On the number of inner iterations per outer iteration of a globally convergent algorithm for optimization with general nonlinear inequality constraints and simple bounds

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

Report 92/14 August 20, 1992

Abstract. This paper considers the number of inner iterations required per outer iteration for the algorithm proposed by Conn *et al.* (1992b). We show that asymptotically, under suitable reasonable assumptions, a single inner iteration suffices.

[1] IBM T.J. Watson Research Center
Yorktown Heights, USA
[2] Rutherford Appleton Laboratory
Chilton, Oxfordshire, England
[3] Department of Mathematics,
Facultés Universitaires ND de la Paix,
Namur, Belgium

Keywords : Nonlinear optimization, inequality constraints, barrier methods, complexity.

1 Introduction

In this paper, we consider the nonlinear programming problem

$$\underset{x \in \Re^n}{\text{minimize}} \quad f(x) \tag{1.1}$$

subject to the *general constraints*

$$c_i(x) \geq 0, \quad i = 1, \ldots, m, \tag{1.2}$$

and the specific *simple bounds*

$$l \leq x \leq u. \tag{1.3}$$

We assume that the region $\mathcal{B} = \{x \in \Re^n \mid l \leq x \leq u\}$ is non-empty and may be infinite. We do not rule out the possibility that further simple bounds on the variables are included amongst the general constraints (1.2) if that is deemed appropriate. Indeed, it is conceivable that all simple bounds should be handled this way. Furthermore, we assume that

AS1. $f(x)$ and the $c_i(x)$ are twice continuously differentiable for all x in $\mathcal{B}$.

Our exposition will be conveniently simplified by taking the lower bounds as identically equal to zero and the upper bound as infinity for a subset of $\mathcal{N} \stackrel{\text{def}}{=} \{1, 2, ..., n\}$ in (1.3) and by assuming that the remaining variables are either not subjected to simple bounds or their simple bounds are treated as general constraints. Thus, in most of what follows, $\mathcal{B} = \{x \in R^n \mid x_j \geq 0 \text{ for all } j \in \mathcal{N}_b\}$, where $\mathcal{N}_b \subseteq \mathcal{N}$ is the index set of *bounded* variables. The modification required to handle more general bounds is indicated at the end of the paper.

The approach we intend to take is that of Conn *et al.* (1992b) and is based upon incorporating the equality constraints via a Lagrangian barrier function whilst handling upper and lower bounds directly. The sequential, approximate minimization of the Lagrangian barrier function is performed in a trust region framework such as that proposed by Conn *et al.* (1988a).

Our aim in this paper is to consider how these two different algorithms mesh together. In particular, we aim to show that ultimately very little work is performed in the iterative sequential minimization algorithm for every iteration of the outer Lagrangian barrier algorithm. This is contrary to most analyses of sequential penalty and barrier function methods in which the effort required to solve the inner iteration subproblems is effectively disregarded, the analysis concentrating on the convergence of the outer iteration (see for instance the books by Fiacco and McCormick, 1968 and Bertsekas, 1982. Exceptions to this are the sequential penalty function method analyzed by Gould, 1989, and the sequential augmented Lagrangian algorithm considered by Conn *et al.*, 1992a).

This work was primarily motivated by observations that the authors made when testing a prototype of their large-scale nonlinear programming package LANCELOT, release B (see Conn *et al.*, to appear 1992 for a description of release A), which includes an implementation of the algorithms discussed in this paper. It was often apparent that only a single iteration of the inner iteration subroutine SBMIN was ultimately required for every outer iteration of our sequential Lagrangian barrier program. While the conditions required in this paper to turn this observation to a proven result are relatively strong (and we feel probably about as weak as is possible), the package frequently exhibits the same behaviour on problems which violate our assumptions.

We define the concepts and notation that we shall need in section 2. Our algorithm is fully described in section 3 and analyzed in sections 4 and 5.

2 Notation

Let $g(x)$ denotes the gradient $\nabla_x f(x)$ of $f(x)$. Similarly, let $A(x)$ denote the Jacobian of $c(x)$, where

$$c(x) = [c_1(x), \cdots, c_m(x)]^T. \tag{2.1}$$

Thus

$$A(x)^T = [\nabla c_1(x), \cdots \nabla c_m(x)]. \tag{2.2}$$

We define the Lagrangian and Lagrangian barrier functions as

$$\ell(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x), \tag{2.3}$$

and

$$\Psi(x, \lambda, s) = f(x) - \sum_{i=1}^{m} \lambda_i s_i \log(c_i(x) + s_i), \tag{2.4}$$

respectively, where the components λ_i of the vector λ are positive and known as Lagrange multiplier estimates and where the elements s_i of the vector s are positive and known as shifts. We note that $\ell(x, \lambda)$ is the Lagrangian with respect to the general constraints only.

Let $g_\ell(x, \lambda)$ and $H_\ell(x, \lambda)$ respectively denote the gradient, $\nabla_x \ell(x, \lambda)$, and Hessian, $\nabla_{xx} \ell(x, \lambda)$, of the Lagrangian. We define the vector $\bar{\lambda}$ by

$$\bar{\lambda}_i(x, \lambda, s) = \frac{\lambda_i s_i}{c_i(x) + s_i}, \tag{2.5}$$

for all $1 \le i \le m$. We note that $\nabla_x \ell(x, \bar{\lambda}) = \nabla_x \Psi(x, \lambda, s)$.

We denote the non-negativity restrictions by

$$x \in \mathcal{B} = \{x \in R^n \mid x_j \ge 0 \quad \text{for all} \quad j \in \mathcal{N}_b\} \tag{2.6}$$

where $\mathcal{N}_b \subseteq \mathcal{N}$. We will make much use of the projection operator defined componentwise by

$$(P[x, l, u])_j = \begin{cases} l_j & \text{if } x_j \le l_j \\ u_j & \text{if } x_j \ge u_j \\ x_i & \text{otherwise.} \end{cases} \tag{2.7}$$

This operator projects the point x onto the region defined by the simple bounds (1.3). Let

$$P(x, v, l, u) = x - P[x - v, l, u]. \tag{2.8}$$

Furthermore, define $P[x] = P[x, \bar{l}, \infty]$ and $P(x, v) = P(x, v, \bar{l}, \infty)$, where $\bar{l}_j = 0$ for $j \in \mathcal{N}_b$ and $-\infty$ otherwise.

Let $x^{(k)} \in \mathcal{B}$ and $\lambda^{(k)}$ be given values of x and λ. If $h(x, \lambda, \ldots)$ is any function of x, $\lambda, \ldots$, we shall write $h^{(k)}$ as a shorthand for $h(x^{(k)}, \lambda^{(k)}, \ldots)$.

For any $x^{(k)}$ we have two possibilities for each component $x_j^{(k)}$, $j = 1, \ldots, n$, namely

(i) $j \in \mathcal{N}_b$ and $0 \le x_j^{(k)} \le (\nabla_x \Psi^{(k)})_j$ or

(ii) $j \in \mathcal{N}_f$ or $(\nabla_x \Psi^{(k)})_j < x_j^{(k)}$,

where $\mathcal{N}_f \stackrel{\text{def}}{=} \mathcal{N} \setminus \mathcal{N}_b$ is the index set of *free* variables. We shall call all $x_j^{(k)}$ that satisfy (i) *dominated* variables while the remaining $x_j^{(k)}$ are *floating* variables. It is important to notice that, as $x^{(k)} \in \mathcal{B}$,

$$(P(x^{(k)}, \nabla_x \Psi^{(k)}))_j = x_j^{(k)} \quad \text{whenever } x_j^{(k)} \text{ is dominated,} \tag{2.9}$$

while

$$(P(x^{(k)}, \nabla_x \Psi^{(k)}))_j = (\nabla_x \Psi^{(k)})_j \quad \text{otherwise.} \tag{2.10}$$

If x^* is the limit point of the (sub-)sequence $\{x^{(k)}\}_{k \in \mathcal{K}}$, we partition $\mathcal{N}$ into four index sets related to the two possibilities (i) and (ii) above and the corresponding x^*. We define

$$\begin{aligned} \mathcal{D}_1 &\stackrel{\text{def}}{=} \{j \in \mathcal{N}_b \mid x_j^{(k)} \text{ is dominated for all } k \in \mathcal{K} \text{ sufficiently large}\}, \\ \mathcal{F}_1 &\stackrel{\text{def}}{=} \mathcal{N}_f \cup \{j \in \mathcal{N}_b \mid x_j^{(k)} \text{ is floating for all } k \in \mathcal{K} \text{ sufficiently large and } x_j^* > 0\}, \\ \mathcal{F}_2 &\stackrel{\text{def}}{=} \{j \in \mathcal{N}_b \mid x_j^{(k)} \text{ is floating for all } k \in \mathcal{K} \text{ sufficiently large but } x_j^* = 0\} \text{ and} \\ \mathcal{F}_3 &\stackrel{\text{def}}{=} \mathcal{N} \setminus \mathcal{D}_1 \bigcup \mathcal{F}_1 \bigcup \mathcal{F}_2. \end{aligned} \tag{2.11}$$

We also define

$$\begin{aligned} \mathcal{I}(x) &\stackrel{\text{def}}{=} \{i \mid c_i(x) > 0\}, \\ \mathcal{A}(x) &\stackrel{\text{def}}{=} \{i \mid c_i(x) \le 0\}, \end{aligned} \tag{2.12}$$

the sets of *inactive* (strictly satisfied) and *active* (violated or just satisfied) constraints at the point x. We develop our algorithm so that the set $\mathcal{A}^* \equiv \mathcal{A}(x^*)$ at any limit point of our generated sequence is precisely the set of constraints for which $c_i(x^*) = 0$. We also write $\mathcal{I}^* \equiv \mathcal{I}(x^*)$.

We will use the notation that if $\mathcal{J}_1$ and $\mathcal{J}_2$ are any subsets of $\mathcal{N}$ and H is an n by n matrix, $H_{[\mathcal{J}_1,\mathcal{J}_2]}$ is the matrix formed by taking the rows and columns of H indexed by $\mathcal{J}_1$ and $\mathcal{J}_2$ respectively. Likewise, if A is an m by n matrix, $A_{[\mathcal{J}_1]}$ is the matrix formed by taking the columns of A indexed by $\mathcal{J}_1$.

We denote the (appropriately dimensioned) identity matrix by I; its j-th column is e_j. A vector of ones is denoted by e.

We will use a variety of vector and subordinate matrix norms. We shall only consider norms $\|\cdot\|_z$ which are *consistent* with the two-norm, that is, norms which satisfy the inequalities

$$\|v\|_z \leq a_0^{\frac{1}{2}}\|v\|_2 \quad \text{and} \quad \|v\|_2 \leq a_0^{\frac{1}{2}}\|v\|_z \tag{2.13}$$

for all vectors v and some constant $a_0 \geq 1$, independent of z. It then follows that, for any pair of two-norm-consistent norms $\|\cdot\|_y$ and $\|\cdot\|_z$,

$$\|v\|_z \leq a_0\|v\|_y \quad \text{and} \quad \|v\|_y \leq a_0\|v\|_z. \tag{2.14}$$

If r is any m-vector whose i-th component is r_i, we use the shorthand $r \equiv [r_i]_{i=1}^m$. Furthermore, if r is as above and $\mathcal{J}$ is a subset of $\{1, 2, \cdots, m\}$, $[r_i]_{i\in\mathcal{J}}$ is just the vector whose components are the r_i, $i \in \mathcal{J}$. Consequently, $\|[r_i]_{i=1}^m\| \equiv \|r\|$.

Following Conn *et al.* (1992b), we now describe an algorithm for solving (1.1), (1.2) and (2.6).

3 Statement of the algorithm

In order to solve the problem (1.1), (1.2) and (2.6), we consider the algorithmic model given in Figure 1. We shall call the vector $P(x^{(k)}, \nabla_x\Psi^{(k)})$ the *projected gradient of the Lagrangian barrier function* or the *projected gradient* for short. The norms $\|\cdot\|_g$ and $\|\cdot\|_c$ are normally chosen to be either two or infinity norms.

Our decreasing sequence of $\mu^{(k)}$'s is given by $\mu^{(k)} = \mu_0(\tau)^{k_j}$, but any monotonic decreasing sequence of $\mu^{(k)}$'s converging to zero if Step 4 is executed an infinite number of times will suffice. It is also irrelevant, in theory, as to how we find a suitable point $x^{(k)}$ satisfying (3.5). However, from a practical perspective, a suitable point is found by an iterative procedure. In our algorithm, it is normal to try to start this inner iteration from, or close to, the solution to the last one. Indeed, from the point of view of the results we are about to establish, this is crucial. Such a starting point is desirable as function and derivative information from the conclusion of one inner iteration may be passed as input to the next. However, we need to bear in mind the requirement (3.6) may preclude us from picking such a starting point as it is possible that $c_i(x^{(k-1)}) + s_i^{(k)} \leq 0$ for some i. This issue is considered in depth in Conn *et al.* (1992b), where it is shown that $c_i(x^{(k)}) + s_i^{(k+1)} > 0$ for all $1 \leq i \leq m$ when Step 3 of Algorithm 3.1 is executed, while techniques for finding a suitable alternative starting point when Step 4 occurs are given.

The main purpose of this paper is to show that asymptotically we take one inner iteration per outer iteration. More specifically, under certain assumptions, we first show that (3.8) is eventually satisfied at each outer iteration. We then show that, under additional assumptions, it

Algorithm 3.1 [Outer-iteration Algorithm]

Step 0 : [Initialization] *The strictly positive constants*

$$\eta_0,\ \omega_0,\ \alpha_\omega,\ \beta_\omega,\ \alpha_\eta,\ \beta_\eta,\ \alpha_\lambda \le 1,\ \tau < 1,\ \rho < 1,\ \gamma_2 < 1,\ \omega_* \ll 1 \text{ and } \eta_* \ll 1 \tag{3.1}$$

for which

$$1 - (1+\alpha_\lambda)^{-1} < \alpha_\eta < \min(1, \alpha_\omega) \quad \text{and} \quad \beta_\eta < \min(1, \beta_\omega). \tag{3.2}$$

are specified. A positive forcing parameter, $\bar{\mu}^{(0)}$, is given. Set

$$\mu^{(0)} = \min(\bar{\mu}^{(0)}, \gamma_2), \quad \omega^{(0)} = \omega_0 (\mu^{(0)})^{\alpha_\omega} \quad \text{and} \quad \eta^{(0)} = \eta_0 (\mu^{(0)})^{\alpha_\eta}. \tag{3.3}$$

An initial estimate of the solution, $x^{est} \in \mathcal{B}$, and vector of positive Lagrange multiplier estimates, $\lambda^{(0)}$, for which $c_i(x^{est}) + \mu^{(0)}(\lambda_i^{(0)})^{\alpha_\lambda} > 0$ are specified. Set $k = 0$.

Step 1 : [Inner iteration] *Compute shifts*

$$s_i^{(k)} = \mu^{(k)} (\lambda_i^{(k)})^{\alpha_\lambda}, \tag{3.4}$$

for $i = 1, ..., m$. Find $x^{(k)} \in \mathcal{B}$ such that

$$\|P(x^{(k)}, \nabla_x \Psi^{(k)})\|_g \le \omega^{(k)} \tag{3.5}$$

and

$$c_i(x^{(k)}) + s_i^{(k)} > 0, \quad \text{for} \quad i = 1, ..., m. \tag{3.6}$$

Step 2 : [Test for convergence] *If*

$$\|P(x^{(k)}, \nabla_x \Psi^{(k)})\|_g \le \omega_* \text{ and } \|[c_i(x^{(k)}) \bar{\lambda}_i(x^{(k)}, \lambda^{(k)}, s^{(k)})]_{i=1}^m\|_c \le \eta_*, \tag{3.7}$$

stop. If

$$\left\| \left[c_i(x^{(k)}) \bar{\lambda}_i(x^{(k)}, \lambda^{(k)}, s^{(k)}) / (\lambda_i^{(k)})^{\alpha_\lambda} \right]_{i=1}^m \right\|_c \le \eta^{(k)}, \tag{3.8}$$

execute Step 3. Otherwise, execute Step 4.

Step 3 : [Update Lagrange multiplier estimates] *Set*

$$\begin{aligned} \lambda^{(k+1)} &= \bar{\lambda}(x^{(k)}, \lambda^{(k)}, s^{(k)}), \\ \bar{\mu}^{(k+1)} &= \bar{\mu}^{(k)}, \\ \mu^{(k+1)} &= \min(\bar{\mu}^{(k+1)}, \gamma_2), \\ \omega^{(k+1)} &= \omega^{(k)} (\mu^{(k+1)})^{\beta_\omega}, \\ \eta^{(k+1)} &= \eta^{(k)} (\mu^{(k+1)})^{\beta_\eta}. \end{aligned} \tag{3.9}$$

Increase k by one and go to Step 1.

Step 4 : [Reduce the forcing parameter] *Set*

$$\begin{aligned} \lambda^{(k+1)} &= \lambda^{(k)}, \\ \bar{\mu}^{(k+1)} &= \tau \bar{\mu}^{(k)}, \\ \mu^{(k+1)} &= \min(\bar{\mu}^{(k+1)}, \gamma_2), \\ \omega^{(k+1)} &= \omega_0 (\mu^{(k+1)})^{\alpha_\omega}, \\ \eta^{(k+1)} &= \eta_0 (\mu^{(k+1)})^{\alpha_\eta}. \end{aligned} \tag{3.10}$$

Increase k by one and go to Step 1.

End of Algorithm 3.1

Figure 1: Outer-iteration algorithm

is possible to satisfy the convergence test (3.5) after a single iteration of the algorithm given in Conn *et al.* (1988a).

The specific inner iteration algorithm we shall consider is given in Figure 2.

There are a number of possible ways of choosing $\gamma_0^{(k,j)}$ and $\gamma_3^{(k,j)}$ in Step 4. The simplest is merely to pick $\gamma_0^{(k,j)} = \gamma_0$ and $\gamma_3^{(k,j)} = \gamma_3$; other alternatives are discussed by Conn *et al.* (to appear 1992).

It remains to give a description of the starting point, initial trust region radius and approximation to the Hessian of the Lagrangian, and of the calculation that is performed in Step 2 of Algorithm 3.2.

Let $0 < \theta < 1$. We let

$$\hat{x}_j^{(k-1)} = \begin{cases} 0 & \text{if } 0 \le x_j^{(k-1)} \le \theta(\nabla_x \Psi^{(k-1)})_j \quad \text{and} \quad j \in \mathcal{N}_b \\ x_j^{(k-1)} & \text{otherwise,} \end{cases} \tag{3.17}$$

and choose

$$x^{(k,0)} = \begin{cases} \hat{x}^{(k-1)} & \text{if } c(\hat{x}^{(k-1)}) + s^{(k)} > 0 \\ x^{(k-1)} & \text{otherwise.} \end{cases} \tag{3.18}$$

Thus variables which are significantly dominated at the end of the $(k-1)$-st iteration are set to their bounds while the remainder are left unaltered. This choice is made since, under a suitable non-degeneracy assumption (AS7 in section 4), the set of dominated variables is asymptotically the same as the set of variables which lie on their bounds (see, Conn *et al.*, 1992b, Theorem 5.4). Furthermore, under a second non-degeneracy assumption (AS5 in section 4), the assignment $x^{(k,0)} = \hat{x}^{(k-1)}$ is guaranteed for k sufficiently large. Our choice of $x^{(k,0)}$ then encourages subsequent iterates to encounter their asymptotic state as soon as possible.

We also pick $\Delta^{(k,0)}$ so that

$$\Delta^{(k,0)} \ge \kappa \| P(x^{(k,0)}, \nabla_x \Psi^{(k,0)}) \|_g^{\zeta} \tag{3.19}$$

for some positive constants κ and $\zeta < 1$ (typical values might be $\kappa = 1$ and $\zeta = 0.9$). This value is chosen so that the trust region does not interfere with the asymptotic convergence of the algorithm, while providing a reasonable starting value in the earlier stages of the method.

Finally $B^{(k,0)}$ is taken to be any sufficiently good symmetric approximation to the Hessian of the Lagrangian function at $x^{(k)}$. We qualify what we mean by "sufficiently good" in the next section but suffice it to say that exact second derivatives satisfy this property and are often to be recommended.

The calculation in Step 2 is performed in two stages.

1. Firstly, the so-called *generalized Cauchy point*, $x^{C(k,j)} \equiv x^{(k,j)} + p^{C(k,j)}$, is determined. This is merely an approximation to the first local minimizer of the quadratic model, $m^{(k,j)}(x^{(k,j)} + p)$, along the Cauchy arc. The *Cauchy arc* is the path $x^{(k,j)} + p$, where

$$p = p^{(k,j)}(t) \stackrel{\text{def}}{=} P[x^{(k,j)} - t\nabla_x \Psi(x^{(k,j)}, \lambda^{(k)}, \mu^{(k)}), l, u] - x^{(k,j)}, \tag{3.20}$$

Algorithm 3.2 [Inner-iteration Algorithm]

Step 0 : [Initialization] *The positive constants* $\mu < \eta < 1$ *and* $\gamma_0 \leq \gamma_2 < 1 \leq \gamma_3$ *are given. The starting point,* $x^{(k,0)}$*, a nonnegative convergence tolerance,* $\omega^{(k)}$*, an initial trust region radius,* $\Delta^{(k,0)}$*, a symmetric approximation,* $B^{(k,0)}$*, to the Hessian of the Lagrangian,* $H_\ell(x^{(k,0)}, \lambda^{(k)})$*, and a two-norm-consistent norm* $\|\cdot\|_g$ *are specified. Compute* $\Psi(x^{(k,0)}, \lambda^{(k)}, s^{(k)})$ *and its gradient. Set the inner iteration counter* $j = 0$*.*

Step 1 : [Test for convergence] *If*

$$\|P(x^{(k,j)}, \nabla_x \Psi^{(k,j)})\|_g \leq \omega^{(k)} \tag{3.11}$$

set $x^{(k)} = x^{(k,j)}$ *and stop.*

Step 2 : [Significantly reduce a model of the Lagrangian barrier function] *Construct a quadratic model,*

$$\begin{aligned} m^{(k,j)}(x^{(k,j)} + p) \stackrel{\text{def}}{=} \quad & \Psi(x^{(k,j)}, \lambda^{(k)}, \mu^{(k)}) + p^T \nabla_x \Psi(x^{(k,j)}, \lambda^{(k)}, \mu^{(k)}) \\ & + \tfrac{1}{2} p^T (B^{(k,j)} + A(x^{(k,j)})^T D^{(k)}(x^{(k,j)}) A(x^{(k,j)})) p, \end{aligned} \tag{3.12}$$

of $\Psi(x + p, \lambda^{(k)}, \mu^{(k)})$*, where* $D^{(k)}(x)$ *is the* m *by* m *diagonal matrix whose* i*-th diagonal is*

$$D^{(k)}(x)_{i,i} = \frac{\lambda_i^{(k)} s_i^{(k)}}{(c_i(x) + s_i^{(k)})^2}. \tag{3.13}$$

Compute a step $p^{(k,j)}$ *which significantly reduces the value of the model,* $m^{(k,j)}(x^{(k,j)} + p)$*.*

Step 3 : [Compute a measure of the effectiveness of the step] *Compute* $\Psi(x^{(k,j)} + p^{(k,j)}, \lambda^{(k)}, s^{(k)})$ *and the ratio*

$$\rho^{(k,j)} = \frac{\Psi(x^{(k,j)}, \lambda^{(k)}, s^{(k)}) - \Psi(x^{(k,j)} + p^{(k,j)}, \lambda^{(k)}, s^{(k)})}{m^{(k,j)}(x^{(k,j)}) - m^{(k,j)}(x^{(k,j)} + p^{(k,j)})}. \tag{3.14}$$

Step 4 : [Accept or reject the step] *Set*

$$x^{(k,j+1)} = \begin{cases} x^{(k,j)} + p^{(k,j)} & \text{if } \rho^{(k,j)} > \mu \\ x^{(k,j)} & \text{otherwise,} \end{cases} \tag{3.15}$$

and

$$\Delta^{(k,j+1)} = \begin{cases} \gamma_0^{(k,j)} \Delta^{(k,j)} & \text{if } \rho^{(k,j)} \leq \mu \\ \Delta^{(k)} & \text{if } \mu < \rho^{(k,j)} < \eta \\ \gamma_3^{(k,j)} \Delta^{(k,j)} & \text{otherwise,} \end{cases} \tag{3.16}$$

where $\gamma_0^{(k,j)} \in [\gamma_0, 1)$ *and* $\gamma_3^{(k,j)} \in [1, \gamma_3]$*.*

Step 5 : [Updating] *If necessary, compute the gradient of* $\Psi(x^{(k,j+1)}, \lambda^{(k)}, \mu^{(k)})$ *and a further approximation to the Hessian of the Lagrangian* $B^{(k,j+1)}$*. Increment the inner iteration counter* j *by one and go to Step 1.*

End of Algorithm 3.2

Figure 2: Inner-iteration algorithm

as the parameter t increases from 0, which finishes when the path first intersects the boundary of the trust region,

$$\|p\|_t \leq \Delta^{(k,j)}, \tag{3.21}$$

for some two-norm-consistent norm $\|\cdot\|_t$. Thus the Cauchy arc is simply the path which starts in the steepest descent direction for the model but which is subsequently "bent" to follow the boundary of the "box" region defined by the feasible region (2.6) (or, in general, (1.3)) and which stops on the boundary of the trust region (3.21). The two or infinity norm is normally chosen, the latter having some advantages as the trust region is then aligned with the feasible region (2.6). (Indeed, it is possible to extend the Cauchy arc along the boundary of the trust region when the infinity norm is used. Further reduction of the quadratic model along this extended Cauchy arc may prove beneficial.)

The method proposed by Conn *et al.* (1988a) calculates the exact generalized Cauchy point by marching along the Cauchy arc until either the trust region boundary is encountered or the model starts to increase. An alternative method by Moré (1988) finds an approximation $p^{C(k,j)} = p^{(k,j)}(t^{C(k,j)})$ which is required to lie within the trust-region and to satisfy the Goldstein-type conditions

$$m^{(k,j)}(x^{(k,j)}+p^{(k,j)}(t^{C(k,j)})) \leq m^{(k,j)}(x^{(k,j)})+\mu_1 p^{(k,j)}(t^{C(k,j)})^T \nabla_x \Psi(x^{(k,j)}, \lambda^{(k)}, s^{(k)}) \tag{3.22}$$

and

$$t^{C(k,j)} \geq \nu_1 \quad \text{or} \quad t^{C(k,j)} \geq \nu_2 t^{L(k,j)}, \tag{3.23}$$

where $t^{L(k,j)} > 0$ is any value for which

$$m^{(k,j)}(x^{(k,j)}+p^{(k,j)}(t^{L(k,j)})) \geq m^{(k,j)}(x^{(k,j)})+\mu_2 p^{(k,j)}(t^{L(k,j)})^T \nabla_x \Psi(x^{(k,j)}, \lambda^{(k)}, s^{(k)}) \tag{3.24}$$

or

$$\|p^{(k,j)}(t^{L(k,j)})\|_t \geq \nu_3 \Delta^{(k,j)}, \tag{3.25}$$

and the positive constants μ_1, μ_2, ν_1, ν_2 and ν_3 satisfy the restrictions $\mu_1 < \mu_2 < 1$, $\nu_2 < 1$ and $\nu_3 < 1$. Condition (3.22) ensures that a sufficient reduction in the model takes place at each iteration while condition (3.23) is needed to guarantee that every step taken is non-negligible. Moré shows that it is always possible to pick such a value of $t^{C(k,j)}$ using a backtracking linesearch, starting on or near to the trust region boundary. Similar methods have been proposed by Calamai and Moré (1987), Burke and Moré (1988), Toint (1988) and Burke *et al.* (1990).

2. Secondly, we pick $p^{(k,j)}$ so that $x^{(k,j)} + p^{(k,j)}$ lies within (2.6), $\|p^{(k,j)}\|_t \leq \beta_2 \Delta^{(k,j)}$ and

$$\begin{aligned} & m^{(k,j)}(x^{(k,j)}) - m^{(k,j)}(x^{(k,j)} + p^{(k,j)}) \\ & \quad \geq \beta_3 [m^{(k,j)}(x^{(k,j)}) - m^{(k,j)}(x^{(k,j)} + p^{C(k,j)})] \geq 0 \end{aligned} \tag{3.26}$$

for some positive $\beta_2 \geq 1$ and $\beta_3 \leq 1$. In fact, we typically choose $\beta_2 = \beta_3 = 1$, in which case we are merely requiring that the computed step gives a value of the model which is no larger than the value at the generalized Cauchy point.

In order to accelerate the convergence of the method, it is normal to try to bias the computed step towards the Newton direction.

The convergence analysis given by Conn *et al.* (1988a) for Algorithm 3.1 indicates that it is desirable to construct improvements beyond the Cauchy point only in the subspace of variables which are free from their bounds at the Cauchy point. In particular, with such a restriction and with a suitable non-degeneracy assumption, it is then shown that the set of variables which are free from their bounds at the solution is determined after a finite number of iterations. This has the advantage of allowing one to analyze the asymptotic convergence rate of the method purely as if it were an unconstrained calculation, merely by focusing on the set of free variables.

Let $\mathcal{F}$ be a subset of $\mathcal{N}$ and let $\mathcal{D} = \mathcal{N} \setminus \mathcal{F}$. Furthermore, let

$$H^{(k,j)} \stackrel{\text{def}}{=} B^{(k,j)} + A(x^{(k,j)})^T D^{(k)}(x^{(k,j)}) A(x^{(k,j)}) \tag{3.27}$$

denote the composite approximation to the Hessian of the Lagrangian barrier function.

The specific algorithm we shall consider is summarized in Figure 3. In Step 2 of this method, the value of $p_{[\mathcal{F}]}$ would normally be computed as the aggregate step after a number of Conjugate Gradient (CG) iterations, where CG is applied to minimize the model in the subspace defined by the free variables. The CG process will end when either a new bound is encountered or the convergence test (3.30) is satisfied. Algorithm 3.3 is itself finite as the number of free variables at each pass of Step 2 is strictly monotonically decreasing. See the paper by Conn *et al.* (1988b) for further details.

4 Convergence analysis

We wish to analyze the asymptotic behaviour of Algorithm 3.1, that is in the case where $\omega_* = \eta_* = 0$. We require the following additional assumptions.

AS2. The matrix $A(x^*)_{[\mathcal{A}^*, \mathcal{F}_1]}$ is of full rank at any limit point x^* of the sequence $\{x^{(k)}\}$ generated by Algorithm 3.1 with the set $\mathcal{F}_1$ defined by (2.11).

Under these assumptions we have the following result.

Theorem 4.1 [Conn *et al.*, 1992b, Theorem 4.4] *Assume that AS1 and AS2 hold, that x^* is a limit point of the sequence $\{x^{(k)}\}$ generated by Algorithm 3.1 and that*

$$\bar{\lambda}_i^{(k)} \stackrel{\text{def}}{=} \frac{\lambda_i^{(k)} s_i^{(k)}}{c_i(x^{(k)}) + s_i^{(k)}}, \tag{4.1}$$

for $i = 1, \cdots, m$. Then x^* is a Kuhn-Tucker (first order stationary) point for (1.1), (1.2) and (2.6) and the corresponding subsequences of $\{\bar{\lambda}^{(k)}\}$ and $\{\nabla_x \Psi^{(k)}\}$ converge to a set of Lagrange multipliers, λ^*, and the gradient of the Lagrangian, $g_\ell(x^*, \lambda^*)$, for the problem, respectively.

Now consider the following further assumptions.

Algorithm 3.3 **[Algorithm to significantly reduce the model]**

Step 0 : [Initialization] *Select positive constants* $\nu < 1$, $\xi < 1$, $\beta_2 \geq 1$ *and* $\beta_3 \leq 1$.

Step 1 : [Calculate the generalized Cauchy point] *Calculate an approximation to the generalized Cauchy point* $x^{C(k,j)} = x^{(k,j)} + p^{C(k,j)}$ *using one of the previously mentioned techniques. Compute the set of variables,* $\mathcal{F}^{C(k,j)}$, *which are free from their bounds at* $x^{C(k,j)}$. *Set* $x = x^{C(k,j)}$, $s = p^{C(k,j)}$ *and* $\mathcal{F} = \mathcal{F}^{C(k,j)}$.

Step 2 : [Further improve the model] *Let* $\mathcal{C}(\beta_2) = \mathcal{S} \bigcap \mathcal{T}(\beta_2)$, *where*

$$\mathcal{S} = \{p_{[\mathcal{F}]} \mid x^{(k,j)} + p \in \mathcal{B} \quad and \quad p_{[\mathcal{D}]} = p_{[\mathcal{D}]}^{C(k,j)}\} \tag{3.28}$$

and

$$\mathcal{T}(\beta_2) = \{p_{[\mathcal{F}]} \mid \|p\|_t \leq \beta_2 \Delta^{(k,j)} \quad and \quad p_{[\mathcal{D}]} = p_{[\mathcal{D}]}^{C(k,j)}\}. \tag{3.29}$$

If $p_{[\mathcal{F}]}$ *lies on the boundary of* $\mathcal{T}(\beta_2)$, *set* $p^{(k,j)} = p$ *and stop. (If* $\|\cdot\|_t$ *is the infinity norm, it is possible to transfer components of* $\mathcal{F}$ *which lie on the trust-region boundary to* $\mathcal{D}$ *and to continue.) Otherwise, recompute* $p_{[\mathcal{F}]}$ *so that (3.26) is satisfied and either* $p_{[\mathcal{F}]}$ *lies strictly interior to* $\mathcal{C}(\beta_2)$ *with*

$$\begin{aligned} &\|H^{(k,j)}_{[\mathcal{F},\mathcal{F}]} p_{[\mathcal{F}]} + (\nabla_x \Psi^{(k,j)}_{[\mathcal{F}]} + H^{(k,j)}_{[\mathcal{F},\mathcal{D}]} p_{[\mathcal{D}]})\|_g \\ \leq \quad & \min(\nu, \|P(x^{(k,j)}, \nabla_x \Psi^{(k,j)})\|_g^{\xi}) \cdot \|P(x^{(k,j)}, \nabla_x \Psi^{(k,j)})\|_g \end{aligned} \tag{3.30}$$

or $p_{[\mathcal{F}]}$ *lies on the boundary of* $\mathcal{C}(\beta_2)$. *Reset* $x_{[\mathcal{F}]}$ *to* $x_{[\mathcal{F}]} + p_{[\mathcal{F}]}$.

Step 3 : [Test for convergence] *If* $p_{[\mathcal{F}]}$ *lies strictly interior to* $\mathcal{C}(\beta_2)$ *and (3.30) is satisfied or if it is decided that sufficient passes have been made, set* $p^{(k,j)} = p$ *and stop. Otherwise remove all of the indices in* $\mathcal{F}$ *for which* $p_{[\mathcal{F}]i}$ *lies on the boundary of* $\mathcal{S}$ *and perform another pass by returning to Step 2.*

End of Algorithm 3.3

Figure 3: Algorithm to significantly reduce the model

AS3. The second derivatives of the functions $f(x)$ and the $c_i(x)$ are Lipschitz continuous at all points within an open set containing $\mathcal{B}$.

AS4. Suppose that (x^*, λ^*) is a Kuhn-Tucker point for the problem (1.1), (1.2) and (2.6), and

$$\begin{array}{lcl} \mathcal{A}_1^* & \stackrel{\mathrm{def}}{=} & \{i | c_i(x^*) = 0 \quad \text{and} \quad \lambda_i^* > 0\} \\ \mathcal{A}_2^* & \stackrel{\mathrm{def}}{=} & \{i | c_i(x^*) = 0 \quad \text{and} \quad \lambda_i^* = 0\} \end{array} \tag{4.2}$$

and

$$\begin{array}{lcl} \mathcal{J}_1 & \stackrel{\mathrm{def}}{=} & \mathcal{N}_f \cup \{j \in \mathcal{N}_b | (g_\ell(x^*, \lambda^*))_j = 0 \quad \text{and} \quad x_j^* > 0\} \\ \mathcal{J}_2 & \stackrel{\mathrm{def}}{=} & \{j \in \mathcal{N}_b | (g_\ell(x^*, \lambda^*))_j = 0 \quad \text{and} \quad x_j^* = 0\}. \end{array} \tag{4.3}$$

Then we assume that the matrix

$$\begin{pmatrix} H_\ell(x^*, \lambda^*)_{[\mathcal{J},\mathcal{J}]} & (A(x^*)_{[\mathcal{A},\mathcal{J}]})^T \\ A(x^*)_{[\mathcal{A},\mathcal{J}]} & 0 \end{pmatrix} \tag{4.4}$$

is non-singular for all sets $\mathcal{A}$ and $\mathcal{J}$, where $\mathcal{A}$ is any set made up from the union of $\mathcal{A}_1^*$ and any subset of $\mathcal{A}_2^*$ and $\mathcal{J}$ is any set made up from the union of $\mathcal{J}_1$ and any subset of $\mathcal{J}_2$.

AS5. (Strict complementary slackness condition 1) Suppose that (x^*, λ^*) is a Kuhn-Tucker point for problem (1.1), (1.2) and (2.6). Then

$$\mathcal{A}_2^* = \{i | c_i(x^*) = 0 \quad \text{and} \quad \lambda_i^* = 0\} = \emptyset. \tag{4.5}$$

AS6. Algorithm 3.1 has a single limit point, x^*.

Under these additional assumptions, we are able to derive the following result.

Theorem 4.2 [Conn *et al.*, 1992b, Theorems 5.3 and 5.5] *Assume that AS1–AS6 hold. Then there is a constant* $\mu_{\min} > 0$ *such that the penalty parameter* $\mu^{(k)}$ *generated by Algorithm 3.1 satisfies* $\mu^{(k)} = \mu_{\min}$ *for all* k *sufficiently large. Furthermore,* $x^{(k)}$ *and* $\bar{\lambda}^{(k)}_{[\mathcal{A}^*]}$ *satisfy the bounds*

$$\|x^{(k)} - x^*\|_g \leq a_x(\mu_{\min})^{\alpha_\eta + k\alpha_\lambda\beta_\eta} \quad \text{and} \quad \|(\bar{\lambda}^{(k)} - \lambda^*)_{[\mathcal{A}^*]}\|_g \leq a_\lambda(\mu_{\min})^{\alpha_\eta + k\alpha_\lambda\beta_\eta}, \tag{4.6}$$

for the two-norm-consistent norm $\|.\|_g$ *and some positive constants* a_x *and* a_λ*, while each* $|\bar{\lambda}_i^{(k)}|$, $i \in \mathcal{I}^*$*, converges to zero at a Q-superlinear rate.*

We shall now investigate the behaviour of Algorithm 3.1 once the penalty parameter has converged to its asymptotic value, $\mu_{\min}$. There is no loss of generality in assuming that we restart the algorithm from the point which is reached when the penalty parameter is reduced for the last time. We shall call this iteration $k = 0$ and will start with $\mu^{(0)} = \mu_{\min}$. By construction, (3.7) is satisfied for all k and the updates (3.9) are always performed. Moreover,

$$\omega^{(k)} = \omega_0(\mu_{\min})^{\alpha_\omega + k\beta_\omega} \quad \text{and} \quad \eta^{(k)} = \eta_0(\mu_{\min})^{\alpha_\eta + k\beta_\eta}. \tag{4.7}$$

We require the following extra assumptions.

AS7. (Strict complementary slackness condition 2) Suppose that (x^*, λ^*) is a Kuhn-Tucker point for problem (1.1), (1.2) and (2.6). Then

$$\mathcal{J}_2 = \{j \in \mathcal{N}_b | (g_\ell(x^*, \lambda^*))_j = 0 \quad \text{and} \quad x_j^* = 0\} = \emptyset. \tag{4.8}$$

AS8. If $\mathcal{J}_1$ is defined by (4.3), the approximations $B^{(k,0)}$ satisfy

$$\|(B^{(k,0)} - \nabla_{xx}\ell(x^*, \lambda^*))_{[\mathcal{J}_1,\mathcal{J}_1]} p_{[\mathcal{J}_1]}^{(k,0)}\|_g \leq \upsilon \|p_{[\mathcal{J}_1]}^{(k,0)}\|_g^{1+\varsigma}, \tag{4.9}$$

for some positive constants υ and ς and all k sufficiently large.

AS9. Suppose that (x^*, λ^*) is a Kuhn-Tucker point for the problem (1.1), (1.2) and (2.6), and that $\mathcal{J}_1$ is defined by (4.3). Then we assume that the second derivative approximations $B^{(k,0)}$ have a single limit, B^* and that the perturbed Kuhn-Tucker matrix

$$\begin{pmatrix} B^*_{[\mathcal{J}_1,\mathcal{J}_1]} & A(x^*)^T_{[\mathcal{A}^*,\mathcal{J}_1]} \\ A(x^*)_{[\mathcal{A}^*,\mathcal{J}_1]} & -(D^*_{[\mathcal{A}^*,\mathcal{A}^*]})^{-1} \end{pmatrix} \tag{4.10}$$

is non-singular and has precisely m negative eigenvalues, where D^* is the limiting diagonal matrix with entries

$$D^*_{i,i} \equiv \lim_{k\to\infty} D^{(k)}(x^{(k)})_{i,i} = \begin{cases} (\lambda_i^*)^{1-\alpha_\lambda}/\mu_{\min} & \text{if } i \in \mathcal{A}^* \\ 0 & \text{if } i \in \mathcal{I}^* \end{cases} \tag{4.11}$$

Assumptions AS5 and AS7 are often known as strict complementary slackness conditions. We observe that AS8 is closely related to the necessary and sufficient conditions for superlinear convergence of the inner iterates given by Dennis and Moré (1974). We also observe that AS9 is entirely equivalent to requiring that the matrix

$$B^*_{[\mathcal{J}_1,\mathcal{J}_1]} + A(x^*)^T_{[\mathcal{A}^*,\mathcal{J}_1]} D^*_{[\mathcal{A}^*,\mathcal{A}^*]} A(x^*)_{[\mathcal{A}^*,\mathcal{J}_1]} \tag{4.12}$$

is positive definite (see, for instance, Gould, 1986). The uniqueness of the limit point in AS9 can also be relaxed by requiring that (4.12) has its smallest eigenvalue uniformly bounded from below by some positive quantity for all limit points B^* of the sequence $B^{(k,0)}$. Moreover it is easy to show that that AS4, AS5 and AS7 guarantee AS9 provided that $\mu_{\min}$ is sufficiently small and sufficient second-order optimality conditions (see Fiacco and McCormick, 1968, Theorem 4) hold at x^* (see Wright, 1992, Theorem 8, for the essence of a proof of this in our case). Although we shall merely assume that AS9 holds in this paper, it is of course possible to try to encourage this eventuality. We might, for instance, insist that Step 4 of Algorithm 3.1 is executed rather than Step 3 so long as the matrix $H^{(k,0)}$ is not positive definite. This is particularly relevant if exact second derivatives are used.

We now show that if we perform the step calculation for Algorithm 3.2 using Algorithm 3.3, a single iteration of Algorithm 3.2 suffices to complete an iteration of Algorithm 3.1 when k is sufficiently large. Moreover, the solution of one inner-iteration subproblem, $x^{(k-1)}$ and the shifted starting point for the next inner iteration (3.18) are asymptotically identical. We do this by showing that, after a finite number of iterations,

(i) moving to the new starting point does not significantly alter the norms of the projected gradient or constraints. Furthermore, the status of each variable (floating or dominated) is unchanged by the move;

(ii) the generalized Cauchy point $x^{C(k,0)}$ occurs before the first "breakpoint" along the Cauchy arc — the breakpoints are the values of $t > 0$ at which the Cauchy arc changes direction as problem or trust region bounds are encountered. Thus the set of variables which are free at the start of the Cauchy arc $x^{(k,0)}$ and those which are free at the generalized Cauchy point are identical;

(iii) any step which satisfies (3.30) also satisfies $p_{[\mathcal{F}_1]}$ lies strictly interior to $\mathcal{C}(\beta_2)$. Thus a single pass of Step 2 of Algorithm 3.3 is required;

(iv) the step $p^{(k,0)}$ is accepted in Step 4 of Algorithm 3.2;

(v) the new point $x^{(k,1)}$ satisfies the convergence test (3.11); and

(vi) $x^{(k+1,0)} = x^{(k)}$.

We have the following theorem.

Theorem 4.3 *Assume that assumptions AS1–AS9 hold and that the convergence tolerances β_ω and β_η satisfy the extra condition*

$$\beta_\omega < (1 + \min(\xi, \varsigma))\alpha_\lambda \beta_\eta. \tag{4.13}$$

Then for all k sufficiently large, a single inner iteration of Algorithm 3.2, with the step computed from Algorithm 3.3, suffices to complete an iteration of Algorithm 3.1. Moreover, the solution to one inner iteration subproblem provides the starting point for the next without further adjustment, for all k sufficiently large.

Proof. In order to make the proof as readable as possible, we will make frequent use of the following shorthand: the iterates will be abbreviated as

$$x \equiv x^{(k)} \stackrel{(3.15)}{\rightarrow} \hat{x} \equiv \hat{x}^{(k)} \stackrel{(3.17)}{\rightarrow} x^{\oplus} \equiv x^{(k+1,0)} \stackrel{(3.18)}{\rightarrow} x^{+} \equiv x^{(k+1,1)}, \tag{4.14}$$

the shifts as

$$s \equiv s^{(k)} \rightarrow s^{+} \equiv s^{(k+1)}, \tag{4.15}$$

and the Lagrange multiplier estimates as

$$\lambda \equiv \lambda^{(k)} \rightarrow \bar{\lambda} \equiv \bar{\lambda}^{(k)} \equiv \bar{\lambda}(x, \lambda, s) \rightarrow \lambda^{+} \equiv \lambda^{(k+1)} \quad \text{and} \quad \bar{\lambda}^{+} \equiv \bar{\lambda}(x^{\oplus}, \lambda^{+}, s^{+}). \tag{4.16}$$

Other quantities which occur at inner iterations $(k+1, 0)$ and $(k+1, 1)$ will be given suffices $\oplus$ and $+$ respectively. Thus $H^{\oplus} \equiv H^{(k+1,0)}$ and $H^{+} \equiv H^{(k+1,1)}$.

Recall, we have used Theorem 4.2 to relabel the sequence of iterates so that

$$\|P(x^{(k)}, \nabla_x \Psi^{(k)})\|_g \le \omega_0 (\mu_{\min})^{\alpha_\omega + k\beta_\omega} \tag{4.17}$$

and

$$\left\|\left[c_i(x^{(k)})\bar{\lambda}_i^{(k)}/(\lambda_i^{(k)})^{\alpha_\lambda}\right]_{i=1}^m\right\|_c \leq \eta_0(\mu_{\min})^{\alpha_\eta+k\beta_\eta} \tag{4.18}$$

for all $k \geq 0$. Let $\bar{\Omega}$ be any closed, bounded set containing the iterates $x^{(k)}$ and $x^{(k+1,0)}$. We shall follow the outline given above.

(i) Status of the starting point. The strict complementary slackness assumption AS7 ensures that for all k sufficiently large, each variable belongs exclusively to one of the sets $\mathcal{F}_1$ and $\mathcal{D}_1$ (see Conn *et al.*, 1992b, Theorem 5.4); moreover,

$$g_\ell(x^*,\lambda^*)_j = 0 \quad \text{for all} \quad j \in \mathcal{F}_1 \quad \text{and} \quad x_j^* > 0 \quad \text{for all} \quad j \in \mathcal{F}_1 \cap \mathcal{N}_b \tag{4.19}$$

and

$$x_j^* = 0 \quad \text{and} \quad g_\ell(x^*,\lambda^*)_j > 0 \quad \text{for all} \quad i \in \mathcal{D}_1. \tag{4.20}$$

As one of $x_j^{(k)}$ and $\nabla_x\Psi_j^{(k)}$ ($\equiv \nabla_x\ell(x,\bar{\lambda})_j$) converges to zero while its partner converges to a strictly positive limit for each $j \in \mathcal{N}_b$ (assumption AS7), we may define nontrivial regions which separate the two sequences for all k sufficiently large. Let

$$\epsilon_x \stackrel{\text{def}}{=} \frac{\theta}{1+\theta}\min_{j\in\mathcal{N}_b}\max[x_j^*, g_\ell(x^*,\lambda^*)_j] > 0, \tag{4.21}$$

where θ is as in (3.17). Then there is an iteration k_0 such that for variables in $\mathcal{F}_1$,

$$|x_j^{(k)} - x_j^*| \leq \epsilon_x \quad \text{and} \quad |\nabla_x\Psi_j^{(k)}| < \epsilon_x, \tag{4.22}$$

while for those in $\mathcal{D}_1$,

$$|x_j^{(k)}| \leq \epsilon_x \quad \text{and} \quad |\nabla_x\Psi_j^{(k)} - g_\ell(x^*,\lambda^*)_j|, \leq \epsilon_x \tag{4.23}$$

for all $k \geq k_0$. Hence, for those variables in $\mathcal{D}_1$, (4.21) and (4.23) give that

$$x_j^{(k)} \leq \epsilon_x = \theta[\min_{j\in\mathcal{N}_b}\max[x_j^*, g_\ell(x^*,\lambda^*)_j] - \epsilon_x] \leq \theta[g_\ell(x^*,\lambda^*)_j - \epsilon_x] \leq \theta(\nabla_x\Psi^{(k)})_j. \tag{4.24}$$

Thus, by definition (3.17), $\hat{x}_j^{(k)} = 0$ for each $j \in \mathcal{D}_1$ when $k \geq k_0$. Similarly, when $j \in \mathcal{F}_1 \cap \mathcal{N}_b$ and $k \geq k_0$, $x_j^{(k)} > \theta(\nabla_x\Psi^{(k)})_j$ and hence, using (3.17), $\hat{x}_j^{(k)} = x_j$ for all $j \in \mathcal{F}_1$. Thus $\hat{x}^{(k)}$ converges to x^*.

The other strict complementary slackness assumption, AS5, ensures that each constraint belongs exclusively to one of the sets $\mathcal{I}^*$ and $\mathcal{A}^*$, for all k sufficiently large. Moreover,

$$c_i(x^*) = 0 \quad \text{and} \quad \lambda_i^* > 0 \quad \text{for all} \quad i \in \mathcal{A}^* \tag{4.25}$$

and

$$c_i(x^*) > 0 \quad \text{and} \quad \lambda_i^* = 0 \quad \text{for all} \quad i \in \mathcal{I}^*, \tag{4.26}$$

and thus one of $c_i(x^{(k)})$ and $\lambda_i^{(k+1)}$ converges to zero while its partner converges to a strictly positive limit for each i.

Using the shorthand introduced in (4.14)–(4.15), we have that $c_i(x) + s_i^+ > c_i(x) > 0$ for each $i \in \mathcal{I}^*$ and all k sufficiently large. Thus, as $\hat{x}$ converges to x^* and s_i^+ converges to zero,

$2c_i(x^*) > c_i(\hat{x}) + s_i^+ > \frac{1}{2}c_i(x^*) > 0$ for all $i \in \mathcal{I}^*$ and k sufficiently large. On the other hand, if $i \in \mathcal{A}^*$, $c_i(x) + s_i^+ > 0$ for all k (see Conn *et al.*, 1992b, Lemma 3.1). In this case, as s_i^+ converges to $s_i^* \equiv \mu_{\min}(\lambda_i^*)^{\alpha_\lambda} > 0$ and $c_i(x)$ converges to zero, the convergence of $\hat{x}$ to x^* and λ_i^+ to λ^* implies that $2s_i^* > c_i(\hat{x}) + s_i^+ > \frac{1}{2}s_i^* > 0$ for all k sufficiently large. Hence, from (3.18), $x^\oplus = \hat{x}$ and thus there is an integer $k_1 \geq k_0$ for which

$$x_j^\oplus = \begin{cases} x_j & \text{for all } j \in \mathcal{F}_1 \\ 0 & \text{for all } j \in \mathcal{D}_1, \end{cases} \tag{4.27}$$

for all $k \geq k_1$.

We next let r be any real number and consider points on the line

$$x(r) \stackrel{\text{def}}{=} x + r(x^\oplus - x). \tag{4.28}$$

We firstly show that the diagonal matrix $D(x(r))$ is bounded for all $0 \leq r \leq 1$, where D is given by (3.13). As x and $x^\oplus$ both converge to x^*, the definition (3.13) implies that $D(x(r))$ converges to the matrix $D_{i,i}^*$, satisfying (4.11), as k increases. Thus, we have the bound

$$\|D(x(r))\|_2 \leq a_1/\mu_{\min} \tag{4.29}$$

where $a_1 \stackrel{\text{def}}{=} 2\|[(\lambda_i^*)^{1-\alpha_\lambda}]_{i=1}^m\|_2$, for all k sufficiently large. It also follows from the convergence of x and $x^\oplus$ to x^* and that of s_i to s_i^* that there is an integer $k_2 \geq k_1$ for which

$$0 < \tfrac{1}{2}c_i(x^*) < c_i(x(r)) + s_i^{(l)} < 2c_i(x^*) \quad \text{for all} \quad i \in \mathcal{I}^* \tag{4.30}$$

and

$$0 < \tfrac{1}{2}\mu_{\min}(\lambda_i^*)^{\alpha_\lambda} < c_i(x(r)) + s_i^{(l)} < 2\mu_{\min}(\lambda_i^*)^{\alpha_\lambda} \quad \text{for all} \quad i \in \mathcal{A}^*, \tag{4.31}$$

for all k sufficiently large and $l \geq k_2$.

We now consider the starting point $x^\oplus$ for the next inner iteration in detail. Firstly, combining (2.9), (2.13) and (4.27), we have that

$$\|x^\oplus - x\|_z \leq a_0\|P(x, \nabla_x\Psi(x, \lambda, s))\|_g \leq a_0\omega_0(\mu_{\min})^{\alpha_\omega + k\beta_\omega} \tag{4.32}$$

for any two-norm-consistent norm $\|.\|_z$.

We may bound the change in $c(x)$, due to the shifted starting point, using the integral mean value theorem (see, eg, Dennis and Schnabel, 1983, page 74), the boundedness of $A(x)$ (assumption AS1 and the definition of $\bar{\Omega}$) and inequalities (2.14) and (4.32) to obtain

$$\begin{aligned} |c_i(x^\oplus) - c_i(x)| \leq & \ \|\textstyle\int_0^1 A(x(r))dr\|_g \|x^\oplus - x\|_g \\ \leq & \ a_0 a_2 \omega_0(\mu_{\min})^{\alpha_\omega + k\beta_\omega} \end{aligned} \tag{4.33}$$

where $x(r)$ is given by (4.28) and a_2 is an upper bound on $\|A(x)\|_g$ within $\bar{\Omega}$.

We next bound the differences in gradients of the Lagrangian barrier function at x and $x^\oplus$. Using the integral mean value theorem, the convergence of $\bar{\lambda} \equiv \lambda^+$ to λ^* (Theorem 4.1), the boundedness of the Hessian of the Lagrangian (with bounded multiplier estimates) and the constraint Jacobian within $\bar{\Omega}$ (assumption AS1) and the inequalities (2.14), (4.29) and (4.32), we obtain

$$\begin{array}{rcl} |\nabla_x \Psi(x^\oplus, \lambda, s)_j - \nabla_x \Psi(x, \lambda, s)_j| & \leq & \|x^\oplus - x\|_2 \cdot \\ & & \|e_j^T \int_0^1 [H_\ell(x(r), \lambda) + A(x(r))^T D(x(r)) A(x(r))] dr\|_2 \\ & \leq & a_0^2 (a_3 + a_1 a_2^2 / \mu_{\min}) \omega_0 (\mu_{\min})^{\alpha_\omega + k\beta_\omega} \\ & \leq & a_0^2 (a_3 + a_1 a_2^2) \omega_0 (\mu_{\min})^{\alpha_\omega - 1 + k\beta_\omega}, \end{array} \tag{4.34}$$

where a_3 is an upper bound on the two-norm of the Hessian of the Lagrangian function (with bounded multiplier estimates) within $\bar{\Omega}$. We now use the identity

$$\lambda_i^+ = \frac{\lambda_i s_i}{c_i(x) + s_i} \tag{4.35}$$

to derive the relationship

$$\begin{array}{rl} \nabla_x \ \Psi(x^\oplus, \lambda^+, s^+) - \nabla_x \Psi(x^\oplus, \lambda, s) = & \sum_{i=1}^m \left(\frac{\lambda_i s_i}{c_i(x^\oplus) + s_i} - \frac{\lambda_i^+ s_i^+}{c_i(x^\oplus) + s_i^+} \right) a_i(x^\oplus) \\ = & \sum_{i=1}^m \left(\frac{\lambda_i s_i}{c_i(x^\oplus) + s_i} - \frac{\lambda_i s_i}{c_i(x) + s_i} \right) a_i(x^\oplus) + \sum_{i=1}^m \left(\frac{\lambda_i s_i}{c_i(x) + s_i} - \frac{\lambda_i^+ s_i^+}{c_i(x^\oplus) + s_i^+} \right) a_i(x^\oplus) \\ = & \sum_{i=1}^m \left(\frac{\lambda_i^+ (c_i(x) - c_i(x^\oplus))}{c_i(x^\oplus) + s_i} + \frac{\lambda_i^+ c_i(x^\oplus)}{c_i(x^\oplus) + s_i^+} \right) a_i(x^\oplus) \\ = & \sum_{i \in \mathcal{A}^*} \left(\frac{\lambda_i^+ (c_i(x) - c_i(x^\oplus))}{c_i(x^\oplus) + s_i} + \frac{\lambda_i^+ c_i(x^\oplus)}{c_i(x^\oplus) + s_i^+} \right) a_i(x^\oplus) + \\ & \sum_{i \in \mathcal{I}^*} \left(\frac{\lambda_i^+ (c_i(x) - c_i(x^\oplus))}{c_i(x^\oplus) + s_i} + \frac{\lambda_i^+ c_i(x^\oplus)}{c_i(x^\oplus) + s_i^+} \right) a_i(x^\oplus). \end{array} \tag{4.36}$$

But, considering $i \in \mathcal{A}^*$, picking k sufficiently large so that $|\lambda_i^+| \leq 2|\lambda_i^*|$ and using the integral mean value theorem, the relationship $c(x^*)_{[\mathcal{A}^*]} = 0$, the bounds (4.31), (4.32), (4.33) and the inequalities (3.2) and (4.6), we obtain the bounds

$$\left| \frac{\lambda_i^+ (c_i(x) - c_i(x^\oplus))}{c_i(x^\oplus) + s_i} \right| \leq 4 a_0 a_2 \omega_0 (\lambda_i^*)^{1 - \alpha_\lambda} (\mu_{\min})^{\alpha_\omega - 1 + k\beta_\omega} \tag{4.37}$$

and

$$\begin{array}{rl} \left| \frac{\lambda_i^+ c_i(x^\oplus)}{c_i(x^\oplus) + s_i^+} \right| & \leq 4 (\lambda_i^*)^{1 - \alpha_\lambda} (\mu_{\min})^{-1} \| \int_0^1 a_i(x^\oplus + r(x^* - x^\oplus)) dr \|_g \|x^\oplus - x^*\|_g \\ & \leq 4 (\lambda_i^*)^{1 - \alpha_\lambda} (\mu_{\min})^{-1} a_2 (\|x^\oplus - x\|_g + \|x - x^*\|_g) \\ & \leq 4 a_2 (\lambda_i^*)^{1 - \alpha_\lambda} (a_0 \omega_0 (\mu_{\min})^{\alpha_\omega - 1 + k\beta_\omega} + a_x (\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda \beta_\eta}) \end{array} \tag{4.38}$$

and hence

$$\begin{array}{rl} \| & \sum_{i \in \mathcal{A}^*} \left(\frac{\lambda_i^+ (c_i(x) - c_i(x^\oplus))}{c_i(x^\oplus) + s_i} + \frac{\lambda_i^+ c_i(x^\oplus)}{c_i(x^\oplus) + s_i^+} \right) a_i(x^\oplus) \|_z \\ & \leq m a_0 a_2 \left(\max_{i \in \mathcal{A}^*} \left| \frac{\lambda_i^+ (c_i(x) - c_i(x^\oplus))}{c_i(x^\oplus) + s_i} \right| + \max_{i \in \mathcal{A}^*} \left| \frac{\lambda_i^+ c_i(x^\oplus)}{c_i(x^\oplus) + s_i^+} \right| \right) \\ & \leq a_{\mathcal{A}} (\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda \beta_\eta}, \end{array} \tag{4.39}$$

where $a_{\mathcal{A}} = 4 m a_0 a_2^2 (2 a_0 \omega_0 + a_x) \max_{i \in \mathcal{A}^*} (\lambda_i^*)^{1 - \alpha_\lambda}$, for any two-norm-consistent norm $\|.\|_z$. Furthermore, the superlinear convergence of λ_i to zero, $i \in \mathcal{I}^*$, (4.30) and the boundedness of the remaining terms implies a bound

$$\| \sum_{i \in \mathcal{I}^*} \left(\frac{\lambda_i^+ (c_i(x) - c_i(x^\oplus))}{c_i(x^\oplus) + s_i} + \frac{\lambda_i^+ c_i(x^\oplus)}{c_i(x^\oplus) + s_i^+} \right) a_i(x^\oplus) \|_z \leq a_{\mathcal{I}} (\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda \beta_\eta}, \tag{4.40}$$

for some constant $a_{\mathcal{I}}$ (In fact, this term can be made arbitrarily smaller than (4.39) by picking k sufficiently large). Thus, combining (4.36), (4.39) and (4.40), we obtain the componentwise bound

$$|\nabla_x \Psi_j^{\oplus} - \nabla_x \Psi(x^{\oplus}, \lambda, s)_j| \leq (a_{\mathcal{A}} + a_{\mathcal{I}})(\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda\beta_\eta} \tag{4.41}$$

for all $j \in \mathcal{N}$ where we have abbreviated $\nabla_x \Psi(x^{\oplus}, \lambda^+, s^+)$ as $\nabla_x \Psi^{\oplus}$.

Now consider the variables whose indices j lie in $\mathcal{F}_1$ for $k \geq k_2$. Firstly, (4.19), (4.21), (4.22) and (4.27) show that

$$x_j^{\oplus} = x_j \geq \frac{x_j^*}{1+\theta} > 0 \tag{4.42}$$

if $j \in \mathcal{N}_b$. Secondly, combining (4.34) and (4.41), and using (2.10), (2.14), (3.2) and (4.17), we derive the inequality

$$\begin{array}{rl} |\nabla_x \Psi_j^{\oplus}| & \\ \leq & |\nabla_x \Psi_j^{\oplus} - \nabla_x \Psi(x^{\oplus}, \lambda, s)_j| + |\nabla_x \Psi(x^{\oplus}, \lambda, s)_j - \nabla_x \Psi(x, \lambda, s)_j| + |\nabla_x \Psi(x, \lambda, s)_j| \\ \leq & (a_{\mathcal{A}} + a_{\mathcal{I}})(\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda\beta_\eta} + a_0^2(a_3 + a_1 a_2^2)\omega_0(\mu_{\min})^{\alpha_\omega - 1 + k\beta_\omega} + a_0\omega_0(\mu_{\min})^{\alpha_\omega + k\beta_\omega} \\ \leq & a_4(\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda\beta_\eta}, \end{array} \tag{4.43}$$

where $a_4 \stackrel{\text{def}}{=} a_{\mathcal{A}} + a_{\mathcal{I}} + a_0\omega_0(1 + a_0(a_3 + a_1 a_2^2))$. As k increases, the right-hand-side of the inequality (4.43) converges to zero. Thus, from (4.22) and for k sufficiently large, $x_j^{\oplus}$ is floating for each $j \in \mathcal{F}_1$, and (2.10) and (4.43) imply that

$$|P(x^{\oplus}, \nabla_x \Psi_j^{\oplus})| = |\nabla_x \Psi_j^{\oplus}| \leq a_4(\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda\beta_\eta}. \tag{4.44}$$

Conversely, consider the variables which lie in $\mathcal{D}_1$ for $k \geq k_2$. Then, combining (4.34) and (4.41), and using (2.14) and (3.2) we obtain the inequality

$$\begin{array}{rl} |\nabla_x \Psi_j^{\oplus} - \nabla_x \Psi(x, \lambda, s)_j| & \\ \leq & |\nabla_x \Psi_j^{\oplus} - \nabla_x \Psi(x^{\oplus}, \lambda, s)_j| + |\nabla_x \Psi(x^{\oplus}, \lambda, s)_j - \nabla_x \Psi(x, \lambda, s)_j| \\ \leq & (a_{\mathcal{A}} + a_{\mathcal{I}})(\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda\beta_\eta} + a_0^2(a_3 + a_1 a_2^2)\omega_0(\mu_{\min})^{\alpha_\omega - 1 + k\beta_\omega} \\ \leq & a_5(\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda\beta_\eta}, \end{array} \tag{4.45}$$

where $a_5 \stackrel{\text{def}}{=} a_{\mathcal{A}} + a_{\mathcal{I}} + a_0^2\omega_0(a_3 + a_1 a_2^2)$. Thus, for sufficiently large k the right-hand-side of (4.45) can be made arbitrarily small. Combining this result with (4.23) and the identity $x_j^{\oplus} = 0$, we see that $x_j^{\oplus}$ is dominated for each $j \in \mathcal{D}_1$, and (2.9) and (4.45) imply that

$$P(x^{\oplus}, \nabla_x \Psi_j^{\oplus}) = x_j^{\oplus} = 0. \tag{4.46}$$

Therefore, using (2.10), (2.14), (4.44) and (4.46), we have

$$\|P(x^{\oplus}, \nabla_x \Psi^{\oplus})\|_g = \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_g \leq a_6(\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda\beta_\eta}, \tag{4.47}$$

for all k sufficiently large, where $a_6 \stackrel{\text{def}}{=} a_0 a_4 \|e_{[\mathcal{F}_1]}\|_2$.

We also need to be able to bound the Lagrange multiplier estimates $\bar{\lambda}^+ \equiv \bar{\lambda}(x^{\oplus}, \lambda^+, s^+)$. We have, from (2.5), that

$$|\lambda_i^+ - \bar{\lambda}_i^+| = \left| \frac{\bar{\lambda}_i^+ c_i^{\oplus}}{c_i^{\oplus} + s_i^+} \right|. \tag{4.48}$$

But then, recalling (4.38), when $i \in \mathcal{A}^*$, and the superlinear convergence of λ_i^+ to zero, when $i \in \mathcal{I}^*$, together with (3.2), we obtain a bound

$$\|\lambda^+ - \bar{\lambda}^+\|_g \leq a_{\lambda^+} (\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda \beta_\eta}, \tag{4.49}$$

for some constant a_{λ^+}. Thus, combining (4.6) and (4.49), we see that $\bar{\lambda}^+$ converges to λ^*, $i \in \mathcal{A}^*$, and, because λ_i^+ converges superlinearly to zero when $i \in \mathcal{I}^*$,

$$\|\bar{\lambda}^+ - \lambda^*\|_g \leq a_{\lambda^\oplus} (\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda \beta_\eta}, \tag{4.50}$$

for some constant $a_{\lambda^\oplus}$.

(ii) The generalized Cauchy point. We consider the Cauchy arc emanating from $x^\oplus$. We have shown that the variables in $\mathcal{D}_1$ are on their bounds; the relationships (4.20), (4.22), (4.23) and (4.45) imply that $\nabla_x \Psi_j^\oplus > 0$ for all sufficiently large k and hence that $p^\oplus(t)_j = 0$ for all $t > 0$ and $j \in \mathcal{D}_1$. Thus the variables in $\mathcal{D}_1$ remain fixed on the bounds throughout the first inner iteration and

$$p^\oplus_{[\mathcal{D}_1]} = 0 \tag{4.51}$$

for all k sufficiently large.

The remaining variables, those indexed by $\mathcal{F}_1$, are free from their bounds. Because of Assumption 7 the set $\mathcal{J}_1$ in assumption AS9 is identical to $\mathcal{F}_1$ and thus the matrix (4.12) is positive definite with extreme eigenvalues $0 < \pi_{\min} \leq \pi_{\max}$, say. Using (4.27) and inequalities (2.9), (2.10) and (3.5), we deduce that $x^\oplus$ converges to x^*. Thus the matrix

$$H^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]} = B^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]} + A(x^\oplus)^T_{[\mathcal{F}_1]} D^+(x^\oplus) A(x^\oplus)_{[\mathcal{F}_1]} \tag{4.52}$$

is also positive definite with extreme eigenvalues satisfying

$$0 < \tfrac{1}{2}\pi_{\min} \leq \pi^\oplus_{\min} \leq \pi^\oplus_{\max} \leq 2\pi_{\max}, \tag{4.53}$$

say, for all sufficiently large k. Hence the model (3.12) is a strictly convex function in the subspace of free variables during the first inner iteration.

We now show that the set

$$\mathcal{L} \stackrel{\text{def}}{=} \{p_{[\mathcal{F}_1]} \mid m^\oplus(x^\oplus + p) \leq m^\oplus(x^\oplus) \quad \text{and} \quad p_{[\mathcal{D}_1]} = 0\} \tag{4.54}$$

lies strictly interior to the set $\mathcal{C}(1)$ (defined in Algorithm 3.3) for all k sufficiently large. The diameter d of $\mathcal{L}$, the maximum distance between two members of the set (measured in the two norm), can be no larger than twice the distance from the center of the ellipsoid defined by $\mathcal{L}$ to the point on $\bar{\mathcal{L}}$ (the boundary of $\mathcal{L}$) furthest from the center. The center of $\mathcal{L}$ is the Newton point,

$$p^*_{[\mathcal{F}_1]} = -(H^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]})^{-1} \nabla_x \Psi^\oplus_{[\mathcal{F}_1]}. \tag{4.55}$$

Let $p_{[\mathcal{F}_1]} \in \bar{\mathcal{L}}$ and $p^*_{[\mathcal{D}_1]} = 0$ and define $v \stackrel{\text{def}}{=} p - p^*$. Then, combining (3.12), (4.52), (4.54) and (4.55), we have that

$$\begin{aligned} \tfrac{1}{2} v^T_{[\mathcal{F}_1]} H^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]} v_{[\mathcal{F}_1]} &= \tfrac{1}{2} p^{*T}_{[\mathcal{F}_1]} H^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]} p^*_{[\mathcal{F}_1]} + (m^\oplus(x^\oplus + p^* + v) - m^\oplus(x^\oplus)) \\ &\quad - (p^* + v)^T_{[\mathcal{F}_1]} (H^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]} p^*_{[\mathcal{F}_1]} + \nabla_x \Psi^\oplus_{[\mathcal{F}_1]}) \\ &= \tfrac{1}{2} p^{*T}_{[\mathcal{F}_1]} H^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]} p^*_{[\mathcal{F}_1]} = \tfrac{1}{2} \nabla_x \Psi^{\oplus T}_{[\mathcal{F}_1]} (H^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]})^{-1} \nabla_x \Psi^\oplus_{[\mathcal{F}_1]}. \end{aligned} \tag{4.56}$$

Hence, using the extremal properties of the Rayleigh quotient and (4.56), we have

$$\begin{aligned} d^2 \stackrel{\text{def}}{=} & \; 4\|v^*_{[\mathcal{F}_1]}\|_2^2 \leq 4 v^{*T}_{[\mathcal{F}_1]} H^{\oplus}_{[\mathcal{F}_1,\mathcal{F}_1]} v^*_{[\mathcal{F}_1]} / \pi^{\oplus}_{\min} \leq 8 v^{*T}_{[\mathcal{F}_1]} H^{\oplus}_{[\mathcal{F}_1,\mathcal{F}_1]} v^*_{[\mathcal{F}_1]} / \pi_{\min} \\ = & \; 8 \nabla_x \Psi^{\oplus T}_{[\mathcal{F}_1]} (H^{\oplus}_{[\mathcal{F}_1,\mathcal{F}_1]})^{-1} \nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]} / \pi_{\min} \leq 16 \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_2^2 / \pi^2_{\min} \end{aligned} \tag{4.57}$$

where $\|v^*_{[\mathcal{F}_1]}\|_2 = \max_{p^*_{[\mathcal{F}_1]} + v_{[\mathcal{F}_1]} \in \bar{\mathcal{L}}} \|v_{[\mathcal{F}_1]}\|_2$. Thus, using (2.14), (4.47) and (4.57), any step within $\mathcal{L}$ satisfies the bound,

$$\|p_{[\mathcal{F}_1]}\|_2 \leq d \leq 4 \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_2 / \pi_{\min} \leq 4 a_0 a_6 (\mu_{\min})^{\alpha_\eta - 1 + k \alpha_\lambda \beta_\eta} / \pi_{\min}, \tag{4.58}$$

for sufficiently large k.

The inequality (4.42) shows that $x^{\oplus}_j$, $j \in \mathcal{F}_1 \cap \mathcal{N}_b$, is separated from its bound for all k sufficiently large while (4.58) shows that all steps within $\mathcal{L}$ become arbitrarily small. Thus the problem bounds are excluded from $\mathcal{L}$. Moreover (2.13), (3.19), (4.47), (4.51) and (4.58) combine to give

$$\|p\|_t = \|p_{[\mathcal{F}_1]}\|_t \leq a_0^{\frac{1}{2}} \|p_{[\mathcal{F}_1]}\|_2 \leq \Delta^{\oplus} \frac{4 a_0 \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_g^{1-\zeta}}{\pi_{\min} \kappa}. \tag{4.59}$$

for all steps within, or on the boundary of, $\mathcal{L}$. Inequality (4.47) then combines with (4.59) to show that any such step is shorter than the distance to the trust region boundary for all k sufficiently large.

Thus $\mathcal{L}$ lies strictly interior to $\mathcal{C}(1) \subseteq \mathcal{C}(\beta_2)$ for all k sufficiently large. But, as all iterates generated by Algorithm 3.3 satisfy (3.26) and thus lie in $\mathcal{L}$, it follows that both the generalized Cauchy point and any subsequent improvements are not restricted by the boundaries of $\mathcal{C}$ or $\mathcal{C}(\beta_2)$.

It remains to consider the Cauchy step in more detail. The Cauchy arc starts in the steepest descent direction for the variables in $\mathcal{F}_1$. The minimizer of the model in this direction occurs when

$$t = t^* = \frac{\nabla_x \Psi^{\oplus T}_{[\mathcal{F}_1]} \nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}}{\nabla_x \Psi^{\oplus T}_{[\mathcal{F}_1]} H^{\oplus}_{[\mathcal{F}_1,\mathcal{F}_1]} \nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}}. \tag{4.60}$$

and thus, from the above discussion, gives the generalized Cauchy point proposed by Conn *et al.* (1988a). We use the definition of t^*, (2.13), (4.53) and the extremal property of the Rayleigh quotient to obtain

$$m^{\oplus}(x^{\oplus}) - m^{\oplus}(x^{\oplus} + p^{C\oplus}) = \tfrac{1}{2} t^* \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_2^2 \geq \frac{\|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_g^2}{4 a_0 \pi_{\max}} \tag{4.61}$$

for this variant of the generalized Cauchy point. Alternatively, if Moré's (1988) variant is used, the requirement (3.22) and the definition of the Cauchy arc imply that

$$m^{\oplus}(x^{\oplus}) - m^{\oplus}(x^{\oplus} + p^{C\oplus}) \geq \mu_1 t^{C\oplus} \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_2^2. \tag{4.62}$$

If the first alternative of (3.23) holds, (4.62) implies that

$$m^{\oplus}(x^{\oplus}) - m^{\oplus}(x^{\oplus} + p^{C\oplus}) \geq \mu_1 \nu_1 \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_2^2. \tag{4.63}$$

Otherwise, we may use the same arguments as above to show that it is impossible for $t^{L\oplus}$ to satisfy (3.25) when k is sufficiently large. Therefore, $t^{L\oplus}$ must satisfy (3.24). Combining (3.12), (3.24), (4.52) and the definition of the Cauchy arc, we have that

$$\tfrac{1}{2}(t^{L\oplus})^2 \nabla_x \Psi^{\oplus T}_{[\mathcal{F}_1]} H^{\oplus}_{[\mathcal{F}_1,\mathcal{F}_1]} \nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]} \geq (1-\mu_2) t^{L\oplus} \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_2^2. \tag{4.64}$$

Hence, combining (4.53) and (4.64) with the extremal properties of the Rayleigh quotient, we have that $t^{L\oplus} \geq (1-\mu_2)/\pi_{\max}$. Thus, when the second alternative of (3.23) holds, this result and (4.62) give that

$$m^{\oplus}(x^{\oplus}) - m^{\oplus}(x^{\oplus} + p^{C\oplus}) \geq [\mu_1 \nu_2 (1-\mu_2)/\pi_{\max}] \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_2^2. \tag{4.65}$$

Therefore, (2.14), (4.63) and (4.65) give the inequality

$$m^{\oplus}(x^{\oplus}) - m^{\oplus}(x^{\oplus} + p^{C\oplus}) \geq (\mu_1/a_0) \min(\nu_1, \nu_2(1-\mu_2)/\pi_{\max}) \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_g^2. \tag{4.66}$$

We shall make use of these results in (iv) below.

(iii) Improvements beyond the generalized Cauchy point. We have that $x^{\oplus}_{[\mathcal{D}]} = 0$, and, as a consequence of (4.47), $\|P(x^{\oplus}, \nabla_x \Psi^{\oplus})\|_g^{\xi} \leq \nu$ for all k sufficiently large. Hence, because we have shown that any p in $\mathcal{L}$ lies strictly interior to $\mathcal{C}$, a single pass of Step 2 of Algorithm 3.3 is required. We must pick p to satisfy (3.30) and (3.26) by determining $p^{\oplus}_{[\mathcal{F}_1]}$ so that

$$\|H^{\oplus}_{[\mathcal{F}_1,\mathcal{F}_1]} p^{\oplus}_{[\mathcal{F}_1]} + \nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_g \leq \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_g^{1+\xi}. \tag{4.67}$$

and

$$m^{\oplus}(x^{\oplus}) - m^{\oplus}(x^{\oplus} + p^{\oplus}) \geq \beta_3 [m^{\oplus}(x^{\oplus}) - m^{\oplus}(x^{\oplus} + p^{C\oplus})] \tag{4.68}$$

for some $\beta_3 \leq 1$. The set of values which satisfy (4.67) and (4.68) is non-empty as the Newton step (4.55) satisfies both inequalities.

It remains to consider such a step in slightly more detail. Suppose that $p^{\oplus}_{[\mathcal{F}_1]}$ satisfies (4.67). Let

$$r^{\oplus}_{[\mathcal{F}_1]} = H^{\oplus}_{[\mathcal{F}_1,\mathcal{F}_1]} p^{\oplus}_{[\mathcal{F}_1]} + \nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]} \tag{4.69}$$

Then combining (2.13), (4.53), (4.67) and (4.69), we have

$$\begin{aligned} \|p^{\oplus}_{[\mathcal{F}_1]}\|_g \leq\ & a_0 \|H^{\oplus -1}_{[\mathcal{F}_1,\mathcal{F}_1]}\|_2 (\|r^{\oplus}_{[\mathcal{F}_1]}\|_g + \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_g) \\ \leq\ & 2a_0 \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_g (1 + \|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\|_g^{\xi})/\pi_{\min}. \end{aligned} \tag{4.70}$$

Thus, combining (4.47) and (4.70), and picking k sufficiently large so that $\|\nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]}\| \leq 1$, we obtain the bound

$$\|p^{\oplus}_{[\mathcal{F}_1]}\|_g \leq\ 4a_0 a_6 (\mu_{\min})^{\alpha_\eta - 1 + k\alpha_\lambda \beta_\eta}/\pi_{\min}. \tag{4.71}$$

(iv) Acceptance of the new point. We have seen that

$$p^{\oplus}_{[\mathcal{D}_1]} = 0 \tag{4.72}$$

and $p^{\oplus}_{[\mathcal{F}_1]}$ satisfies (4.67). As $p^{\oplus}$ can be made arbitrarily small, it follows (as in (4.30) and (4.31)) from the convergence of $x^{\oplus}$ to x^* and that of s_i^+ to s_i^* that there is an integer k_3 for which

$$0 < \tfrac{1}{2}c_i(x^*) < c_i(x^{\oplus} + p^{\oplus}) + s_i^+ < 2c_i(x^*) \quad \text{for all} \quad i \in \mathcal{I}^* \tag{4.73}$$

and

$$0 < \tfrac{1}{2}\mu_{\min}(\lambda_i^*)^{\alpha_\lambda} < c_i(x^{\oplus} + p^{\oplus}) + s_i^+ < 2\mu_{\min}(\lambda_i^*)^{\alpha_\lambda} \quad \text{for all} \quad i \in \mathcal{A}^*, \tag{4.74}$$

for all k sufficiently large and $l \geq k_3$. Thus

$$c_i(x^{\oplus} + p^{\oplus}) + s_i^+ > 0 \tag{4.75}$$

for all $1 \leq i \leq m$ and k sufficiently large.

We now wish to show that the quantity

$$|\rho^{\oplus} - 1| = \frac{|\Psi(x^{\oplus} + p^{\oplus}, \lambda^+, s^+) - m^{\oplus}(x^{\oplus} + p^{\oplus})|}{|m^{\oplus}(x^{\oplus}) - m^{\oplus}(x^{\oplus} + p^{\oplus})|} \tag{4.76}$$

converges to zero, ensuring that the new point will prove acceptable in Step 4 of Algorithm 3.2.

Consider first the denominator on the right-hand-side of (4.76). Combining (4.61), (4.66) and (4.68), we have

$$m^{\oplus}(x^{\oplus}) - m^{\oplus}(x^{\oplus} + p^{\oplus}) \geq a_7 \|\nabla_x \Psi(x^{\oplus}, \lambda^+, s^+)_{[\mathcal{F}_1]}\|_g^2, \tag{4.77}$$

where $a_7 = \beta_3 \min(1/(4a_0\pi_{\max}), \mu_1 \min(\nu_1, \nu_2(1-\mu_2)/\pi_{\max})/a_0)$. Turning to the numerator on the right-hand-side of (4.76), we use the integral mean value theorem to obtain

$$\begin{array}{rl}
\multicolumn{2}{l}{\Psi(x^{\oplus} + p^{\oplus}, \lambda^+, s^+)} \\
= & \Psi(x^{\oplus}, \lambda^+, s^+) + p^{\oplus T}_{[\mathcal{F}_1]} \nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]} + \frac{1}{2}\int_0^1 p^{\oplus T}_{[\mathcal{F}_1]} \nabla_{xx}\Psi(x^{\oplus}(t), \lambda^+, s^+)_{[\mathcal{F}_1,\mathcal{F}_1]} p^{\oplus}_{[\mathcal{F}_1]} dt \\
= & \Psi(x^{\oplus}, \lambda^+, s^+) + p^{\oplus T}_{[\mathcal{F}_1]} \nabla_x \Psi^{\oplus}_{[\mathcal{F}_1]} \\
 & + \frac{1}{2}\int_0^1 p^{\oplus T}_{[\mathcal{F}_1]} [\nabla_{xx}\Psi(x^{\oplus}(t), \lambda^+, s^+) - \nabla_{xx}\Psi^{\oplus}]_{[\mathcal{F}_1,\mathcal{F}_1]} p^{\oplus}_{[\mathcal{F}_1]} dt \\
 & + \frac{1}{2} p^{\oplus T}_{[\mathcal{F}_1]} [\nabla_{xx}\Psi^{\oplus} - H^{\oplus}]_{[\mathcal{F}_1,\mathcal{F}_1]} p^{\oplus}_{[\mathcal{F}_1]} + \frac{1}{2} p^{\oplus T}_{[\mathcal{F}_1]} H^{\oplus}_{[\mathcal{F}_1,\mathcal{F}_1]} p^{\oplus}_{[\mathcal{F}_1]} \\
= & m^{\oplus}(x^{\oplus} + p^{\oplus}) + \frac{1}{2} p^{\oplus T}_{[\mathcal{F}_1]} [\nabla_{xx}\Psi^{\oplus} - H^{\oplus}]_{[\mathcal{F}_1,\mathcal{F}_1]} p^{\oplus}_{[\mathcal{F}_1]} \\
 & + \frac{1}{2}\int_0^1 p^{\oplus T}_{[\mathcal{F}_1]} [\nabla_{xx}\Psi(x^{\oplus}(t), \lambda^+, s^+) - \nabla_{xx}\Psi^{\oplus}]_{[\mathcal{F}_1,\mathcal{F}_1]} p^{\oplus}_{[\mathcal{F}_1]} dt,
\end{array} \tag{4.78}$$

where $x^{\oplus}(t) = x^{\oplus} + tp^{\oplus}$ and we have abbreviated $\nabla_{xx}\Psi(x^{\oplus}, \lambda^+, s^+)$ as $\nabla_{xx}\Psi^{\oplus}$.

Considering the last two terms in (4.78) in turn, we have the bounds

$$\begin{array}{rl}
\multicolumn{2}{l}{|\frac{1}{2} p^{\oplus T}_{[\mathcal{F}_1]} [\nabla_{xx}\Psi^{\oplus} - H^{\oplus}]_{[\mathcal{F}_1,\mathcal{F}_1]} p^{\oplus}_{[\mathcal{F}_1]}|} \\
\leq & \frac{1}{2} a_0 (v \|p^{\oplus}_{[\mathcal{F}_1]}\|_g^{\varsigma} + \|[\nabla_{xx}\ell(x^{\oplus}, \bar{\lambda}^+) - \nabla_{xx}\ell(x^*, \lambda^*)]_{[\mathcal{F}_1,\mathcal{F}_1]}\|_g) \|p^{\oplus}_{[\mathcal{F}_1]}\|_g^2,
\end{array} \tag{4.79}$$

using (2.13), (3.27), the definition of the Hessian of the Lagrangian barrier function and AS8, and

$$|\tfrac{1}{2}\int_0^1 p^{\oplus T}_{[\mathcal{F}_1]} [\nabla_{xx}\Psi(x^{\oplus}(t), \lambda^+, s^+) - \nabla_{xx}\Psi^{\oplus}]_{[\mathcal{F}_1,\mathcal{F}_1]} p^{\oplus}_{[\mathcal{F}_1]} dt| \leq \tfrac{1}{4} a_0 a_8 \|p^{\oplus}_{[\mathcal{F}_1]}\|_g^3, \tag{4.80}$$

using (2.13), the convergence (and hence boundedness) of the Lagrange multiplier estimates and the Lipschitz continuity of the second derivatives of the problem functions (assumption AS3) with some composite Lipschitz constant a_8. Thus, combining (4.70), (4.76), (4.77), (4.78), (4.79) and (4.80), we obtain

$$|\rho^\oplus - 1| \le 2a_0^3(1 + \|\nabla_x \Psi^\oplus_{[\mathcal{F}_1]}\|^\xi)^2_g \cdot \frac{\frac{1}{2}a_8\|p^\oplus_{[\mathcal{F}_1]}\|_g + \upsilon\|p^\oplus_{[\mathcal{F}_1]}\|^\varsigma_g + \|[\nabla_{xx}\ell(x^\oplus, \bar{\lambda}^+) - \nabla_{xx}\ell(x^*, \lambda^*)]_{[\mathcal{F}_1,\mathcal{F}_1]}\|_g}{a_7 \pi^2_{\min}}. \tag{4.81}$$

As the right-hand-side of (4.81) converges to zero as k increases, $x^+ = x^\oplus + p^\oplus$ for all k sufficiently large.

(v) Convergence of the inner iteration at the new point. The relationship (4.75) ensures that x^+ satisfies the feasibility test (3.6). We now show that x^+ satisfies the inner-iteration convergence test (3.11).

Firstly, in the same vein as (4.34) , for $j \in \mathcal{D}_1$ we have that

$$\begin{array}{ll} |\nabla_x \Psi(x^+, \lambda^+, s^+)_j - \nabla_x \Psi^\oplus_j| & \\ \quad \le \; \|p^\oplus\|_2 \cdot \|e_j^T \int_0^1 [H_\ell(x^\oplus(t), \lambda) + A(x^\oplus(t))^T D^+(x^\oplus(t)) A(x^\oplus(t))]_j dt\|_2 & \\ \quad \le \; a_0(a_3 + a_1 a_2^2/\mu_{\min})\|p^\oplus\|_2, & \end{array} \tag{4.82}$$

where $x^\oplus(t) = x^\oplus + tp^\oplus$ and where we use the bound

$$\|D^+(x^\oplus(t))\| \le a_1/\mu_{\min} \tag{4.83}$$

for all $0 \le t \le 1$. This latter follows from the definition (3.13) and the convergence of $x^\oplus$ and, because of (4.72) and (4.71), the convergence of $x^\oplus + p^\oplus$ to x^*. Thus, as the right-hand-side of (4.82) can be made arbitrarily small, by taking k sufficiently large, (4.23) and the identity $x_j^+ = x_j^\oplus = 0$ for each $j \in \mathcal{D}_1$, imply that x_j^+ is dominated for each $j \in \mathcal{D}_1$ while (2.9) and (4.46) imply that

$$P(x^+, \nabla_x \Psi(x^+, \lambda^+, s^+))_j = x_j^+ = 0. \tag{4.84}$$

We now consider the components of $P(x^+, \nabla_x \Psi(x^+, \lambda^+, s^+))_j$ for $j \in \mathcal{F}_1$. Using the integral mean value theorem, we have

$$\begin{array}{rl} \nabla_x \Psi(x^+, \lambda^+, s^+)_{[\mathcal{F}_1]} = & \nabla_x \Psi^\oplus_{[\mathcal{F}_1]} + \int_0^1 \nabla_{xx} \Psi(x^\oplus(t), \lambda^+, s^+)_{[\mathcal{F}_1,\mathcal{F}_1]} p^\oplus_{[\mathcal{F}_1]} dt \\ = & [H^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]} p^\oplus_{[\mathcal{F}_1]} + \nabla_x \Psi^\oplus_{[\mathcal{F}_1]}] + [\nabla_{xx}\Psi^\oplus - H^\oplus]_{[\mathcal{F}_1,\mathcal{F}_1]} p^\oplus_{[\mathcal{F}_1]} \\ & + \int_0^1 [\nabla_{xx}\Psi(x^\oplus(t), \lambda^+, s^+) - \nabla_{xx}\Psi^\oplus]_{[\mathcal{F}_1,\mathcal{F}_1]} p^\oplus_{[\mathcal{F}_1]} dt \end{array} \tag{4.85}$$

where $x^\oplus(t) = x^\oplus + tp^\oplus$. We observe that each of the three terms on the right-hand-side of (4.85) reflects a different aspect of the approximations made. The first corresponds to the approximation to the Newton direction used, the second to the approximation of a nonlinear function by a quadratic and the third to the particular approximation to the second derivatives used. We now bound each of these terms in turn.

The first term satisfies the bound (4.67). Hence, combining (4.47) and (4.67), we obtain

$$\|H^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]} p^\oplus_{[\mathcal{F}_1]} + \nabla_x \Psi^\oplus_{[\mathcal{F}_1]}\|_g \le a_6^{1+\xi} (\mu_{\min})^{(\alpha_\eta - 1)(1+\xi) + k\alpha_\lambda \beta_\eta (1+\xi)}. \tag{4.86}$$

The same arguments as those used to establish (4.79) imply that the second term on the right-hand-side of (4.85) satisfies the bound

$$\begin{aligned} &\|[\nabla_{xx}\Psi^\oplus - H^\oplus]_{[\mathcal{F}_1,\mathcal{F}_1]} p^\oplus_{[\mathcal{F}_1]}\|_g \\ &\quad\leq\ (v\|p^\oplus_{[\mathcal{F}_1]}\|^\varsigma_g + \|(\nabla_{xx}\ell(x^\oplus,\bar\lambda^+) - \nabla_{xx}\ell(x^*,\lambda^*))_{[\mathcal{F}_1,\mathcal{F}_1]}\|_g)\|p^\oplus_{[\mathcal{F}_1]}\|_g \\ &\quad\leq\ (v\|p^\oplus_{[\mathcal{F}_1]}\|^\varsigma_g + a_9\|x^\oplus - x^*\|_g + a_{10}\|\bar\lambda^+ - \lambda^*\|_g)\|p^\oplus_{[\mathcal{F}_1]}\|_g, \end{aligned} \tag{4.87}$$

for some composite Lipschitz constants a_9 and a_{10}. We may then combine (2.14), (4.6), (4.17), (4.32), (4.50), (4.71) and (4.87) to obtain the bound

$$\begin{aligned} &\|[\nabla_{xx}\Psi^\oplus - H^\oplus]_{[\mathcal{F}_1,\mathcal{F}_1]} p^\oplus_{[\mathcal{F}_1]}\|_g \\ &\quad\leq\ [v[(4a_0a_6/\pi_{\min})(\mu_{\min})^{\alpha_\eta-1+k\alpha_\lambda\beta_\eta}]^\varsigma + a_9[a_x(\mu_{\min})^{\alpha_\eta+k\alpha_\lambda\beta_\eta} + a_0\omega_0(\mu_{\min})^{\alpha_\omega+k\beta_\omega}] \\ &\qquad + a_{10}a_{\lambda^\oplus}(\mu_{\min})^{\alpha_\eta-1+k\alpha_\lambda\beta_\eta}](4a_0a_6/\pi_{\min})(\mu_{\min})^{\alpha_\eta-1+k\alpha_\lambda\beta_\eta} \end{aligned} \tag{4.88}$$

for all sufficiently large k. Lastly, the third term on the right-hand-side of (4.85) satisfies the bound

$$\|\int_0^1 [\nabla_{xx}\Psi(x^\oplus(t),\lambda^+,s^+) - \nabla_{xx}\Psi^\oplus_{[\mathcal{F}_1,\mathcal{F}_1]} p^\oplus_{[\mathcal{F}_1]}]dt\|_g \leq \tfrac{1}{2}a_0a_8\|p^\oplus_{[\mathcal{F}_1]}\|^2_g. \tag{4.89}$$

by the same arguments we used to establish inequality (4.80). We may then combine (4.71) and (4.89) so that

$$\|\int_0^1 [\nabla_{xx}\Psi(x^\oplus(t),\lambda^+,s^+) - \nabla_{xx}\Psi^\oplus]_{[\mathcal{F}_1,\mathcal{F}_1]} p^\oplus_{[\mathcal{F}_1]}dt\| \leq 8a_0^3a_6^2a_8(\mu_{\min})^{2\alpha_\eta-2+k2\alpha_\lambda\beta_\eta}/\pi^2_{\min} \tag{4.90}$$

for all k sufficiently large.

We now combine equation (4.85) with the inequalities (4.86), (4.90) and (4.88), the condition $\xi < 1$ and the definitions of $\alpha_\eta(< 1)$ and $\beta_\eta(> 0)$ to obtain the bound

$$\|\nabla_x\Psi(x^+,\lambda^+,s^+)_{[\mathcal{F}_1]}\| \leq a_{11}(\mu_{\min})^{\bar\alpha+k\bar\beta}, \tag{4.91}$$

where

$$\bar\alpha = (\alpha_\eta - 1)(1+\max(1,\varsigma)), \quad \bar\beta = \alpha_\lambda\beta_\eta(1+\min(\xi,\varsigma)) \tag{4.92}$$

and

$$\begin{aligned} a_{11} =\ & a_6^{1+\xi} + 8a_0^3a_6^2a_8/\pi^2_{\min} + (4a_0a_6/\pi_{\min})(v((4a_0a_6/\pi_{\min})^\varsigma) \\ & + a_9(a_x + a_0\omega_0) + a_{10}a_\lambda). \end{aligned} \tag{4.93}$$

Firstly, observe that the right-hand-side of (4.91) may be made arbitrarily small. Therefore, (2.10), (4.84) and (4.91) imply that

$$\|P(x^+,\nabla_x\Psi(x^+,\lambda^+,s^+))\|_g = \|\nabla_x\Psi(x^+,\lambda^+,s^+)_{[\mathcal{F}_1]}\|_g \leq a_{11}(\mu_{\min})^{\bar\alpha+k\bar\beta}. \tag{4.94}$$

Secondly, define $\delta = \log_{\mu_{\min}}(a_{11}/\omega_0)$. Now let k_1 be any integer for which

$$k_1 \geq \frac{\alpha_\omega + \beta_\omega - \bar\alpha - \delta}{\bar\beta - \beta_\omega}. \tag{4.95}$$

Then (4.13), (4.94) and (4.95) imply that

$$\|P(x^+,\nabla_x\Psi(x^+,\lambda^+,s^+))\|_g \leq a_{11}(\mu_{\min})^{\bar\alpha+k\bar\beta} \leq \omega_0(\mu_{\min})^{\alpha_\omega+(k+1)\beta_\omega} = \omega^+ \tag{4.96}$$

for all sufficiently large $k \geq k_1$. Thus, the iterate x^+ satisfies the inner iteration convergence test (3.5) for all k sufficiently large and we have $x^{(k+1)} = x^{(k+1,1)} \equiv x^+$.

(vi) Redundancy of the shifted starting point. Finally, we observe that all the variables $x_j^{(k)}$, $j \in \mathcal{D}$, lie on their bounds for sufficiently large k. Therefore, $x^{(k+1,0)} = x^{(k)}$ and the perturbed starting point is redundant. ■

5 The general case

We now turn briefly to the more general problem (1.1)—(1.3). The presence of the more general bounds (1.3) does not significantly alter the conclusions that we are able to draw. The algorithms of section 3 are basically unchanged. We now use the region $\mathcal{B} = \{x \in \Re^n \mid l \leq x \leq u\}$ — and hence $\mathcal{N}_b = \mathcal{N}$ — and replace $P(x,v)$ by $P(x,v,l,u)$ where appropriate. The concept of floating and dominated variables stays essentially the same. For each iterate in $\mathcal{B}$ we have three mutually exclusive possibilities, namely, (i) $0 \leq x_j^{(k)} - l_j \leq (\nabla_x \Psi^{(k)})_i$, (ii) $(\nabla_x \Psi^{(k)})_i \leq x_j^{(k)} - u_j \leq 0$ or (iii) $x_j^{(k)} - u_j < (\nabla_x \Psi^{(k)})_i < x_j^{(k)} - l_j$, for each component $x_j^{(k)}$. In case (i) we then have that $P(x^{(k)}, \nabla_x \Psi^{(k)}, l, u)_i = x_j^{(k)} - l_j$ while in case (ii) $P(x^{(k)}, \nabla_x \Psi^{(k)}, l, u)_i = x_j^{(k)} - u_j$ and in case (iii) $P(x^{(k)}, \nabla_x \Psi^{(k)}, l, u)_i = (\nabla_x \Psi^{(k)})_i$. The variables that satisfy (i) and (ii) are said to be the dominated variables, the ones satisfying (i) are *dominated above* while those satisfying (ii) are *dominated below*. Consequently, the sets corresponding to (2.11) are straightforward to define. $\mathcal{D}_1$ is now made up as the union of two sets $\mathcal{D}_{1l}$, whose variables are dominated above for all k sufficiently large, and $\mathcal{D}_{1u}$, whose variables are dominated below for all k sufficiently large. $\mathcal{F}_1$ contains variables which float for all k sufficiently large and which converge to values interior to $\mathcal{B}$. Similarly $\mathcal{F}_2$ is the union of two sets, $\mathcal{F}_{2l}$ and $\mathcal{F}_{2u}$, whose variables are floating for all k sufficiently large but which converge to their lower and upper bounds respectively. We also replace (3.17) by

$$\hat{x}_j^{(k-1)} = \begin{cases} l_j & \text{if } 0 \leq x_i^{(k-1)} - l_j \leq \theta(\nabla_x \Psi^{(k-1)})_i \\ u_j & \text{if } \theta(\nabla_x \Psi^{(k-1)})_i \leq x_i^{(k-1)} - u_j \leq 0 \\ x_i^{(k-1)} & \text{otherwise.} \end{cases} \tag{5.1}$$

With such definitions, we may reprove the results of section 4, extending AS4, AS7—AS9 in the obvious way. The only important new ingredient is that Conn *et al.* (1992b) indicate that the non-degeneracy assumption AS7 ensures that the iterates are asymptotically isolated in the three sets $\mathcal{F}_1$, $\mathcal{D}_{1l}$ and $\mathcal{D}_{1u}$.

6 Conclusions

We have shown that, under suitable assumptions, a single inner iteration is needed for each outer iteration of the Lagrangian barrier algorithm. We anticipate that such an algorithm may prove to be an important ingredient of release B of the LANCELOT package.

7 Acknowledgement

The work reported here has been partly supported by the NATO travel grant CRG 890867.

References

[Bertsekas, 1982] D. P. Bertsekas. *Constrained Optimization and Lagrange Multipliers Methods*. Academic Press, London, 1982.

[Burke and Moré, 1988] J. V. Burke and J. J. Moré. On the identification of active constraints. *SIAM Journal on Numerical Analysis*, 25(5):1197–1211, 1988.

[Burke *et al.*, 1990] J. V. Burke, J. J. Moré, and G. Toraldo. Convergence properties of trust region methods for linear and convex constraints. *Mathematical Programming, Series A*, 47(3):305–336, 1990.

[Calamai and Moré, 1987] P. H. Calamai and J. J. Moré. Projected gradient methods for linearly constrained problems. *Mathematical Programming*, 39:93–116, 1987.

[Conn *et al.*, 1988a] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM Journal on Numerical Analysis*, 25:433–460, 1988. See also same journal 26:764–767, 1989.

[Conn *et al.*, 1988b] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. Testing a class of methods for solving minimization problems with simple bounds on the variables. *Mathematics of Computation*, 50:399–430, 1988.

[Conn *et al.*, 1992a] 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 equality constraints and simple bounds. In D.F Griffiths and G.A. Watson, editors, *Proceedings of the 14th Biennal Numerical Analysis Conference Dundee 1991*. Longmans, 1992.

[Conn *et al.*, 1992b] A. R. Conn, Nick Gould, and Ph. L. Toint. Convergence properties of minimization algorithms for convex constraints using a structured trust region. Technical Report 92/11, Department of Mathematics, FUNDP, Namur, Belgium, 1992.

[Conn *et al.*, to appear 1992] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. LANCELOT*: a Fortran package for large-scale nonlinear optimization (Release A)*. Springer Verlag, Heidelberg, Berlin, New York, (to appear), 1992.

[Dennis and Moré, 1974] J. E. Dennis and J. J. Moré. A characterization of superlinear convergence and its application to quasi-Newton methods. *Mathematics of Computation*, 28(126):549–560, 1974.

[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.

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

[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.

[Moré, 1988] J. J. Moré. Trust regions and projected gradients. In M. Iri and K. Yajima, editors, *System Modelling and Optimization*, volume 113, pages 1–13, Berlin, 1988. Springer Verlag. Lecture Notes in Control and Information Sciences.

[Toint, 1988] Ph. L. Toint. Global convergence of a class of trust region methods for nonconvex minimization in Hilbert space. *IMA Journal of Numerical Analysis*, 8:231–252, 1988.

[Wright, 1992] M. H. Wright. Interior methods for constrained optimization. volume 1 of *Acta Numerica*, pages 241–407. Cambridge University Press, new York, 1992.