com

[2] CERFACS, 42 Avenue Gustave Coriolis, 31057 Toulouse Cedex, France, EC
Email : gould@cerfacs.fr or nimg@directory.rl.ac.uk

[3] Department of Mathematics, Facultés Universitaires ND de la Paix,
61, rue de Bruxelles, B-5000 Namur, Belgium, EC
Email : pht@math.fundp.ac.be

Keywords : Large-scale problems, nonlinear optimization, numerical algorithms.

Mathematics Subject Classifications : 65K05, 90C30

A note on using alternative second-order models for the subproblems arising in barrier function methods for minimization

A. R. Conn, Nick Gould and Ph. L. Toint

October 13, 1993

Abstract

Inequality constrained minimization problems are often solved by considering a sequence of parameterized barrier functions. Each barrier function is approximately minimized and the relevant parameters subsequently adjusted. It is common for the estimated solution to one barrier function problem to be used as a starting estimate for the next. However, this has unfortunate repercussions for the standard Newton-like methods applied to the barrier subproblem. In this note, we consider a class of alternative Newton methods which attempt to avoid such difficulties. Such schemes have already proved of use in the Harwell Subroutine Library quadratic programming codes `VE14` and `VE19`.

This paper is dedicated to Professor J. Stoer on the occasion of his sixtieth birthday.

1 Introduction

In this note, we consider solving the problem

$$\underset{x \in \Re^n}{\text{minimize}} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad c_i(\boldsymbol{x}) \geq 0 \quad \text{for} \quad i = 1, \ldots, m \tag{1.1}$$

using a sequence of (logarithmic) barrier functions

$$\Psi(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s}) = f(\boldsymbol{x}) - \sum_{i=1}^{m} w_i \log(c_i(\boldsymbol{x}) + s_i), \tag{1.2}$$

where the w_i are termed *weights* and the s_i called *shifts*. Here both f and the c_i are assumed to be at least twice-continuously differentiable. Traditional, unshifted (i.e., $\boldsymbol{s} = 0$), barrier functions of this form were first considered by Frisch (1955), popularized by Fiacco and McCormick (1968) and extensively studied by Wright (1976), Murray and Wright (1978), Karmarkar (1984), Gill *et al.* (1986), Gould (1986), McCormick (1991), Nash and Sofer (1993) and Wright (1992b) amongst others. Variations on the theme include the modified (unshifted) barrier function of Jittorntrum and Osborne (1980), the shifted barrier functions of Gill *et al.* (1988) and Freund (1991), the modified (shifted) barrier function of Polyak (1992) and the Lagrangian barrier function of Conn *et al.* (1992a).

A typical barrier function method attempts to solve (1.1) by (approximately) minimizing a sequence of barrier functions $\Psi(\boldsymbol{x}, \boldsymbol{w}^{(k)}, \boldsymbol{s}^{(k)})$ for appropriate sequences of weights $\{\boldsymbol{w}^{(k)}\}$ and shifts $\{\boldsymbol{s}^{(k)}\}$. The approximate minimizer of $\Psi(\boldsymbol{x}, \boldsymbol{w}^{(k)}, \boldsymbol{s}^{(k)})$ is generally found by applying an iterative unconstrained minimization method — the inner iteration – to Ψ. Usually, the minimization is terminated when an *approximate stationary point*, $\boldsymbol{x}^{(k)}$, of Ψ is determined. Such a point is required to satisfy the inner-iteration stopping rule

$$\|\nabla_x \Psi(\boldsymbol{x}^{(k)}, \boldsymbol{w}^{(k)}, \boldsymbol{s}^{(k)})\| \leq \omega^{(k)} \tag{1.3}$$

for some sequence of positive tolerances $\{\omega^{(k)}\}$ which converge to zero. Of course, this does not guarantee that $\boldsymbol{x}^{(k)}$ is actually a minimizer of $\Psi(\boldsymbol{x}, \boldsymbol{w}^{(k)}, \boldsymbol{s}^{(k)})$ even if $\{\omega^{(k)}\}$ is zero — second order sufficiency assumptions would be needed to ensure this — but tests of the form (1.3) are commonplace (see, for example, Polyak, 1992 and Conn *et al.*, 1992a).

The bulk of the work is performed in the inner iteration. As each inner iteration is clearly influenced by the choice of starting point, there is some interest in trying to determine good starting points. Since, under relatively mild conditions (see, for example, Wright, 1992b, Theorem 8), it can be shown that the sequence $\{\boldsymbol{x}^{(k)}\}$ converges to a first-order stationary point for (1.1), one might imagine that $\boldsymbol{x}^{(k)}$ provides a good starting point for the $k+1$-st inner iteration, especially as the solution is approached. In some sense this is true. However, Wright (1993) has shown that if one naively uses Newton's method to solve the inner iteration subproblem starting from this point, difficulties may arise. In particular, she shows that it is highly likely that a full Newton step will be impossible as this step crosses the constraint boundary.

In this note we show that this difficulty arises because the standard Newton method is actually an asymptotically inappropriate member of a whole class of Newton methods for the subproblem. A different member of the class is then proposed which aims to alleviate the aforementioned difficulty as the solution is approached. We indicate that the proposed alternative is effective in practice when applied to three different barrier function methods for solving bound-constrained quadratic programs.

After stating our notation in Section 2, we consider the difficulties associated with the standard Newton method and propose some alternatives in Section 3. These alternatives are considered in detail for a number of common barrier functions in Section 4 and numerical results indicating their effectiveness are provided in Section 5.

Our intention here is not to produce a completely rigorous theory, merely to indicate the possible advantages of viewing Newton's method from different perspectives. To these ends, we assume, where necessary, that all exhibited square systems of equations are nonsingular. While it is easy to derive examples which defeat such assumptions, we believe that such non-singular problems occur sufficiently frequently in practice to make the conclusions drawn here useful.

2 Notation

We let $\boldsymbol{g}(\boldsymbol{x})$ denote the gradient $\nabla_x f(\boldsymbol{x})$, $\boldsymbol{a}_i(\boldsymbol{x})$ denote the gradient $\nabla_x c_i(\boldsymbol{x})$, $\boldsymbol{A}(\boldsymbol{x})$ be the Jacobian matrix whose rows are $\boldsymbol{a}_i(\boldsymbol{x})^T$ and $\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla_{xx} f(\boldsymbol{x}) - \sum_{i=1}^m \lambda_i \nabla_{xx} c_i(\boldsymbol{x})$ be the Hessian matrix of the Lagrangian function associated with (1.1). We shall denote the diagonal matrix whose i-th diagonal component is d_i by $\boldsymbol{D}[d_i]$; the dimension of $\boldsymbol{D}$ should be obvious from the context. We then have that

$$\nabla_x \Psi(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s}) = \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{A}(\boldsymbol{x})^T \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s}) \tag{2.1}$$

and

$$\nabla_{xx} \Psi(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s}) = \boldsymbol{H}(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s})) + \boldsymbol{A}(\boldsymbol{x})^T \boldsymbol{D}[u_i(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s})/(c_i(\boldsymbol{x}) + s_i)] \boldsymbol{A}(\boldsymbol{x}), \tag{2.2}$$

where the Lagrange multiplier function $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s})$ is defined, componentwise, as

$$u_i(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s}) \stackrel{\text{def}}{=} \frac{w_i}{c_i(\boldsymbol{x}) + s_i} \quad \text{for} \quad i = 1, \ldots, m. \tag{2.3}$$

We also need to compare the relative rates of convergence of sequences of numbers. If $\{\omega^{(k)}\}$ and $\{\eta^{(k)}\}$ are two sequences of positive numbers converging to zero as k tends to infinity, we say that $\omega^{(k)} = O(\eta^{(k)})$ if there exists a constant κ such that $\omega^{(k)} \leq \kappa \eta^{(k)}$ for all k. If $\omega^{(k)} = O(\eta^{(k)})$ and $\eta^{(k)} = O(\omega^{(k)})$, we say that $\omega^{(k)} = \Theta(\eta^{(k)})$. We also say that $\omega^{(k)} = o(\eta^{(k)})$ if there is a third sequence $\{\kappa^{(k)}\}$ of positive scalars converging to zero as k tends to infinity such that $\omega^{(k)} \leq \kappa^{(k)} \eta^{(k)}$ for all k.

3 Motivation

Suppose that we have obtained $\boldsymbol{x}^{(k)}$ satisfying (1.3) and now wish to solve the $k+1$-st inner iteration subproblem (with some given $\boldsymbol{w}^{(k+1)}$ and $\boldsymbol{s}^{(k+1)}$) starting from $\boldsymbol{x}^{(k)}$. Furthermore, suppose also that we intend using Newton's method — or, more precisely, a globally convergent Newton method — to solve the inner iteration subproblem. That is, we obtain our first correction to $\boldsymbol{x} = \boldsymbol{x}^{(k)}$ by determining the Newton search direction $\delta \boldsymbol{x}^N$ from the *standard Newton equations*,

$$\nabla_{xx} \Psi(\boldsymbol{x}, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) \delta \boldsymbol{x}^N = -\nabla_x \Psi(\boldsymbol{x}, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) \tag{3.1}$$

and subsequently performing a linesearch in this direction. We refer to the solution, $\delta \boldsymbol{x}^N$, as the *standard Newton direction* or *correction* and an iteration based on this direction as a *standard Newton iteration*. There are a number of potential dangers.

1. The Newton step may not be possible as the value $\boldsymbol{x}^{(k)} + \delta \boldsymbol{x}^N$ may violate one or more of the "shifted" constraints $\boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{s}^{(k)} > 0$. Thus a restricted step will be necessary and the rapid convergence of Newton's method will be thwarted.

2. $\nabla_{xx} \Psi(\boldsymbol{x}^{(k)}, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)})$ may be badly conditioned making an accurate solution of the Newton equations difficult.

3. There is no guarantee that, just because (1.3) ensures that $\nabla_x \Psi(\boldsymbol{x}^{(k)}, \boldsymbol{w}^{(k)}, \boldsymbol{s}^{(k)})$ is small for large k, the same is true for $\nabla_x \Psi(\boldsymbol{x}^{(k)}, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)})$. Hence, an undamped (unit stepsize) Newton process may require a significant number of steps before a suitable $\boldsymbol{x}^{(k+1)}$ satisfying (1.3) is determined.

We consider the following alternative, based on the method proposed by Gould (1989) for obtaining superior starting points for the quadratic penalty function. Ideally, we wish to determine a value $\boldsymbol{x}$ for which

$$\nabla_x \Psi(\boldsymbol{x}, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) = \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{A}(\boldsymbol{x})^T \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) = 0, \tag{3.2}$$

where

$$u_i(\boldsymbol{x}, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) = \frac{w_i^{(k+1)}}{c_i(\boldsymbol{x}) + s_i^{(k+1)}}, \quad \text{for} \quad i = 1, \ldots, m, \tag{3.3}$$

for then the inner-iteration stopping rule (1.3) will be satisfied for any choice of $\omega^{(k+1)}$. The Newton direction (3.1) is obtained by computing the Newton correction for (3.2) while directly eliminating $\boldsymbol{u}(\boldsymbol{x})$ via (3.3).

Suppose instead that we consider *independent* variables $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ which are required to satisfy the equations

$$\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{A}(\boldsymbol{x})^T \boldsymbol{\lambda} = 0, \tag{3.4}$$

and

$$(c_i(\boldsymbol{x}) + s_i^{(k+1)})\lambda_i - w_i^{(k+1)} = 0, \quad \text{for} \quad i = 1, \ldots, m. \tag{3.5}$$

Clearly, if $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ satisfy (3.5), then $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) = \boldsymbol{\lambda}$. Moreover, the two systems of nonlinear equations (3.2)–(3.3) and (3.4)–(3.5) have identical solutions $\boldsymbol{x}$.

Suppose that $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ are estimates of the solution of (3.4)–(3.5). If we write down the Newton equations for the corrections $\delta \boldsymbol{x}$ and $\delta \boldsymbol{\lambda}$ to these estimates, the corrections satisfy the equations

$$\begin{pmatrix} \boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\lambda}) & -\boldsymbol{A}(\boldsymbol{x})^T \\ \boldsymbol{D}[\lambda_i]\boldsymbol{A}(\boldsymbol{x}) & \boldsymbol{D}[c_i(\boldsymbol{x}) + s_i^{(k+1)}] \end{pmatrix} \begin{pmatrix} \delta \boldsymbol{x} \\ \delta \boldsymbol{\lambda} \end{pmatrix} = - \begin{pmatrix} \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{A}(\boldsymbol{x})^T \boldsymbol{\lambda} \\ \boldsymbol{D}[(c_i(\boldsymbol{x}) + s_i^{(k+1)})\lambda_i - w_i^{(k+1)}]\boldsymbol{e} \end{pmatrix}, \tag{3.6}$$

where $\boldsymbol{e}$ is a vector of ones. Eliminating the variables $\delta \boldsymbol{\lambda}$ and rearranging, we obtain

$$\Big(\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\lambda}) + \boldsymbol{A}(x)^T \boldsymbol{D}[\lambda_i/(c_i(\boldsymbol{x}) + s_i^{(k+1)})]\boldsymbol{A}(\boldsymbol{x})\Big)\, \delta \boldsymbol{x} = -\nabla_x \Psi(\boldsymbol{x}, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}). \tag{3.7}$$

We see that, although the choice of $\boldsymbol{\lambda}$ does not affect the right-hand-side of (3.7), it most certainly influences the left-hand-side. Moreover, the effect on the right-hand-side of (3.6) may be significant. We note that we are not proposing that $\boldsymbol{\lambda}$ should be changed using the correction $\delta \boldsymbol{\lambda}$ from (3.6), but we are merely using the equations (3.6) to derive an alternative Newton correction to $\boldsymbol{x}^{(k)}$. We also note that, although they are mathematically equivalent, there may sometimes be good numerical reasons why one might prefer to solve (3.6) rather than (3.7) (see, for instance, Gould, 1986).

The standard Newton equations (3.1) corresponds to the choice $\boldsymbol{x} = \boldsymbol{x}^{(k)}$ and

$$\lambda_i = \lambda_i^N \stackrel{\text{def}}{=} \frac{w_i^{(k+1)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k+1)}} \quad \text{for} \quad i = 1, \ldots, m \tag{3.8}$$

(c.f. (2.1)–(2.3)). Notice that, with this choice, the first term on the right-hand-side of (3.6) may be quite large while the second term vanishes.

An alternative, and potentially better, choice is obtained by selecting $\boldsymbol{x} = \boldsymbol{x}^{(k)}$ and

$$\lambda_i = \lambda_i^A \stackrel{\text{def}}{=} \frac{w_i^{(k)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}} \quad \text{for} \quad i = 1, \ldots, m. \tag{3.9}$$

For then, we see from (1.3) that the first term on the right-hand-side of (3.6) is arbitrarily small for sufficiently large k. Furthermore, each component of the second,

$$\begin{aligned} r_i^{(k+1)} &\stackrel{\text{def}}{=} w_i^{(k)} \frac{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k+1)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}} - w_i^{(k+1)} \\ &= \frac{w_i^{(k)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}} \left(c_i(\boldsymbol{x}^{(k)}) \left(1 - \frac{w_i^{(k+1)}}{w_i^{(k)}} \right) + \left(\frac{s_i^{(k+1)}}{s^{(k)}}{}_i - \frac{w_i^{(k+1)}}{w_i^{(k)}} \right) s_i^{(k)} \right), \end{aligned} \tag{3.10}$$

for $i = 1, \ldots, m$, will be small if $\{s_i^{(k)}\}$ and $\{w_i^{(k)}\}$ converge while the remaining terms remain finite. Thus, in this case, one would expect the Newton iteration (3.6) to converge rapidly. We refer to the solution, $\delta \boldsymbol{x}^A$, of (3.7), when $\boldsymbol{x} = \boldsymbol{x}^{(k)}$ and $\boldsymbol{\lambda}$ is given by (3.9), as the *alternative* Newton direction at $\boldsymbol{x}^{(k)}$.

We note that the unshifted variant of the equations (3.6) form the basis of a whole class of nonlinear primal-dual methods for convex optimization problems proposed by McCormick (1991). Unlike the methods considered here, such methods explicitly use the the corrections $\delta\boldsymbol{\lambda}$ to construct improved Lagrange multiplier estimates but may require extra precautions to ensure that such estimates are strictly positive.

We must also consider the value of the shifted constraints after such a Newton correction, $\boldsymbol{c}(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x}) + \boldsymbol{s}^{(k+1)}$. A Taylor's expansion around $\boldsymbol{x}^{(k)}$ yields

$$\boldsymbol{c}(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x}) + \boldsymbol{s}^{(k+1)} = \boldsymbol{c}(\boldsymbol{x}^{(k)}) + \boldsymbol{s}^{(k+1)} + \boldsymbol{A}(\boldsymbol{x}^{(k)})\delta \boldsymbol{x} + O(\|\delta \boldsymbol{x}\|^2). \tag{3.11}$$

In order to assess the (shifted) feasibility (or otherwise) of $\boldsymbol{x}^{(k)} + \delta \boldsymbol{x}$, it is thus important to determine the size of $\boldsymbol{A}(\boldsymbol{x}^{(k)})\delta \boldsymbol{x}$. We now consider various methods in detail.

4 Shifts and weights

A large variety of shifts and weights have been proposed for barrier function methods. Our concern here is the relationship between the shifts and weights for one inner iteration and the next and its implication for the ease of solving successive inner-iteration subproblems.

In this section, we consider a number of different barrier function methods. For each, we analyze the size of the right-hand-side of the expanded Newton system (3.6). We next consider the feasibility of the constraints after taking Newton steps (3.7) corresponding to the two choices

(3.8) and (3.9) for $\boldsymbol{\lambda}$. We then show that the latter choice does not suffer from this drawback. An analysis of the size of the gradient of the barrier function at the new point following such a step indicates that the new point is a good one to start a standard Newton iteration. Indeed, for the last class of methods considered, this new point will asymptotically satisfy the inner iteration stopping rule and thus a single inner iteration will eventually suffice for each outer iteration.

4.1 Traditional barrier function methods

In these methods the shifts are zero. The weights are typically given as

$$w_i^{(k+1)} = \mu^{(k+1)} \quad \text{for} \quad i = 1, \ldots, m, \tag{4.1}$$

where the positive sequence of penalty parameters $\{\mu^{(k+1)}\}$ monotonically converge to zero. In this case, (3.10) is

$$r_i^{(k+1)} = \left(1 - \frac{\mu^{(k+1)}}{\mu^{(k)}}\right) \mu^{(k)}, \quad \text{for} \quad i = 1, \ldots, m, \tag{4.2}$$

and hence one would expect the Newton iteration (3.6)/(3.9) to yield a significant improvement for sufficiently large k.

In particular, if, as is normal, $\omega^{(k)} = \kappa_1 \mu^{(k)}$ for some constant κ_1, (1.3), (3.10) and (4.2) imply that the right-hand-side of (3.6) is $O(\mu^{(k)})$. Consequently, in this case, provided the coefficient matrix of (3.6) has a bounded inverse, one would expect that $\delta \boldsymbol{x} = O(\mu^{(k)})$ and $\delta \boldsymbol{\lambda} = O(\mu^{(k)})$ and that

$$\boldsymbol{g}(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x}) - \boldsymbol{A}(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x})^T (\boldsymbol{\lambda} + \delta \boldsymbol{\lambda}) = O(\mu^{(k)2}) \tag{4.3}$$

and

$$c_i(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x})(\lambda_i + \delta \lambda_i) = w_i^{(k+1)} + O(\mu^{(k)2}), \tag{4.4}$$

for $i = 1, \ldots, m$. It then follows from (2.3) and (4.4) that

$$\begin{aligned} \lambda_i + \delta \lambda_i &= \frac{w_i^{(k+1)}}{c_i(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x})} + O\left(\frac{\mu^{(k)2}}{c_i(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x})}\right) \\ &= u_i(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x}, \boldsymbol{w}^{(k+1)}, 0) + O\left(\frac{\mu^{(k)2}}{c_i(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x})}\right), \end{aligned} \tag{4.5}$$

for $i = 1, \ldots, m$.

Now consider $\boldsymbol{c}(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x})$. For the inactive constraints at a limit point $\boldsymbol{x}^*$ of $\{\boldsymbol{x}^{(k)}\}$, that is, those constraints for which $c_i(\boldsymbol{x}^*) > 0$, it follows, from the previous observation that $\delta \boldsymbol{x} = O(\mu^{(k)})$ and the continuity of c_i, that $c_i(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x})$ is bounded away from zero for all k sufficiently large. We thus have that

$$\tfrac{1}{2} c_i(\boldsymbol{x}^{(k)}) \leq c_i(\boldsymbol{x}^{(k)} + \delta \boldsymbol{x}) \leq 2 c_i(\boldsymbol{x}^{(k)}), \tag{4.6}$$

for all inactive constraints and all sufficiently large k. It remains to consider the active constraints, that is those for which $c_i(\boldsymbol{x}^*) = 0$. We argue in the same way as Wright (1993).

Consider (3.7) as $\mu^{(k)}$ converges to zero. The coefficient matrix will be dominated by

$$\boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})^T \boldsymbol{D}_{\mathcal{A}}[\lambda_i / c_i(\boldsymbol{x}^{(k)})] \boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)}), \tag{4.7}$$

where $\boldsymbol{A}_{\mathcal{A}}$ is the matrix whose rows are the $\boldsymbol{a}_i^T(\boldsymbol{x}^{(k)})$ corresponding to the active constraints and $\boldsymbol{D}_{\mathcal{A}}$ is the diagonal matrix whose entries are those of $\boldsymbol{D}$ for the active constraints. Likewise, because of the relationship (1.3), the right-hand-side of (3.7) is dominated by

$$-\boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})^T\boldsymbol{D}_{\mathcal{A}}[(w_i^{(k)}-w_i^{(k+1)})/c_i(\boldsymbol{x}^{(k)})]\boldsymbol{e}, \tag{4.8}$$

for small $\mu^{(k)}$ and thus (3.7) is (approximately)

$$\boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})^T\boldsymbol{D}_{\mathcal{A}}[\lambda_i/c_i(\boldsymbol{x}^{(k)})]\boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})\delta\boldsymbol{x} \approx -\boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})^T\boldsymbol{D}_{\mathcal{A}}[(w_i^{(k)}-w_i^{(k+1)})/c_i(\boldsymbol{x}^{(k)})]\boldsymbol{e}. \tag{4.9}$$

Now assume that $\boldsymbol{A}(x)_{\mathcal{A}}$ is full-rank. Then it follows from (4.9) that

$$\boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})\delta\boldsymbol{x} \approx -\boldsymbol{D}_{\mathcal{A}}[(w_i^{(k)}-w_i^{(k+1)})/\lambda_i]\boldsymbol{e}. \tag{4.10}$$

As Wright (1993) observes, the relationships (3.8), (4.1) and (4.10) imply that

$$\boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})\delta\boldsymbol{x}^N \approx (1-\mu^{(k)}/\mu^{(k+1)})\boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}^{(k)}), \tag{4.11}$$

when $\lambda_i = \lambda_i^N$, where $\delta\boldsymbol{x}^N$ is the standard Newton correction and where $\boldsymbol{c}_{\mathcal{A}}$ denotes the vector of active constraints, In this case, (3.11) and (4.11) then give that

$$\boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}^{(k)}+\delta\boldsymbol{x}^N) \approx (2-\mu^{(k)}/\mu^{(k+1)})\boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}^{(k)}), \tag{4.12}$$

which will be negative if $\mu^{(k+1)}$ is significantly smaller than $\frac{1}{2}\mu^{(k)}$. As it is normal to reduce $\mu^{(k)}$ by significantly more than a half, it must be expected, as Wright (1993) indicated, that a full Newton step (3.1) will be infeasible.

If, on the other hand, we choose $\lambda_i = \lambda_i^A$ from (3.9), the relationships (3.9), (4.1) and (4.10) imply that

$$\boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})\delta\boldsymbol{x}^A \approx -(1-\mu^{(k+1)}/\mu^{(k)})\boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}^{(k)}), \tag{4.13}$$

where $\delta\boldsymbol{x}^A$ is the alternative Newton correction. Combining (3.11) and (4.13), we see that

$$\boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}^{(k)}+\delta\boldsymbol{x}^A) \approx \frac{\mu^{(k+1)}}{\mu^{(k)}}\boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}^{(k)}), \tag{4.14}$$

which indicates that a step in the alternative direction will be safely feasible. Moreover, in this case, as both the inactive and active constraints are bounded away from zero at $\boldsymbol{x}^{(k)}+\delta\boldsymbol{x}^A$ for fixed, but sufficiently large k, we have from (4.14) and the convergence of $\mu^{(k)}/c_i(\boldsymbol{x}^{(k)})$ to the Lagrange multiplier λ_i^* under appropriate assumptions (see, e.g. Wright, 1992b, Theorem 8) that

$$\frac{\mu^{(k)}}{c_i(\boldsymbol{x}^{(k)}+\delta\boldsymbol{x}^A)} \approx \frac{\mu^{(k)}}{c_i(\boldsymbol{x}^{(k)})}\frac{\mu^{(k)}}{\mu^{(k+1)}} \approx \lambda_i^*\frac{\mu^{(k)}}{\mu^{(k+1)}}, \tag{4.15}$$

for all active constraints. Likewise, from (4.6),

$$\frac{\mu^{(k)}}{2c_i(\boldsymbol{x}^{(k)})} \leq \frac{\mu^{(k)}}{c_i(\boldsymbol{x}^{(k)}+\delta\boldsymbol{x}^A)} \leq 2\frac{\mu^{(k)}}{c_i(\boldsymbol{x}^{(k)})} \tag{4.16}$$

for the inactive constraints, and all such terms converge to zero as the Lagrange multipliers for these are zero. Thus, combining (4.5), (4.15) and (4.16), we see that

$$\lambda_i + \delta\lambda_i = u_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A, \boldsymbol{w}^{(k+1)}, 0) + O(\mu^{(k)2}/\mu^{(k+1)}), \tag{4.17}$$

for $i = 1, \ldots, m$, and hence, from (2.1) and (4.3), that

$$\nabla_x\Psi(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A, \boldsymbol{w}^{(k+1)}, 0) = O(\mu^{(k)2}/\mu^{(k+1)}). \tag{4.18}$$

In view of the estimate (4.18), one now expects the standard Newton iteration to converge fast when started from $\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A$ so long as the sequence $\{\mu^{(k)}\}$ does not converge to zero too fast. In particular, following (4.18), one would expect that the gradient of the barrier function after a single standard Newton iteration from this point would be asymptotically $O(\mu^{(k)4}/\mu^{(k+1)2})$ and thus the stopping rule (1.3) to be satisfied at such a point so long as $\mu^{(k)} = o(\mu^{(k+1)\frac{3}{4}})$. An analogous result for the quadratic penalty function was given by Gould (1989). Similar means of avoiding the poor behaviour of Newton's method following a reduction in the penalty parameter in traditional barrier function methods have proposed by Fiacco and McCormick (1968) and Jarre *et al.* (1988) in addition to the aforementioned work by Wright (1993). Murray (1969) and Bartholomew-Biggs (1972) give related methods for the quadratic penalty function.

Although the Hessian of the barrier function is likely to be ill-conditioned near the constraint boundaries, a number of schemes have been proposed for accurately solving the Newton equations. We refer the interested reader to the papers by Wright (1976), Murray and Wright (1978), Gould (1986), McCormick (1991) and Wright (1992a).

4.2 Jittorntrum and Osborne's modified barrier function method

In this method, the shifts are, once again, zero and the weights satisfy the relationship

$$w_i^{(k+1)} = \frac{\mu^{(k+1)}}{c_i(\boldsymbol{x}^{(k)})} w_i^{(k)}, \quad \text{for} \quad i = 1, \ldots, m, \tag{4.19}$$

where the positive sequence of penalty parameters $\{\mu^{(k+1)}\}$ monotonically converge to zero. In this case, (3.10) is

$$r_i^{(k+1)} = w_i^{(k)} - \mu^{(k+1)} \frac{w_i^{(k)}}{c_i(\boldsymbol{x}^{(k)})} \quad \text{for} \quad i = 1, \ldots, m. \tag{4.20}$$

Moreover, under mild conditions, one has that each $w_i^{(k)} = O(\mu^{(k)})$ and $\{w_i^{(k)}/c_i(\boldsymbol{x}^{(k)})\}$ converges to a Lagrange multiplier λ_i^* (see, Jittorntrum and Osborne, 1980). Therefore $r_i^{(k+1)} = O(\mu^{(k)})$ and, once again, one expects the Newton iteration (3.6)/(3.9) to yield a significant improvement for large k.

If, as in the previous section, $\omega^{(k)} = \kappa_1\mu^{(k)}$ for some constant κ_1, much of the analysis of that section remains valid. In particular, (4.3)–(4.6) and, under the same full-rank assumption, (4.10) hold. Combining (4.10) and (4.19), we obtain

$$\boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})\delta\boldsymbol{x} \approx -\boldsymbol{D}_{\mathcal{A}} \left[\left(1 - \frac{\mu^{(k+1)}}{c_i(\boldsymbol{x}^{(k)})} \right) \frac{w_i^{(k)}}{\lambda_i} \right] \boldsymbol{e}. \tag{4.21}$$

Hence, if $\lambda_i = \lambda_i^N$ from (3.8), the relationships (4.19) and (4.21) imply that

$$\begin{aligned} \boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})\delta\boldsymbol{x}^N \quad &\approx -\boldsymbol{D}_{\mathcal{A}}\left[c_i(\boldsymbol{x}^{(k)})\left(1 - \frac{\mu^{(k+1)}}{c_i(\boldsymbol{x}^{(k)})}\right)\frac{w_i^{(k)}}{w_i^{(k+1)}}\right]\boldsymbol{e} \\ &= \boldsymbol{c}(\boldsymbol{x}^{(k)}) - \boldsymbol{D}_{\mathcal{A}}[c_i(\boldsymbol{x}^{(k)})/\mu^{(k+1)}])\boldsymbol{c}(\boldsymbol{x}^{(k)}), \end{aligned} \tag{4.22}$$

where $\delta\boldsymbol{x}^N$ is the standard Newton correction. As before, (3.11) and (4.22) then give that

$$\boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^N) \approx 2\boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}^{(k)}) - \frac{\mu^{(k)}}{\mu^{(k+1)}}\boldsymbol{D}_{\mathcal{A}}[c_i(\boldsymbol{x}^{(k)})/\mu^{(k)}]\boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}^{(k)}). \tag{4.23}$$

Dividing both sides of the relationship (4.19) by $c_i(\boldsymbol{x}^{(k+1)})$ and using the convergence of the sequence $\{w_i^{(k)}/c_i(\boldsymbol{x}^{(k)})\}$ to λ_i^*, we see that $c_i(\boldsymbol{x}^{(k)})/\mu^{(k)}$ converges to one provided λ_i^* is not zero. Thus (4.23) will be negative if $\mu^{(k+1)}$ is significantly smaller than $\frac{1}{2}\mu^{(k)}$. But, as before, it is normal to reduce $\mu^{(k)}$ by significantly more than a half and therefore a full Newton step (3.1) will be infeasible.

On the other hand, if we choose $\lambda_i = \lambda_i^A$ from (3.9), the relationships (4.19) and (4.21) imply that

$$\boldsymbol{A}_{\mathcal{A}}(\boldsymbol{x}^{(k)})\delta\boldsymbol{x}^A \approx -\boldsymbol{D}_{\mathcal{A}}[c_i(\boldsymbol{x}^{(k)})(1 - \mu^{(k+1)}/c_i(\boldsymbol{x}^{(k)}))]\boldsymbol{e} = -\boldsymbol{c}(\boldsymbol{x}^{(k)}) + \mu^{(k+1)}\boldsymbol{e}, \tag{4.24}$$

for the alternative Newton correction, $\delta\boldsymbol{x}^A$. Thus, combining (3.11) and (4.24), we have that

$$\boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A) \approx \mu^{(k+1)}\boldsymbol{e} + O(\mu^{(k)2}), \tag{4.25}$$

which indicates that a step in the alternative direction will be safely feasible so long as $\mu^{(k)} = o(\mu^{(k+1)\frac{1}{2}})$. Moreover, in this case, as both the inactive and active constraints are bounded away from zero at $\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A$ for fixed, but sufficiently large k, we have from (4.25) that

$$\frac{\mu^{(k)}}{c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A)} \approx \frac{\mu^{(k)}}{\mu^{(k+1)}}, \tag{4.26}$$

for all active constraints. Furthermore (4.16) holds for the inactive constraints. Thus, combining (4.5), (4.16) and (4.26), we see once again that (4.17) and (4.18) hold.

The comments at the end of Section (4.1) then apply equally here, namely that one would expect that the gradient of the barrier function after a single standard Newton iteration from the point $\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A$ will be $O(\mu^{(k)4}/\mu^{(k+1)2})$. Again, the stopping rule (1.3) will be satisfied at such a point so long as $\mu^{(k)} = o(\mu^{(k+1)\frac{3}{4}})$. Moreover, care should be taken to avoid the possible effects of ill-conditioning in the Newton systems and the remedies suggested in the final paragraph of Section (4.1) are equally appropriate.

4.3 The shifted barrier function methods

In this method, it is intended that the shifts and weights are chosen so that

$$w_i^{(k)}/s_i^{(k)} \to \lambda_i^* \quad \text{for} \quad i = 1, \ldots, m, \tag{4.27}$$

where $\boldsymbol{\lambda}^*$ are a set of Lagrange multipliers associated with the problem (1.1). In this case, in the framework of Gill *et al.* (1986), it follows that

$$\frac{w_i^{(k)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}} \to \lambda_i^* \quad \text{for} \quad i = 1, \ldots, m. \tag{4.28}$$

Thus, using (4.27) and (4.28), (3.10) gives

$$\begin{aligned} r_i^{(k+1)} &= w_i^{(k)} \frac{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k+1)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}} - w_i^{(k+1)} \\ &\approx (c_i(\boldsymbol{x}^{(k)}) + s_i^{(k+1)})\lambda_i^* - w_i^{(k+1)} \\ &\approx c_i(\boldsymbol{x}^{(k)})\lambda_i^*, \end{aligned} \tag{4.29}$$

which tends to zero because of the complementary slackness condition $c_i(\boldsymbol{x}^*)\lambda_i^* = 0$. It is difficult to say more about this method without more specific information on the shifts and weights.

4.4 Polyak's modified- and the Lagrangian barrier function methods

For the point of discussion here, the methods of Polyak (1992) and Conn *et al.* (1992a) may be considered to work in two phases, the non-asymptotic and asymptotic phases. The purpose of the non-asymptotic phase is to locate a $(\boldsymbol{x}, \lambda)$-neighbourhood of a first-order stationary point, $(\boldsymbol{x}^*, \lambda^*)$ while the asymptotic phase moves from this neighbourhood to $(\boldsymbol{x}^*, \lambda^*)$. Control of the non-asymptotic phase is primarily achieved by reducing the penalty parameter, while the penalty parameter remains fixed and estimates of the Lagrange multipliers are adjusted in the asymptotic phase. If the problem is degenerate, the non-uniqueness of the optimal Lagrange multipliers may prevent the algorithm from ever entering its asymptotic phase. None the less, convergence will still occur so long as $\boldsymbol{A}(x)_{\mathcal{A}}$ is full-rank.

If the algorithm has not entered its asymptotic phase, or perhaps if the problem is degenerate, the shifts and weights satisfy the relationships

$$s_i^{(k+1)} = \frac{\mu^{(k+1)}}{\mu^{(k)}} s_i^{(k)} \quad \text{and} \quad w_i^{(k+1)} = \frac{\mu^{(k+1)}}{\mu^{(k)}} w_i^{(k)}, \quad \text{for} \quad i = 1, \ldots, m, \tag{4.30}$$

where the adjacent penalty parameter values are such that $0 < \mu^{(k+1)}/\mu^{(k)} \le \tau < 1$. In this case, (3.10) is

$$r_i^{(k+1)} = \frac{w_i^{(k)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}} c_i(\boldsymbol{x}^{(k)}) \left(1 - \frac{\mu^{(k+1)}}{\mu^{(k)}}\right) \quad \text{for} \quad i = 1, \ldots, m. \tag{4.31}$$

Moreover, under mild conditions one has that each $\{w_i^{(k)}/(c_i(\boldsymbol{x}^{(k)})+s_i^{(k)})\}$ converges to a Lagrange multiplier λ_i^* and thus, because of the complementary slackness condition $c_i(\boldsymbol{x}^*)\lambda_i^* = 0$, (4.31) may be made arbitrarily small. We need to be cautious here as there is no guarantee that $\boldsymbol{x}^{(k)}$ is feasible for the shifted constraints once the updates (4.30) have been applied. It may then be necessary to find an alternative starting point for the $k+1$-st inner iteration. Suitable methods are given by Conn *et al.* (1992a).

If the asymptotic phase of the algorithm is reached, the penalty parameter $\mu^{(k)}$ remains fixed at some value $\overset{*}{\mu} > 0$ and the Lagrange multiplier estimates $\boldsymbol{\lambda}^{(k+1)}$ are defined by

$$\lambda_i^{(k+1)} = \frac{w_i^{(k)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}}, \quad \text{for} \quad i = 1, \ldots, m \tag{4.32}$$

(c.f. (3.9)). Here, the shifts and weights are defined to be

$$s_i^{(k+1)} = \overset{*}{\mu} (\lambda_i^{(k+1)})^{\alpha_\lambda} \quad \text{and} \quad w_i^{(k+1)} = \lambda_i^{(k+1)} s_i^{(k+1)} \quad \text{for} \quad i = 1, \ldots, m, \tag{4.33}$$

and some constant $0 \leq \alpha_\lambda \leq 1$ — the choice $\alpha_\lambda = 0$ gives Polyak's method while any $0 < \alpha_\lambda \leq 1$ defines a Lagrangian barrier function. We note that the theory given by Conn *et al.* (1992a) does not hold for the case $\alpha_\lambda = 0$. Both Polyak (1992) and Conn *et al.* (1992a) indicate that this asymptotic behaviour will occur under certain non-degeneracy assumptions. The main advantage of using nonzero shifts is that the ill-conditioning associated with traditional barrier function methods will no longer be present in a neighbourhood of a non-degenerate first-order stationary point for (1.1).

In this case (3.10), (4.32) and (4.33) give

$$\begin{aligned} r_i^{(k+1)} &= w_i^{(k)} \frac{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k+1)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}} - \lambda_i^{(k+1)} s_i^{(k+1)} \\ &= w_i^{(k)} \frac{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k+1)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}} - w_i^{(k)} \frac{s_i^{(k+1)}}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}} \\ &= w_i^{(k)} \frac{c_i(\boldsymbol{x}^{(k)})}{c_i(\boldsymbol{x}^{(k)}) + s_i^{(k)}} \\ &= \lambda_i^{(k+1)} c_i(\boldsymbol{x}^{(k)}), \end{aligned} \tag{4.34}$$

for each $i = 1, \ldots, m$. Under mild conditions, Polyak (1992) and Conn *et al.* (1992a) show that limit points of $\{\boldsymbol{x}^{(k)}\}$ are Kuhn-Tucker points and that the corresponding $\lambda_i^{(k+1)}$ converge to Lagrange multipliers. Hence, one expects $\lambda_i^{(k+1)} c_i(\boldsymbol{x}^{(k)})$ to be small because of the limiting complementary slackness condition at a Kuhn-Tucker point.

In particular, the asymptotic phase of the Lagrangian barrier function algorithm of Conn *et al.* (1992a) is entered whenever the condition

$$\| D[\overset{*}{\mu}\, \lambda_i^{(k+1)} c_i(\boldsymbol{x}^{(k)}) / s_i^{(k)}] \boldsymbol{e} \| \leq \eta^{(k)}, \tag{4.35}$$

for another positive sequence $\{\eta^{(k)}\}$ whose limit is zero. In this case, (4.33) and (4.35) imply that

$$\lambda_i^{(k+1)} c_i(\boldsymbol{x}^{(k)}) \leq \eta^{(k)} s_i^{(k)} / \overset{*}{\mu} \leq \kappa_2 \eta^{(k)}, \tag{4.36}$$

where $\kappa_2 = 2 \max_{i=1,\ldots,m} (\lambda_i^*)^{\alpha_\lambda}$ for each $i = 1, \ldots, m$ and all k sufficiently large. To proceed further, we need to consider the exact form of the sequences $\{\omega^{(k)}\}$ and $\{\eta^{(k)}\}$. These are given by

$$\omega^{(k)} = \kappa_\omega\, \overset{*}{\mu}^{\,\beta_\omega k} \quad \text{and} \quad \eta^{(k)} = \kappa_\eta\, \overset{*}{\mu}^{\,\beta_\eta k}, \tag{4.37}$$

where the constants β_η and β_ω satisfy the relationship $0 < \beta_\eta < \min(1, \beta_\omega)$, while κ_ω and κ_ω are strictly positive.

Consider first the choice (3.8), $\lambda_i = \lambda_i^N$. In this case, Conn *et al.* (1992b, equation (4.47)), ensures that the first component of the right-hand-side of (3.6) is $O(\overset{*}{\mu}{}^{\beta_\eta k-1})$. As the second is, by definition, zero, we would then expect that, so long as the coefficient matrix of (3.6) has a bounded inverse, $\delta\boldsymbol{x}^N = O(\overset{*}{\mu}{}^{\beta_\eta k-1})$ and $\delta\boldsymbol{\lambda}^N = O(\overset{*}{\mu}{}^{\beta_\eta k-1})$ and that

$$\boldsymbol{g}(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^N) - \boldsymbol{A}(\boldsymbol{x}^{(k)} + \delta^N\boldsymbol{x})^T(\boldsymbol{\lambda} + \delta\boldsymbol{\lambda}^N) = O(\overset{*}{\mu}{}^{2\beta_\eta k-2}) \tag{4.38}$$

and

$$(c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^N) + s_i^{(k+1)})(\lambda_i + \delta\lambda_i^N) = w_i^{(k+1)} + O(\overset{*}{\mu}{}^{2\beta_\eta k-2}), \tag{4.39}$$

for $i = 1, \ldots, m$. It then follows from (2.3) and (4.39) that

$$\begin{aligned} \lambda_i + \delta\lambda_i^N &= \frac{w_i^{(k+1)}}{c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^N) + s_i^{(k+1)}} + O\left(\frac{\overset{*}{\mu}{}^{2\beta_\eta k-2}}{c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^N) + s_i^{(k+1)}}\right) \\ &= u_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^N, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) + O\left(\frac{\overset{*}{\mu}{}^{2\beta_\eta k-2}}{c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^N) + s_i^{(k+1)}}\right), \end{aligned} \tag{4.40}$$

for $i = 1, \ldots, m$.

But, now consider the values of the shifted constraints $\boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{s}^{(k+1)}$ at the perturbed point $\boldsymbol{x} = \boldsymbol{x}^{(k)} + \delta\boldsymbol{x}$, where $\delta\boldsymbol{x} = o(1)$. Suppose, furthermore that strict complementary slackness holds at limit points of $\{\boldsymbol{x}^{(k)}\}$. For the inactive constraints, $\boldsymbol{c}(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x})$ is bounded away from zero for all k sufficiently large and thus $c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}) + s_i^{(k+1)} = \Theta(1)$. For the active constraints, the $s_i^{(k)}$ converge to $\overset{*}{\mu}\,(\lambda_i^*)^{\alpha_\lambda}$ and thus $c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}) + s_i^{(k+1)} = \Theta(\overset{*}{\mu})$.

Hence, as $\delta\boldsymbol{x}^N = o(1)$, (4.40) shows that

$$\boldsymbol{\lambda} + \delta\boldsymbol{\lambda}^N = \boldsymbol{u}(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^N, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) + O(\overset{*}{\mu}{}^{2\beta_\eta k-3}), \tag{4.41}$$

and thus (2.1), (4.38) and (4.41) yield that

$$\nabla_x \Psi(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^N, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) = O(\overset{*}{\mu}{}^{2\beta_\eta k-3}). \tag{4.42}$$

But then, as $\overset{*}{\mu}{}^{2\beta_\eta k-3} = o(\omega^{(k+1)})$ for all k sufficiently large, one would eventually expect a single iteration of Newton's method to suffice for each inner iteration. This is made rigorous by Conn *et al.* (1992b).

Now consider the choice (3.9), $\lambda_i = \lambda_i^A$. In this case, (1.3), (4.34) and (4.36) imply that the right-hand-side of (3.6) is $O(\max(\omega^{(k)}, \eta^{(k)})) = O(\overset{*}{\mu}{}^{\beta_\eta k})$. Consequently, provided the coefficient matrix of (3.6) has a bounded inverse, one would expect that $\delta\boldsymbol{x}^A = O(\overset{*}{\mu}{}^{\beta_\eta k})$ and $\delta\boldsymbol{\lambda}^A = O(\overset{*}{\mu}{}^{\beta_\eta k})$, and that

$$\boldsymbol{g}(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A) - \boldsymbol{A}(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A)^T(\boldsymbol{\lambda} + \delta\boldsymbol{\lambda}^A) = O(\overset{*}{\mu}{}^{2\beta_\eta k}) \tag{4.43}$$

and

$$(c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A) + s_i^{(k+1)})(\lambda_i + \delta\lambda_i^A) = w_i^{(k+1)} + O(\overset{*}{\mu}{}^{2\beta_\eta k}), \tag{4.44}$$

for $i = 1, \ldots, m$. It then follows from (2.3) and (4.39) that

$$\begin{aligned} \lambda_i + \delta\lambda_i^A &= \frac{w_i^{(k+1)}}{c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A) + s_i^{(k+1)}} + O\left(\frac{\overset{*}{\mu}{}^{2\beta_\eta k}}{c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A) + s_i^{(k+1)}}\right) \\ &= u_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) + O\left(\frac{\overset{*}{\mu}{}^{2\beta_\eta k}}{c_i(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A) + s_i^{(k+1)}}\right), \end{aligned} \tag{4.45}$$

for $i = 1, \ldots, m$.

Arguing as before that $\boldsymbol{c}(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A) + \boldsymbol{s}^{(k+1)} = \Theta(\overset{*}{\mu})$, (4.45) shows that

$$\boldsymbol{\lambda} + \delta\boldsymbol{\lambda}^A = \boldsymbol{u}(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) + O(\overset{*}{\mu}{}^{2\beta_\eta k-1}), \tag{4.46}$$

and thus (2.1), (4.38) and (4.46) yield that

$$\nabla_x \Psi(\boldsymbol{x}^{(k)} + \delta\boldsymbol{x}^A, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)}) = O(\overset{*}{\mu}{}^{2\beta_\eta k-1}). \tag{4.47}$$

But then again, as $\overset{*}{\mu}{}^{2\beta_\eta k-1} = o(\omega^{(k+1)})$ for all k sufficiently large, one would eventually expect a single iteration of Newton's method to suffice for each inner iteration.

Thus, we see that it is not crucial to use the alternative model for the methods considered in this section to achieve a reasonable second inner iterate. None the less, there are differences in the estimates (4.42) and (4.47) and the slightly better asymptotic estimate provided by (4.47) sometime manifests itself in practice.

5 Numerical experiments

In this section, we indicate the effect of using the alternative initial model discussed in the previous sections. We illustrate the effect using an algorithm for solving bound-constrained quadratic programming problems that incorporates a variety of shifts and weights, as implemented in Nick Gould's Harwell Subroutine Library (1993) code `VE14`.

method	standard			alternative		
	inner iterations	outer iterations	time (secs)	inner iterations	outer iterations	time (secs)
Trad	96	8	27.03	87	8	25.18
J and O	80	5	21.73	75	5	21.26
LBF	76	8	21.57	63	8	19.17

method	standard		alternative	
	number of factorizations	number of backtracks	number of factorizations	number of backtracks
Trad	49	159	46	93
J and O	41	216	39	143
LBF	41	148	34	115

Table 1: `BQPGAUSS` ($n = 2003$), optimal value $= -0.36258$

We consider three variants which are included in `VE14`. These are the traditional barrier function method ("Trad", see Section 4.1), the proposal by Jittorntrum and Osborne (1980) ("J and O", see Section 4.2) and the Lagrangian barrier function method ("LBF", see Section 4.4) with the parameter choice $\alpha_\lambda = 0.5$. As we have suggested, each method comprises an outer

method	standard			alternative		
	inner iterations	outer iterations	time (secs)	inner iterations	outer iterations	time (secs)
Trad	82	9	356.96	72	9	303.62
J and O	61	5	261.44	55	5	231.45
LBF	51	9	223.91	35	8	159.50

method	standard		alternative	
	number of factorizations	number of backtracks	number of factorizations	number of backtracks
Trad	45	81	38	31
J and O	33	135	29	81
LBF	28	47	20	32

Table 2: `JNLBRNGA` ($n = 15625$), optimal value -0.26851

iteration in which the shifts, weights and tolerances are adjusted according to predefined rules (see Section 4) and a sequence of inner iterations which conclude as soon as a value $x^{(k)}$ satisfying (1.3) is obtained. The inner iteration subproblem is solved using a simple backtracking linesearch method. In this, a search direction is computed by minimizing a quadratic model of the barrier function; the model is such that the gradients of the model and barrier function agree and the Hessian of the model is the matrix

$$\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\lambda}) + \boldsymbol{A}(x)^T \boldsymbol{D}[\lambda_i/(c_i(\boldsymbol{x}) + s_i^{(k+1)})]\boldsymbol{A}(\boldsymbol{x}) \tag{5.1}$$

of (3.7), modified (if necessary) to be positive definite. The model is minimized by solving the linear system which define its stationary point, using the sparse, multifrontal code `MA27` (Duff and Reid, 1982) from the Harwell Subroutine Library (1990) and, if necessary, modifying the factorization to ensure a convex model using the techniques described by Gill *et al.* (1992). Then, a step along this direction is found as the smallest non-negative power of 0.5 which is both feasible for the "shifted" constraints $\boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{s}^{(k)} > 0$ and satisfies a loose Armijo sufficient-decrease condition (see, for example, Dennis and Schnabel, 1983 or Fletcher, 1987). We appreciate that a more sophisticated linesearch, such as those specifically proposed for barrier functions by Lasdon *et al.* (1973) or Murray and Wright (1992), may be beneficial, but note that the simple backtracking strategy performed reasonably well in practice.

Before the minimization commences, a good symbolic ordering is found for the rows of the Hessian matrix. The Hessian of the model may remain fixed for a number of inner iterations. In tests, we have found that changing (and consequently refactorizing) the matrix every couple of inner iterations achieves a good compromise between the cost of the factorization and the effectiveness of an outdated model, although we also choose to delay refactorization if the ratio of norms of successive gradients of the barrier function is decreased by more than a fixed factor

method	standard			alternative		
	inner iterations	outer iterations	time (secs)	inner iterations	outer iterations	time (secs)
Trad	81	9	335.95	81	9	323.17
J and O	95	6	369.22	86	6	343.81
LBF	56	10	250.66	49	10	219.55

method	standard		alternative	
	number of factorizations	number of backtracks	number of factorizations	number of backtracks
Trad	45	110	43	50
J and O	49	223	46	146
LBF	32	45	29	19

Table 3: `OBSTCLBM` (n = 15625), optimal value = 7.2958

(0.1 in the tests performed here).

We consider two possible choices for $\boldsymbol{\lambda}$ in (5.1), those given by $\boldsymbol{\lambda} = \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{w}^{(k+1)}, \boldsymbol{s}^{(k+1)})$ from (2.3), denoted "standard", and those from (3.9), denoted "alternative". We only choose $\boldsymbol{\lambda}$ as (3.9) for the first model/step of each inner iteration in the "alternative" method, reverting to (2.3) for the second and subsequent models/steps. Both choices of $\boldsymbol{\lambda}$ are implemented as options within `VE14`, the default being to use (3.9) for the first model/step of each inner iteration. Further experiments, where additional steps were performed with $\boldsymbol{\lambda}$ chosen as (3.9), were less successful and are not reported here. The lack of success here is not really that unreasonable since the single "alternative" (primal-dual) step is intended to account for the change in shifts. Once the recovery is made, there is no reason not to continue with the "standard" (primal) Newton method given by (3.1).

In Tables 1–4 we give the numbers of outer and inner iterations and the cpu times required to solve four large examples from the `CUTE` test collection (Bongartz *et al.*, 1993) — all of the remaining large examples in the collection for which direct methods are appropriate[1] are variants of these and similar performances were observed. We note that the problem `BQPGAUSS` is non-convex, while the remaining problems are convex. We also report the number of factorizations that are required to solve the problems and the total number of times the stepsize was reduced in the backtracking linesearch. All tests were performed on a SUN Sparc 10 workstation in double precision and were stopped when the norm of the projected gradient of the objective function within the feasible region was smaller than 10^{-6}.

We draw the following conclusions from these experiments:

- The alternative choice of $\boldsymbol{\lambda}$ pays respectable dividends in both the number of iterations and

[1] The Hessian of the one other large problem from the collection, `ODNAMUR`, is too dense to assemble and store on our machine.

method	standard			alternative		
	inner iterations	outer iterations	time (secs)	inner iterations	outer iterations	time (secs)
Trad	84	9	304.71	74	9	251.54
J and O	75	5	256.41	55	5	192.65
LBF	60	9	220.88	58	9	206.74

method	standard		alternative	
	number of factorizations	number of backtracks	number of factorizations	number of backtracks
Trad	46	83	38	36
J and O	39	138	29	72
LBF	33	79	28	115

Table 4: `TORSION1` ($n = 14884$), optimal value -0.42570

the required cpu times to solve the problems. In some of the cases, as much as a twenty five percent improvement is possible. For the two unshifted methods, the number of backtracks performed is significantly reduced indicating that the alternative choice helps in producing good initial search directions — a closer examination of the runs indeed reveals that this is so. For the shifted method, the payoff is not as high but this may be explained by the theory of Conn *et al.* (1992b) which indicates that the standard Newton correction also provides acceptable steps in many cases. However, there is a slight, but noticeable, improvement in the "close-to" asymptotics, in that the gradient of the barrier function after the first Newton step of each inner iteration is almost always slightly smaller in the alternative method and this appears to be beneficial for the second and, if required, subsequent Newton steps.

- As these are the first reported results for Lagrangian barrier function methods, we also observe that such methods outperform the unshifted barrier function methods in almost all of our tests. While we cannot infer that this is a general trend, it is at least an indication that the theory provided by Conn *et al.* (1992a) is of use in producing good algorithms for bound-constrained quadratic programs.

6 Conclusions

We have presented a class of alternatives to the usual Newton direction for calculating an initial improvement to each of a sequence of barrier function minimizations. The method has proved to be effective in practice within the Harwell Subroutine Library (1993) bound constrained quadratic programming subroutine `VE14` and shows similar signs within a (as yet, unfinished) related, general quadratic programming code `VE19`.

7 Acknowledgement

The authors would like to thank two anonymous referees whose useful comments have improved the presentation of this paper.

References

[Bartholomew-Biggs, 1972] M. C. Bartholomew-Biggs. Constrained minimization using recursive quadratic programming. In F. A. Lootsma, editor, *Numerical Methods for Nonlinear Optimization*, pages 411–428, London, 1972. Academic Press.

[Bongartz *et al.*, 1993] I. Bongartz, A. R. Conn, N. I. M. Gould, and Ph. L. Toint. CUTE: Constrained and Unconstrained Testing Environment. Technical Report TR/PA/93/10, CERFACS, Toulouse, France, 1993.

[Conn *et al.*, 1992a] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. Convergence properties of minimization algorithms for convex constraints using a structured trust region. Technical Report 92-069, Rutherford Appleton Laboratory, Chilton, England, 1992.

[Conn *et al.*, 1992b] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. On the number of inner iterations per outer iteration of a globally convergent algorithm for optimization with general nonlinear inequality constraints and simple bounds. Technical Report 92-068, Rutherford Appleton Laboratory, Chilton, Oxfordshire, England, 1992.

[Dennis and Schnabel, 1983] J. E. Dennis and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations.* Prentice-Hall, Englewood Cliffs, USA, 1983.

[Duff and Reid, 1982] I. S. Duff and J. K. Reid. MA27: A set of Fortran subroutines for solving sparse symmetric sets of linear equations. Report R-10533, AERE Harwell Laboratory, Harwell, UK, 1982.

[Fiacco and McCormick, 1968] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* J. Wiley and Sons, New-York, 1968. Reprinted as *Classics in Applied Mathematics 4*, SIAM, 1990.

[Fletcher, 1987] R. Fletcher. *Practical Methods of Optimization.* J. Wiley and Sons, Chichester, second edition, 1987.

[Freund, 1991] R. M. Freund. Theoretical efficiency of a shifted-barrier-function algorithm for linear programming. *Linear Algebra and Applications*, 152:19–41, 1991.

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

[Gill *et al.*, 1986] P. E. Gill, W. Murray, M. A. Saunders, J. A. Tomlin, and M. H. Wright. On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method. *Mathematical Programming*, 36:183–209, 1986.

[Gill *et al.*, 1988] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Shifted barrier methods for linear programming. Technical Report SOL88-9, Department of Operations Research, Stanford University, Stanford, California 94305, USA, 1988.

[Gill *et al.*, 1992] P. E. Gill, W. Murray, D. B. Ponceléon, and M. A. Saunders. Preconditioners for indefinite systems arising in optimization. *SIAM Journal on Matrix Analysis and Applications*, 13:292–311, 1992.

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

[Gould, 1989] N. I. M. Gould. On the convergence of a sequential penalty function method for constrained minimization. *SIAM Journal on Numerical Analysis*, 26:107–128, 1989.

[Harwell Subroutine Library, 1990] Harwell Subroutine Library. *A catalogue of subroutines (release 10).* Advanced Computing Department, Harwell Laboratory, Harwell, UK, 1990.

[Harwell Subroutine Library, 1993] Harwell Subroutine Library. *A catalogue of subroutines (release 11).* Advanced Computing Department, Harwell Laboratory, Harwell, UK, 1993.

[Jarre *et al.*, 1988] F. Jarre, G. Sonnevend, and J. Stoer. An implementation of the method of analytic centers. In A. Benoussan and J. L. Lions, editors, *Analysis and Optimization of Systems*, pages 297–307, Berlin, 1988. Springer Verlag. Lecture Notes in Control and Information Sciences 111.

[Jittorntrum and Osborne, 1980] K. Jittorntrum and M. Osborne. A modified barrier function method with improved rate of convergence for degenerate problems. *Journal of the Australian Mathematical Society (Series B)*, 21:305–329, 1980.

[Karmarkar, 1984] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[Lasdon *et al.*, 1973] L. S. Lasdon, R. L. Fox, and M. W. Ratner. An efficient one-dimensional search procedure for barrier functions. *Mathematical Programming*, 4(3):279–296, 1973.

[McCormick, 1991] G. P. McCormick. The superlinear convergence of a nonlinear primal-dual algorithm. Technical Report OR T-550/91, Department of Operations Research, George Washington University, Washington DC 20052, 1991.

[Murray and Wright, 1978] W. Murray and M. H. Wright. Project Lagrangian methods based on the trajectories of penalty and barrier functions. Technical Report SOL78-23, Department of Operations Research, Stanford University, Stanford, California 94305, USA, 1978.

[Murray and Wright, 1992] W. Murray and M. H. Wright. Line search procedures for the logarithmic barrier function. Numerical Analysis Manuscript 92-01, AT&T Bell Laboratories, Murray Hill, USA, 1992.

[Murray, 1969] W. Murray. An algorithm for constrained minimization. In R. Fletcher, editor, *Optimization*, pages 189–196, London and New York, 1969. Academic Press.

[Nash and Sofer, 1993] S. G. Nash and A. Sofer. A barrier method for large-scale constrained optimization. *ORSA Journal of Computing*, 5(1):40–53, 1993.

[Polyak, 1992] R. Polyak. Modified barrier functions (theory and methods). *Mathematical Programming*, 54(2):177–222, 1992.

[Wright, 1976] M. H. Wright. *Numerical methods for nonlinearly constrained optimization.* PhD thesis, Department of Computer Science, Stanford University, Stanford, California 94305, USA, 1976.

[Wright, 1992a] M. H. Wright. Determining subspace information from the Hessan of a barrier function. Numerical Analysis Manuscript 92-02, AT&T Bell Laboratories, Murray Hill, USA, 1992.

[Wright, 1992b] M. H. Wright. Interior methods for constrained optimization. *Acta Numerica*, 1:341–407, 1992.

[Wright, 1993] M. H. Wright. Why a pure primal Newton barrier step may be infeasible. Numerical Analysis Manuscript 93-02, AT&T Bell Laboratories, Murray Hill, USA, 1993.