On the Number of Inner Iterations Per Outer Iteration of a Globally Convergent Algorithm for Optimization with General Nonlinear Inequality Constraints and Simple Bounds

Conn, Andy; Gould, Nick; Toint, Philippe

Published in:
Computational Optimization and Applications

Publication date:
1997

Document Version
Early version, also known as pre-print

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

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

Report 91/9 1991

Abstract. This paper considers the number of inner iterations required per outer iteration for the algorithm proposed by Conn *et al.*1991a (). We show that asymptotically, under suitable reasonable assumptions, a single inner iteration suffices.

[1] **IBM T.J. Watson Research Centre**
Yorktown Heights, USA

[2] **Rutherford Appleton Laboratory Oxfordshire, England**

[3] **Department of Mathematics,**
Facultés Universitaires ND de la Paix, Namur, Belgium

Invited paper at the Dundee Conference on Numerical Analysis, 1991.

1 Introduction

In this paper, we consider the nonlinear programming problem

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

subject to the general constraints

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

and the simple bounds

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

We assume that the region $\mathcal{B} = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$ is non-empty and may be infinite. Furthermore, we assume that

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

The exposition is conveniently simplified by taking the lower bounds as identically equal to zero and the upper bound as infinity. Thus, in most of what follows, $\mathcal{B} = \{x \in \mathbb{R}^n \mid x \geq 0\}$. 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 () and is based upon incorporating the equality constraints via an augmented Lagrangian function whilst handling upper and lower bounds directly. The sequential, approximate minimization of the augmented Lagrangian 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 augmented Lagrangian algorithm. This is contrary to most analyses of sequential penalty function and augmented Lagrangian 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 book by Bertsekas1982, . An exception to this is the sequential penalty function method analyzed by Gould1989,).

This work was primarily motivated by observations that the authors made when testing their large-scale nonlinear programming package LANCELOT (see, Conn *et al.*1991b,), which is 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 the sequential augmented Lagrangian program AUGLG. 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 augmented Lagrangian functions as

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

and

$$\Phi(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i c_i(x) + \frac{1}{2\mu} \sum_{i=1}^{m} c_i(x)^2 \tag{2.4}$$

respectively. We note that $\ell(x, \lambda)$ is the Lagrangian with respect to the c_i 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 denote the non-negativity restrictions by

$$x \in \mathcal{B} = \{x \in \mathbb{R}^n \mid x \geq 0\}. \tag{2.5}$$

We will make much use of the projection operator defined componentwise by,

$$(P[x, l, u])_i = \begin{cases} l_i & \text{if } x_i \leq l_i \\ u_i & \text{if } x_i \geq u_i \\ x_i & \text{otherwise.} \end{cases} \tag{2.6}$$

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

$$P(x, v, l, u) = x - P[x - v, l, u]. \tag{2.7}$$

Furthermore, define $P[x] = P[x, 0, \infty]$ and $P(x, v) = P(x, v, 0, \infty)$.

Let $\in \mathcal{B}$ and 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 $h(, , \ldots)$.

Now, let $\mathcal{N} = \{1, \ldots, n\}$. For any $x^{(k)}$ we have two possibilities for each component $x_i^{(k)}, i = 1, \ldots, n$, namely (i) $0 \leq x_i^{(k)} \leq (\nabla_x \Phi^{(k)})_i$ or (ii) $(\nabla_x \Phi^{(k)})_i < x_i^{(k)}$. We shall call all $x_i^{(k)}$ that satisfy (i) *dominated* variables while the remaining $x_i^{(k)}$ are *floating* variables. It is important to notice that, as $\in \mathcal{B}$

$$(P(, \nabla_x))_i =_i \quad \text{whenever } _i \text{ is dominated} \tag{2.8}$$

while

$$(P(, \nabla_x))_i = (\nabla_x)_i \quad \text{otherwise.} \tag{2.9}$$

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

$$\begin{array}{lcl} \mathcal{D}_1 & \stackrel{\text{def}}{=} & \{i \mid x_i^{(k)} \text{ are dominated for all } k \in K \text{ sufficiently large }\}, \\ \mathcal{F}_1 & \stackrel{\text{def}}{=} & \{i \mid x_i^{(k)} \text{ are floating for all } k \in K \text{ sufficiently large and } x_i^* > 0\}, \\ \mathcal{F}_2 & \stackrel{\text{def}}{=} & \{i \mid x_i^{(k)} \text{ are floating for all } k \in K \text{ sufficiently large but } x_i^* = 0\} \text{ and} \\ \mathcal{F}_3 & \stackrel{\text{def}}{=} & \mathcal{N} \setminus (\mathcal{D}_1 \cup \mathcal{F}_1 \cup \mathcal{F}_2)\,. \end{array} \tag{2.10}$$

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 i-th column is e_i. 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

$$\|v\|_z \leq a_0^{\frac{1}{2}} \|v\|_2 \quad \text{and} \quad \|v\|_2 \leq a_0^{\frac{1}{2}} \|v\|_z \tag{2.11}$$

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 $\|\cdot\|_z$,

$$\|v\|_z \leq a_0 \|v\|_y \quad \text{and} \quad \|v\|_y \leq a_0 \|v\|_z. \tag{2.12}$$

Following (), we now describe an algorithm for solving (1.1), (1.2) and (2.5).

3 Statement of the algorithm

In order to solve the problem (1.1), (1.2) and (2.5), we consider the following algorithmic model.

Algorithm 3.1 [**Outer Iteration Algorithm**]

step 0 : [Initialization] *The positive constants* η_0, ω_0, μ_0, $\tau < 1$, $\gamma_0 < 1$, α_ω, β_ω, α_η, β_η, *small positive convergence tolerances* ω_* *and* η_*, *and the vector* $\lambda^{(0)} \in I\!R^m$ *are chosen. The two-norm-consistent norms* $\|\cdot\|_p$ *and* $\|\cdot\|_c$ *are specified. We require that*

$$\alpha_\eta < \min(1, \alpha_\omega) \quad \textit{and} \quad \beta_\eta < \min(1, \beta_\omega). \tag{3.1}$$

Set $\mu^{(0)} = \mu_0$, $\alpha^{(0)} = \min(\mu^{(0)}, \gamma_0)$, $\omega^{(0)} = \omega_0(\alpha^{(0)})^{\alpha_\omega}$, $\eta^{(0)} = \eta_0(\alpha^{(0)})^{\alpha_\eta}$ *and* $k = 0$.

step 1 : [Inner Iteration] *Find* $x^{(k)} \in \mathcal{B}$ *such that*

$$\|P(x^{(k)}, \nabla_x \Phi^{(k)})\|_p \leq \omega^{(k)} \tag{3.2}$$

If

$$\|c(x^{(k)})\|_c \leq \eta^{(k)} \tag{3.3}$$

execute step 2. Otherwise, execute step 3.

step 2 : [Test for convergence and update Lagrange multipliers] *If*

$$\|P(x^{(k)}, \nabla_x \Phi^{(k)})\|_p \leq \omega_* \quad \textit{and} \quad \|c(x^{(k)})\|_c \leq \eta_*, \tag{3.4}$$

stop. Otherwise, set

$$\begin{array}{rcl} \alpha^{(k+1)} & = & \min(\mu^{(k+1)}, \gamma_0), \\ \eta^{(k+1)} & = & \eta^{(k)}(\alpha^{(k+1)})^{\beta_\eta}, \\ \omega^{(k+1)} & = & \omega^{(k)}(\alpha^{(k+1)})^{\beta_\omega}, \\ \lambda^{(k+1)} & = & \lambda^{(k)} + c(x^{(k)})/\mu^{(k)}, \\ \mu^{(k+1)} & = & \mu^{(k)}, \end{array} \tag{3.5}$$

increment k *by one and go to step 1.*

step 3 : [Decrease penalty parameter if constraints too large]

$$\begin{array}{rcl} \mu^{(k+1)} & = & \tau\mu^{(k)}, \\ \alpha^{(k+1)} & = & \min(\mu^{(k+1)}, \gamma_0), \\ \eta^{(k+1)} & = & \eta_0(\alpha^{(k+1)})^{\alpha_\eta}, \\ \omega^{(k+1)} & = & \omega_0(\alpha^{(k+1)})^{\alpha_\omega}, \end{array} \tag{3.6}$$

increment k *by one and go to step 1.*

end of Algorithm 3.1

We shall call the vector $P(x^{(k)}, \nabla_x \Phi^{(k)})$ the *projected gradient of the augmented Lagrangian* or the *projected gradient* for short. The norms $\|\cdot\|_p$ 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 3 is executed an infinite number of times, will suffice. It is also irrelevant, in theory, as to how we find a suitable point satisfying (3.2). However, from a practical perspective, a suitable point is found by an iterative procedure. In our algorithm, it is normal 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.

The main purpose of this article is to show that asymptotically we take one inner iteration per outer iteration. More specifically, under certain assumptions, we first show that (3.3) is eventually satisfied at each outer iteration. We then show that, under additional assumptions, it is possible to satisfy the convergence test (3.2) after a single iteration of the algorithm given in ().

The specific inner iteration algorithm we shall consider is as follows:

Algorithm 3.2 [**Inner Iteration Algorithm**]

step 0 : [Initialization] *The positive constants $\mu < \eta < 1$ and $\gamma_0 \leq \gamma_1 < 1 \leq \gamma_2$ 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 and a two-norm-consistent norm $\|\cdot\|_p$ are specified. Compute $\Phi(x^{(k,0)},,)$ and its gradient. Set the inner iteration counter $j = 0$.*

step 1 : [Test for convergence] *If*

$$\|P(x^{(k,j)}, \nabla_x \Phi^{(k,j)})\|_p \leq \omega^{(k)} \tag{3.7}$$

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

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

$$\begin{aligned} m^{(k,j)}(x+s) \stackrel{\text{def}}{=} \quad & \Phi(x,,) + s^T \nabla_x \Phi(x,,) \\ & + \tfrac{1}{2} s^T (B^{(k,j)} + \tfrac{1}{\mu} A(x)^T A(x)) s, \end{aligned} \tag{3.8}$$

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

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

$$\rho^{(k,j)} = \frac{\Phi(x^{(k,j)},,) - \Phi(x^{(k,j)} + s^{(k,j)},,)}{m^{(k,j)}(x^{(k,j)}) - m^{(k,j)}(x^{(k,j)} + s^{(k,j)})}. \tag{3.9}$$

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

$$x^{(k,j+1)} = \begin{cases} x^{(k,j)} + s^{(k,j)} & \text{if } \rho^{(k)} > \mu \\ x^{(k,j)} & \text{otherwise,} \end{cases} \tag{3.10}$$

and

$$\Delta^{(k,j+1)} = \begin{cases} \gamma_0^{(k,j)} \Delta^{(k,j)} & \text{if } \rho^{(k,j)} \leq \mu \\ & \text{if } \mu < \rho^{(k,j)} < \eta \\ \gamma_2^{(k,j)} \Delta^{(k,j)} & \text{otherwise,} \end{cases} \tag{3.11}$$

where $\gamma_0^{(k,j)} \in [\gamma_0, 1)$ and $\gamma_2^{(k,j)} \in [1, \gamma_2]$.

step 5 : [Updating] *If necessary, compute the gradient of $\Phi(x^{(k,j+1)},,)$ 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

There are a number of possible ways of choosing $\gamma_0^{(k,j)}$ and $\gamma_2^{(k,j)}$ in step 4. The simplest is merely to pick $\gamma_0^{(k,j)} = \gamma_0$ and $\gamma_2^{(k,j)} = \gamma_2$; other alternatives are discussed by ().

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 choose

$$x_i^{(k,0)} = \begin{cases} 0 & \text{if } 0 \leq x_i^{(k-1)} \leq \theta(\nabla_x \Phi^{(k-1)})_i \\ x_i^{(k-1)} & \text{otherwise.} \end{cases} \tag{3.12}$$

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, , , Theorem 5.4). 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

$$\Delta^{(k,0)} \geq \kappa \| P(x^{(k,0)}, \nabla_x \Phi^{(k,0)}) \|_p^\zeta \tag{3.13}$$

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

$$s = s^{(k,j)}(t) \stackrel{\text{def}}{=} P[x^{(k,j)} - t\nabla_x \Phi(x^{(k,j+1)}, ,), l, u] - x^{(k,j)}, \tag{3.14}$$

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

$$\|s\|_t \leq \Delta^{(k,j)}, \tag{3.15}$$

   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 (2.5) (or, in general, (1.3)) and which stops on the boundary of the trust region (3.15). The two or infinity norm is

normally picked, the latter having some advantages as the trust region is then aligned with the feasible region (2.5).

The method proposed by () 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 Moré1988 () finds an approximation $s^{C(k,j)} = s^{(k,j)}(t^{C(k,j)})$ which is required to lie within the trust-region and to satisfy the Goldstein-Armijo type conditions

$$m^{(k,j)}(x^{(k,j)}+s^{(k,j)}(t^{C(k,j)})) \leq m^{(k,j)}(x^{(k,j)})+\mu_1 s^{(k,j)}(t^{C(k,j)})^T\nabla_x\Phi(x^{(k,j)},\lambda^{(k)},\mu^{(k)}) \tag{3.16}$$

and

$$t^{C(k,j)} \geq \nu_1 \quad \text{or} \quad t^{C(k,j)} \geq \nu_2 t^{L(k,j)}, \tag{3.17}$$

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

$$m^{(k,j)}(x^{(k,j)}+s^{(k,j)}(t^{L(k,j)})) \geq m^{(k,j)}(x^{(k,j)})+\mu_2 s^{(k,j)}(t^{L(k,j)})^T\nabla_x\Phi(x^{(k,j)},\lambda^{(k)},\mu^{(k)}) \tag{3.18}$$

or

$$\|s^{(k,j)}(t^{L(k,j)})\| \geq \nu_3\Delta^{(k,j)}, \tag{3.19}$$

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 (3.16) ensures that a sufficient reduction in the model takes place at each iteration while condition (3.17) 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 Moré1987 (), Burke and Moré1988 (), Toint1988 () and Burke *et al.*1990 ().

2. Secondly, we pick $s^{(k,j)}$ so that $x^{(k,j)} + s^{(k,j)}$ lies within (2.5), $\|s^{(k,j)}\|_t \leq \beta_1\Delta^{(k,j)}$ and

$$\begin{aligned} &m^{(k,j)}(x^{(k,j)}) - m^{(k,j)}(x^{(k,j)} + s^{(k,j)}) \\ &\quad \geq \beta_2(m^{(k,j)}(x^{(k,j)}) - m^{(k,j)}(x^{(k,j)} + s^{C(k,j)})) \geq 0 \end{aligned} \tag{3.20}$$

for some positive $\beta_1 \geq 1$ and $\beta_2 \leq 1$. In fact, we typically choose $\beta_1 = \beta_2 = 1$, 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 () 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

$$H^{(k,j)} \stackrel{\text{def}}{=} B^{(k,j)} + \frac{1}{\mu^{(k)}} A(x^{(k,j)})^T A(x^{(k,j)}) \tag{3.21}$$

denote the composite approximation to the Hessian of the augmented Lagrangian. The specific algorithm we shall consider may be summarized as follows:

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

step 0 : [Initialization] *Select positive constants $\nu < 1$, $\xi < 1$, $\beta_1 \geq 1$ and $\beta_2 \leq 1$.*

step 1 : [Calculate the generalized Cauchy point] *Calculate an approximation to the the generalized Cauchy point $x^{C(k,j)} = x^{(k,j)} + s^{C(k,j)}$ 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 $x = x^{C(k,j)}$, $s = s^{C(k,j)}$ and $\mathcal{F} = \mathcal{F}^{C(k,j)}$.*

step 2 : [Further improve the model] *Let $\mathcal{C}(\beta_1) = \mathcal{S} \bigcap \mathcal{T}(\beta_1)$, where*

$$\mathcal{S} = \{s_{[\mathcal{F}]} \mid x^{(k,j)} + s \in \mathcal{B} \quad \text{and} \quad s_{[\mathcal{D}]} = s_{[\mathcal{D}]}^{C(k,j)}\} \tag{3.22}$$

and

$$\mathcal{T}(\beta_1) = \{s_{[\mathcal{F}]} \mid \|s\|_t \leq \beta_1 \Delta^{(k,j)} \quad \text{and} \quad s_{[\mathcal{D}]} = s_{[\mathcal{D}]}^{C(k,j)}\}. \tag{3.23}$$

If $s_{[\mathcal{F}]}$ lies on the boundary of $\mathcal{T}(\beta_1)$, set $s^{(k,j)} = s$ and stop[1]. Otherwise, recompute $s_{[\mathcal{F}]}$ so that (3.20) is satisfied and either $s_{[\mathcal{F}]}$ lies strictly interior to $\mathcal{C}(\beta_1)$ with

$$\begin{aligned} &\|H_{[\mathcal{F},\mathcal{F}]}^{(k,j)} s_{[\mathcal{F}]} + (\nabla_x \Phi_{[\mathcal{F}]}^{(k,j)} + H_{[\mathcal{F},\mathcal{D}]}^{(k,j)} s_{[\mathcal{D}]})\|_p \\ &\quad \leq \min(\nu, \|P(x^{(k,j)}, \nabla_x \Phi^{(k,j)})\|_p^{\xi}) \cdot \|P(x^{(k,j)}, \nabla_x \Phi^{(k,j)})\|_p \end{aligned} \tag{3.24}$$

or $s_{[\mathcal{F}]}$ lies on the boundary of $\mathcal{C}(\beta_1)$. Reset $x_{[\mathcal{F}]}$ to $x_{[\mathcal{F}]} + s_{[\mathcal{F}]}$.

step 3 : [Test for convergence] *If $s_{[\mathcal{F}]}$ lies strictly interior to $\mathcal{C}(\beta_1)$ and (3.24) is satisfied or if it is decided that sufficient passes have been made, set $s^{(k,j)} = s$ and stop. Otherwise remove all of the indices in $\mathcal{F}$ for which $|s_{[\mathcal{F}]i}|$ lies on the boundary of $\mathcal{S}$ and perform another pass by returning to step 2.*

end of Algorithm 3.3

[1] 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.

In step 2 of this method, the value of $s_{[\mathcal{F}]}$ 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 (3.24) 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 iterates generated by Algorithm 3.1 all lie within a closed bounded domain Ω.

AS3. The matrix $A()_{[\mathcal{F}_1]}$ is of full rank at any limit point of the sequence generated by Algorithm 3.1.

Under these assumptions we have the following result.

Theorem 4.1 [, , Theorem 4.4] *Assume that AS1–AS3 hold, that is a limit point of the sequence {} generated by Algorithm 3.1 and that*

$$\stackrel{\text{def}}{=} +c()/. \tag{4.1}$$

Then is a Kuhn-Tucker (first order stationary) point for (1.1), (1.2) and (2.5) and the corresponding subsequences of {} and $\{\nabla_x\}$ converge to a set of Lagrange multipliers, , and the gradient of the Lagrangian, $g_\ell(,)$, for the problem, respectively.

Now consider the following further assumptions.

AS4. The second derivatives of the functions $f(x)$ and $c_i(x)$ are Lipschitz continuous at all points within Ω.

AS5. Suppose that $(,)$ is a Kuhn-Tucker point for the problem (1.1), (1.2) and (2.5), and

$$\begin{array}{rcl} \mathcal{J}_1 & = & \{i \mid (g_\ell(,))_i = 0 \text{ and } \ _i > 0\}, \\ \mathcal{J}_2 & = & \{i \mid (g_\ell(,))_i = 0 \text{ and } \ _i = 0\}. \end{array} \tag{4.2}$$

Then we assume that the Kuhn-Tucker matrix

$$\left[\begin{array}{cc} H_\ell(,)_{[\mathcal{J},\mathcal{J}]} & A()^T_{[\mathcal{J}]} \\ A()_{[\mathcal{J}]} & 0 \end{array} \right] \tag{4.3}$$

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

AS6. Algorithm 3.1 has a single limit point, .

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

Theorem 4.2 [, , Theorems 5.3 and 5.5] *Assume that AS1, AS3—AS6 hold. Then there is a constant* > 0 *such that the penalty parameter generated by Algorithm 3.1 satisfies* $=$ *for all* k *sufficiently large. Furthermore, and satisfy the bounds*

$$\| - \|_p \leq a_x()^{+k} \quad \textit{and} \quad \| - \|_p \leq a_\lambda()^{+k}, \tag{4.4}$$

where

$$\stackrel{\text{def}}{=} \min(, \gamma_0) \leq, \tag{4.5}$$

for the two-norm-consistent norm $||.||_p$ *and some positive constants* a_x *and* a_λ.

We shall now investigate the behaviour of Algorithm 3.1 once the penalty parameter has converged to its asymptotic value, . 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 $\mu^{(0)} =$. By construction, (3.3) is satisfied for all k and the updates (3.5) are always performed. Moreover,

$$= ()^{+k} \quad \text{and} \quad = ()^{+k}. \tag{4.6}$$

We require the following extra assumptions.

AS7. The set

$$\mathcal{J}_2 = \{i \mid g_\ell(,)_i = 0 \quad \text{and} \quad _i = 0\} = \emptyset. \tag{4.7}$$

AS8. If $\mathcal{J}_1$ is defined by (4.2), the approximations $B^{(k,0)}$ satisfy

$$\|(B^{(k,0)} - \nabla_{xx}\ell(,))_{[\mathcal{J}_1,\mathcal{J}_1]} s^{(k,0)}_{[\mathcal{J}_1]}\|_p \leq \upsilon \|s^{(k,0)}_{[\mathcal{J}_1]}\|_p^{1+\varsigma}, \tag{4.8}$$

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

AS9. Suppose that $(,)$ is a Kuhn-Tucker point for the problem (1.1), (1.2) and (2.5), and that $\mathcal{J}_1$ is defined by (4.2). Then we assume that the second derivative approximations $B^{(k,0)}$ have a single limit, and that the perturbed Kuhn-Tucker matrix

$$\begin{bmatrix} _{[\mathcal{J}_1,\mathcal{J}_1]} & A()^T_{[\mathcal{J}_1]} \\ A()_{[\mathcal{J}_1]} & -I \end{bmatrix} \tag{4.9}$$

is non-singular and has precisely m negative eigenvalues.

Assumption AS7 is often known as the strict complementary slackness condition. We observe that AS8 is closely related to the necessary and sufficient conditions for superlinear convergence of the inner iterates given by Dennis and Moré1974 (). We also observe that AS9 is entirely equivalent to requiring that the matrix

$$_{[\mathcal{J}_1,\mathcal{J}_1]} + \frac{1}{-} A()^T_{[\mathcal{J}_1]} A()_{[\mathcal{J}_1]} \tag{4.10}$$

is positive definite (see, for instance, Gould1986,). The uniqueness of the limit point in AS9 can also be relaxed by requiring that (4.10) has its smallest eigenvalue uniformly bounded from below by some positive quantity for all limit points of the sequence $B^{(k,0)}$. Moreover, (, Proposition 2.4) has shown that AS5 and AS7 guarantee AS9 provided that is sufficiently small. 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 3 of Algorithm (3.1) is executed rather than step 2 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 (3.12) 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 (3.24) also satisfies $s_{[\mathcal{F}_1]}$ lies strictly interior to $\mathcal{C}(\beta_1)$. Thus a single pass of step 2 of Algorithm 3.3 is required;

(iv) the step $s^{(k,0)}$ is accepted in step 4 of Algorithm 3.1;

(v) the new point $x^{(k,1)}$ satisfies the convergence test (3.7); and

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

We have the following theorem.

Theorem 4.3 *Assume that assumptions AS1,AS3–AS9 hold and that the convergence tolerances and satisfy the extra condition*

$$< (1 + \min(\xi, \varsigma)). \tag{4.11}$$

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

$$\|P(,\nabla_x\Phi(,,))\|_p \leq ()^{+k} \tag{4.12}$$

and

$$\|c()\|_c \leq ()^{+k} \tag{4.13}$$

for all $k \geq 0$.

We shall follow the outline given above.

(i) Status of the starting point. The non-degeneracy 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 , , Theorem 5.4); moreover,

$$g_\ell(,)_i = 0 \quad \text{and} \quad {}_i > 0 \quad \text{for all} \quad i \in \mathcal{F}_1 \tag{4.14}$$

and

$${}_i = 0 \quad \text{and} \quad g_\ell(,)_i > 0 \quad \text{for all} \quad i \in \mathcal{D}_1. \tag{4.15}$$

As one of ${}_i$ and ∇_{xi} converges to zero while its partner converges to a strictly positive limit for each i (assumption AS7), we may define nontrivial regions which separate the two sequences for all k sufficiently large. Let

$$\epsilon = \frac{\theta}{1+\theta}\min_{j\in\mathcal{N}}\max(_j, g_\ell(,)_j) > 0, \tag{4.16}$$

where θ is as in (3.12). Then there is an iteration k_0 such that for variables in $\mathcal{F}_1$,

$$|_i -_i| \leq \epsilon \quad \text{and} \quad |\nabla_{xi}| \leq \epsilon, \tag{4.17}$$

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

$$|_i| \leq \epsilon \quad \text{and} \quad |\nabla_{xi} - g_\ell(,)_i|, \leq \epsilon \tag{4.18}$$

for all $k \geq k_0$. Hence, for those variables in $\mathcal{D}_1$, (4.16) and (4.18) give that

$$\begin{aligned} {}_i &\leq \epsilon \leq \theta[\min_{j\in\mathcal{N}}\max[_j, g_\ell(,)_j] - \epsilon] \\ &\leq \theta[g_\ell(,)_i - \epsilon] \leq \theta(\nabla_x)_i. \end{aligned} \tag{4.19}$$

Thus, by definition (3.12), ${}_i = 0$ for each $i \in \mathcal{D}_1$ when $k \geq k_0$. Similarly, when $i \in \mathcal{F}_1$ and $k \geq k_0$, ${}_i > \theta(\nabla_x)_i$ and hence, using (3.12), ${}_i =_i$.

We now consider the starting point for the next inner iteration in detail. Firstly, combining (2.8), (2.11) and (3.12), we have that

$$\| - \|_z \leq a_0\|P(,\nabla_x\Phi(,,))\|_p \tag{4.20}$$

for any two-norm-consistent norm $\|.\|_z$. We may bound $\|c()\|_p$ using the integral mean value theorem (see, eg, Dennis and Schnabel1983, , page 74), the boundedness of $A(x)$ (assumptions AS1 and AS2) and inequalities (2.12), (3.1), (4.12), (4.13) and (4.20) to obtain

$$\begin{aligned} \|c()\|_p \leq & \quad a_0\|c()\|_c + \|\int_0^1 A(x(s))ds\|_p\| - \|_p \\ \leq & \quad a_0()^{+k} + a_1a_0()^{+k} \\ \leq & \quad a_0(1+a_1/)()^{+k} \end{aligned} \tag{4.21}$$

where $x(s) \stackrel{\text{def}}{=} +s(-)$ and a_1 is an upper bound on $\|A(x)\|$ within Ω.

Now consider the variables whose indices i lie in $\mathcal{F}_1$ for $k \geq k_0$. Firstly, (3.12), (4.16) and (4.17) show that

$$_i =_i \geq \frac{i}{1+\theta} > 0. \tag{4.22}$$

We next bound $|\nabla_x \Phi(,,)_i|$. Again using the integral mean value theorem, the convergence of $\equiv \lambda^{(k+1)}$ to (Theorem 4.1), the boundedness of the Hessian of the Lagrangian (with bounded multiplier estimates) and the constraint Jacobian within Ω (assumptions AS1 and AS2) and the inequalities (2.12), (4.12) and (4.20), we obtain

$$\begin{aligned} |\nabla_x \Phi(,,)_i| \leq & \; |\nabla_x \Phi(,,)_i| + \| - \|_2 \cdot \\ & |e_i^T \int_0^1 (H_\ell(x(s),) + \tfrac{1}{}A(x(s))^T A(x(s)))ds| \\ \leq & \; a_0(1 + a_2 + a_1^2/)()^{+k}, \end{aligned} \tag{4.23}$$

where a_2 is an upper bound on the norm of the Hessian of the Lagrangian (with bounded multiplier estimates) within Ω. We now combine the identity

$$\nabla_x \Phi(, \lambda^{(k+1)},) = \nabla_x \Phi(, \lambda^{(k)},) + A()^T c()/ \tag{4.24}$$

with (2.12), (3.1), (4.5), (4.21) and (4.23) to derive the inequality

$$\begin{aligned} |\nabla_x \Phi(, \lambda^{(k+1)},)_i| \leq & \; |\nabla_x \Phi(,,)_i| \\ & + a_0 \|\nabla_x c_i()\|_p \|c()\|_p/ \\ \leq & \; a_0(1 + a_2 + a_1^2/)()^{+k} \\ & + a_0^2 a_1 (1 + a_1 \omega_0/\eta_0)()^{-1+k} \\ \leq & \; a_3()^{-1+k}, \end{aligned} \tag{4.25}$$

where $a_3 \stackrel{\text{def}}{=} a_0(1+a_2+a_1^2(1+a_0))+a_0^2 a_1$. As k increases, the right-hand-side of the inequality (4.25) converges to zero. Thus for k sufficiently large, $_i$ is floating for each $i \in \mathcal{F}_1$ and (2.9) and (4.25) imply that

$$|P(, \nabla_x \Phi(, \lambda^{(k+1)},))_i| = |\nabla_x \Phi(, \lambda^{(k+1)},)_i| \leq a_3()^{-1+k}. \tag{4.26}$$

Conversely, consider the variables which lie in $\mathcal{D}_1$ for $k \geq k_0$. We then have that

$$\begin{aligned} & |\nabla_x \Phi(,,)_i - \nabla_x \Phi(,,)_i| \\ & \quad \leq \; \| - \|_2 \cdot | \int_0^1 (H_\ell(x(s),) + \tfrac{1}{}A(x(s))^T A(x(s)))_i ds| \\ & \quad \leq \; a_0(a_2 + a_1^2/)()^{+k} \end{aligned} \tag{4.27}$$

using the same tools that we used to obtain (4.23). Then, combining (2.12), (3.1), (4.5), (4.21), (4.24) and (4.27) we obtain the inequality

$$\begin{aligned} & |\nabla_x \Phi(, \lambda^{(k+1)},)_i - \nabla_x \Phi(,,)_i| \\ & \quad \leq \; a_0(a_2 + a_1^2/)()^{+k} + a_0 a_1 (1 + a_1 \omega_0/\eta_0)()^{-1+k} \leq a_4()^{-1+k}, \end{aligned} \tag{4.28}$$

where $a_4 \stackrel{\text{def}}{=} a_0(a_2 + 2a_1^2) + a_0 a_1$. Thus, for sufficiently large k the right-hand-side of (4.28) can be made arbitrarily small. Combining this result with (4.18)

and the identity $_i = 0$, we see that $_i$ is dominated for each $i \in \mathcal{D}_1$ and (2.8) and (4.28) imply that

$$P(, \nabla_x \Phi(, \lambda^{(k+1)},))_i =_i = 0. \tag{4.29}$$

Therefore, using (2.9), (2.12), (4.26) and (4.29), we have

$$\|P(, \nabla_x \Phi(, \lambda^{(k+1)},))\|_p = \|\nabla_x \Phi(, \lambda^{(k+1)},)_{[\mathcal{F}_1]}\|_p \leq a_5()^{-1+k}, \tag{4.30}$$

for all k sufficiently large, where $a_5 \stackrel{\text{def}}{=} a_0 a_3 \|e_{[\mathcal{F}_1]}\|_2$.

(ii) The generalized Cauchy point. We consider the Cauchy arc emanating from . We have shown that the variables in $\mathcal{D}_1$ are on their bounds; the relationships (4.15), (4.18) and (4.28) imply that $\nabla_x \Phi(, \lambda^{(k+1)},)_i > 0$ and hence that $s^{(k+1,0)}(t)_i = 0$ for all $t > 0$ and $i \in \mathcal{D}_1$. Thus the variables in $\mathcal{D}_1$ remain fixed on the bounds throughout the first inner iteration and

$$s^{(k+1,0)}_{[\mathcal{D}_1]} = 0 \tag{4.31}$$

for all k sufficiently large.

The remaining variables, those indexed by $\mathcal{F}_1$, are free from their bounds. The set $\mathcal{J}_1$ in assumption AS9 is identical to $\mathcal{F}_1$ and under this assumption the matrix (4.10) is positive definite with extreme eigenvalues $0 < \lambda_{\min} \leq \lambda_{\max}$, say. The definition (3.12) and inequalities (2.8), (2.9) and (3.2) imply that converges to . Thus the matrix

$$H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]} = B^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]} + \frac{1}{-} A()^T_{[\mathcal{F}_1]} A()_{[\mathcal{F}_1]} \tag{4.32}$$

is also positive definite with extreme eigenvalues satisfying

$$0 < \tfrac{1}{2}\lambda_{\min} \leq \lambda^{(k+1,0)}_{\min} \leq \lambda^{(k+1,0)}_{\max} \leq 2\lambda_{\max}, \tag{4.33}$$

say, for all sufficiently large k. Hence the model (3.8) is a strictly convex function in the subspace of free variables during the first inner iteration.

We now show that the set

$$\mathcal{L} \stackrel{\text{def}}{=} \{s_{[\mathcal{F}_1]} \mid m^{(k+1,0)}(+s) \leq m^{(k+1,0)}() \quad \text{and} \quad s_{[\mathcal{D}_1]} = 0\} \tag{4.34}$$

lies strictly interior to the set $\mathcal{C}(1)$ 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,

$$s^*_{[\mathcal{F}_1]} = -H^{(k+1,0)-1}_{[\mathcal{F}_1,\mathcal{F}_1]} \nabla_x \Phi(, \lambda^{(k+1)},)_{[\mathcal{F}_1]}. \tag{4.35}$$

Let $s_{[\mathcal{F}_1]} \stackrel{\text{def}}{=} s^*_{[\mathcal{F}_1]} + v_{[\mathcal{F}_1]} \in \bar{\mathcal{L}}$. Then, combining (3.8), (4.32), (4.34) and (4.35), we have that

$$\begin{aligned}
&\tfrac{1}{2} v^T_{[\mathcal{F}_1]} H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]} v_{[\mathcal{F}_1]} \\
&\quad = \tfrac{1}{2} s^{*T}_{[\mathcal{F}_1]} H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]} s^*_{[\mathcal{F}_1]} + (m^{(k+1,0)}(+s^* + v) - m^{(k+1,0)}()) \\
&\qquad -(s^* + v)^T_{[\mathcal{F}_1]} (H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]} s^*_{[\mathcal{F}_1]} + \nabla_x \Phi(, \lambda^{(k+1)},)_{[\mathcal{F}_1]}) \\
&\quad = \tfrac{1}{2} s^{*T}_{[\mathcal{F}_1]} H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]} s^*_{[\mathcal{F}_1]} \\
&\quad = \tfrac{1}{2} \nabla_x \Phi(, \lambda^{(k+1)},)^T_{[\mathcal{F}_1]} H^{(k+1,0)-1}_{[\mathcal{F}_1,\mathcal{F}_1]} \nabla_x \Phi(, \lambda^{(k+1)},)_{[\mathcal{F}_1]}.
\end{aligned} \tag{4.36}$$

Hence, using the extremal properties of the Rayleigh quotient and (4.36), we have

$$\begin{aligned} d^2 \stackrel{\rm def}{=} & 4\|v^*_{[\mathcal{F}_1]}\|_2^2 \leq 4v^{*T}_{[\mathcal{F}_1]} H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]} v^*_{[\mathcal{F}_1]}/\lambda^{(k+1,0)}_{\min} \leq 8v^{*T}_{[\mathcal{F}_1]} H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]} v^*_{[\mathcal{F}_1]}/\lambda_{\min} \\ = & 8\nabla_x \Phi(\,,\lambda^{(k+1)},\,)^T_{[\mathcal{F}_1]} H^{(k+1,0)-1}_{[\mathcal{F}_1,\mathcal{F}_1]} \nabla_x \Phi(\,,\lambda^{(k+1)},\,)_{[\mathcal{F}_1]}/\lambda_{\min} \\ \leq & 16\|\nabla_x \Phi(\,,\lambda^{(k+1)},\,)_{[\mathcal{F}_1]}\|_2^2/\lambda^2_{\min} \end{aligned} \tag{4.37}$$

where $\|v^*_{[\mathcal{F}_1]}\|_2 = \max_{s^*_{[\mathcal{F}_1]}+v_{[\mathcal{F}_1]} \in \bar{\mathcal{L}}} \|v_{[\mathcal{F}_1]}\|_2$. Thus, using (2.12), (4.30) and (4.37), any step within $\mathcal{L}$ satisfies the bound

$$\begin{aligned} \|s_{[\mathcal{F}_1]}\|_2 \leq & \; d \leq 4\|\nabla_x \Phi(\,,\lambda^{(k+1)},\,)_{[\mathcal{F}_1]}\|_2/\lambda_{\min} \\ \leq & \; 4a_0 a_5()^{-1+k}/\lambda_{\min}. \end{aligned} \tag{4.38}$$

The inequality (4.22) shows that $_i$, $i \in \mathcal{F}_1$, is separated from its bound for all k sufficiently large while (4.38) and the two-norm consistency of the infinity norm shows that all steps within $\mathcal{L}$ become arbitrarily small. Thus the problem bounds are excluded from $\mathcal{L}$. Moreover (2.11), (3.13), (4.30), (4.31) and (4.38) combine to give

$$\|s\|_t = \|s_{[\mathcal{F}_1]}\|_t \leq a_0^{\frac{1}{2}} \|s_{[\mathcal{F}_1]}\|_2 \leq \Delta^{(k+1,0)} \frac{4a_0 \|\nabla_x \Phi(\,,\lambda^{(k+1)},\,)_{[\mathcal{F}_1]}\|_p^{1-\zeta}}{\lambda_{\min}\kappa}. \tag{4.39}$$

for all steps on or within $\mathcal{L}$. Inequality (4.30) then combines with (4.39) 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 $\mathcal{C}(1) \subseteq \mathcal{C}(\beta_1)$ for all k sufficiently large. But, as all iterates generated by Algorithm 3.3 satisfy (3.20) 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 $\mathcal{C}(\beta_1)$.

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

$$t = t^* = \frac{\nabla_x \Phi(\,,\lambda^{(k+1)},\,)^T_{[\mathcal{F}_1]} \nabla_x \Phi(\,,\lambda^{(k+1)},\,)_{[\mathcal{F}_1]}}{\nabla_x \Phi(\,,\lambda^{(k+1)},\,)^T_{[\mathcal{F}_1]} H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]} \nabla_x \Phi(\,,\lambda^{(k+1)},\,)_{[\mathcal{F}_1]}}. \tag{4.40}$$

and thus, from the above discussion, gives the generalized Cauchy point proposed by (). We use the definition of t^*, (2.11) and the extremal property of the Rayleigh quotient to obtain

$$\begin{aligned} m^{(k,0)}(x^{(k+1,0)}) - m^{(k+1,0)}(x^{(k+1,0)} + s^{C(k+1,0)}) = & \; \tfrac{1}{2} t^* \|\nabla_x \Phi(\,,\lambda^{(k+1)},\,)_{[\mathcal{F}_1]}\|_2^2 \\ \geq & \; \frac{\|\nabla_x \Phi(\,,\lambda^{(k+1)},\,)_{[\mathcal{F}_1]}\|_p^2}{4a_0 \lambda_{\max}} \end{aligned} \tag{4.41}$$

for this variant of the generalized Cauchy point. Alternatively, if Moré's (1988) variant is used, the requirement (3.16) and the definition of the Cauchy arc imply that

$$m^{(k+1,0)}(x^{(k+1,0)}) - m^{(k+1,0)}(x^{(k+1,0)} + s^{C(k+1,0)}) \geq \mu_1 t^{C(k,0)} \|\nabla_x \Phi(\,,\lambda^{(k+1)},\,)_{[\mathcal{F}_1]}\|_2^2. \tag{4.42}$$

If the first alternative of (3.17) holds, (4.42) implies that

$$m^{(k+1,0)}(x^{(k+1,0)}) - m^{(k+1,0)}(x^{(k+1,0)} + s^{C(k+1,0)}) \geq \mu_1\nu_1\|\nabla_x\Phi(,\lambda^{(k+1)},)_{[\mathcal{F}_1]}\|_2^2. \tag{4.43}$$

Otherwise, we may use the same arguments as above to show that it is impossible for $t^{L(k+1,0)}$ to satisfy (3.19) when k is sufficiently large. Therefore, $t^{L(k+1,0)}$ must satisfy (3.18). Combining (3.8), (3.18), (4.32) and the definition of the Cauchy arc, we have that

$$\begin{aligned} &\tfrac{1}{2}(t^{L(k+1,0)})^2\nabla_x\Phi(,\lambda^{(k+1)},)^T_{[\mathcal{F}_1]}H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]}\nabla_x\Phi(,\lambda^{(k+1)},)_{[\mathcal{F}_1]} \\ &\quad \geq (1-\mu_2)t^{L(k+1,0)}\|\nabla_x\Phi(,\lambda^{(k+1)},)_{[\mathcal{F}_1]}\|_2^2. \end{aligned} \tag{4.44}$$

Hence, combining (4.33) and (4.44) with the extremal properties of the Rayleigh quotient, we have that $t^{L(k,j)} \geq (1-\mu_2)/\lambda_{\max}$. Thus, when the second alternative of (3.17) holds, this result and (4.42) give that

$$\begin{aligned} &m^{(k+1,0)}(x^{(k+1,0)}) - m^{(k+1,0)}(x^{(k+1,0)} + s^{C(k+1,0)}) \\ &\quad \geq \quad [\mu_1\nu_2(1-\mu_2)/\lambda_{\max}]\|\nabla_x\Phi(,\lambda^{(k+1)},)_{[\mathcal{F}_1]}\|_2^2. \end{aligned} \tag{4.45}$$

Therefore, (2.12), (4.43) and (4.45) give the inequality

$$\begin{aligned} &m^{(k,0)}(x^{(k,0)}) - m^{(k,0)}(x^{(k,0)} + s^{C(k,0)}) \\ &\quad \geq (\mu_1/a_0)\min(\nu_1, \nu_2(1-\mu_2)/\lambda_{\max})\|\nabla_x\Phi(,\lambda^{(k+1)},)_{[\mathcal{F}_1]}\|_p^2. \end{aligned} \tag{4.46}$$

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

(iii) Improvements beyond the generalized Cauchy point. We have that $x^{(k+1,0)}_{[\mathcal{D}]} = 0$, and, as a consequence of (4.30), $\|P(,\nabla_x\Phi^{(k+1,0)})\|_p^\xi \leq \nu$ for all k sufficiently large. Hence, because we have shown that any s in $\mathcal{L}$ lies strictly interior to $\mathcal{C}$, a single pass of step 2 of Algorithm 3.3 is required. We must pick s to satisfy (3.24) and (3.20) by determining $s^{(k+1,0)}_{[\mathcal{F}_1]}$ so that

$$\|H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]}s^{(k+1,0)}_{[\mathcal{F}_1]} + \nabla_x\Phi^{(k+1,0)}_{[\mathcal{F}]}\|_p \leq \|\nabla_x\Phi^{(k+1,0)}_{[\mathcal{F}]}\|_p^{1+\xi}. \tag{4.47}$$

and

$$\begin{aligned} &m^{(k,j)}(x^{(k+1,0)}) - m^{(k+1,0)}(x^{(k+1,0)} + s^{(k+1,0)}) \\ &\quad \geq \beta_2(m^{(k+1,0)}(x^{(k+1,0)}) - m^{(k+1,0)}(x^{(k+1,0)} + s^{C(k+1,0)})) \end{aligned} \tag{4.48}$$

for some $\beta_2 \leq 1$. The set of values which satisfy (4.47) and (4.48) is non-empty as the Newton step (4.35) satisfies both inequalities.

It remains to consider such a step in slightly more detail. Suppose that $s^{(k+1,0)}_{[\mathcal{F}_1]}$ satisfies (4.47). Let

$$r^{(k+1,0)}_{[\mathcal{F}_1]} = H^{(k+1,0)}_{[\mathcal{F}_1,\mathcal{F}_1]}s^{(k+1,0)}_{[\mathcal{F}_1]} + \nabla_x\Phi^{(k+1,0)}_{[\mathcal{F}]} \tag{4.49}$$

Then combining (2.11), (4.33), (4.47) and (4.49), we have

$$\begin{aligned} \|s^{(k+1,0)}_{[\mathcal{F}_1]}\|_p \leq &\quad a_0\|H^{(k+1,0)-1}_{[\mathcal{F}_1,\mathcal{F}_1]}\|_2(\|r^{(k+1,0)}_{[\mathcal{F}_1]}\|_p + \|\nabla_x\Phi^{(k+1,0)}_{[\mathcal{F}]}\|_p) \\ \leq &\quad 2a_0\|\nabla_x\Phi^{(k+1,0)}_{[\mathcal{F}]}\|_p(1 + \|\nabla_x\Phi^{(k+1,0)}_{[\mathcal{F}]}\|_p^\xi)/\lambda_{\min}. \end{aligned} \tag{4.50}$$

(iv) Acceptance of the new point. We have seen that $s_{[\mathcal{D}_1]}^{(k+1,0)} = 0$ and $s_{[\mathcal{F}_1]}^{(k+1,0)}$ satisfies (4.47). We now wish to show that the quantity

$$|\rho^{(k+1,0)} - 1| = \frac{|\Phi(+s^{(k+1,0)}, \lambda^{(k+1)},) - m^{(k+1,0)}(+s^{(k+1,0)})|}{|m^{(k+1,0)}() - m^{(k+1,0)}(+s^{(k+1,0)})|} \tag{4.51}$$

converges to zero. For then the new point will prove acceptable in step 4 of Algorithm 3.1.

Consider first the denominator on the right-hand-side of (4.51). Combining (4.41), (4.46) and (4.48), we have

$$m^{(k+1,0)}() - m^{(k+1,0)}(+s^{(k+1,0)}) \geq a_6 \|\nabla_x \Phi(, \lambda^{(k+1)},)_{[\mathcal{F}_1]}\|_p^2, \tag{4.52}$$

where $a_6 = \beta_2 \min(1/(4a_0\lambda_{\max}), \mu_1 \min(\nu_1, \nu_2(1-\mu_2)/\lambda_{\max})/a_0)$. Turning to the numerator on the right-hand-side of (4.51), we use the integral mean value theorem to obtain

$$\begin{aligned}
&\Phi(+s^{(k+1,0)}, \lambda^{(k+1)},) \\
&= \Phi(, \lambda^{(k+1)},) + s_{[\mathcal{F}_1]}^{(k+1,0)T} \nabla_x \Phi(, \lambda^{(k+1)},)_{[\mathcal{F}_1]} \\
&\quad + \tfrac{1}{2} \textstyle\int_0^1 s_{[\mathcal{F}_1]}^{(k+1,0)T} \nabla_{xx} \Phi(x(t), \lambda^{(k+1)},)_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)} dt \\
&= \Phi(, \lambda^{(k+1)},) + s_{[\mathcal{F}_1]}^{(k+1,0)T} \nabla_x \Phi(, \lambda^{(k+1)},)_{[\mathcal{F}_1]} \\
&\quad + \tfrac{1}{2} \textstyle\int_0^1 s_{[\mathcal{F}_1]}^{(k+1,0)T} (\nabla_{xx} \Phi(x(t), \lambda^{(k+1)},) - \nabla_{xx} \Phi(, \lambda^{(k+1)},))_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)} dt \\
&\quad + \tfrac{1}{2} s_{[\mathcal{F}_1]}^{(k+1,0)T} (\nabla_{xx} \Phi(, \lambda^{(k+1)},) - H^{(k+1,0)})_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)} \\
&\quad + \tfrac{1}{2} s_{[\mathcal{F}_1]}^{(k+1,0)T} H_{[\mathcal{F}_1,\mathcal{F}_1]}^{(k+1,0)} s_{[\mathcal{F}_1]}^{(k+1,0)} \\
&= m^{(k+1,0)}(+s^{(k+1,0)}) \\
&\quad + \tfrac{1}{2} \textstyle\int_0^1 s_{[\mathcal{F}_1]}^{(k+1,0)T} (\nabla_{xx} \Phi(x(t), \lambda^{(k+1)},) - \nabla_{xx} \Phi(, \lambda^{(k+1)},))_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)} dt \\
&\quad + \tfrac{1}{2} s_{[\mathcal{F}_1]}^{(k+1,0)T} (\nabla_{xx} \Phi(, \lambda^{(k+1)},) - H^{(k+1,0)})_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)},
\end{aligned} \tag{4.53}$$

where $x(t) = +ts^{(k+1,0)}$. Considering the last two terms in (4.53) in turn, we have the bounds

$$\begin{aligned}
&|\tfrac{1}{2} \textstyle\int_0^1 s_{[\mathcal{F}_1]}^{(k+1,0)T} (\nabla_{xx} \Phi(x(t), \lambda^{(k+1)},) - \nabla_{xx} \Phi(, \lambda^{(k+1)},))_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)} dt| \\
&\quad \leq \tfrac{1}{4} a_0 a_7 \|s_{[\mathcal{F}_1]}^{(k+1,0)}\|_p^3
\end{aligned} \tag{4.54}$$

using (2.11), the convergence (and hence boundedness) of the Lagrange multiplier estimates and the Lipschitz continuity of the second derivatives of the problem functions (assumption AS4) with some composite Lipschitz constant a_7, and

$$\begin{aligned}
&|\tfrac{1}{2} s_{[\mathcal{F}_1]}^{(k+1,0)T} (\nabla_{xx} \Phi(, \lambda^{(k+1)},) - H^{(k+1,0)})_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)}| \\
&\quad \leq a_0 (\tfrac{1}{2} \upsilon \|s_{[\mathcal{F}_1]}^{(k+1,0)}\|_p^{\varsigma} + \|(\nabla_{xx} \ell(, \lambda^{(k+1)}) - \nabla_{xx} \ell(,))_{[\mathcal{F}_1,\mathcal{F}_1]}\|_p) \|s_{[\mathcal{F}_1]}^{(k+1,0)}\|_p^2
\end{aligned} \tag{4.55}$$

using (2.11), (3.21), the definition of the Hessian of the augmented Lagrangian and AS8. Thus, combining (4.50), (4.51), (4.52), (4.53), (4.54) and (4.55), we obtain

$$|\rho^{(k+1,0)} - 1| \leq 4a_0^3(1 + \|\nabla_x \Phi_{[\mathcal{F}]}^{(k+1,0)}\|^{\xi})_p^2 \cdot \frac{\frac{1}{4}a_7\|s_{[\mathcal{F}_1]}^{(k+1,0)}\|_p + \frac{1}{2}\upsilon\|s_{[\mathcal{F}_1]}^{(k+1,0)}\|_p^{\varsigma} + \|(\nabla_{xx}\ell(,\lambda^{(k+1)}) - \nabla_{xx}\ell(,))_{[\mathcal{F}_1,\mathcal{F}_1]}\|_p}{a_6\lambda_{\min}^2}. \tag{4.56}$$

As the right-hand-side of (4.56) converges to zero as k increases, $= +s^{(k+1,0)}$ for all k sufficiently large.

(v) Convergence of the inner iteration at the new point. We now show that satisfies the inner-iteration convergence test (3.7).

Firstly, in the same vein as (4.27) , for $i \in \mathcal{D}_1$ we have that

$$\begin{array}{rcl} |\nabla_x \Phi(,\lambda^{(k+1)},)_i - \nabla_x \Phi(,\lambda^{(k+1)},)_i| & & \\ \leq & \|s^{(k+1,0)}\|_2 \cdot |\int_0^1 (H_\ell(x(t),) + \frac{1}{}A(x(t))^T A(x(t)))_i dt| & \\ \leq & (a_2 + a_1^2/)\|s^{(k+1,0)}\|_2, & \end{array} \tag{4.57}$$

where $x(t) = +ts^{(k+1,0)}$. Thus, as the right-hand-side of (4.57) can be made arbitrarily small, by taking k sufficiently large, (4.18) and the identity $_i =_i = 0$ for each $i \in \mathcal{D}_1$, imply that $_i$ is dominated for each $i \in \mathcal{D}_1$ and (2.8) and (4.25) imply that

$$P(,\nabla_x \Phi(,\lambda^{(k+1)},))_i =_i = 0. \tag{4.58}$$

We now consider the components of $P(,\nabla_x \Phi(,\lambda^{(k+1)},))_i$ for $i \in \mathcal{F}_1$. Using the integral mean value theorem, we have

$$\begin{array}{rcl} \nabla_x \Phi(,\lambda^{(k+1)},)_{[\mathcal{F}_1]} & & \\ = & \nabla_x \Phi(,\lambda^{(k+1)},)_{[\mathcal{F}_1]} + \int_0^1 \nabla_{xx}\Phi(x(t),\lambda^{(k+1)},)_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)} dt & \\ = & (H_{[\mathcal{F}_1,\mathcal{F}_1]}^{(k+1,0)} s_{[\mathcal{F}_1]}^{(k+1,0)} + \nabla_x \Phi(,\lambda^{(k+1)},)_{[\mathcal{F}_1]}) & \\ & + \int_0^1 (\nabla_{xx}\Phi(x(t),\lambda^{(k+1)},) - \nabla_{xx}\Phi(,\lambda^{(k+1)},))_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)} dt & \\ & + (\nabla_{xx}\Phi(,\lambda^{(k+1)},) - H^{(k+1,0)})_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)} & \end{array} \tag{4.59}$$

where $x(t) = +ts^{(k+1,0)}$. We observe that each of the three terms on the right-hand-side of (4.59) 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 (4.47). Hence, combining (4.30) and (4.47), we obtain

$$\|H_{[\mathcal{F}_1,\mathcal{F}_1]}^{(k+1,0)} s_{[\mathcal{F}_1]}^{(k+1,0)} + \nabla_x \Phi_{[\mathcal{F}]}^{(k+1,0)}\|_p \leq a_5^{1+\xi}()^{(-1)(1+\xi)+k(1+\xi)}. \tag{4.60}$$

The second term satisfies the bound

$$\begin{array}{c} \|\int_0^1 (\nabla_{xx}\Phi(x(t),\lambda^{(k+1)},) - \nabla_{xx}\Phi(,\lambda^{(k+1)},))_{[\mathcal{F}_1,\mathcal{F}_1]} s_{[\mathcal{F}_1]}^{(k+1,0)} dt\|_p \\ \leq \frac{1}{2}a_0 a_7 \|s_{[\mathcal{F}_1]}^{(k+1,0)}\|_p^2. \end{array} \tag{4.61}$$

by the same arguments we used to establish inequality (4.54). Picking k sufficiently large that $\|\nabla_x \Phi^{(k+1,0)}_{[\mathcal{F}]}\| \leq 1$, we may combine (4.30), (4.50) and (4.61) so that

$$\begin{aligned} \| \int_0^1 (\nabla_{xx}\Phi(x(t), \lambda^{(k+1)},) - \nabla_{xx}\Phi(, \lambda^{(k+1)},))_{[\mathcal{F}_1,\mathcal{F}_1]} s^{(k+1,0)}_{[\mathcal{F}_1]} dt\| \\ \leq 8 a_0^3 a_5^2 a_7()^{2-2+k2}/\lambda^2_{\min}. \end{aligned} \tag{4.62}$$

Lastly, using the same arguments as those used to establish (4.55), the definitions (3.21) and of the Hessian of the augmented Lagrangian, the Lipschitz continuity of the second derivatives of the problem functions (assumption AS4) and the accuracy of the second derivative approximations (assumption AS8) imply that

$$\begin{aligned} &\|(\nabla_{xx}\Phi(, \lambda^{(k+1)},) - H^{(k+1,0)})_{[\mathcal{F}_1,\mathcal{F}_1]} s^{(k+1,0)}_{[\mathcal{F}_1]}\|_p \\ &\quad \leq \quad (\upsilon \|s^{(k+1,0)}_{[\mathcal{F}_1]}\|^{\varsigma}_p + \|(\nabla_{xx}\ell(, \lambda^{(k+1)}) - \nabla_{xx}\ell(,))_{[\mathcal{F}_1,\mathcal{F}_1]}\|_p) \|s^{(k+1,0)}_{[\mathcal{F}_1]}\|_p \\ &\quad \leq \quad (\upsilon \|s^{(k+1,0)}_{[\mathcal{F}_1]}\|^{\varsigma}_p + a_8 \| - \|_p + a_9 \|\lambda^{(k+1)} - \|_p) \|s^{(k+1,0)}_{[\mathcal{F}_1]}\|_p, \end{aligned} \tag{4.63}$$

for some composite Lipschitz constants a_8 and a_9. Again picking k sufficiently large that $\|\nabla_x \Phi^{(k+1,0)}_{[\mathcal{F}]}\|_p \leq 1$ and recalling that $\lambda^{(k+1)} =$, we may combine (2.12), (4.4), (4.12), (4.20), (4.30), (4.50) and (4.63) so that

$$\begin{aligned} &\|(\nabla_{xx}\Phi(, \lambda^{(k+1)},) - H^{(k+1,0)})_{[\mathcal{F}_1,\mathcal{F}_1]} s^{(k+1,0)}_{[\mathcal{F}_1]}\|_p \\ &\quad \leq \quad [\upsilon((4a_0a_5/\lambda_{\min})()^{-1+k})^{\varsigma} + a_8(a_x()^{+k} + a_0()^{+k}) \\ &\qquad +a_9 a_\lambda()^{+k}](4a_0a_5/\lambda_{\min})()^{-1+k}. \end{aligned} \tag{4.64}$$

We now combine equation (4.59) with the inequalities (4.60), (4.62) and (4.64), the condition $\xi < 1$ and the definitions of < 1 and > 0 to obtain the bound

$$\|\nabla_x \Phi(, \lambda^{(k+1)},)_{[\mathcal{F}_1]}\| \leq a_{10}()^{+k}, \tag{4.65}$$

where

$$\begin{aligned} a_{10} = \quad & a_5^{1+\xi} + 8a_0^3 a_5^2 a_6/\lambda^2_{\min} + (4a_0a_5/\lambda_{\min})(\upsilon((4a_0a_5/\lambda_{\min})^{\varsigma}) + a_8(a_x + a_0) + a_9 a_\lambda), \\ = \quad & (-1)(1 + \max(1, \varsigma)) \quad \text{and} \\ = \quad & (1 + \min(\xi, \varsigma)). \end{aligned} \tag{4.66}$$

Firstly, observe that the right-hand-side of (4.65) may be made arbitrarily small. Therefore, (2.9), (4.58) and (4.65) imply that

$$\|P(, \nabla_x \Phi(, \lambda^{(k+1)},))\|_p = \|\nabla_x \Phi(, \lambda^{(k+1)},)_{[\mathcal{F}_1]}\|_p \leq a_{10}()^{+k}. \tag{4.67}$$

Secondly, define $\delta = \log(a_{10}/\omega_0)$. Now let k_1 be any integer for which

$$k_1 \geq \frac{+ - -\delta}{-}. \tag{4.68}$$

Then (4.11), (4.67) and (4.68) imply that

$$\|P(, \nabla_x \Phi(, \lambda^{(k+1)},))\|_p \leq a_{10}()^{+k} \leq \omega_0()^{+(k+1)} = \omega^{(k+1)} \tag{4.69}$$

for all sufficiently large $k \geq k_1$. Thus, the iterate satisfies the inner iteration convergence test (3.2) for all k sufficiently large and we have $x^{(k+1)} =$.
(vi) Redundancy of the shifted starting point. Finally, we observe that all the variables $x_i^{(k+1)}$, $i \in \mathcal{D}$, lie on their bounds for sufficiently large k. Therefore, $x^{(k+2,0)} = x^{(k+1)}$ 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 $\mathcal{B} = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$ and replace $P(x, v)$ by $P(x, v, l, u)$ where appropriate. The concept of floating and dominated variables stays essentially the same. Now for each iterate in $\mathcal{B}$ we have three mutually exclusive possibilities, namely, (i) $0 \leq_i -l_i \leq (\nabla_x \Phi^{(k)})_i$, (ii) $(\nabla_x \Phi^{(k)})_i \leq_i -u_i \leq 0$ or (iii) $_i - u_i < (\nabla_x \Phi^{(k)})_i <_i -l_i$, for each component $_i$. In case (i) we then have that $(\nabla_x \Phi^{(k)})_i =_i -l_i$ while in case (ii) $(\nabla_x \Phi^{(k)})_i =_i -u_i$ and in case (iii) $(\nabla_x \Phi^{(k)})_i = (\nabla_x \Phi^{(k)})_i$. 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 (2.10) 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 (3.12) by

$$x_i^{(k,0)} = \begin{cases} l_i & \text{if } 0 \leq x_i^{(k-1)} - l_i \leq \theta(\nabla_x \Phi^{(k-1)})_i \\ u_i & \text{if } \theta(\nabla_x \Phi^{(k-1)})_i \leq x_i^{(k-1)} - u_i \leq 0 \\ x_i^{(k-1)} & \text{otherwise.} \end{cases} \tag{5.1}$$

With such definitions, we may reprove the results of section 4, extending AS5, AS7—AS9 in the obvious way. The only important new ingredient is that () indicate that the non-degeneracy assumption AS7 ensures that the iterates are asymptotically isolated in 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 augmented Lagrangian algorithm which lies at the heart of the LANCELOT package. This then places the algorithm in the class of diagonal multiplier methods whose asymptotic behaviour has been studied by Tapia1977 ().

References

[Bertsekas1982] D.P. Bertsekas. *Constrained Optimization and Lagrange multiplier methods*. Academic Press, London and New York, 1982.

[Burke and Moré1988] J. V. Burke and J. J. Moré. On the identification of active constraints. *SIAM Journal on Numerical Analysis*, 25:1197–1211, 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*, 47:305–336, 1990.

[Calamai and Moré1987] 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 *SIAM Journal on Numerical Analysis,* 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.*1991a] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, 28(1):545–572, 1991.

[Conn *et al.*1991b] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. LANCELOT *User's guide*, 1991.

[Dennis and Moré1974] J. E. Dennis and J. J. Moré. A characterization of superlinear convergence and its application to quasi-Newton methods. *Mathematics of Computation*, 28:549–560, 1974.

[Dennis and Schnabel1983] J. E. Dennis and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

[Gould1986] 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.

[Gould1989] 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.

[Moré1988] J. J. Moré. Trust regions and projected gradients. In M. Iri and K. Yajima, editors, *Systems Modelling and Optimization: Proceedings of the*

13th IFIP conference on Systems Modelling and Optimization, August 31 - September 4, 1987, Lecture Notes in Control and Information Science, Vol. 113, Springer Verlag, Berlin, pages 1–13, 1988.

[Tapia1977] R. A. Tapia. Diagonalized multiplier methods and quasi-newton methods for constrained optimization. *Journal of Optimization Theory and Applications*, 22:135–194, 1977.

[Toint1988] 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.

A. R. Conn,
T. J. Watson IBM Research Center,
Yorktown Heights, New York 10598, USA.

N. I. M. Gould,
Atlas Centre, Rutherford Appleton Laboratory,
Chilton, Oxfordshire, England.

Ph. L. Toint,
Department of Mathematics, FUNDP,
Namur, B-5000, Belgium.