A Globally Convergent Lagrangian Barrier Algorithm for Optimization with General Inequality Constraints and Simple Bounds

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

Report 92/07(2nd revision) September 5, 1995

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

[2] CERFACS, 42 Avenue Gustave Coriolis, 31057 Toulouse Cedex, France, EC
Email : gould@cerfacs.fr or nimg@directory.rl.ac.uk
Current reports available by anonymous ftp from the directory
"pub/reports" on camelot.cc.rl.ac.uk (internet 130.246.8.61)

[3] Department of Mathematics, Facultés Universitaires ND de la Paix, 61, rue de Bruxelles,
B-5000 Namur, Belgium, EC
Email : pht@math.fundp.ac.be
Current reports available by anonymous ftp from the directory
"pub/reports" on thales.math.fundp.ac.be (internet 138.48.4.14)

Keywords : Constrained optimization, barrier methods, inequality constraints, convergence theory.

A Globally Convergent Lagrangian Barrier Algorithm for Optimization with General Inequality Constraints and Simple Bounds

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

September 5, 1995

Abstract

We consider the global and local convergence properties of a class of Lagrangian barrier methods for solving nonlinear programming problems. In such methods, simple bound constraints may be treated separately from more general constraints. The objective and general constraint functions are combined in a Lagrangian barrier function. A sequence of such functions are approximately minimized within the domain defined by the simple bounds. Global convergence of the sequence of generated iterates to a first-order stationary point for the original problem is established. Furthermore, possible numerical difficulties associated with barrier function methods are avoided as it is shown that a potentially troublesome penalty parameter is bounded away from zero. This paper is a companion to our previous work (see, Conn *et al.*, 1991) on augmented Lagrangian methods.

1 Introduction

In this paper, we consider the problem of finding a local minimizer of the function

$$f(x), \tag{1.1}$$

where x is required to satisfy the general inequality constraints

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

and specific simple bounds

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

Here f and c_i map $\Re^n$ into $\Re$ and the inequalities (1.3) are considered component-wise. We shall assume that the region $\mathcal{B} = \{x \mid l \leq x \leq u\}$ is non-empty and may be infinite. We do not rule

[0]The research of Conn and Toint was supported in part by the Advanced Research Projects Agency of the Departement of Defense and was monitored by the Air Force Office of Scientific Research under Contract No F49620-91-C-0079. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

out the possibility that further simple bounds on the variables are included amongst the general constraints (1.2) if that is deemed appropriate. We further assume that

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

We shall attempt to solve our problem by sequential minimization of the *Lagrangian barrier function*

$$\Psi(x, \lambda, s) = f(x) - \sum_{i=1}^{m} \lambda_i s_i \log(c_i(x) + s_i), \tag{1.4}$$

where the components λ_i of the vector λ are positive and are known as Lagrange multiplier estimates and where the elements s_i of the vector s are positive and are known as shifts. Notice that we *do not* include the simple bounds (1.3) in the Lagrangian barrier function. The intention is that the sequential minimization will automatically ensure that the simple bound constraints are always satisfied.

1.1 Background

The logarithmic-barrier function method for finding a local minimizer of (1.1) subject to a set of inequality constraints (1.2) was first introduced by Frisch (1955). The method was put in a sound theoretical framework by Fiacco and McCormick (1968), who also provide an interesting history of such techniques up until then. The basic idea is quite simple.

A composite function, the *barrier* function, is constructed by combining the objective and constraint functions in such a way as to introduce a "barrier" — an infinite penalty — along the constraint boundary. A typical barrier function is the *logarithmic* barrier function

$$f(x) - \mu \sum_{i=1}^{m} \log(c_i(x)), \tag{1.5}$$

where μ is a positive *penalty* parameter. Fiacco and McCormick (1968) show that, under extremely modest conditions, the sequence of minimizers of (1.5) converges to the solution of the original problem whenever the sequence of penalty parameters converge to zero. In particular, under a strict complementary slackness assumption, the error in solving (1.5), that is, the difference between the minimizer of (1.5) and the solution to the original problem, is of order μ as μ tends to zero. (Mifflin, 1975, shows an order $\sqrt{\mu}$ error in the absence of the complementary slackness assumption and a weakening of the assumption that (1.5) be solved exactly.) For further discussion, see the recent survey by Wright (1992).

It was originally envisaged that each of the sequence of barrier functions be minimized using standard methods for unconstrained minimization. However Lootsma (1969) and Murray (1971) painted a less optimistic picture by showing that, under most circumstances, the spectral condition number of the Hessian matrix of the barrier function increases without bound as μ shrinks. This has important repercussions as it indicates that a simple-minded sequential minimization is likely to encounter numerical difficulties. Consequently, the initial enthusiasm for barrier function methods declined. Methods which alleviate these difficulties have been proposed (see, e.g., Murray, 1971, Wright, 1976, Murray and Wright, 1978, Gould, 1986, and McCormick, 1991) that are immediately applicable to smaller dense problems. Nash and Sofer (1993) have recently discusssed an approach that is applicable to the large-scale, nonlinear problems although there experience is only with simple bounds.

Interest in the use of barrier functions was rekindled by the seminal paper of Karmarkar (1984) on polynomial-time interior point algorithms for linear programming and by the intimate connection between these methods and barrier function methods observed by Gill *et al.* (1986). The ill-conditioning problems described above do *not* occur for primal and dual non-degenerate linear programs as the solutions to such problems occur at vertices of the constraint boundary. Furthermore, even in the presence of degeneracy, stable numerical methods may be used to solve the problems (Murray, 1992). Moreover, and most significantly, these methods have turned out to be most effective in practice (see the excellent bibliography of Kranich, 1991).

However, it is quite surprising how the lessons of the early 1970s seem to have been forgotten in the rush to extend interior point methods to solving general constrained optimization problems. The most significant advance seems to us to be the observation that, although the ill-conditioning difficulties are present in most nonlinear programs, the effects may be benign provided sufficient care is taken. In particular Ponceleón (1990) has shown that if the only constraints that are handled by logarithmic terms are simple bounds, the ill-conditioning manifests itself solely on the diagonal of the Hessian matrix of the barrier function. She then shows by a sensitivity analysis that such terms are ultimately irrelevant in assessing the sensitivity of the Newton equations for the problem to numerical perturbations in the data. Methods of this sort have been successfully applied to the minimization of nonlinear functions whose only constraints are simple bounds (1.3) on the variables (see, for instance, Nash and Sofer, 1993).

It is interesting to recall the parallel development of a second class of methods for constrained minimization, penalty function methods. These methods were designed for the case where one wishes to minimize (1.1) subject to a set of equality constraints

$$c_i(x) = 0, \quad 1 \leq i \leq m. \tag{1.6}$$

A composite function, the *penalty* function, is constructed by a suitable combination of the objective and constraint functions. A typical example is the *quadratic* penalty function

$$f(x) + \frac{1}{2\mu}\sum_{i=1}^{m}(c_i(x))^2, \tag{1.7}$$

where as before μ is a positive penalty parameter. One then minimizes a sequence of penalty functions for a given set of penalty parameter values. Fiacco and McCormick (1968) again showed that, under extremely modest conditions, the sequence of minimizers of (1.7) converge to the solution of the original problem whenever the sequence of penalty parameters converge to zero. However, the analysis of Lootsma (1969) and Murray (1971) again had serious ramifications for a simple-minded sequential minimization of (1.7). This time, though, there was almost immediately a way around the ill-conditioning difficulty, the development of augmented Lagrangian methods.

These methods were introduced by Arrow and Solow (1958), Hestenes (1969), Powell (1969) and Rockafellar (1976). The *augmented Lagrangian* function (corresponding to the quadratic penalty function (1.7)) for the above problem is

$$f(x) + \frac{1}{2\mu} \sum_{i=1}^{m} (c_i(x) + s_i))^2, \tag{1.8}$$

where the shifts $s_i = \mu \lambda_i$ and the λ_i are known as Lagrange multiplier estimates. As before, one could fix λ and solve the required problem by sequential minimization of (1.8) as μ converges to zero. However, by adjusting λ so that the Lagrange multiplier estimates converge to the Lagrange multipliers at the solution, it is possible to avoid the need for μ to tend to zero and thus circumvent the conditioning problems inherent in the simple penalty function approach. See Bertsekas (1982) and Conn *et al.* (1991) for further details.

It seems rather strange that such devices were not immediately applied to circumvent the conditioning difficulties associated with traditional barrier function methods, but this appears to be the case. To our knowledge, the first move in this direction was the work by Jittorntrum and Osborne (1980) in which the authors consider a sequential minimization of the modified barrier function

$$f(x) - \mu \sum_{i=1}^{m} \lambda_i \log(c_i(x)) \tag{1.9}$$

for appropriate Lagrange multiplier estimates λ_i. They show that it is possible to get better than linear error estimates of the solution as μ converges to zero merely by choosing the Lagrange multiplier estimates carefully.

The methods which are closest in spirit to the algorithm considered in this paper are the shifted-barrier method analysed for linear programs by Gill *et al.* (1988) and the class of modified barrier methods proposed by Polyak (1982) and analysed in Polyak (1992). Gill *et al.* consider the *shifted* barrier function

$$f(x) - \sum_{i=1}^{m} w_i \log(c_i(x) + s_i), \tag{1.10}$$

where the w_i are termed *weights* and the s_i called *shifts.* A sequence of shifted barrier functions are minimized subject to the restriction that the ratios w_i/s_i converge to the Lagrange multipliers associated with the solution of the original problem. The authors prove convergence of such a scheme under mild conditions for linear programming problems. Polyak (1982) considers the *modified* barrier function

$$f(x) - \mu \sum_{i=1}^{m} \lambda_i \log(1 + c_i(x)/\mu). \tag{1.11}$$

He motivates such a function by noting the equivalence of the constraints (1.2) and

$$\mu \log(1 + c_i(x)/\mu) \geq 0 \quad \text{for} \quad i = 1, \cdots, m. \tag{1.12}$$

The function (1.11) is then merely the classical Lagrangian function for the problem of minimizing (1.1) subject to the constraints (1.12). It is shown in Polyak (1992) that, provided μ is sufficiently small and other reasonable assumptions are satisfied, a sequential minimization of (1.11) *in which μ remains fixed but the Lagrange multipliers are adjusted* will converge to a solution of the original problem. This has the desirable effect of limiting the size of the condition number of the Hessian matrix of (1.11). Breitfeld and Shanno (1993) point out that this additional flexibility allows the encorporation of equality constraints via two (shifted) inequalities. Finally, Freund (1991), Jensen *et al.* (1992)) and Powell (1992a) have analysed and implemented shifted and modified barrier function methods for linear programming. Jensen and Polyak (1993) extend this work to convex problems.

1.2 Motivation

In this paper, we consider the Lagrangian barrier function (1.4). This function is of the form (1.10) when the weights satisfy $w_i = \lambda_i s_i$. As above, we can motivate the its form by observing that the constraints (1.2) are equivalent to

$$s_i \log(1 + c_i(x)/s_i) \geq 0 \quad \text{for} \quad i = 1, \cdots, m \tag{1.13}$$

provided that $s_i > 0$. The classical Lagrangian function for the problem of minimizing (1.1) subject to (1.13) is then

$$f(x) - \sum_{i=1}^{m} \lambda_i s_i \log(1 + c_i(x)/s_i), \tag{1.14}$$

which differs from (1.4) by the constant $\sum_{i=1}^{m} \lambda_i s_i \log(s_i)$. Notice, also, the similarity between (1.4) and (1.8), particularly the shifting of the constraint values. .[1] We aim to show that using (1.4) is an appropriate analogue of (1.8) for inequality constrained optimization by obtaining complementary results to those contained in our previous paper on augmented Lagrangian function methods (see, Conn *et al.*, 1991).

We are interested in considering Lagrangian barrier functions methods for a number of reasons.

1. Much of the work on interior point methods for linear programming appears to indicate that barrier functions provide a way of determining which constraints are active at the solution in a significantly more efficient fashion than active set methods (see, for example, Lustig *et al.*, 1991). More recent experience with the same class of problems using shifted barrier methods has been equally encouraging (see Jensen *et al.*, 1992). The theory presented below for the Lagrangian barrier function indicates that that a similar effect may be expected, as the asymptotically inactive constraints are quickly identified (see Theorem 4.2 (iv)).

2. The experiences with interior point methods for linear programming also suggests that these methods are less sensitive to degeneracy than methods that attempt to identify activities explicitly (see, for instance, Güler *et al.*, 1993, and Rendl *et al.*, 1993). Based on this experience, there is hope that a shifted barrier-function method will be less sensitive to degeneracy for the general nonlinear problem than, for example, the implementation of a gradient-projection augmented Lagrangian method given by Conn *et al.* (1992b) in LANCELOT A.

[1] It is also rather curious to note the strong similarity between (1.8) and the first two terms of a Taylor's expansion of (1.14) for small $c_i(x)/s_i$. Thus one might assume that for small $c_i(x)/s_i$, which is the case asymptotically, the two functions may behave similarly

3. There is numerical evidence that the method presented here is superior to standard barrier function methods when applied to problems with simple bounds (see Conn *et al.*, 1993 and Nash *et al.*, 1993) and preliminary evidence suggests that the same is true for more general constraints (see Breitfeld and Shanno, 1993, and Breifeld and Shanno, 1993).

4. Unlike many modern interior point approaches that are based upon extensions of methods for linear and quadratic programming, our approach makes no convexity or self-concordancy assumptions (see Jarre and Saunders, 1991 and Nesterov and Nemirovsky, 1993).

5. A positive shift enables us to avoid the inherent ill-conditioning present in the classical barrier function methods, thus preventing significant numerical difficulties that may occur for nonlinear problems. An advantage usually ascribed to barrier function methods is that iterates stay feasible. While this is not true for shifted barrier methods, the shifts offer some control over infeasibility. In particular, the shifts allow the algorithm to start an infeasible point.

6. The coherency between the theory developed here and that developed in Conn *et al.* (1991) allows an obvious combination of the two approaches, in the common case where both equality and inequality constraints are present (see Section 8.2).

7. Our current interest is in solving large-scale problems. We have recently developed an algorithm for large-scale nonlinear programming based on sequential minimization of the augmented Lagrangian function (1.8) within a region defined by the simple bounds (1.3) (see Conn *et al.*, 1992b). One disadvantage to such an approach when inequality constraints of the form (1.2) are present is the need to introduce slack variables (see, e.g., Fletcher, 1987, page 146) to convert the inequalities to the form (1.6). Although any slack variables might be treated specially, (see, Conn *et al.*, 1992e), there is still likely to be an overhead incurred from the increase in the number of unknowns. Thus, it would seem to be preferable to avoid slack variables if at all possible.

The combination of reasons outlined above gives ample justification, both from a practical and theoretical point of view, for considering Lagrangian barrier functions.

1.3 Outline

Our exposition will be considerably simplified if we consider the special case where $l_i = 0$ and $u_i = \infty$ for a subset of $\mathcal{N} \stackrel{\text{def}}{=} \{1, 2, ..., n\}$ in (1.3) and where the remaining variables are either not subjected to simple bounds or their simple bounds are treated as general constraints (that is, are encorporated into the barrier function). Indeed, it might sometimes pay to handle all simple bounds as general constraints. For example, this is what is usually done for interior point methods for linear programming. However, there are also circumstances where it is necessary to ensure that certain bounds are always satisfied. For instance, a bound on a variable may have been imposed to ensure that a problem function is well-defined — such a constraint is commonly called a *hard* bound. Although straightforward, the modification required to handle more general bound constraints will be indicated at the end of the paper. Thus we consider the problem:

$$\underset{x \in \Re^n}{\text{minimize}} \quad f(x) \tag{1.15}$$

subject to the constraints

$$c_i(x) \geq 0, \quad 1 \leq i \leq m, \tag{1.16}$$

and the non-negativity restrictions

$$x \in \mathcal{B} = \{x \in R^n \mid x_i \geq 0 \quad \text{for all} \quad i \in \mathcal{N}_b\}, \tag{1.17}$$

where $\mathcal{N}_b \subseteq \mathcal{N}$ is the index set of *bounded* variables.

The paper is organised as follows. In Section 2 we introduce concepts and definitions and then state a generic algorithm for solving (1.15)–(1.17) in Section 3. Global convergence is established in Section 4, while issues of asymptotic convergence follow in Section 5. In Section 6 the consequences of satisfying second order conditions are given. The calculation of good starting points for the inner iteration is considered in Section 7. We conclude in Section 8 by indicating how this theory applies to the original problem (1.1)–(1.3). In order to keep the development of the theory as coherent as possible, we place many of the proofs of the results given in Sections 4 and 5 in appendices. Furthermore, in order to simplify these proofs, we give them for a particular, but generic, instance of the algorithm of Section 3.

We have intentionally kept our development as close as possible to that of Conn *et al.* (1991) in order to emphasize the unity between our approaches to shifting an exterior and interior penalty function (in this case the quadratic penalty function and the logarithmic barrier function, respectively). Moreover this coherency is a *strength* of the approach allowing us to combine both methods in a straightforward way to handle mixtures of inequality and equality constraints, see, for example, Section 8.2. However, if the overall framework for our development is close to that of Conn *et al.* (1991), substantial differences appear in the proofs of the results.

2 Notation

In this section we introduce the notation that will be used throughout the paper.

2.1 Derivatives

Let $g(x)$ denote the gradient, $\nabla_x f(x)$, of $f(x)$ and let $H(x)$ denote its Hessian matrix, $\nabla_{xx} f(x)$. Let $A(x)$ denote the m by n Jacobian of $c(x)$, where

$$c(x)^T \stackrel{\text{def}}{=} (c_1(x), ..., c_m(x)), \tag{2.1}$$

and let $H_i(x)$ denote the Hessian matrix, $\nabla_{xx} c_i(x)$, of $c_i(x)$. Finally, let $g^\ell(x,\lambda)$ and $H^\ell(x,\lambda)$ denote the gradient and Hessian matrix (taken with respect to x) of the Lagrangian function

$$\ell(x,\lambda) \stackrel{\text{def}}{=} f(x) - \sum_{i=1}^m \lambda_i c_i(x). \tag{2.2}$$

We note that $\ell(x,\lambda)$ is the Lagrangian function with respect to the general inequality constraints only.

2.2 Lagrange multiplier estimates

If we define *first-order Lagrange multiplier (dual variable) estimates* $\bar{\lambda}(x,\lambda,s)$ for which

$$\bar{\lambda}_i(x,\lambda,s) \stackrel{\text{def}}{=} \frac{\lambda_i s_i}{c_i(x)+s_i}, \tag{2.3}$$

we shall make much use of the identities

$$\begin{aligned} \nabla_x \Psi(x,\lambda,s) = & \; \nabla_x f(x) - \textstyle\sum_{i=1}^m \frac{\lambda_i s_i}{c_i(x)+s_i} \nabla_x c_i(x) \\ = & \; \nabla_x f(x) - A(x)^T \bar{\lambda}(x,\lambda,s) \\ = & \; g^\ell(x, \bar{\lambda}(x,\lambda,s)) \end{aligned} \tag{2.4}$$

and

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

2.3 Shorthand

Now suppose that $\{x_k \in \mathcal{B}\}$, $\{\lambda_k > 0\}$ and $\{s_k > 0\}$ are infinite sequences of n-vectors, m-vectors and m-vectors respectively. For any function F, we shall use the notation that F_k denotes F evaluated with arguments x_k, λ_k or s_k as appropriate. So, for instance, using the identity (2.4), we have

$$\nabla_x \Psi_k = \nabla_x \Psi(x_k, \lambda_k, s_k) = g^\ell(x_k, \bar{\lambda}_k), \tag{2.6}$$

where we have written

$$\bar{\lambda}_k = \bar{\lambda}(x_k, \lambda_k, s_k). \tag{2.7}$$

If x_* is a limit point of $\{x_k \in \mathcal{B}\}$, we shall write F_* as a shorthand for the quantity F evaluated with argument x_*. We will also sometimes (as we have already done) write F as shorthand for $F(x)$. If we wish to consider the i-th component of, for example, s_k we will write $s_{k,i}$

If r is any m-vector whose i-th component is r_i, we sometimes write $r \equiv [r_i]_{i=1}^m$. 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, $i \in \mathcal{J}$. We denote any vector norm (or its subordinate matrix norm) by $\|.\|$. Consequently, $\|[r_i]_{i=1}^m\| \equiv \|r\|$.

We will use the notation that if $\mathcal{J}_1$ and $\mathcal{J}_2$ are any subsets of integer indices 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 *rows* of A indexed by $\mathcal{J}_1$.

2.4 A projection operator

We will use the projection operator, defined component-wise by

$$(P[x])_i \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } x_i \le 0 \quad \text{and} \quad i \in \mathcal{N}_b \\ x_i & \text{otherwise.} \end{cases} \tag{2.8}$$

This operator projects the point x onto the region $\mathcal{B}$. Furthermore, we will make use of the 'projection' (which is essentially a translation by v of $P[x]$)

$$P(x, v) \stackrel{\text{def}}{=} x - P[x - v]. \tag{2.9}$$

2.5 Dominated and floating variables

For any $x_k \in \mathcal{B}$, there are two possibilities for each component $x_{k,i}$, namely

$$\begin{array}{lll} \text{(i)} & i \in \mathcal{N}_b \quad \text{and} \quad 0 \le x_{k,i} \le \nabla_x \Psi_{k,i}, & \text{or} \\ \text{(ii)} & i \in \mathcal{N}_f \quad \text{or} \quad \nabla_x \Psi_{k,i} < x_{k,i}, & \end{array} \tag{2.10}$$

where $\mathcal{N}_f \stackrel{\text{def}}{=} \mathcal{N} \backslash \mathcal{N}_b$ is the index set of *free* variables. In case (i) we then have

$$P(x_k, \nabla_x \Psi_k)_i = x_{k,i}, \tag{2.11}$$

whereas in case (ii) we have

$$P(x_k, \nabla_x \Psi_k)_i = \nabla_x \Psi_{k,i}. \tag{2.12}$$

We shall refer to an $x_{k,i}$ which satisfies (i) as a *dominated* variable; a variable which satisfies (ii) is known as a *floating* variable. The algorithm which we are about to develop constructs iterates which force $P(x_k, \nabla_x \Psi_k)$ to zero as k increases. The dominated variables are thus pushed to zero, while the floating variables are allowed to find their own levels.

If, in addition, there is a convergent subsequence $\{x_k\}, k \in \mathcal{K}$, with limit point x_*, we shall partition the set $\mathcal{N}$ into the following four subsets, relating to the two possibilities (i) and (ii) above and to the corresponding x_*:

$$\begin{array}{ll} \mathcal{D}_1 & \stackrel{\text{def}}{=} \{i \in \mathcal{N}_b \mid x_{k,i} \text{ is dominated for all } k \in \mathcal{K} \text{ sufficiently large}\}, \\ \mathcal{F}_1 & \stackrel{\text{def}}{=} \{i \in \mathcal{N}_b \mid x_{k,i} \text{ is floating for all } k \in \mathcal{K} \text{ sufficiently large and } x_{*,i} > 0\} \cup \mathcal{N}_f, \\ \mathcal{F}_2 & \stackrel{\text{def}}{=} \{i \in \mathcal{N}_b \mid x_{k,i} \text{ is floating for all } k \in \mathcal{K} \text{ sufficiently large but } x_{*,i} = 0\} \text{ and} \\ \mathcal{F}_3 & \stackrel{\text{def}}{=} \mathcal{N} \setminus (\mathcal{D}_1 \bigcup \mathcal{F}_1 \bigcup \mathcal{F}_2). \end{array} \tag{2.13}$$

From time to time we will slightly abuse notation by saying that a variable x_i belongs to (for instance) $\mathcal{F}_1$, when strictly we should say that the index of the variable belongs to $\mathcal{F}_1$. We will also mention the components of a (given) vector in the set $\mathcal{F}_1$ when strictly we mean the components of the vector whose indices lie in $\mathcal{F}_1$.

If the iterates are chosen so that $P(x_k, \nabla_x \Psi_k)$ approaches zero as k increases, we have the following analog of Conn *et al.* (1991, Lemma 2.1).

Lemma 2.1 *Suppose that* $\{x_k\}, k \in \mathcal{K}$, *is a convergent subsequence with limit point* x_*, *that* λ_k, s_k, $\mathcal{D}_1$, $\mathcal{F}_1$, $\mathcal{F}_2$ *and* $\mathcal{F}_3$ *are as above and that* $P(x_k, \nabla_x \Psi_k)$ *approaches zero as* $k \in \mathcal{K}$ *increases. Then*

(i) the variables in the sets $\mathcal{D}_1$, $\mathcal{F}_2$ *and* $\mathcal{F}_3$ *all converge to their bounds;*

(ii) the components of $\nabla_x \Psi_{k,i}$ *in the sets* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *converge to zero; and*

(iii) if a component of $\nabla_x \Psi_{k,i}$ *in the set* $\mathcal{F}_3$ *converges to a finite limit, the limit is zero.*

Proof. (i) The result is true for variables in $\mathcal{D}_1$ from (2.11) for those in $\mathcal{F}_2$ by definition and for those in $\mathcal{F}_3$ as, again from (2.11), there must be a subsequence of the $k \in \mathcal{K}$ for which $x_{k,i}$ converges to zero.
(ii) The result follows for i in $\mathcal{F}_1$ and $\mathcal{F}_2$, from (2.12).
(iii) This is true for i in $\mathcal{F}_3$ as there must be a subsequence of the $k \in \mathcal{K}$ for which, from (2.12), $\nabla_x \Psi_{k,i}$ converges to zero. ■

It will sometimes be convenient to group the variables in sets $\mathcal{F}_2$ and $\mathcal{F}_3$ together and call the resulting set

$$\mathcal{F}_4 \stackrel{\text{def}}{=} \mathcal{F}_2 \bigcup \mathcal{F}_3. \tag{2.14}$$

As we see from Lemma 2.1, $\mathcal{F}_4$ gives variables which lie on their bounds at the solution and which may correspond to zero components of the gradient of the Lagrangian barrier function. These variables are potentially (dual) degenerate at the solution of the nonlinear programming problem.

2.6 Inactive and active constraints

As well as being concerned with which variables are fixed to, and which free from, their bounds at a limit point of a generated sequence $\{x_k\}$, we are also interested in knowing which of the nonlinear constraints (1.16) are *inactive* (strictly satisfied), and which are *active* (violated or just satisfied), at such a point. We define

$$\begin{array}{rcl} \mathcal{I}(x) & \stackrel{\text{def}}{=} & \{i \mid c_i(x) > 0\}, \\ \mathcal{A}(x) & \stackrel{\text{def}}{=} & \{i \mid c_i(x) \leq 0\}. \end{array} \tag{2.15}$$

We intend to develop our algorithm so that the set $\mathcal{A}^* \equiv \mathcal{A}(x_*)$ at any limit point of our generated sequence is precisely the index set of the set of constraints for which $c_{*,i} = 0$.

2.7 Kuhn-Tucker points

A point x_* is said to be a *Kuhn-Tucker* (first-order stationary) point for the problem (1.1)–(1.3) if there is an associated vector of Lagrange multipliers λ_* for which the *Kuhn-Tucker conditions*,

$$\begin{array}{c} x_{*,\mathcal{N}_b} \geq 0, \quad g^\ell_{*,\mathcal{N}_b} \geq 0, \quad c_* \geq 0, \quad \lambda_* \geq 0, \\ \\ g^\ell_{*,\mathcal{N}_f} = 0, \quad x_*^T g^\ell_* = 0 \text{ and } c_*^T \lambda_* = 0, \end{array} \tag{2.16}$$

hold. Under a suitable constraint qualification, these conditions are necessary if x_* is to solve (1.1)–(1.3) (see, for example, Fletcher, 1987, Theorem 9.1.1).

At any point x and for any scalar ω, we define the set

$$\mathcal{L}(x,\omega;x_*,\mathcal{F}) \stackrel{\text{def}}{=} \{\lambda_{\mathcal{A}^*} | \, \lambda_{\mathcal{A}^*} \geq 0 \quad \text{and} \quad \|(g(x) - A(x)^T_{\mathcal{A}^*}\lambda_{\mathcal{A}^*})_{\mathcal{F}}\| \leq \omega\} \tag{2.17}$$

relative to the point x_* and set $\mathcal{F} \subseteq \mathcal{N}$. Our intention is to construct a sequence $\{x_k\}$ so that for a specific $\mathcal{F}$ (the index set for floating variables x_i), $\mathcal{L}(x_k,\bar{\omega}_k;x_*,\mathcal{F})$ is non-empty for some $\bar{\omega}_k$ converging to zero. Under a suitable boundedness assumption, this will then ensure that the Kuhn-Tucker conditions are satisfied at all limit points of $\{x_k\}$.

We are now in a position to describe the algorithm we propose to use in more detail.

3 The algorithm

3.1 Statement of the algorithm

In order to solve problem (1.15)–(1.17), we consider the following general algorithmic framework.

Algorithm 3.1 [**General Outer Iteration Algorithm**]

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

$$\eta_s,\ \omega_s,\ \alpha_\omega,\ \beta_\omega,\ \alpha_\eta,\ \beta_\eta,\ \alpha_\lambda \leq 1,\ \tau < 1,\ \rho < 1,\ \omega_* \ll 1 \text{ and } \eta_* \ll 1 \tag{3.1}$$

for which

$$\alpha_\eta + (1+\alpha_\lambda)^{-1} > 1 \tag{3.2}$$

are specified. A positive penalty parameter, $\mu_0 < 1$, is given. Set

$$\omega_0 = \omega_s \mu_0^{\alpha_\omega} \quad \text{and} \quad \eta_0 = \eta_s \mu_0^{\alpha_\eta}. \tag{3.3}$$

An initial estimate of the solution, $x_{-1} \in B$, and vector of positive Lagrange multiplier estimates, λ_0, for which $c_i(x_{-1}) + \mu_0 \lambda_{0,i}^{\alpha_\lambda} > 0$ are specified. Set $k = 0$.

step 1 : [Inner iteration] *Compute shifts*

$$s_{k,i} = \mu_k \lambda_{k,i}^{\alpha_\lambda}, \tag{3.4}$$

for $i = 1, ..., m$. Find $x_k \in B$ such that

$$\|P(x_k, \nabla_x \Psi_k)\| \leq \omega_k \tag{3.5}$$

and

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

step 2 : [Test for convergence] *If*

$$\|P(x_k, \nabla_x \Psi_k)\| \le \omega_* \text{ and } \|[c_i(x_k)\bar{\lambda}_i(x_k, \lambda_k, s_k)]_{i=1}^m\| \le \eta_*, \tag{3.7}$$

stop. If

$$\left\| \left[\frac{c_i(x_k)\bar{\lambda}_i(x_k, \lambda_k, s_k)}{\lambda_{k,i}^{\alpha_\lambda}} \right]_{i=1}^m \right\| \le \eta_k, \tag{3.8}$$

execute step 3. Otherwise, execute step 4.

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

$$\begin{aligned} \lambda_{k+1} &= \bar{\lambda}(x_k, \lambda_k, s_k), \\ \mu_{k+1} &= \mu_k, \\ \omega_{k+1} &= \omega_k \mu_{k+1}^{\beta_\omega}, \\ \eta_{k+1} &= \eta_k \mu_{k+1}^{\beta_\eta}. \end{aligned} \tag{3.9}$$

Increase k by one and go to step 1.

step 4 : [Reduce the penalty parameter] *Set*

$$\begin{aligned} \lambda_{k+1} &= \lambda_k, \\ \mu_{k+1} &= \tau \mu_k, \\ \omega_{k+1} &= \omega_s \mu_{k+1}^{\alpha_\omega}, \\ \eta_{k+1} &= \eta_s \mu_{k+1}^{\alpha_\eta}. \end{aligned} \tag{3.10}$$

Increase k by one and go to step 1.

end of Algorithm 3.1

Although it might appear quite complicated, the idea behind Algorithm 3.1 is rather simple. We wish the algorithm to converge to a point for which the Kuhn-Tucker conditions (2.16) are satisfied. The whole algorithm is driven by the value of the *penalty* parameter, μ_k. The inner-iteration convergence test (3.5) is intended to ensure that these conditions hold at any limit point. The algorithm is designed to be locally convergent if the penalty parameter is fixed at a sufficiently small value and the Lagrange multipliers estimates are updated using the first-order formula (2.3). As a last resort, we can guarantee that the penalty parameter is sufficiently small by driving it to zero while at the same time ensuring that the Lagrange multiplier estimates are well behaved. The test (3.8) is merely to detect when the penalty parameter is small enough for us to move from a globally convergent to a locally convergent regime. The remaining details of the algorithm are concerned with picking two sequences of tolerances, $\{\omega_k\}$ to limit the accuracy required of the inner-iteration algorithm and $\{\eta_k\}$ to measure whether we have entered the asymptotic phase of the calculation. The exact relationship between the two sequences is designed to allow a complete analysis of the algorithm.

3.2 Starting points

Before we analyse Algorithm 3.1, we need to comment on the crucial Step 1 in the algorithm. One might reasonably expect to try to satisfy the convergence test (3.5) by (approximately) minimizing (1.4) within (1.17). However, this relies on ensuring that $c(x) + s_k > 0$ for all iterates generated during the inner iteration. In particular, it is essential from a practical point of view that this condition is satisfied at the starting point for the inner iteration. In one important case, this is trivially so. For we have,

Lemma 3.1 *The iterates generated by Algorithm 3.1 satisfy the condition*

$$c_i(x_k) + s_{k+1,i} > 0, \text{ for } i = 1, ..., m, \quad x_k \in \mathcal{B} \tag{3.11}$$

for $k = -1$ and all iterations $k \geq 0$ for which (3.8) is satisfied.

Proof. The result is true for $k = -1$ by choice of the initial Lagrange multiplier estimates and shifts in Steps 0 and 1 of the algorithm.

The k-th inner iteration (Step 1) of the algorithm ensures that (3.6) is satisfied. If (3.8) is satisfied, the updates (3.4) and (3.9) apply. For each constraint, there are two possibilities. If $c_i(x_k) > 0$, (3.11) follows immediately, as the algorithm ensures that the shifts are always positive. If, on the other hand, $c_i(x_k) \leq 0$,

$$\frac{s_{k,i}}{c_i(x_k) + s_{k,i}} \geq 1. \tag{3.12}$$

In this case, the definition (2.3) of the multiplier update ensures that

$$\lambda_{k+1,i} = \bar{\lambda}_i(x_k, \lambda_k, s_k) \geq \lambda_{k,i}. \tag{3.13}$$

Hence $s_{k+1,i} \geq s_{k,i}$ follows from (3.4) and (3.9), and thus (3.6) gives (3.11). ■

Thus, so long as we are able to update the multiplier estimates rather than reducing the penalty parameter, the terminating iterate from one inner iteration gives a suitable starting point for the next. We shall consider what to do in other cases in due course.

In any case, it is useful to try to start a new inner iteration from, or close to, the solution of the last one, as function and derivative information from the conclusion of one inner iteration may be passed as input to the next. This is not specific to the algorithm discussed in this paper, but also applies, for instance, to augmented Lagrangian methods, as discussed by Conn *et al.* (1991) and Conn *et al.* (1992c). We also note that a result similar to that of the latter reference can also be shown, under certain additional assumptions, for the algorithm considered here, namely that (3.8) is eventually satisfied at each outer iteration, and that a single Newton-like inner iteration is guaranteed to yield a point satisfying both (3.5) and (3.6) for large enough k. The details of this analysis are presented by Conn *et al.* (1992d).

3.3 The inner iteration

In order to satisfy the inner-iteration termination test (3.5), one may in theory apply any algorithm for solving *simple-bound constrained minimization* problems — problems in which the minimizer of an objective function within a region defined by simple bounds on the variables is sought — to the problem of minimizing (1.4) within (1.17). Indeed, as the condition

$$P(x, \nabla_x \Psi(x, \lambda_k, s_k)) = 0 \tag{3.14}$$

is required at optimality for such a problem, (3.5) can be viewed as an inexact stopping rule for such iterative algorithms. We merely mention here that the projected gradient methods of Calamai and Moré (1987), Burke and Moré (1988), Conn *et al.* (1988a), Conn *et al.* (1988b) and Burke *et al.* (1990) and the interior point method of Nash and Sofer (1993) are all appropriate, but that methods which take special account of the nature of (1.4) may yet be preferred.

3.4 Further discussion

We should also comment on the rather peculiar test (3.8) in Algorithm 3.1. In our previous work on solving problems with general equality constraints $c_i(x) = 0$, $i = 1, ..., m$ (see, Conn *et al.*, 1991), we measure the success or failure of an outer iterate x_k by the size of the norm of the constraint violation

$$||c_k|| \equiv ||[c_i(x_k)]_{i=1}^m|| \, . \tag{3.15}$$

Specifically, we ask whether

$$||c_k|| \leq \eta_k, \tag{3.16}$$

for some convergence tolerance η_k (see Conn *et al.*, 1991, Test (3.6)). In the current algorithm, we employ a similar test. As one would not expect all of the general inequality constraint functions to converge to zero for the problem under consideration in this paper, the test (3.16) is inappropriate. However, one would expect the *complementary slacknesses* $c_i(x)\lambda_i$, $i = 1, ..., m$, to converge to zero for suitable Lagrange multiplier estimates λ_i. The test (3.8) is designed with this in mind.

In fact, there is a stronger similarity between Conn *et al.* (1991, Test (3.6)) and (3.8) than is directly apparent. For the former test may be rewritten as

$$||\bar{\lambda}_k - \lambda_k|| \leq \eta_k / \mu_k, \tag{3.17}$$

using the first-order multiplier update proposed in Conn *et al.*, 1991. The test (3.8) may likewise be written as (3.17) because of the definition of the multiplier estimates (2.3) and shifts (3.4).

Our primary aim is now to analyse the convergence behaviour of Algorithm 3.1.

4 Global convergence analysis

In this section, we shall consider the convergence of Algorithm 3.1 from arbitrary starting points. We aim to show that all finite limit points of the iterates generated by the algorithm are Kuhn-Tucker points. We shall analyse the convergence of Algorithm 3.1 in the case where the convergence tolerances ω_* and η_* are both zero.

We shall make use of the the following assumption.

AS2: Suppose that the subsequence $\{x_k\}, k \in \mathcal{K}$, generated by Algorithm 3.1, converges to x_*, and that $\mathcal{F}_1$ is defined by (2.13). Then we assume that the set $\mathcal{L}(x_*, 0; x_*, \mathcal{F}_1)$ is bounded.

Note that AS2 excludes the possibility that $\mathcal{F}_1$ is empty unless there are no general constraints active at x_*. In view of Lemma 2.1, this seems reasonable as otherwise we are allowing the possibility that there are more than n active constraints at x_*.

As a consequence of AS2 we have the following result that essentially says that a perturbation of a bounded set of optimal dual variables for the problem remains bounded.

Lemma 4.1 *Suppose that AS2 holds. Then $\mathcal{L}(x, \omega; x_*, \mathcal{F}_1)$ is bounded for all (x, ω) sufficiently close to $(x_*, 0)$.*

Proof. The result follows directly from the analysis given by Fiacco (1983, Theorem 2.2.9).
We now give our most general global convergence result. This is in the spirit of Conn *et al.* (1991, Theorem 4.4).

Theorem 4.2 *Suppose that AS1 holds. Let $\{x_k\} \in B, k \in \mathcal{K}$, be any sequence generated by Algorithm 3.1 which converges to the point x_* and for which AS2 holds. Then*

(i) x_* is a Kuhn-Tucker (first-order stationary) point for the problem (1.15)–(1.17).

(ii) The sequence $\{\bar{\lambda}_k\}$ remains bounded for $k \in \mathcal{K}$ and any limit point of this sequence is a set of Lagrange multipliers λ_* corresponding to the Kuhn-Tucker point at x_*.

(iii) the gradients $\nabla_x \Psi_k$ converge to g_*^ℓ for $k \in \mathcal{K}$, and

(iv) the Lagrange multiplier estimates $\bar{\lambda}(x_k, \lambda_k, s_k)_i$ satisfy

$$\bar{\lambda}(x_k, \lambda_k, s_k)_i = \sigma_{k,i} \lambda_{k,i}, \tag{4.1}$$

where $\sigma_{k,i}$ converges to zero for all $i \in \mathcal{I}^*$ as $k \in \mathcal{K}$ tends to infinity.

Proof. See Appendix B, Section B.2.
We remark that the proof of Theorem 4.2 remains true regardless of the actual choice of $\{\omega_k\}$ provided the sequence converges to zero. We also note that the method of proof given here requires that $\alpha_\lambda > 0$ and, unfortunately, therefore excludes the modified barrier function method of Polyak (1992).

Significantly, the superlinear convergence of the Lagrange multiplier estimates $\bar{\lambda}_k$ for inactive constraints in Theorem 4.2, (iv), is a consequence of our shift strategy. Only a linear convergence rate is proved by Polyak (1992) for his modified barrier method and Powell (1992b) indicates that this can be a disadvantage when inactive constraints are relatively small at a limit point. We note that the modified barrier method of Jittorntrum and Osborne (1980) also exhibits the same superlinear convergence property without benefiting from the advantages associated with shifts, given in the next section.

Now suppose we replace AS2 by the following stronger assumption:

AS3: For any convergent subsequence of the sequence $\{x_k\}$ with limit point x_*, say, the matrix $A_{*,\mathcal{A}^*,\mathcal{F}_1}$ is of full rank for the corresponding set $\mathcal{F}_1$ defined by (2.13).

Furthermore, we define the *least-squares Lagrange multiplier estimates* (corresponding to the sets $\mathcal{F}_1$ and $\mathcal{A}^*$)

$$\lambda(x)_{\mathcal{A}^*} \stackrel{\text{def}}{=} -(A(x)^+_{\mathcal{A}^*,\mathcal{F}_1})^T g(x)_{\mathcal{F}_1} \tag{4.2}$$

at all points where the right generalized inverse

$$A(x)^+_{\mathcal{A}^*,\mathcal{F}_1} \stackrel{\text{def}}{=} A(x)^T_{\mathcal{A}^*,\mathcal{F}_1}(A(x)_{\mathcal{A}^*,\mathcal{F}_1} A(x)^T_{\mathcal{A}^*,\mathcal{F}_1})^{-1} \tag{4.3}$$

of $A(x)_{\mathcal{A}^*,\mathcal{F}_1}$ is well defined. We note that these estimates are differentiable functions of x whenever $A(x)_{\mathcal{A}^*,\mathcal{F}_1}$ is of full rank (see, for example, Conn *et al.*, 1991, Lemma 2.2).

With this stronger assumption and definition, we are able to obtain error bounds on dual variables for active constraints and deviations on complementary slackness which depend upon the errors in the primal variables and Lagrange multipliers for inactive constraints. This result has the same flavour as Conn *et al.* (1991, Lemma 4.3).

Theorem 4.3 *Suppose that the assumptions of Theorem 4.2 hold excepting that AS2 is replaced by AS3. Then the conclusions of Theorem 4.2 remain true and, in addition, we have that*

(v) the vector of Lagrange multipliers λ_* corresponding to the Kuhn-Tucker point at x_* are unique, and

(vi) there are positive constants a_1, a_2, a_3 and an integer k_0 such that

$$\|(\bar{\lambda}(x_k, \lambda_k, s_k) - \lambda_*)_{\mathcal{A}^*}\| \leq a_1\omega_k + a_2\|x_k - x_*\| + a_3\sigma_k\|\lambda_{k,\mathcal{I}^*}\|, \tag{4.4}$$

$$\|(\lambda(x_k) - \lambda_*)_{\mathcal{A}^*}\| \leq a_2\|x_k - x_*\|, \tag{4.5}$$

$$\begin{aligned}\left\|\left[c_i(x_k)\bar{\lambda}_{k,i}/\lambda_{k,i}^{\alpha_\lambda}\right]_{i=1}^m\right\| \leq \;& \mu_k\left[a_1\omega_k + a_2\|x_k - x_*\| + \right.\\ & \left.(1 + \sigma_k(1 + a_3))\|\lambda_{k,\mathcal{I}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\|\right]\end{aligned} \tag{4.6}$$

and

$$\begin{aligned}\|c_{k,\mathcal{A}}\| \leq \; \mu_k\left\|\left[\lambda_{k,i}^{\alpha_\lambda}/\bar{\lambda}_{k,i}\right]_{i\in\mathcal{A}}\right\| \;& \left[a_1\omega_k + a_2\|x_k - x_*\| + \right.\\ & \left. a_3\sigma_k\|\lambda_{k,\mathcal{I}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}}\|\right]\end{aligned} \tag{4.7}$$

for all $k \geq k_0, (k \in \mathcal{K})$, and any subset $\mathcal{A} \subseteq \mathcal{A}^*$ and where

$$\sigma_k \stackrel{\text{def}}{=} \max_{i\in\mathcal{I}^*} \sigma_{k,i} \tag{4.8}$$

converges to zero as $k \in \mathcal{K}$ tends to infinity.

Proof. See Appendix B, Section B.3.

5 Asymptotic convergence analysis

We now give our first rate-of-convergence result, which is in the spirit of Conn *et al.* (1991, Lemma 5.1). As a preliminary, we need to make two additional (second-order) assumptions.

AS4: The second derivatives of the functions $f(x)$ and the c_i are Lipschitz continuous at all points within an open set containing $\mathcal{B}$.

AS5: Suppose that (x_*, λ_*) is a Kuhn-Tucker point for problem (1.15)–(1.17) and that

$$\begin{aligned} \mathcal{A}_1^* &\stackrel{\text{def}}{=} \{i \mid c_{*,i} = 0 \quad \text{and} \quad \lambda_{*,i} > 0\} \\ \mathcal{A}_2^* &\stackrel{\text{def}}{=} \{i \mid c_{*,i} = 0 \quad \text{and} \quad \lambda_{*,i} = 0\} \end{aligned} \tag{5.1}$$

and

$$\begin{aligned} \mathcal{J}_1 &\stackrel{\text{def}}{=} \{i \in \mathcal{N}_b \mid g^{\ell}_{*,i} = 0 \quad \text{and} \quad x_{*,i} > 0\} \cup \mathcal{N}_f \\ \mathcal{J}_2 &\stackrel{\text{def}}{=} \{i \in \mathcal{N}_b \mid g^{\ell}_{*,i} = 0 \quad \text{and} \quad x_{*,i} = 0\}. \end{aligned} \tag{5.2}$$

Then we assume that the matrix

$$\begin{pmatrix} H^{\ell}_{*,\mathcal{J},\mathcal{J}} & A^T_{*,\mathcal{A},\mathcal{J}} \\ A_{*,\mathcal{A},\mathcal{J}} & 0 \end{pmatrix} \tag{5.3}$$

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 $\mathcal{J}_2$.

We note that assumption AS5 implies AS3.

It is inconvenient that the estimates (4.4)-(4.6) depend upon the error in the primal variables, $x_k - x_*$, as this term, unlike the other terms in the estimates, depends on *a posteriori* information. The next lemma removes this dependence and gives a result similar to the previous theory in which the errors in x are bounded by the errors in the dual variables $\lambda_k - \lambda_*$ (see Polyak, 1992, Theorem 1). However, as an inexact minimization of the Lagrangian barrier function is made, a term reflecting this is also present in the bound. Once again, the result allows for our handling of simple bound constraints.

Lemma 5.1 *Suppose that AS1 holds. Let $\{x_k\} \in B, k \in \mathcal{K}$, be any sequence generated by Algorithm 3.1 which converges to the point x_* for which AS5 holds. Let λ_* be the corresponding vector of Lagrange multipliers. Furthermore, suppose that AS4 holds and that the condition*

$$\left\| \left[\lambda_{k,i}^{\alpha_\lambda} / \bar{\lambda}_{k,i} \right]_{i \in \mathcal{A}_1^*} \right\| \leq a_4 \mu_k^{\zeta - 1} \tag{5.4}$$

is satisfied for some strictly positive constants a_4 and ζ and all $k \in \mathcal{K}$. Let χ be any constant satisfying

$$0 < \chi \leq \zeta. \tag{5.5}$$

Then there are positive constants $\mu_{\max}$, $a_5, \cdots, a_{13}$, and an integer value k_0 so that, if $\mu_{k_0} \leq \mu_{\max}$,

$$\begin{aligned} \|x_k - x_*\| \leq \quad & a_5 \omega_k + a_6 \mu_k^{\chi} \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + \\ & a_7 \mu_k^{1-\chi} \left\| \left[\lambda_{k,i}^{\alpha_\lambda} \right]_{i \in \mathcal{A}_2^*} \right\| + a_8 \sigma_k \|\lambda_{k,\mathcal{I}^*}\|, \end{aligned} \tag{5.6}$$

$$\|(\bar{\lambda}(x_k, \lambda_k, s_k) - \lambda_*)_{\mathcal{A}^*}\| \leq \begin{array}{l} a_9\omega_k + a_{10}\mu_k^{\chi}\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + \\ a_{11}\mu_k^{1-\chi}\left\|\left[\lambda_{k,i}^{\alpha_\lambda}\right]_{i\in\mathcal{A}_2^*}\right\| + a_{12}\sigma_k\|\lambda_{k,\mathcal{I}^*}\|, \end{array} \tag{5.7}$$

and

$$\left\|\left[c_i(x_k)\bar{\lambda}_{k,i}/\lambda_{k,i}^{\alpha_\lambda}\right]_{i=1}^m\right\| \leq \mu_k\left[\begin{array}{l} a_9\omega_k + a_{13}\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + \\ a_{11}\mu_k^{1-\chi}\left\|\left[\lambda_{k,i}^{\alpha_\lambda}\right]_{i\in\mathcal{A}_2^*}\right\| + (1 + (1 + a_{12})\sigma_k)\|\lambda_{k,\mathcal{I}^*}\| \end{array}\right], \tag{5.8}$$

for all $k \geq k_0, (k \in \mathcal{K})$ and where the scalar σ_k, as defined by (4.8), converges to zero as $k \in \mathcal{K}$ tends to infinity.

Proof. See Appendix B, Section C.1.

In order for Lemma 5.1 to be useful, we need to ensure that the condition (5.4) holds. There is at least one case where this is automatic. Thus, we consider the following additional assumption.

AS6: The iterates $\{x_k\}$ generated by Algorithm 3.1 have a single limit point x_*.

We then have:

Lemma 5.2 *Suppose that AS1 holds and that the iterates $\{x_k\}$ generated by Algorithm 3.1 satisfy AS6 and converges to the point x_* for which AS3 holds. Let λ_* be the corresponding vector of Lagrange multipliers. Now we know that (i) $\{\lambda_k\}$ converges to λ_*, (ii)*

$$\sigma_k \leq \mu_k\theta_k, \tag{5.9}$$

where θ_k converges to zero, as k increases, and (iii) inequality (5.4) is satisfied for all k. Moreover, if AS4 and AS5 replace AS3, (iv) the conclusions of Lemma 5.1 hold for all k, and any $0 < \chi \leq 1$.

Proof. See Appendix C, Section C.2.

We now give our second major result of the paper. Namely, under suitable assumptions, the penalty parameter will be bounded away from zero in Algorithm 3.1. This is important as many methods for solving the inner iteration subproblem will encounter difficulties if the current iterate is allowed to approach the boundary of the shifted constraints $c(x) + s_k \geq 0$. These manifest themselves through the increasing ill-conditioning of the Hessian of the Lagrangian barrier function and the subsequent difficulty of performing the inner-iteration. These problems will certainly arise in the neighbourhood of a first-order stationary point if μ_k converges to zero. The result is analogous to Theorem 5.3 of Conn *et al.* (1991).

We need to consider the following extra assumption.

AS7: (Strict complementary slackness condition 1) Suppose that (x_*, λ_*) is a Kuhn-Tucker point for problem (1.15)–(1.17). Then

$$\mathcal{A}_2^* = \{i \mid c_{*,i} = 0 \quad \text{and} \quad \lambda_{*,i} = 0\} = \emptyset. \tag{5.10}$$

Theorem 5.3 *Suppose that the iterates $\{x_k\}$ generated by Algorithm 1 of Section 3 satisfy AS6 and that AS4 and AS5 hold. Furthermore, suppose that either*

(i) $\alpha_\lambda = 1$ hold and we define

$$\alpha \stackrel{\text{def}}{=} \min(\tfrac{1}{2}, \alpha_\omega) \quad \textit{and} \quad \beta \stackrel{\text{def}}{=} \min(\tfrac{1}{2}, \beta_\omega) \tag{5.11}$$

or

(ii) AS7 holds and we define

$$\alpha \stackrel{\text{def}}{=} \min(1, \alpha_\omega) \quad \textit{and} \quad \beta \stackrel{\text{def}}{=} \min(1, \beta_\omega). \tag{5.12}$$

Then, whenever α_η and β_η satisfy the conditions

$$\alpha_\eta < \min(1, \alpha_\omega) \tag{5.13}$$

$$\beta_\eta < \beta \tag{5.14}$$

and

$$\alpha_\eta + \beta_\eta < \alpha + 1, \tag{5.15}$$

there is a constant $\mu_{\min} > 0$ such that $\mu_k \geq \mu_{\min}$ for all k.

Proof. See Appendix C, Section C.3.

It is unclear how Algorithm 3.1 behaves when $\alpha_\lambda < 1$, in the absence of AS7. The inequalities from Lemma 5.1 appear not to be strong enough to guarantee at least a linear improvement in the error of the Lagrange multiplier estimates λ_k because of the presence of the term $a_{11}\mu_k^{1-\chi}\|[\lambda_{k,i}^{\alpha_\lambda}]_{i\in\mathcal{A}_2^*}\|$ in the bound (5.7).

We should also point out that it is indeed possible to find values α_ω, α_η, β_ω and β_η which satisfy the requirements (3.2), (5.13), (5.14) and (5.15) for any $0 < \alpha_\lambda \leq 1$. For instance, the values $\alpha_\omega = 1$, $\alpha_\eta = 0.75$, $\beta_\omega = 1$ and $\beta_\eta = 0.25$ suffice.

We caution the reader that, although the result of Theorem 5.3 is an important ingredient in overcoming the numerical difficulties normally associated with barrier function methods, ensuring that μ_k is bounded away from zero is not sufficient. The numerical difficulties arise because of the singularity of the barrier function when $c_i(x) + s_{k,i} = 0$ for any $1 \leq i \leq m$. The algorithm is designed so that $c_i(x) + s_{k,i} > 0$ for all $1 \leq i \leq m$. If, in addition, AS7 holds, Theorem 5.3 ensures that limit

$$\lim_{x\to x_*, k\to\infty} c_i(x) + s_{k,i} = c_{*,i} + \mu_{\min}\lambda_{*,i}^{\alpha_\lambda} > 0 \tag{5.16}$$

for all $1 \leq i \leq m$, and thus numerical difficulties will not arise as the limit is approached. In the absence of AS7, $c_{*,i} + \mu_{\min}\lambda_{*,i}^{\alpha_\lambda} = 0$ for all $i \in \mathcal{A}_2^*$ and thus numerical problems are possible in a small neighbourhood of the limit.

We are also interested in the behaviour of Algorithm 3.1 in the case when the generated sequence of iterates has more than one limit point. We know that, under the assumptions of Theorem 4.2, each limit point will be a Kuhn-Tucker point. We show, by way of a concrete example that, in the absence of AS6, the conclusion of Theorem 5.3 is false.

Theorem 5.4 *There is a problem of the form (1.15)—(1.17), satisfying AS4, AS5 and AS7, for which Algorithm 3.1 of Section 3 generates a sequence of iterates $\{x_k\}$ with a pair of limit points, while the penalty parameter μ_k converges to zero as k tends to infinity.*

Proof. See Appendix C, Section C.4.

If we make the following additional assumption, our definition of floating variables completely characterizes the set of inactive, and hence active, bounds at a limit point.

AS8: (Strict complementary slackness condition 2) Suppose that (x_*, λ_*) is a Kuhn-Tucker point for problem (1.15)–(1.17). Then

$$\mathcal{J}_2 = \{i \in \mathcal{N}_b | g^\ell_{*,i} = 0 \quad \text{and} \quad x_{*,i} = 0\} = \emptyset. \tag{5.17}$$

We then have the following direct analog of Conn *et al.* (1991, Theorem 5.4).

Theorem 5.5 *Suppose that the iterates x_k, $k \in \mathcal{K}$, converge to the the limit point x_* with corresponding Lagrange multipliers λ_*, that AS1, AS2 and AS8 hold. Then for k sufficiently large, the set of floating variables are precisely those which lie away from their bounds, if present, at x_*.*

Proof. From Theorem 4.2, $\nabla_x \Psi_k$ converges to g^ℓ_* and from Lemma 2.1, the variables in the set $\mathcal{F}_4$ then converge to zero and the corresponding components of g^ℓ_* are zero. Hence, under AS8, $\mathcal{F}_4$ is null. Therefore, each variable ultimately remains tied to one of the sets $\mathcal{F}_1$ or $\mathcal{D}_1$ for all k sufficiently large; a variable in $\mathcal{F}_1$ is, by definition, floating and, whenever the variable is bounded, converges to a value away from its bound. Conversely, a variable in $\mathcal{D}_1$ is dominated and converges to its bound. ∎

We conclude the section by giving a rate-of-convergence result for our algorithm in the spirit of Conn *et al.* (1991, Theorem 5.5). For a comprehensive discussion of convergence, the reader is referred to Ortega and Rheinboldt (1970).

Theorem 5.6 *Suppose that the iterates $\{x_k\}$ generated by Algorithm 3.1 of Section 3 satisfy AS6, that AS1 and AS3 hold and that λ_* are the corresponding vector of Lagrange multipliers. Then, if (3.8) holds for all $k \geq k_0$,*

(i) *the Lagrange multiplier estimates for the inactive constraints, $\lambda_{k,\mathcal{I}^*}$, generated by Algorithm 3.1 converge Q-superlinearly to zero;*

(ii) *the Lagrange multiplier estimates for the active constraints, $\lambda_{k,\mathcal{A}^*}$, converge at least R-linearly to λ_*. The R-factor is at most $\mu_{\min}^{\beta_\eta}$, where $\mu_{\min}$ is the smallest value of the penalty parameter generated by the algorithm; and*

(iii) *if AS4 and AS5 replace AS3, x_k converges to x_* at least R-linearly, with R-factor at most $\mu_{\min}^{\min(1,\beta_\omega,\alpha_\lambda\beta_\eta)}$.*

Proof. See Appendix C, Section C.5.

As an immediate corollary we have

Corollary 5.7 *Under the assumptions of Theorem 5.3, the results of Theorem 5.6 follow, with the R-factor governing the convergence of* $\{x_k\}$ *being at most* $\mu_{\min}^{\alpha_\lambda \beta_\eta}$.

Proof. See Appendix C, Section C.6.

Note that the rate of convergence is effectively as fast as one can reasonably wish, because it is always possible to reduce the penalty parameter if the current rate is not judged fast enough and the conditioning of the unconstrained problem permits such a reduction.

6 Second-order conditions

It is useful to know how our algorithms behave if we impose further conditions on the iterates generated by the inner iteration. In particular, suppose that the sequence $\{x_k\}$ satisfies the following second-order sufficiency condition:

AS9: Suppose that the iterates x_k and Lagrange multiplier estimates $\bar{\lambda}_k$, generated by Algorithm 3.1, converge to the Kuhn-Tucker point (x_*, λ_*) for $k \in \mathcal{K}$ and that $\mathcal{J}_1$ and $\mathcal{J}_2$ are as defined by (5.2). Then we assume that $\nabla_{xx}\Psi_{k,\mathcal{J},\mathcal{J}}$ is uniformly positive definite (that is, its smallest eigenvalue is uniformly bounded away from zero) for all $k \in \mathcal{K}$ sufficiently large and all sets $\mathcal{J}$, where $\mathcal{J}$ is any set made up from the union of $\mathcal{J}_1$ and any subset of $\mathcal{J}_2$.

With such a condition we have the following result.

Theorem 6.1 *Under AS1, AS2, AS7 and AS9, the iterates* x_k, $k \in \mathcal{K}$, *generated by Algorithm 3.1 converge to an isolated local solution of (1.15)–(1.17).*

Proof. Let $\mathcal{J}$ be any set as described in AS9. Then

$$\nabla_{xx}\Psi_{k,\mathcal{J},\mathcal{J}} = H^{\ell}_{k,\mathcal{J},\mathcal{J}} + A^T_{k,\mathcal{A}^*,\mathcal{J}} D_{k,\mathcal{A}^*,\mathcal{A}^*} A_{k,\mathcal{A}^*,\mathcal{J}} + A^T_{k,\mathcal{I}^*,\mathcal{J}} D_{k,\mathcal{I}^*,\mathcal{I}^*} A_{k,\mathcal{I}^*,\mathcal{J}} \tag{6.1}$$

where D_k is a diagonal matrix with entries

$$D_{k,i,i} = \frac{\lambda_{k,i} s_{k,i}}{(c_i(x_k) + s_{k,i})^2} = \frac{\bar{\lambda}_{k,i}}{c_i(x_k) + s_{k,i}} \tag{6.2}$$

for $1 \leq i \leq m$. Let $s_{\mathcal{J}}$ be any non-zero vector satisfying

$$A_{k,\mathcal{A}^*,\mathcal{J}} s_{\mathcal{J}} = 0. \tag{6.3}$$

Then for any such vector,

$$s_{\mathcal{J}}^T \nabla_{xx}\Psi_{k,\mathcal{J},\mathcal{J}} s_{\mathcal{J}} \geq 2\epsilon s_{\mathcal{J}}^T s_{\mathcal{J}} \tag{6.4}$$

for some $\epsilon > 0$, under AS9. We note that the diagonal entries $D_{k,i,i}$, $i \in \mathcal{I}^*$, converge to zero. Hence, for k sufficiently large

$$s_{\mathcal{J}}^T A^T_{k,\mathcal{I}^*,\mathcal{J}} D_{k,\mathcal{I}^*,\mathcal{I}^*} A_{k,\mathcal{I}^*,\mathcal{J}} s_{\mathcal{J}} \leq \epsilon s_{\mathcal{J}}^T s_{\mathcal{J}} \tag{6.5}$$

and thus combining (6.1)–(6.5),

$$s_{\mathcal{J}}^T H^\ell(x_k, \bar{\lambda}_k)_{\mathcal{J},\mathcal{J}} s_{\mathcal{J}} \geq \epsilon s_{\mathcal{J}}^T s_{\mathcal{J}}. \tag{6.6}$$

By continuity of H as x_k and $\bar{\lambda}_k$ approach their limits, this gives that

$$s_{\mathcal{J}}^T H^\ell_{*,\mathcal{J},\mathcal{J}} s_{\mathcal{J}} \geq \epsilon s_{\mathcal{J}}^T s_{\mathcal{J}} \tag{6.7}$$

for all non-zero $s_{\mathcal{J}}$ satisfying

$$A_{*,\mathcal{A}^*,\mathcal{J}} s_{\mathcal{J}} = 0, \tag{6.8}$$

which, given AS7, implies that x_* is an isolated local solution to (1.15)–(1.17) (see, for example, Avriel, 1976, Theorem 3.11). ∎

We would be able to relax the reliance on AS7 in Theorem 6.1 if it were clear that the elements $D_{k,i,i}$, $i \in \mathcal{A}_2^*$, converged to zero for some subsequence of $\mathcal{K}$. However, it is not known if such a result holds in general.

We remark that AS9 may be ensured by tightening the inner iteration termination test (Step 2 of the algorithm) so that, in addition to (3.5), $\nabla_{xx}\Psi_{k,\mathcal{J},\mathcal{J}}$ is required to be uniformly positive definite, for all floating variables $\mathcal{J}$ and all k sufficiently large. If the strict complementary slackness condition AS8 holds at x_*, Theorem 5.5 ensures that the set $\mathcal{J}_2$ is empty and $\mathcal{J}_1$ identical to the set of floating variables after a finite number of iterations and thus, under this tighter termination test, AS9 and Theorem 6.1 holds.

There is a weaker version of this result, proved in the same way, that if the assumption of uniform positive-definiteness in AS9 is replaced by an assumption of positive semi-definiteness, the limit point then satisfies second-order necessary conditions (Avriel, 1976, Theorem 3.10) for a minimizer. This weaker version of AS9 is easier to ensure in practice as certain methods for solving the inner iteration subproblem, for instance that of Conn *et al.* (1988a), guarantee that the second derivative matrix at the limit point of a sequence of generated inner iterates will be positive semi-definite.

7 Feasible starting points

We now return to the issue raised in Section 3.2, namely, how to find a point for which

$$c(x) + s_{k+1} > 0 \quad \text{and} \quad x \in \mathcal{B} \tag{7.1}$$

from which to start the $k+1$-st inner iteration of Algorithm 3.1. We saw in Lemma 3.1 that this is trivial whenever (3.8) holds as the current estimate of the solution, x_k, satisfies (3.11). Furthermore, under the assumptions of Theorem 5.3, we know that (3.8) will hold for all sufficiently large k. The main difficulty we face is that, when (3.8) fails to hold, the updates (3.10) do not guarantee that (3.11) holds and thus we may need a different starting point for the $k+1$-st inner iteration.

There is, of course, one case where satisfying (7.1) is trivial. In certain circumstances, we may know of a *feasible point*, that is a point $x_{\rm feas}$ which satisfies (1.16) and (1.17). This may be because we have *a priori* knowledge of our problem, or because we encounter such a point as the algorithm progresses. Any feasible point automatically satisfies (7.1) as $s_{k+1} > 0$. One could start the $k+1$-st inner iteration from x_{feas} whenever (3.11) is violated.

There is, however, a disadvantage to this approach in that a "poor" feasible point may result in considerable expense when solving the inner-iteration subproblem. Ideally, one would like a feasible point "close" to x_k or x_* as there is then some likelihood that solving the inner-iteration will be inexpensive. It may, of course, be possible to find a "good" interpolatory point between x_k and x_{feas} satisfying (7.1). This could indeed be easy if the general constraints are linear. Finding a feasible starting point is also easy to do when the (shifted) feasible domain is such that projecting x_k onto it can be achieved at little cost, as is the case for simple bounds for example.

We consider the following alternative. Suppose that the k-th iteration of Algorithm 3.1 involves the execution of Step 4. Consider the *auxiliary* problem

$$\underset{x \in \Re^n, \xi \in \Re}{\text{minimize}} \quad \xi \tag{7.2}$$

subject to the constraints

$$c(x) + \xi s_{k+1} \geq 0, \quad \xi \geq 0, \quad x \in \mathcal{B}. \tag{7.3}$$

Then it follows that if we can find suitable values $x = \hat{x}$ and $\xi = \hat{\xi} < 1$ to satisfy (7.3), the same values $x = \hat{x}$ satisfy (7.1) and thus give an appropriate starting point for the $k+1$-st inner iteration. Furthermore, the problem (7.2)–(7.3) has a solution value zero if and only if the solution is a feasible point for the original constraint set (1.16)–(1.17). Thus we can guarantee that there are suitable values $x = \hat{x}$ and $\xi = \hat{\xi}$ whenever the original problem (1.15)–(1.17) has a solution.

Turning to the auxiliary problem (7.2)–(7.3), we first observe from (3.6) and (3.10) that the values $x = x_k$ and $\xi = \tau^{-1}$ give a feasible point for the constraint set (7.3). We may then solve (7.2)–(7.3) using a traditional barrier function or interior point method (see, for instance, Fiacco and McCormick, 1968, or Wright, 1992) or by a Lagrangian barrier function method such as that proposed in this paper.

If we attempt to solve (7.2)–(7.3) using a traditional barrier function / interior point method, we need not be overly concerned with the conditioning dangers often associated with these methods (see, for instance, Murray, 1971). For we only need an approximation to the solution for which $\xi = \hat{\xi} < 1$. Therefore, we can stop the minimization at the first point for which $\xi < 1$ and the method need never enter its potentially dangerous asymptotic phase.

If, on the other hand, we chose to solve the auxiliary problem using the algorithm given in Section 3, the presence of an initial feasible point for this problem means that we avoid the need to solve a further auxiliary point problem. The introduction of additional shifts means that it is less apparent how early to stop the minimization in order to satisfy (7.1) - the requirements (7.1) will have to be carefully monitored - but nonetheless early termination will still be possible.

The problem (7.2)–(7.3) involves one more variable, ξ, than the original problem (1.15)–(1.17). Thus the data structures for solving both problems may be effectively shared. There are alternatives to (7.2)–(7.3). For instance, if w is a vector of strictly positive weights, one might consider the auxiliary problem

$$\underset{x\in\Re^n, s\in\Re^m}{\text{minimize}} \quad w^T s \tag{7.4}$$

subject to the constraints

$$c(x) + s \geq 0, \quad s \geq 0, \quad x \in \mathcal{B} \tag{7.5}$$

and stop when $s < s_{k+1}$. Again, an initial feasible point is available for this problem but it now involves m additional variables, which is likely to add a significant overhead to the computational burden. Alternatively, if we partition $\{1, 2, \cdots, m\}$ into disjoint sets $\mathcal{C}_1$ and $\mathcal{C}_2$ for which

$$c_i(x_k) + s_{k+1,i} \leq 0 \quad i \in \mathcal{C}_1 \tag{7.6}$$

and

$$c_i(x_k) + s_{k+1,i} > 0 \quad i \in \mathcal{C}_2 \tag{7.7}$$

and let $0 < \hat{s}_{k+1,i} < s_{k+1,i}$ for $i \in \mathcal{C}_2$, we might consider the third alternative auxiliary problem

$$\underset{x\in\Re^n, s_i\in\Re}{\text{minimize}} \quad \sum_{i\in\mathcal{C}_1} w_i s_i \tag{7.8}$$

subject to the constraints

$$c_i(x) + s_i \geq 0, \quad s_i \geq 0, \quad i \in \mathcal{C}_1, \tag{7.9}$$

$$c_i(x) + \hat{s}_{k+1,i} \geq 0, \quad i \in \mathcal{C}_2 \tag{7.10}$$

and (1.17) and stop when $s_i < s_{k+1,i}$ for all $i \in \mathcal{C}_1$. Once again, an initial feasible point is available for this problem and this time the problem involves $|\mathcal{C}_1|$ additional variables. If $|\mathcal{C}_1|$ is small, solving (7.8)–(7.10) may be preferable to (7.2)–(7.3).

8 Further comments

8.1 The general problem

We now briefly turn to the more general problem (1.1)–(1.3). As we indicated in our introduction, the presence of the more general constraints (1.3) do not significantly alter the conclusions that we have drawn so far. If we define the appropriate generalization of the projection (2.8) by

$$(P[x])_i \stackrel{\text{def}}{=} \begin{cases} l_i & \text{if} \quad x_i \leq l_i \\ u_i & \text{if} \quad x_i \geq u_i \\ x_i & \text{otherwise} \end{cases} \tag{8.1}$$

and let $\mathcal{B} = \{x | \; l \leq x \leq u\}$, we may then use the algorithm of Section 3 without further significant modification. Our concept of floating and dominated variables stays essentially the same; for any iterate x_k in $\mathcal{B}$ we have three mutually exclusive possibilities for each component $x_{k,i}$, namely

$$\begin{array}{ll} \text{(i)} & 0 \leq x_{k,i} - l_i \leq \nabla_x \Psi_{k,i} \\ \text{(ii)} & \nabla_x \Psi_{k,i} \leq x_{k,i} - u_i \leq 0 \\ \text{(iii)} & x_{k,i} - u_i < \nabla_x \Psi_{k,i} < x_{k,i} - l_i. \end{array} \tag{8.2}$$

In case (i) we then have

$$P(x_k, \nabla_x \Psi_k)_i = x_{k,i} - l_i \tag{8.3}$$

whereas in case (ii) we have

$$P(x_k, \nabla_x \Psi_k)_i = x_{k,i} - u_i \tag{8.4}$$

and in case (iii)

$$P(x_k, \nabla_x \Psi_k)_i = \nabla_x \Psi_{k,i}. \tag{8.5}$$

The $x_{k,i}$ which satisfy (i) or (ii) are now the dominated variables (the ones satisfying (i) are said to be *dominated above* and those satisfying (ii) *dominated below*); those which satisfy (iii) are the floating variables. As a consequence, the sets corresponding to those given in (2.13) are straightforward to define. Now $\mathcal{F}_1$ contains variables which float for all $k \in \mathcal{K}$ sufficiently large and converge to the interior of $\mathcal{B}$. Furthermore $\mathcal{D}_1$ is the union of the two sets — $\mathcal{D}_{1l}$, made up of variables which are dominated above for all $k \in \mathcal{K}$ sufficiently large, and $\mathcal{D}_{1u}$, made up of variables which are dominated below for all $k \in \mathcal{K}$ sufficiently large. Likewise $\mathcal{F}_2$ is the union of the two sets $\mathcal{F}_{2l}$, made up of variables which are floating for all sufficiently large $k \in \mathcal{K}$ but converge to their lower bounds, and $\mathcal{F}_{2u}$, made up of variables which are floating for all sufficiently large $k \in \mathcal{K}$ but converge to their upper bounds. With such definitions, we may reprove all of the results of sections 3 to 7, assumptions AS5 and AS8 being extended in the obvious way and Theorem 5.5 being strengthened to say that, for all $k \in \mathcal{K}$ sufficiently large, $\mathcal{F}_{1l}$ and $\mathcal{F}_{1u}$ are precisely the variables which lie at their lower and upper bounds (respectively) at x_*.

8.2 Equality constraints

It may happen that we wish to solve a problem in which there are *equality* constraints

$$c_i(x) = 0 \quad m+1 \leq i \leq m_t \tag{8.6}$$

in addition to the constraints (1.2) and (1.3). In this case, we may construct a composite Lagrangian barrier/augmented Lagrangian function

$$\Theta(x, \lambda, s, \mu) = f(x) - \sum_{i=1}^{m} \lambda_i s_i \log(c_i(x) + s_i) + \sum_{i=m+1}^{m_t} \lambda_i c_i(x) + \frac{1}{2\mu} \sum_{i=m+1}^{m_t} c_i(x)^2 \tag{8.7}$$

and solve the general problem (1.1)-(1.3) and (8.6) by a sequential minimization of (8.7) within the region defined by (1.3).

The only change we need to make to the Algorithm 3.1 is to replace the test (3.8) by the alternative

$$\left\| \left[c_i(x_k) \bar{\lambda}_i(x_k, \lambda_k, s_k) / \lambda_{k,i}^{\alpha_\lambda} \right]_{i=1}^{m} \right\| + \left\| [c_i(x_k)]_{i=m+1}^{m_t} \right\| \leq \eta_k, \tag{8.8}$$

and to use the definition $\bar{\lambda}_i = \lambda_i + c_i(x)/\mu$ for $m+1 \le i \le m_t$. It is obvious that replacing (3.8) by (8.8) in Algorithm 3.1 makes no difference if there are no equality constraints. Moreover, if, instead, there are no inequality constraints, the above modification to Algorithm 3.1 gives Algorithm 1 of Conn *et al.* (1991).

A careful examination of the present paper and that by Conn *et al.* (1991) reveals that the exact form of the test (8.8) only plays a role in Lemmas B.1 and 5.2 and Theorems 5.3 and 5.6 in this paper and Lemmas 4.1 and Theorems 5.3 and 5.5 in its predecessor. We now briefly consider what can be deduced about the composite algorithm.

In the first relevant lemma in each paper, one merely needs to obtain an upper bound on $\left\| \left[c_i(x_k)\bar{\lambda}_i(x_k,\lambda_k,s_k)/\lambda_{k,i}^{\alpha_\lambda} \right]_{i=1}^m \right\|$ or $\left\| [c_i(x_k)]_{i=m+1}^{m_t} \right\|$ as appropriate, when the Lagrange multipliers are updated. But satisfaction of (8.8) yields both that

$$\left\| \left[c_i(x_k)\bar{\lambda}_i(x_k,\lambda_k,s_k)/\lambda_{k,i}^{\alpha_\lambda} \right]_{i=1}^m \right\| \le \eta_k \tag{8.9}$$

and

$$\left\| [c_i(x_k)]_{i=m+1}^{m_t} \right\| \le \eta_k. \tag{8.10}$$

Thus the conclusions of both lemmas are true when the composite algorithm is used. Furthermore, if we replace the set $\mathcal{A}^*$ in AS3 from this paper by the union of $\mathcal{A}^*$ and $\{m+1,\cdots,m_t\}$, it is straightforward to deduce that Theorem 4.3 remains true and the error estimates provided by the present Theorem 4.3 and Theorem 4.3 of Conn *et al.* (1991) are valid.

These estimates are sufficient to ensure that if the test (8.8) were to fail for all $k \ge k_1$, one would obtain, corresponding to (C.36) in the proof of Theorem 5.2, Appendix C,

$$\left\| \left[c_i(x_k)\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}} \right]_{i=1}^m \right\| + \left\| [c_i(x_k)]_{i=m+1}^{m_t} \right\| \le a_{26}\mu_k, \tag{8.11}$$

for some constant a_{26} for all $k \ge k_2 \ge k_1$. This is sufficient to ensure that Lemma 5.2 remains true for the composite algorithm provided we replace the set $\mathcal{A}_1^*$ in AS5 from this paper by the union of $\mathcal{A}_1^*$ and $\{m+1,\cdots,m_t\}$. The direct analogue of the error estimates provided by Lemma 5.1 suffice to enable one to establish Theorems 5.3 and 5.6 for the composite algorithm.

Thus the convergence properties of the composite algorithm are no worse that those predicted for the specific algorithms analyzed in sections 4 and 5 of Conn *et al.* (1991) and the same sections of the present paper.

8.3 Final comments

We note that the results given here are unaltered if the convergence tolerance (3.5) is replaced by

$$\|D_k P(x_k, \nabla_x \Psi_k)\| \le \omega_k. \tag{8.12}$$

for any sequence of positive diagonal matrices $\{D_k\}$ with uniformly bounded condition number. This is important as the method of Conn *et al.* (1988a), which we would consider using to solve the inner iteration problem, allows for different scalings for the components of the gradients to cope with variables of differing magnitudes.

Although the rules for how the convergence tolerances η_k and ω_k are updated have been made rather rigid in this paper and although the results contained here may be proved under more general updating rules, we have refrained from even stating these as the resulting conditions on the updates seemed rather complicated and are unlikely to provide more practical updates.

We have made no attempt in this paper to consider how algorithms for solving the inner-iteration subproblem (see Section 3.3) mesh with Algorithm 3.1. Nor have we provided any detailed numerical evidence that the approach taken here is effective on general problems. In particular, it may sometimes be inefficient to determine a feasible starting point after a penalty parameter update using the techniques proposed in Section 7. We are currently considering the first issue and consequently cannot yet report on the second, except in special cases (see Conn *et al.*, 1992e). However, it is perhaps worthwhile adding that a rather rudimentary implementation of a Lagrangian barrier method solved over ninety percent of a non-trivial set of about one thousand problems — sometimes considerably more efficiently than the more sophisticated implementation of an augmented Lagrangian algorithm in our Fortran package, LANCELOT A. Moreover, considerable success was reported in the Harwell Subroutine Library (1993) code `VE14`, a Lagrangian Barrier algorithm for bound-constrained quadratic programming problems.

9 Acknowledgements

The authors are grateful to Roman Polyak for the stimulating conversations, concerning modified barrier functions, during our meetings at Yorktown Heights. Theywould also like to thank two anonymous referees whose useful comments have substantially improved the presentation.

Appendix

A Simplified statement of the Algorithm

In order to simplify the proofs given in the appendix, we give them in the case of a particular instance of Algorithm 3.1. This simplified algorithm corresponds to choosing parameters $\eta_s = \omega_s = \alpha_\omega = \beta_\omega = 1$, $\alpha_\eta = \frac{5}{6}, \beta_\eta = \frac{1}{6}$ and $\alpha_\lambda = \frac{1}{2}$.

For proofs that correspond to the more general statements the interested reader is referred to the technical report, Conn *et al.* (1992a). However, the choice of parameters in practice appears to be significantly more critical than for augmented Lagrangian approaches. Thus the choice given here, which is just for convenience of exposition, should not be taken as an indication of suitable values.

Algorithm A.1 [Outer Iteration Algorithm]

step 0 : [Initialization] *Choose the strictly positive constants*

$$\omega_s, \eta_s, \mu_0 < 1, \tau < 1, \textit{ and define } \omega_0 = \omega_s \mu_0 \textit{ and } \eta_0 = \eta_s \mu_0^{\frac{5}{6}}. \tag{A.1}$$

An initial estimate of the solution, $x_{-1} \in B$, *and vector of positive Lagrange multiplier estimates,* λ_0, *for which* $c_i(x_{-1}) + \mu_0\sqrt{\lambda_{0,i}} > 0$ *are specified. Set* $k = 0$.

In addition set

$$\omega_* \ll 1 \textit{ and } \eta_* \ll 1 \tag{A.2}$$

step 1 : [Inner iteration] *Compute shifts*

$$s_{k,i} = \mu_k \sqrt{\lambda_{k,i}}, \tag{A.3}$$

for $i = 1, ..., m$. *Find* $x_k \in B$ *such that*

$$\|P(x_k, \nabla\Psi_k)\| \le \omega_k \tag{A.4}$$

and

$$c_i(x_k) + s_{k,i} > 0, \quad \textit{for} \quad i = 1, ..., m. \tag{A.5}$$

step 2 : [Test for convergence] *If*

$$\|P(x_k, \nabla\Psi_k)\| \le \omega_* \textit{ and } \|[c_i(x_k)\bar{\lambda}_{k,i}]_{i=1}^m\| \le \eta_*, \tag{A.6}$$

stop. If

$$\left\| \left[\frac{c_i(x_k)\bar{\lambda}_{k,i}}{\sqrt{\lambda_{k,i}}} \right]_{i=1}^m \right\| \le \eta_k, \tag{A.7}$$

execute step 3. Otherwise, execute step 4.

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

$$\begin{aligned}\lambda_{k+1} &= \bar{\lambda}_k,\\ \mu_{k+1} &= \mu_k,\\ \omega_{k+1} &= \mu_{k+1}\omega_k,\\ \eta_{k+1} &= \mu_{k+1}^{\frac{1}{6}}\eta_k.\end{aligned} \tag{A.8}$$

Increase k by one and go to step 1.

step 4 : [Reduce the penalty parameter] *Set*

$$\begin{aligned}\lambda_{k+1} &= \lambda_k,\\ \mu_{k+1} &= \tau\mu_k,\\ \omega_{k+1} &= \omega_s\mu_{k+1},\\ \eta_{k+1} &= \eta_s\mu_{k+1}^{\frac{5}{6}}.\end{aligned} \tag{A.9}$$

Increase k by one and go to step 1.

end of Algorithm A.1

B Details of proofs from Section 4

B.1 An auxilliary lemma

We require the following lemma in the proof of global convergence of our algorithm. The lemma is the analog of Conn *et al.* (1991, Lemma 4.1). In essence, the result shows that the Lagrange multiplier estimates generated by the algorithm cannot behave too badly.

Lemma B.1 *Suppose that μ_k converges to zero as k increases when Algorithm A.1 is executed. Then the product $\mu_k(\lambda_{k,i})^{\frac{3}{2}}$ converges to zero for each $1 \le i \le m$.*

Proof. If μ_k converges to zero, Step 4 of the algorithm must be executed infinitely often. Let $K = \{k_0, k_1, k_2, ...\}$ be the set of the indices of the iterations in which Step 4 of the algorithm is executed and for which

$$\mu_k \le (\tfrac{1}{2})^6. \tag{B.1}$$

We consider how the i-th Lagrange multiplier estimate changes between two successive iterations indexed in the set $\mathcal{K}$. Firstly note that $\lambda_{k_p+1,i} = \lambda_{k_p,i}$. At iteration $k_p + j$, for $k_p + 1 < k_p + j \le k_{p+1}$, we have

$$\lambda_{k_p+j,i} = \lambda_{k_p+j-1,i} - \left(\frac{c_{k_p+j-1,i}\lambda_{k_p+j,i}}{\sqrt{\lambda_{k_p+j-1,i}}}\right)\frac{1}{\mu_{k_p+j-1}}, \tag{B.2}$$

from (2.5), (A.3) and (A.8) and

$$\mu_{k_{p+1}} = \mu_{k_p+j} = \mu_{k_p+1} = \tau\mu_{k_p}, \tag{B.3}$$

Hence summing (B.2) and using the fact that $\lambda_{k_p+1,i} = \lambda_{k_p,i}$,

$$\lambda_{k_p+j,i} = \lambda_{k_p,i} - \sum_{l=1}^{j-1} \left(\frac{c_{k_p+l,i}\lambda_{k_p+l+1,i}}{\sqrt{\lambda_{k_p+l,i}}} \right) \frac{1}{\mu_{k_p+1}} \tag{B.4}$$

where the summation in (B.4) is null if $j = 1$.

Now suppose that $j > 1$. Then for the set of iterations $k_p + l, 1 \leq l < j$, Step 3 of the algorithm must have been executed and hence, from (A.5), (B.3) and the recursive definition of η_k, we must also have

$$\left\| \left[\frac{c_{k_p+l,i}\lambda_{k_p+l+1,i}}{\sqrt{\lambda_{k_p+l,i}}} \right]_{i=1}^{m} \right\| \leq \eta_s \mu_{k_p+1}^{\frac{5}{6}+\frac{1}{6}(l-1)} \tag{B.5}$$

Combining equations (B.1) to (B.5), we obtain the bound

$$\begin{aligned} \|\lambda_{k_p+j}\| &\leq \|\lambda_{k_p}\| + \textstyle\sum_{l=1}^{j-1} \left\| \left[\frac{c_{k_p+l,i}\lambda_{k_p+l+1,i}}{\sqrt{\lambda_{k_p+l,i}}} \right]_{i=1}^{m} \right\| \cdot \frac{1}{\mu_{k_p+1}} \\ &\leq \|\lambda_{k_p}\| + 2\eta_s / \mu_{k_p+1}^{\frac{1}{6}} \end{aligned} \tag{B.6}$$

Thus, multiplying (B.6) by $\mu_{k_p+j}^{\frac{2}{3}}$ and using (B.3), we obtain that

$$\mu_{k_p+j}^{\frac{2}{3}} \|\lambda_{k_p+j}\| \leq (\tau\mu_{k_p})^{\frac{2}{3}} \|\lambda_{k_p}\| + 2\eta_s\sqrt{\tau\mu_{k_p}}. \tag{B.7}$$

Equation (B.7) is also satisfied when $j = 1$ as equations (A.8) and (B.3) give

$$\mu_{k_p+j}^{\frac{2}{3}} \|\lambda_{k_p+j}\| = (\tau\mu_{k_i})^{\frac{2}{3}} \|\lambda_{k_p}\|. \tag{B.8}$$

Hence from (B.7),

$$\mu_{k_{p+1}}^{\frac{2}{3}} \|\lambda_{k_{p+1}}\| \leq (\tau\mu_{k_p})^{\frac{2}{3}} \|\lambda_{k_p}\| + 2\eta_s\sqrt{\tau\mu_{k_p}}. \tag{B.9}$$

We now show that (B.9) implies that $\mu_{k_p}^{\frac{2}{3}} \|\lambda_{k_p}\|$ converges to zero as k increases. For, if we define

$$\alpha_p \stackrel{\text{def}}{=} \mu_{k_p}^{\frac{2}{3}} \|\lambda_{k_p}\| \text{ and } \beta_p \stackrel{\text{def}}{=} 2\eta_s\sqrt{\mu_{k_p}}, \tag{B.10}$$

equations (B.3), (B.9) and (B.10) give that

$$\alpha_{p+1} \leq \tau^{\frac{2}{3}}\alpha_p + \sqrt{\tau}\beta_p \text{ and } \beta_{p+1} = \sqrt{\tau}\beta_p \tag{B.11}$$

and hence that

$$0 \leq \alpha_p \leq \tau^{\frac{2}{3}p}\alpha_0 + \tau^{\frac{p}{2}} \sum_{l=0}^{p-1} \tau^{\frac{l}{6}}\beta_0. \tag{B.12}$$

It now follows that

$$0 \leq \alpha_p \leq \tau^{\frac{2}{3}p}\alpha_0 + \frac{\tau^{\frac{p}{2}}}{1-\tau^{\frac{l}{6}}}\beta_0. \tag{B.13}$$

But, both α_0 and β_0 are finite. Thus, as p increases, α_p converges to zero; the second part of equation (B.11) implies that β_p converges to zero. Therefore, as the right-hand side of (B.7) converges to zero, so does $\mu_k^{\frac{2}{3}}\|\lambda_k\|$ for all k. The truth of the lemma is finally established by raising $\mu_k^{\frac{2}{3}}\|\lambda_{k,i}\|$ to the power $\frac{3}{2}$. ■

We note that Lemma B.1 may be proved under much weaker conditions on the sequence $\{\eta_k\}$ than those imposed in Algorithm A.1. All that is needed is that, in the proof just given,

$$\sum_{l=1}^{j-1}\left\|\left[\frac{c_{k_p+l,i}\lambda_{k_p+l+1,i}}{\sqrt{\lambda_{k_p+l,i}}}\right]_{i=1}^{m}\right\|$$

in (B.6) should be bounded by some multiple of a positive power of $\mu_{k_{p+1}}$.

B.2 Proof of Theorem 4.2

In order to prove *(i)*, *(ii)* and *(iii)* we consider each constraint in turn and distinguish two cases:

1. constraints for which $c_{*,i} \neq 0$; and
2. constraints for which $c_{*,i} = 0$.

For the first of these cases, we need to consider the possibility that

a. the penalty parameter μ_k is bounded away from zero; and

b. the penalty parameter μ_k converges to zero.

Case 1a. As μ_k is bounded away from zero, test (A.7) must be satisfied for all k sufficiently large and hence $|c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_k}|$ converges to zero. Thus as $\{c_{k,i}\}$ converges to $c_{*,i} \neq 0$, for $k \in \mathcal{K}$, $\bar{\lambda}_{k,i}/\sqrt{\lambda_k}$ converges to zero. Hence, using (2.3) and (A.3),

$$\frac{\bar{\lambda}_{k,i}}{\sqrt{\lambda_{k,i}}} \equiv \frac{\mu\lambda_{k,i}}{c_{k,i}+\mu\sqrt{\lambda_{k,i}}} = \sqrt{\lambda_{k,i}}\frac{\mu\sqrt{\lambda_{k,i}}}{c_{k,i}+\mu\sqrt{\lambda_{k,i}}} \to 0. \tag{B.14}$$

We aim to show that $\bar{\lambda}_{k,i}$ converges to zero and that $c_{*,i} > 0$.

Suppose first that $\lambda_{k,i}$ does not converge to zero. It follows directly from (2.5) and (A.3) that

$$c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}} = \mu_k(\lambda_{k,i} - \bar{\lambda}_{k,i}). \tag{B.15}$$

Then, as the left-hand side of (B.15) converges to zero and μ_k and $\lambda_{k,i}$ are bounded away from zero, we deduce that

$$\bar{\lambda}_{k,i} = \lambda_{k,i}(1-\epsilon_{k,i}), \tag{B.16}$$

for some $\{\epsilon_{k,i}\}$, $k \in \mathcal{K}$, converging to zero. But then, by definition (2.3),

$$\frac{\mu\sqrt{\lambda_{k,i}}}{c_{k,i}+\mu\sqrt{\lambda_{k,i}}} = 1-\epsilon_{k,i}. \tag{B.17}$$

However, as $\lambda_{k,i}$ is bounded away from zero, (B.17) contradicts (B.14). Thus $\lambda_{k,i}$ converges to zero, for $k \in \mathcal{K}$.

It now follows that, as $\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}}$ converges to zero, so does $\bar{\lambda}_{k,i}$. It also follows from (A.5) that $c_{k,i}+\mu_k\sqrt{\lambda_{k,i}}>0$. As μ_k is bounded and $\lambda_{k,i}$ converges to zero, we have that $c_{*,i}\geq 0$. But as $c_{*,i}\neq 0$, we conclude that $c_{*,i}>0$, $\bar{\lambda}_{k,i}$ converges to $\lambda_{*,i}=0$, for $k\in\mathcal{K}$, and $c_{*,i}\lambda_{*,i}=0$.

Case 1b. As μ_k converges to zero, Lemma B.1 shows that $\mu_k(\lambda_{k,i})^{\frac{3}{2}}$ and hence $\mu_k\lambda_{k,i}$ and $\mu_k\sqrt{\lambda_{k,i}}$ converges to zero. It follows immediately that the numerator of (2.3) converges to zero while the denominator converges to $c_{*,i}$ and hence that $\bar{\lambda}_{k,i}$ converges to zero for $k\in\mathcal{K}$. Furthermore, it follows from (A.5) that $c_{k,i}+\mu_k\sqrt{\lambda_{k,i}}>0$: as $\mu_k\sqrt{\lambda_{k,i}}$ converges to zero, we have that $c_{*,i}\geq 0$. But as $c_{*,i}$ is, by assumption, nonzero, $c_{*,i}>0$. Hence we may conclude that that $c_{*,i}>0$, $\bar{\lambda}_{k,i}$ converges to $\lambda_{*,i}=0$, for $k\in\mathcal{K}$, and $c_{*,i}\lambda_{*,i}=0$.

We note from (2.15) that the set $\mathcal{I}^*\equiv\mathcal{I}(x_*)$ is precisely the set of constraints covered in Case 1. Having thus identified the constraints in $\mathcal{A}^*\equiv\mathcal{A}(x_*)$ as those in Case 2 above, we consider Case 2 in detail.

Case 2. By construction, at every iteration of the algorithm, $\bar{\lambda}_k>0$. Moreover, from (2.6), (2.12), (A.4) and Case 1 above,

$$\begin{aligned}
&\|(g_k-A_{k,\mathcal{A}^*}^T\bar{\lambda}_{k,\mathcal{A}^*})_{\mathcal{F}_1}\| \\
&\quad\leq \|(A_{k,\mathcal{I}^*}^T\bar{\lambda}_{k,\mathcal{I}^*})_{\mathcal{F}_1}\|+\|P(x_k,\nabla_x\Psi_k)_{\mathcal{F}_1}\| \\
&\quad\leq \|(A_{k,\mathcal{I}^*}^T\bar{\lambda}_{k,\mathcal{I}^*})_{\mathcal{F}_1}\|+\omega_k\leq\bar{\omega}_k
\end{aligned}\tag{B.18}$$

for some $\bar{\omega}_k$ converging to zero. Thus using AS2 and Lemma 4.1, the Lagrange multiplier estimates $\bar{\lambda}_{k,\mathcal{A}^*}$ are bounded and, as $\mathcal{L}(x_k,\bar{\omega}_k;x_*,\mathcal{F}_1)$ is non-empty, these multipliers have at least one limit point. If $\lambda_{\mathcal{A}}$ is such a limit, AS1, (B.18) and the identity $c_{*,\mathcal{A}^*}=0$ ensure that $(g_*-A_{*,\mathcal{A}^*}^T\lambda_{\mathcal{A}})_{\mathcal{F}_1}=0$, $c_{*,\mathcal{A}^*}^T\lambda_{\mathcal{A}}=0$ and $\lambda_{\mathcal{A}}\geq 0$.

Thus, from AS2, there is a subsequence $\mathcal{K}'\subseteq\mathcal{K}$ for which $\{x_k\}$ converges to x_* and $\{\bar{\lambda}_k\}$ converges to λ_* as $k\in\mathcal{K}'$ tends to infinity and hence, from (2.4), $\nabla_x\Psi_k$ converges to g_*^ℓ. We also have that

$$c_*^T\lambda_*=0\tag{B.19}$$

with both $c_{*,i}$ and $\lambda_{*,i}$ $(i=1,\ldots,m)$ non-negative and at least one of the pair equal to zero. We may now invoke Lemma 2.1 and the convergence of $\nabla_x\Psi_k$ to g_*^ℓ to see that

$$g_{*,\mathcal{F}_1}^\ell=0 \quad\text{and}\quad x_*^Tg_*^\ell=0.\tag{B.20}$$

The variables in the set $\mathcal{F}_1\cap\mathcal{N}_b$ are, by definition, positive at x_*. The components of g_*^ℓ indexed by $\mathcal{D}_1$ are non-negative from (2.10) as their corresponding variables are dominated. This then gives the conditions

$$\begin{array}{rlll}
x_{*,i}>0 & \text{and}\ \ g_{*,i}^\ell=0 & \text{for} & i\in\mathcal{F}_1\cap\mathcal{N}_b, \\
& \phantom{\text{and}}\ \ g_{*,i}^\ell=0 & \text{for} & i\in\mathcal{F}_1\cap\mathcal{N}_f, \\
x_{*,i}=0 & \text{and}\ \ g_{*,i}^\ell\geq 0 & \text{for} & i\in\mathcal{D}_1\ \ \text{and} \\
x_{*,i}=0 & \text{and}\ \ g_{*,i}^\ell=0 & \text{for} & i\in\mathcal{F}_4.
\end{array}\tag{B.21}$$

Thus we have shown that x_* is a Kuhn-Tucker point and hence we have established results *(i)*, *(ii)* and *(iii)*. It remains to prove *(iv)*.

If μ_k is bounded away from zero, we have established in Case 1a above that $\lambda_{k,i}$ converges to zero. Hence, as μ_k is finite, $s_{k,i}$ also converges to zero. On the other hand, if μ_k converges to zero, we have established in Case 1b that $\mu_k\sqrt{\lambda_{k,i}}$ and hence, once again, $s_{k,i}$ converge to zero. But as $i \in \mathcal{I}^*$, $c_{k,i}$ is bounded away from zero for all $k \in \mathcal{K}$ sufficiently large, and therefore $\sigma_{k,i}$ converges to zero for all $i \in \mathcal{I}^*$ which establishes *(iv)*.

B.3 Proof of Theorem 4.3

Assumption AS3 implies that there is at most one point in $\mathcal{L}(x_*, 0; x_*, \mathcal{F}_1)$ and thus AS2 holds. The conclusions of Theorem 4.2 then follow. The conclusion *(v)* of the current theorem is a direct consequence of AS3.

We have already identified the set of constraints for which $c_i(x_*) = 0$ with $\mathcal{A}^*$. Let

$$\sigma_{k,i} \stackrel{\text{def}}{=} \frac{s_{k,i}}{c_{k,i} + s_{k,i}}. \tag{B.22}$$

Then (2.3) shows that $\bar{\lambda}_{k,i} = \sigma_{k,i}\lambda_{k,i}$. We now prove that $\sigma_{k,i}$ converges to zero for all $i \in \mathcal{I}^*$ as $k \in \mathcal{K}$ tends to infinity.

To prove *(vi)*, we let $\bar{\Omega}$ be any closed, bounded set containing the iterates x_k, $k \in \mathcal{K}$. We note that, as a consequence of AS1 and AS3, for $k \in \mathcal{K}$ sufficiently large, $A^+_{k,\mathcal{A}^*,\mathcal{F}_1}$ exists, is bounded and converges to $A^+_{*,\mathcal{A}^*,\mathcal{F}_1}$. Thus we may write

$$\|A^+_{k,\mathcal{A}^*,\mathcal{F}_1}\| \leq a_1 \tag{B.23}$$

for some constant $a_1 > 0$. As the variables in the set $\mathcal{F}_1$ are floating, equations (2.6), (2.7), (2.12) and the inner iteration termination criterion (A.4) give that

$$\|g_{k,\mathcal{F}_1} + A^T_{k,\mathcal{A}^*,\mathcal{F}_1}\bar{\lambda}_{k,\mathcal{A}^*} + A^T_{k,\mathcal{I}^*,\mathcal{F}_1}\bar{\lambda}_{k,\mathcal{I}^*}\| \leq \omega_k. \tag{B.24}$$

By assumption, $\lambda(x)_{\mathcal{A}^*}$ is bounded for all x in a neighbourhood of x_*. Thus we may deduce from (4.2), (B.23) and (B.24) that

$$\begin{aligned} \|\bar{\lambda}_{k,\mathcal{A}^*} - \lambda_{k,\mathcal{A}^*}\| &= \|A^{+\,T}_{k,\mathcal{A}^*,\mathcal{F}_1} g_{k,\mathcal{F}_1} + \bar{\lambda}_{k,\mathcal{A}^*}\| \\ &= \|A^{+\,T}_{k,\mathcal{A}^*,\mathcal{F}_1}(g_{k,\mathcal{F}_1} + A^T_{k,\mathcal{A}^*,\mathcal{F}_1}\bar{\lambda}_{k,\mathcal{A}^*})\| \\ &\leq \|A^{+\,T}_{k,\mathcal{A}^*,\mathcal{F}_1}\|(\omega_k + \|A^T_{k,\mathcal{I}^*,\mathcal{F}_1}\|\|\bar{\lambda}_{k,\mathcal{I}^*}\| \\ &\leq a_1\omega_k + a_3\|\bar{\lambda}_{k,\mathcal{I}^*}\|, \end{aligned} \tag{B.25}$$

where $a_3 \stackrel{\text{def}}{=} a_1 \max_{x\in\bar{\Omega}} \|A(x)^T_{\mathcal{I}^*,\mathcal{F}_1}\|$. Moreover, from the integral mean-value theorem and the (local) differentiability of the least-squares Lagrange multiplier estimates (see, for example, Conn *et al.*, 1991, Lemma 2.2) we have that

$$\lambda_{k,\mathcal{A}^*} - \lambda_{*,\mathcal{A}^*} = \left(\int_0^1 \nabla_x\lambda(x(t))_{\mathcal{A}^*}dt\right)(x_k - x_*), \tag{B.26}$$

where $\nabla_x \lambda(x)_{\mathcal{A}^*}$ is given by Conn *et al.* (1991), equation (2.17), and where $x(t) = x_k + t(x_* - x_k)$. Now the terms within the integral sign are bounded for all x sufficiently close to x_* and hence (B.26) gives

$$\|\lambda_{\mathcal{A}^*,k} - \lambda_{\mathcal{A}^*,*}\| \leq a_2 \|x_k - x_*\| \tag{B.27}$$

for all $k \in \mathcal{K}$ sufficiently large, for some constant $a_2 > 0$, which is just the inequality (4.5). We then have that $\lambda_{\mathcal{A}^*,k}$ converges to $\lambda_{\mathcal{A}^*,*}$. Combining (4.1), (B.25) and (B.27), we obtain

$$\begin{aligned} \|\bar{\lambda}_{k,\mathcal{A}^*} - \lambda_{*,\mathcal{A}^*}\| \leq & \ \|\bar{\lambda}_{k,\mathcal{A}^*} - \lambda_{k,\mathcal{A}^*}\| + \|\lambda_{k,\mathcal{A}^*} - \lambda_{*,\mathcal{A}^*}\| \\ \leq & \ a_1 \omega_k + a_2 \|x_k - x_*\| + a_3 \|\bar{\lambda}_{k,\mathcal{I}^*}\| \\ \leq & \ a_1 \omega_k + a_2 \|x_k - x_*\| + a_3 \sigma_k \|\lambda_{k,\mathcal{I}^*}\|, \end{aligned} \tag{B.28}$$

the required inequality (4.4). It remains to establish (4.6) and (4.7).

The relationships (2.5) and (A.3) imply that

$$c_{k,i} = \mu_k (\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i})(\lambda_{k,i} - \bar{\lambda}_{k,i}), \tag{B.29}$$

and

$$c_{k,i} \bar{\lambda}_{k,i} / \sqrt{\lambda_{k,i}} = \mu_k (\lambda_{k,i} - \bar{\lambda}_{k,i}) \tag{B.30}$$

for $1 \leq i \leq m$. Bounding (B.29) and using the triangle inequality and the inclusion $\mathcal{A} \subseteq \mathcal{A}^*$, we obtain

$$\begin{aligned} \|c_{k,\mathcal{A}}\| \leq & \ \mu_k \left\| \left[\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i} \right]_{i \in \mathcal{A}} \right\| \|(\bar{\lambda}_k - \lambda_k)_{\mathcal{A}}\| \\ \leq & \ \mu_k \left\| \left[\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i} \right]_{i \in \mathcal{A}} \right\| \left[\|(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}}\| \right] \\ \leq & \ \mu_k \left\| \left[\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i} \right]_{i \in \mathcal{A}} \right\| \left[\|(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}}\| \right]. \end{aligned} \tag{B.31}$$

But then, combining (B.28) and (B.31), we see that (4.7) holds for all $k \in \mathcal{K}$ sufficiently large. Furthermore, the triangle inequality, the relationships (4.1), (4.4) and

$$\lambda_{*,\mathcal{I}^*} = 0 \tag{B.32}$$

yield the bound

$$\begin{aligned} \|\bar{\lambda}_k - \lambda_k\| \leq & \ \|\bar{\lambda}_k - \lambda_*\| + \|\lambda_k - \lambda_*\| \\ \leq & \ \|(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + \|\bar{\lambda}_{k,\mathcal{I}^*}\| + \|\lambda_{k,\mathcal{I}^*}\| \\ \leq a_1 \omega_k + & \ a_2 \|x_k - x_*\| + (1 + (1 + a_3) \sigma_k) \|\lambda_{k,\mathcal{I}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| \end{aligned} \tag{B.33}$$

Hence taking norms of (B.30) and using (B.33), we see that (4.6) holds for all $k \in \mathcal{K}$ sufficiently large.

C Details of proofs from Section 5

C.1 Proof of Lemma 5.1

We first need to make some observations concerning the status of the variables as the limit point is approached. We pick k sufficiently large that the sets $\mathcal{F}_1$ and $\mathcal{D}_1$, defined in (2.13), have been determined. Then, for $k \in \mathcal{K}$, the remaining variables either float (variables in $\mathcal{F}_2$) or oscillate between floating and being dominated (variables in $\mathcal{F}_3$). Now recall the definition (2.14) of $\mathcal{F}_4$ and pick an infinite subsequence, $\tilde{\mathcal{K}}$, of $\mathcal{K}$ such that:

(i) $\mathcal{F}_4 = \mathcal{F}_5 \cup \mathcal{D}_2$ with $\mathcal{F}_5 \cap \mathcal{D}_2 = \emptyset$;

(ii) variables in $\mathcal{F}_5$ are floating for all $k \in \tilde{\mathcal{K}}$; and

(iii) variables in $\mathcal{D}_2$ are dominated for all $k \in \tilde{\mathcal{K}}$.

Notice that the set $\mathcal{F}_2$ of (2.13) is contained within $\mathcal{F}_5$. Note, also, that there are only a finite number ($\leq 2^{|\mathcal{F}_4|}$) of such subsequences $\tilde{\mathcal{K}}$ and that for k sufficiently large, each $k \in \mathcal{K}$ is in one such subsequence. It is thus sufficient to prove the lemma for $k \in \tilde{\mathcal{K}}$.

Now, for $k \in \tilde{\mathcal{K}}$, define

$$\mathcal{F} \stackrel{\text{def}}{=} \mathcal{F}_1 \cup \mathcal{F}_5 \quad \text{and} \quad \mathcal{D} \stackrel{\text{def}}{=} \mathcal{D}_1 \cup \mathcal{D}_2. \tag{C.1}$$

So, the variables in $\mathcal{F}$ are floating while those in $\mathcal{D}$ are dominated.

We also need to consider the status of the constraints in $\mathcal{A}_2^*$. We choose a χ satisfying (5.5) and pick an infinite subsequence, $\bar{\mathcal{K}}$, of $\tilde{\mathcal{K}}$ such that

(a) $\mathcal{A}_2^* = \mathcal{A}_s^* \cup \mathcal{A}_b^*$ with $\mathcal{A}_s^* \cap \mathcal{A}_b^* = \emptyset$ where $\mathcal{A}_s^*$ and $\mathcal{A}_b^*$ are defined below;

(b) the Lagrange multiplier estimates satisfy

$$\bar{\lambda}_{k,i} \leq \mu_k^{1-\chi} \sqrt{\lambda_{k,i}} \tag{C.2}$$

for all constraints $i \in \mathcal{A}_s^*$ and all $k \in \bar{\mathcal{K}}$; and

(c) the Lagrange multiplier estimates satisfy

$$\bar{\lambda}_{k,i} > \mu_k^{1-\chi} \sqrt{\lambda_{k,i}} \tag{C.3}$$

for all constraints $i \in \mathcal{A}_b^*$ and all $k \in \bar{\mathcal{K}}$.

We note that there are only a finite number ($\leq 2^{|\mathcal{A}_2^*|}$) of such subsequences $\bar{\mathcal{K}}$ and that for k sufficiently large, each $k \in \mathcal{K}$ is in one such subsequence. It is thus sufficient to prove the lemma for $k \in \bar{\mathcal{K}}$.

We define

$$\mathcal{A} = \mathcal{A}_1^* \cup \mathcal{A}_b^* \tag{C.4}$$

and note that this set is consistent with the set $\mathcal{A}$ described by AS5. It then follows from (5.1) and (C.4) that

$$\mathcal{A}^* = \mathcal{A} \cup \mathcal{A}_s^* \quad \text{with} \quad \mathcal{A} \cap \mathcal{A}_s^* = \emptyset. \tag{C.5}$$

We note that, if $i \in \mathcal{A}_b^*$, (C.3) gives

$$\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i} < \mu_k^{\chi-1} \tag{C.6}$$

for all $k \in \bar{\mathcal{K}}$. Moreover, inequalities (5.4) and (5.5) imply

$$\left\| \left[\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i} \right]_{i \in \mathcal{A}_1^*} \right\| \leq a_4 \mu_k^{\zeta-1} \leq a_4 \mu_k^{\chi-1}. \tag{C.7}$$

It then follows directly from (C.6) and (C.7) that

$$\left\| \left[\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i} \right]_{i \in \mathcal{A}} \right\| \leq a_{14} \mu_k^{\chi-1} \tag{C.8}$$

for some positive constants χ, satisfying (5.5), and a_{14} and for all $k \in \mathcal{K}$. Furthermore

$$\lambda_{*,\mathcal{A}_s^*} = 0, \tag{C.9}$$

as $\mathcal{A}_s^* \subseteq \mathcal{A}_2^*$. Finally, the same inclusion and (C.2) imply that

$$\|\bar{\lambda}_{k,\mathcal{A}_s^*}\| \leq \mu_k^{1-\chi} \left\| \left[\sqrt{\lambda_{k,i}}\right]_{i \in \mathcal{A}_s^*} \right\| \leq \mu_k^{1-\chi} \left\| \left[\sqrt{\lambda_{k,i}}\right]_{i \in \mathcal{A}_2^*} \right\| \tag{C.10}$$

for all $k \in \bar{\mathcal{K}}$.

We may now invoke Theorem 4.3, part (vi), the bound (C.8) and the inclusion $\mathcal{A} \subseteq \mathcal{A}^*$ to obtain the inequalities

$$\|(\bar{\lambda}(x_k, \lambda_k, s_k) - \lambda_*)_{\mathcal{A}}\| \leq a_1 \omega_k + a_2 \|x_k - x_*\| + a_3 \sigma_k \|\lambda_{k,\mathcal{I}^*}\|, \tag{C.11}$$

and

$$\begin{array}{rl} \|c_{k,\mathcal{A}}\| \leq & a_{14} \mu_k^{\chi} \left[a_1 \omega_k + a_2 \|x_k - x_*\| + \right. \\ & \left. a_3 \sigma_k \|\lambda_{k,\mathcal{I}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\|\right] \end{array} \tag{C.12}$$

for all sufficiently large $k \in \bar{\mathcal{K}}$. Moreover, $\bar{\lambda}_k$ converges to λ_* and hence (2.4) implies that $\nabla_x \Psi_k$ converges to g_*^ℓ. Therefore, from Lemma 2.1,

$$x_{*,i} = 0 \quad \text{for all} \quad i \in \mathcal{D} \quad \text{and} \quad g_{*,i}^\ell = 0 \quad \text{for all} \quad i \in \mathcal{F}. \tag{C.13}$$

Using Taylor's theorem and the identities (B.32), (C.5) and (C.9), we have

$$\begin{array}{rl} \nabla_x \Psi_k = & g_k + A_k^T \bar{\lambda}_k \\ = & g_* + H_*(x_k - x_*) + A_*^T \bar{\lambda}_k + \\ & \sum_{j=1}^m \bar{\lambda}_{k,j} H_{*,j}(x_k - x_*) + r_1(x_k, x_*, \bar{\lambda}_k) \\ = & g_*^\ell + H_*^\ell(x_k - x_*) + A_{*,\mathcal{A}}^T(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}} + A_{*,\mathcal{A}_s^*}^T \bar{\lambda}_{k,\mathcal{A}_s^*} + \\ & A_{*,\mathcal{I}^*}^T \bar{\lambda}_{k,\mathcal{I}^*} + r_1(x_k, x_*, \bar{\lambda}_k) + r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*), \end{array} \tag{C.14}$$

where

$$r_1(x_k, x_*, \bar{\lambda}_k) = \int_0^1 (H^\ell(x_k + t(x_* - x_k), \bar{\lambda}_k) - H^\ell(x_*, \bar{\lambda}_k))(x_k - x_*) dt \tag{C.15}$$

and

$$r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*) = \sum_{j=1}^m (\bar{\lambda}_{k,j} - \lambda_{j,*}) H_j(x_*)(x_k - x_*). \tag{C.16}$$

The boundedness and Lipschitz continuity of the Hessian matrices of f and the c_i in a neighbourhood of x_* along with the convergence of $\bar{\lambda}_k$ to λ_* for which the relationships (4.8) and (B.32) hold then give that

$$\begin{array}{rl} \|r_1(x_k, x_*, \bar{\lambda}_k)\| & \leq a_{15} \|x_k - x_*\|^2 \\ \text{and} & \\ \|r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*)\| & \leq a_{16} \|x_k - x_*\| \|\bar{\lambda}_k - \lambda_*\| \\ & \leq a_{16} \|x_k - x_*\| (\|(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}^*}\| + \sigma_k \|\lambda_{k,\mathcal{I}^*}\|) \end{array} \tag{C.17}$$

for some positive constants a_{15} and a_{16}, using (4.1). In addition, again using Taylor's theorem and that $c_{*,\mathcal{A}} = 0$,

$$c_{k,\mathcal{A}} = A_{*,\mathcal{A}}(x_k - x_*) + r_3(x_k, x_*)_{\mathcal{A}}, \tag{C.18}$$

where

$$(r_3(x_k, x_*))_i = \int_0^1 t_2 \int_0^1 (x_k - x_*)^T H_i(x_* + t_1 t_2(x_k - x_*))(x_k - x_*) dt_1 dt_2 \tag{C.19}$$

for $i \in \mathcal{A}$ (see Gruver and Sachs, 1980, p.11)). The boundedness of the Hessian matrices of the c_i in a neighbourhood of x_* then gives that

$$\|r_3(x_k, x_*)_{\mathcal{A}}\| \leq a_{17}\|x_k - x_*\|^2 \tag{C.20}$$

for some constant $a_{17} > 0$. Combining (C.14) and (C.18), we obtain

$$\begin{gathered}\begin{pmatrix} H_*^\ell & A_{*,\mathcal{A}}^T \\ A_{*,\mathcal{A}} & 0 \end{pmatrix}\begin{pmatrix} x_k - x_* \\ (\bar{\lambda}_k - \lambda_*)_{\mathcal{A}} \end{pmatrix} \\ = \begin{pmatrix} \nabla_x \Psi_k - g_*^\ell - A_{*,\mathcal{I}^*}^T \bar{\lambda}_{k,\mathcal{I}^*} - A_{*,\mathcal{A}_s^*}^T \bar{\lambda}_{k,\mathcal{A}_s^*} \\ c_{k,\mathcal{A}} \end{pmatrix} - \begin{pmatrix} r_1 + r_2 \\ (r_3)_{\mathcal{A}} \end{pmatrix},\end{gathered} \tag{C.21}$$

where we have suppressed the arguments of r_1, r_2 and r_3 for brevity. We may then use (C.13) to rewrite (C.21) as

$$\begin{gathered}\begin{pmatrix} H_{*,\mathcal{F},\mathcal{F}}^\ell & H_{*,\mathcal{F},\mathcal{D}}^\ell & A_{*,\mathcal{A},\mathcal{F}}^T \\ H_{*,\mathcal{D},\mathcal{F}}^\ell & H_{*,\mathcal{D},\mathcal{D}}^\ell & A_{*,\mathcal{A},\mathcal{D}}^T \\ A_{*,\mathcal{A},\mathcal{F}} & A_{*,\mathcal{A},\mathcal{D}} & 0 \end{pmatrix}\begin{pmatrix} (x_k - x_*)_{\mathcal{F}} \\ x_{k,\mathcal{D}} \\ (\bar{\lambda}_k - \lambda_*)_{\mathcal{A}} \end{pmatrix} \\ = \begin{pmatrix} \nabla_x \Psi_{k,\mathcal{F}} - A_{*,\mathcal{A}_s^*,\mathcal{F}}^T \bar{\lambda}_{k,\mathcal{A}_s^*} - A_{*,\mathcal{I}^*,\mathcal{F}}^T \bar{\lambda}_{k,\mathcal{I}^*} \\ \nabla_x \Psi_{k,\mathcal{F}} - g_{*,\mathcal{D}}^\ell - A_{*,\mathcal{A}_s^*,\mathcal{D}}^T \bar{\lambda}_{k,\mathcal{A}_s^*} - A_{*,\mathcal{I}^*,\mathcal{D}}^T \bar{\lambda}_{k,\mathcal{I}^*} \\ c_{k,\mathcal{A}} \end{pmatrix} - \begin{pmatrix} (r_1 + r_2)_{\mathcal{F}} \\ (r_1 + r_2)_{\mathcal{D}} \\ (r_3)_{\mathcal{A}} \end{pmatrix}.\end{gathered} \tag{C.22}$$

Then, rearranging (C.22) and removing the middle horizontal block we obtain

$$\begin{gathered}\begin{pmatrix} H_{*,\mathcal{F},\mathcal{F}}^\ell & A_{*,\mathcal{A},\mathcal{F}}^T \\ A_{*,\mathcal{A},\mathcal{F}} & 0 \end{pmatrix}\begin{pmatrix} (x_k - x_*)_{\mathcal{F}} \\ (\bar{\lambda}_k - \lambda_*)_{\mathcal{A}} \end{pmatrix} = \\ \begin{pmatrix} \nabla_x \Psi_{k,\mathcal{F}} - H_{*,\mathcal{F},\mathcal{D}}^\ell x_{k,\mathcal{D}} - A_{*,\mathcal{A}_s^*,\mathcal{F}}^T \bar{\lambda}_{k,\mathcal{A}_s^*} - A_{*,\mathcal{I}^*,\mathcal{F}}^T \bar{\lambda}_{k,\mathcal{I}^*} \\ c_{k,\mathcal{A}} - A_{*,\mathcal{A},\mathcal{D}} x_{k,\mathcal{D}} \end{pmatrix} - \begin{pmatrix} (r_1 + r_2)_{\mathcal{F}} \\ (r_3)_{\mathcal{A}} \end{pmatrix}.\end{gathered} \tag{C.23}$$

Roughly, the rest of the proof proceeds by showing that the right-hand side of (C.23) is $O(\omega_k)+ O(\sigma_k\|\lambda_{k,\mathcal{I}^*}\|)+ O(\mu_k\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\|)$. This will then ensure that the vector on the left-hand side is of the same size, which is the result we require. Firstly observe that

$$\|x_{k,\mathcal{D}}\| \leq \omega_k, \tag{C.24}$$

from (2.11) and (A.4) and

$$\|\nabla_x \Psi_{k,\mathcal{F}}\| \leq \omega_k, \tag{C.25}$$

from (2.12). Consequently, using (C.13) and (C.24),

$$\|x_k - x_*\| \leq \|(x_k - x_*)_{\mathcal{F}}\| + \omega_k. \tag{C.26}$$

Let $\Delta x_k = \|(x_k - x_*)_{\mathcal{F}}\|$. Combining (4.8), (B.32), (C.11) and (C.26), we obtain

$$\|(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}}\| \leq a_{18}\omega_k + a_2 \Delta x_k + a_3 \sigma_k \|\lambda_{k,\mathcal{I}^*}\|, \tag{C.27}$$

where $a_{18} \stackrel{\text{def}}{=} a_1 + a_2$. Furthermore, from (C.17), (C.20), (C.26) and (C.27),

$$\left\| \begin{pmatrix} (r_1 + r_2)_{\mathcal{F}} \\ (r_3)_{\mathcal{A}} \end{pmatrix} \right\| \leq \begin{array}{c} a_{19}(\Delta x_k)^2 + a_{20}\Delta x_k \omega_k + a_{21}\omega_k^2 + \\ a_{22}\sigma_k \|\lambda_{k,\mathcal{I}^*}\|(\omega_k + \Delta x_k) \end{array} \tag{C.28}$$

where $a_{19} \stackrel{\text{def}}{=} a_{15} + a_{17} + a_{16}a_2$, $a_{20} \stackrel{\text{def}}{=} 2(a_{15} + a_{17}) + a_{16}(a_{18} + a_2)$, $a_{21} \stackrel{\text{def}}{=} a_{15} + a_{17} + a_{16}a_{18}$ and $a_{22} \stackrel{\text{def}}{=} a_{16}(1 + a_3)$. Moreover, from (C.10), (C.12), (C.24), (C.25) and (C.26),

$$\begin{array}{r} \left\| \begin{pmatrix} \nabla_x \Psi_{k,\mathcal{F}} - H^{\ell}_{*,\mathcal{F},\mathcal{D}} x_{k,\mathcal{D}} - A^T_{*,\mathcal{A}^*_s,\mathcal{F}} \bar{\lambda}_{k,\mathcal{A}^*_s} - A^T_{*,\mathcal{I}^*,\mathcal{F}} \bar{\lambda}_{k,\mathcal{I}^*} \\ c_{k,\mathcal{A}} - A_{*,\mathcal{A},\mathcal{D}} x_{k,\mathcal{D}} \end{pmatrix} \right\| \leq \\ a_{23}\omega_k + a_{24}\sigma_k \|\lambda_{k,\mathcal{I}^*}\| + a_{25}\mu_k^{1-\chi} \left\| [\sqrt{\lambda_{k,i}}]_{i \in \mathcal{A}^*_2} \right\| + \\ a_{14}\mu_k^{\chi} \left[a_{18}\omega_k + a_2 \Delta x_k + a_3 \sigma_k \|\lambda_{k,\mathcal{I}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| \right], \end{array} \tag{C.29}$$

where

$$a_{23} \stackrel{\text{def}}{=} 1 + \left\| \begin{pmatrix} H^{\ell}_{*,\mathcal{F},\mathcal{D}} \\ A_{*,\mathcal{A},\mathcal{D}} \end{pmatrix} \right\| \quad a_{24} \stackrel{\text{def}}{=} \|A^T_{*,\mathcal{I}^*,\mathcal{F}}\| \text{and} \quad a_{25} \stackrel{\text{def}}{=} \|A^T_{*,\mathcal{A}^*_s,\mathcal{F}}\|. \tag{C.30}$$

By assumption AS5, the coefficient matrix on the left-hand side of (C.23) is non-singular. Let its inverse have norm M. Multiplying both sides of the equation by this inverse and taking norms, we obtain

$$\begin{array}{rcl} \left\| \begin{pmatrix} (x_k - x_*)_{\mathcal{F}} \\ (\bar{\lambda}_k - \lambda_*)_{\mathcal{A}} \end{pmatrix} \right\| & \leq & M[a_{19}(\Delta x_k)^2 + a_{20}\Delta x_k \omega_k + a_{21}\omega_k^2 + \\ & & a_{22}\sigma_k \|\lambda_{k,\mathcal{I}^*}\|(\omega_k + \Delta x_k) + a_{23}\omega_k + a_{24}\sigma_k \|\lambda_{k,\mathcal{I}^*}\| + \\ & & a_{25}\mu_k^{1-\chi} \left\| [\sqrt{\lambda_{k,i}}]_{i \in \mathcal{A}^*_2} \right\| + a_{14}\mu_k^{\chi}(a_{18}\omega_k + a_2 \Delta x_k + \\ & & \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + a_3 \sigma_k \|\lambda_{k,\mathcal{I}^*}\|)] \\ & = & (Ma_{19}\Delta x_k + Ma_{20}\omega_k + Ma_2 a_{14}\mu_k^{\chi})\Delta x_k + \\ & & (Ma_{21}\omega_k + Ma_{14}a_{18}\mu_k^{\chi} + Ma_{23})\omega_k + \\ & & Ma_{14}\mu_k^{\chi}\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + Ma_{25}\mu_k^{1-\chi} \left\| [\sqrt{\lambda_{k,i}}]_{i \in \mathcal{A}^*_2} \right\| + \\ & & (Ma_{24} + Ma_{22}(\omega_k + \Delta x_k) + Ma_3 a_{14}\mu_k^{\chi})\sigma_k \|\lambda_{k,\mathcal{I}^*}\|. \end{array} \tag{C.31}$$

The mechanisms of Algorithm A.1 ensures that ω_k converges to zero. Moreover, Theorem 4.2 guarantees that Δx_k also converges to zero for $k \in \bar{\mathcal{K}}$. Thus, there is a k_0 for which

$$\omega_k \leq \min(1, 1/(4Ma_{20})) \tag{C.32}$$

and

$$\Delta x_k \leq \min(1, 1/(4Ma_{19})) \tag{C.33}$$

for all $k \geq k_0$ ($k \in \bar{\mathcal{K}}$). Furthermore, let

$$\mu_{\max} \equiv \min(1, 1/(4Ma_2 a_{14})^{1/\chi}). \tag{C.34}$$

Then, if $\mu_k \leq \mu_{\max}$, (C.31), (C.32), (C.34) and (C.33) give

$$\begin{aligned}\Delta x_k \leq\ & \tfrac{3}{4}\Delta x_k + M(a_{21} + a_{14}a_{18} + a_{23})\omega_k + \\ & M a_{14}\mu_k^{\chi}\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + M a_{25}\mu_k^{1-\chi}\left\|\left[\sqrt{\lambda_{k,i}}\right]_{i\in\mathcal{A}_2^*}\right\| + \\ & M(a_{24} + 2a_{22} + a_3 a_{14})\sigma_k\|\lambda_{k,\mathcal{I}^*}\|.\end{aligned} \tag{C.35}$$

Cancelling the Δx_k terms in (C.35), multipling the resulting inequality by four and substituting into (C.26), we obtain the desired inequality (5.6), where $a_5 \stackrel{\text{def}}{=} 1 + 4M(a_{21} + a_{14}a_{18} + a_{23})$, $a_6 \stackrel{\text{def}}{=} 4Ma_{14}$, $a_7 \stackrel{\text{def}}{=} 4Ma_{25}$ and $a_8 \stackrel{\text{def}}{=} 4M(a_{24} + 2a_{22} + a_3a_{14})$. The remaining inequalities (5.7) and (5.8) follow directly by substituting (5.6) into (4.4) and (4.6), the required constants being $a_9 \stackrel{\text{def}}{=} a_1 + a_2a_5$, $a_{10} \stackrel{\text{def}}{=} a_2a_6$, $a_{11} \stackrel{\text{def}}{=} a_2a_7$, $a_{12} \stackrel{\text{def}}{=} a_3 + a_2a_8$ and $a_{13} \stackrel{\text{def}}{=} 1 + a_2a_6$.

C.2 Proof of Lemma 5.2

We have, from Theorem 4.3 and AS6, that the complete sequence of Lagrange multiplier estimates $\{\bar\lambda_k\}$ generated by Algorithm A.1 converges to λ_*. We now consider the sequence $\{\lambda_k\}$.

There are three possibilities. Firstly, μ_k may be bounded away from zero. In this case, Step 3 of Algorithm A.1 must be executed for all k sufficiently large which ensures that $\{\lambda_k\}$ and $\{\bar\lambda_{k-1}\}$ are identical for all large k. As the latter sequence converges to λ_*, so does the former.

Secondly, μ_k may converge to zero but nonetheless there may be an infinite number of iterates for which (A.7) is satisfied. In this case, the only time adjacent members of the sequence $\{\lambda_k\}$ differ, $\lambda_k = \bar\lambda_{k-1}$ and we have already observed that the latter sequence $\{\bar\lambda_{k-1}\}$ converges to λ_*.

Finally, if the test (A.7) were to fail for all $k > k_1$, $\|\lambda_{k,\mathcal{I}^*}\|$ and $\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\|$ will remain fixed for all $k \geq k_1$, as Step 4 would then be executed for all subsequent iterations. But then (4.6) implies that

$$\left\|\left[c_{k,i}\bar\lambda_{k,i}/\sqrt{\lambda_{k,i}}\right]_{i=1}^{m}\right\| \leq a_{26}\mu_k \tag{C.36}$$

for some constant a_{26} for all $k \geq k_2 \geq k_1$. As μ_k converges to zero as k increases,

$$a_{26}\mu_k \leq \eta_s\mu_k^{\frac{5}{6}} = \eta_k \tag{C.37}$$

for all k sufficiently large. But then inequality (A.7) must be satisfied for some $k \geq k_1$ contradicting the supposition. Hence, this latter possibility proves to be impossible. Thus $\{\lambda_k\}$ converges to λ_*.

Inequality (5.9) then follows immediately for $i \in \mathcal{I}^*$ by considering the definitions (A.3), (4.8) and (B.22) and using the convergence of $\lambda_{k,\mathcal{I}^*}$ to $\lambda_{*,\mathcal{I}^*} = 0$; a suitable representation of θ_k would be

$$\theta_k = \max_{i\in\mathcal{I}^*}\left(\frac{\sqrt{\lambda_{k,i}}}{c_{k,i} + \mu_k\sqrt{\lambda_{k,i}}}\right). \tag{C.38}$$

Now $\bar\lambda_{k,i}$ converges to $\lambda_{*,i} > 0$ and is thus bounded away from zero for all k, for each $i \in \mathcal{A}_1^*$. But this and the convergence of $\{\bar\lambda_k\}$ to λ_* implies that $\sqrt{\lambda_{k,i}}/\bar\lambda_{k,i}$ is bounded and hence inequality (5.4), with $\zeta = 1$, holds for all k. The remaining results follow directly from Lemma 5.1 on substituting $\zeta = 1$ into (5.5).

C.3 Proof of Theorem 5.3

The appropriate version of Theorem 5.3 for the simplified Algorithm 2 is now stated:

Theorem C.1 *Suppose that the iterates $\{x_k\}$ generated by Algorithm 3.1 of Section 3 satisfy AS6 and that AS4 and AS5 hold. Furthermore, suppose that AS7 holds. Then there is a constant $\mu_{\min} > 0$ such that $\mu_k \geq \mu_{\min}$ for all k.*

Proof. Suppose, otherwise, that μ_k tends to zero. Then, Step 4 of the algorithm must be executed infinitely often. We aim to obtain a contradiction to this statement by showing that Step 3 is always executed for k sufficiently large. We note that our assumptions are sufficient for the conclusions of Theorem 4.3 to hold.

Lemma 5.2, part (i), ensures that $\{\lambda_k\}$ converges to λ_*. We note that, by definition,

$$\mu_k < 1. \tag{C.39}$$

Consider the convergence tolerance ω_k as generated by the algorithm. By construction

$$\omega_k \leq \omega_s \mu_k \tag{C.40}$$

for all k. (This follows by definition if Step 4 of the algorithm occurs and because the penalty parameter is unchanged while ω_k is reduced when Step 3 occurs.) As Lemma 5.2, part (iii), ensures that (5.4) is satisfied for all k, we may apply Lemma 5.1 to the iterates generated by the algorithm. We identify the set $\mathcal{K}$ with the complete set of integers. As we are currently assuming that μ_k converges to zero, we can ensure that μ_k is sufficiently small so that Lemma 5.1 applies to Algorithm A.1 and thus that there is an integer k_1 and constants $a_9, \cdots, a_{13}$ so that (5.7) and (5.8) hold for all $k \geq k_1$. In particular, if we choose

$$\chi = \chi_0 \stackrel{\text{def}}{=} 1. \tag{C.41}$$

we obtain the bounds

$$\|(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}^*}\| \leq a_9 \omega_k + (a_{10} + a_{11}) \mu_k \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + + a_{12} \sigma_k \|\lambda_{k,\mathcal{I}^*}\| \tag{C.42}$$

and

$$\begin{aligned} \| [c_{k,i} \bar{\lambda}_{k,i} / \sqrt{\lambda_{k,i}}]_{i=1}^m \| \leq \quad & \mu_k [a_9 \omega_k + (a_{11} + a_{13}) \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + \\ & (1 + (1 + a_{12}) \sigma_k) \|\lambda_{k,\mathcal{I}^*}\|] \end{aligned} \tag{C.43}$$

for all $k \geq k_1$, from (C.39) and the inclusion $\mathcal{A}_2^* \subseteq \mathcal{A}^*$. Moreover, as Lemma 5.2, part (ii), ensures that θ_k converges to zero, there is an integer k_2 for which

$$\sigma_k \leq \mu_k \tag{C.44}$$

for all $k \geq k_2$. Thus, combining (C.39), (C.42), (C.43) and (C.44), we have that

$$\|(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}^*}\| \leq a_9 \omega_k + a_{27} \mu_k \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + + a_{12} \mu_k \|\lambda_{k,\mathcal{I}^*}\| \tag{C.45}$$

and

$$||[c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}}]_{i=1}^{m}|| \leq \quad \mu_k\,[a_9\omega_k + a_{28}\|(\lambda_k-\lambda_*)_{\mathcal{A}^*}\|+ \\ a_{29}\|\lambda_{k,\mathcal{I}^*}\|] \tag{C.46}$$

for all $k \geq \max(k_1, k_2)$, where $a_{27} \stackrel{\text{def}}{=} a_{10} + a_{11}$, $a_{28} \stackrel{\text{def}}{=} a_{11} + a_{13}$ and $a_{29} \stackrel{\text{def}}{=} 2 + a_{12}$.

Now, let k_3 be the smallest integer such that

$$\mu_k^{\frac{1}{6}} \leq \frac{\eta_s}{\omega_s a_{30}}, \tag{C.47}$$

$$\mu_k^{\frac{5}{6}} \leq \min\left(1, \frac{1}{a_{31}}\right), \tag{C.48}$$

$$\mu_k \leq \frac{\eta_s}{2\omega_s a_9} \tag{C.49}$$

and

$$\mu_k \leq \frac{\eta_s}{2\omega_s(a_{29} + a_{28}a_{31})} \tag{C.50}$$

for all $k \geq k_3$, where $a_{30} \stackrel{\text{def}}{=} a_9 + a_{28} + a_{29}$ and $a_{31} \stackrel{\text{def}}{=} a_9 + a_{12} + a_{27}$. Furthermore, let k_4 be such that

$$\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| \leq \omega_s \quad \text{and} \quad \|\lambda_{k,\mathcal{I}^*}\| \leq \omega_s \tag{C.51}$$

for all $k \geq k_4$.

Finally define $k_5 = \max(k_1, k_2, k_3, k_4)$, let Γ be the set $\{k|$ Step 4 is executed at iteration $k-1$ and $k \geq k_5\}$ and let k_0 be the smallest element of Γ. By assumption, Γ has an infinite number of elements.

For iteration k_0, $\omega_{k_0} = \omega_s\mu_{k_0}$ and $\eta_{k_0} = \eta_s\mu_{k_0}$. Then (C.46) gives

$$\begin{aligned} &||[c_{k_0,i}\bar{\lambda}_{k_0,i}/\sqrt{\lambda_{k_0,i}}]_{i=1}^{m}|| \\ &\quad\leq \mu_{k_0}\,[a_9\omega_{k_0} + a_{28}\|(\lambda_{k_0}-\lambda_*)_{\mathcal{A}^*}\| + a_{29}\|\lambda_{k_0,\mathcal{I}^*}\|] \\ &\quad\leq \omega_s(a_9 + a_{28} + a_{29})\mu_{k_0} = \omega_s a_{30}\mu_{k_0} \quad \text{[from (C.51)]} \\ &\quad\leq \eta_s \mu_{k_0}^{\frac{5}{6}} = \eta_{k_0} \quad \text{[from (C.47)].} \end{aligned} \tag{C.52}$$

Thus, from (C.52), Step 3 of Algorithm A.1 will be executed with $\lambda_{k_0+1} = \bar{\lambda}(x_{k_0}, \lambda_{k_0}, s_{k_0})$. Inequality (C.45), in conjunction with (C.40) and (C.51) guarantees that

$$\begin{aligned} \|(\lambda_{k_0+1}-\lambda_*)_{\mathcal{A}^*}\| &\leq a_9\omega_{k_0} + a_{27}\mu_{k_0}\|(\lambda_{k_0}-\lambda_*)_{\mathcal{A}^*}\| + a_{12}\mu_{k_0}\|\lambda_{k_0,\mathcal{I}^*}\| \\ &\leq a_9\omega_s\mu_{k_0} + a_{27}\omega_s\mu_{k_0} + a_{12}\omega_s\mu_{k_0} \\ &\leq \omega_s a_{31}\mu_{k_0}. \end{aligned} \tag{C.53}$$

Furthermore, inequality (4.1), in conjunction with (4.8), (C.44), and (C.51) ensures that

$$\|\lambda_{k_0+1,\mathcal{I}^*}\| \leq \sigma_{k_0}\|\lambda_{k_0,\mathcal{I}^*}\| \leq \omega_s\mu_{k_0}. \tag{C.54}$$

We shall now suppose that Step 3 is executed for iterations $k_0 + i$, $(0 \leq i \leq j)$, and show that

$$\|(\lambda_{k_0+i+1}-\lambda_*)_{\mathcal{A}^*}\| \leq \omega_s a_{31}\mu_{k_0}^{1+\frac{1}{6}i} \tag{C.55}$$

and

$$\|\lambda_{k_0+i+1,\mathcal{I}^*}\| \leq \omega_s \mu_{k_0}^{1+\frac{1}{6}i}. \tag{C.56}$$

Inequalities (C.53) and (C.54) show that this is true for $j = 0$. We aim to show that the same is true for $i = j + 1$. Under our supposition, we have, for iteration $k_0 + j + 1$, that $\mu_{k_0+j+1} = \mu_{k_0}$, $\omega_{k_0+j+1} = \omega_s \mu_{k_0}^{j+2}$ and $\eta_{k_0+j+1} = \eta_s \mu_{k_0}^{1+\frac{1}{6}(j+1)}$. Then (C.46) gives

$$\begin{aligned}
&||[c_{k_0+j+1,i}\bar{\lambda}_{k_0+j+1,i}/\sqrt{\lambda_{k_0+j+1,i}}]_{i=1}^{m}|| \\
&\quad \leq \mu_{k_0}\left[a_9\omega_s\mu_{k_0}^{j+2} + a_{28}\|(\lambda_{k_0+j+1} - \lambda_*)_{\mathcal{A}^*}\| + a_{29}\|\lambda_{k_0+j+1,\mathcal{I}^*}\|\right] \\
&\quad \leq \mu_{k_0}\left[a_9\omega_s\mu_{k_0}^{j+2} + a_{28}a_{31}\omega_s\mu_{k_0}^{1+\frac{1}{6}j} + a_{29}\omega_s\mu_{k_0}^{1+\frac{1}{6}j}\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{[from (C.55)-(C.56)]} \\
&\quad \leq \omega_s(a_9 + (a_{29} + a_{28}a_{31})\mu_{k_0}^{2+\frac{1}{6}j} \\
&\quad \leq \eta_s\mu_{k_0}^{1+\frac{1}{6}j} = \eta_{k_0+j+1} \qquad\qquad \text{[from (C.49)-(C.50)]}.
\end{aligned} \tag{C.57}$$

Thus, from (C.57), Step 3 of Algorithm A.1 will be executed with $\lambda_{k_0+j+2} = \bar{\lambda}(x_{k_0+j+1}, \lambda_{k_0+j+1}, s_{k_0+j+1})$. Inequality (C.45) then guarantees that

$$\begin{aligned}
&\|(\lambda_{k_0+j+2} - \lambda_*)_{\mathcal{A}^*}\| \\
&\quad \leq a_9\omega_{k_0+j+1} + a_{27}\mu_{k_0+j+1}\|(\lambda_{k_0+j+1} - \lambda_*)_{\mathcal{A}^*}\| + a_{12}\mu_{k_0+j+1}\|\lambda_{k_0+j+1,\mathcal{I}^*}\| \\
&\quad \leq a_9\omega_s\mu_{k_0}^{j+2} + \omega_s(a_{27}a_{31}\mu_{k_0}^{\frac{5}{6}} + a_{12})\mu_{k_0}^{1+\frac{1}{6}(j+1)} \qquad \text{[from (C.55)-(C.56)]} \\
&\quad \leq \omega_s a_{31}\mu_{k_0}^{1+\frac{1}{6}(j+1)} \qquad\qquad \text{[from (C.48)]}
\end{aligned} \tag{C.58}$$

which establishes (C.55) for $i = j + 1$.

Furthermore, inequalities (4.1) and (C.44) ensure that

$$\begin{aligned}
\|\lambda_{k_0+j+2,\mathcal{I}^*}\| \leq{}& \sigma_{k_0+j+1}\|\lambda_{k_0+j+1,\mathcal{I}^*}\| \leq \mu_{k_0+j+1}\|\lambda_{k_0+j+1,\mathcal{I}^*}\| \qquad \text{[from (4.8)]} \\
\leq{}& \omega_s\mu_{k_0}^{1+\frac{1}{6}(j+1))} \qquad\qquad \text{[from (C.56)]}
\end{aligned} \tag{C.59}$$

which establishes (C.56) for $i = j+1$. Hence, Step 3 of the algorithm is executed for all iterations $k \geq k_0$. But this implies that Γ is finite which contradicts the assumption that Step 4 is executed infinitely often. Hence the theorem is proved. ■

C.4 Proof of Theorem 5.4

We proceed by considering an example which has more than one Kuhn-Tucker point and for which the optimal Lagrange multipliers are distinct. We consider a sequence of iterates which is converging satisfactorily to a single Kuhn-Tucker point $(x_{*,1}, \lambda_{*,1})$ (and thus the penalty parameter has settled down to a single value). We now introduce an "extra" iterate x_k near to a different Kuhn-Tucker point $(x_{*,2}, \lambda_{*,2})$. We make use of the identity

$$c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}} = \mu_k(\lambda_{k,i} - \bar{\lambda}_{k,i}), \tag{C.1}$$

derived from (2.5) and (A.3), to show that if the Lagrange multiplier estimate $\bar{\lambda}_{k,i}$ calculated at x_k is a sufficiently "accurate" approximation of $\lambda_{*,2}$ (while $\lambda_{k,i}$ is an "accurate" representation of $\lambda_{*,1}$), the acceptance test (A.7) will fail and the penalty parameter will be reduced. Moreover, we show that this behaviour can be indefinitely repeated.

To be specific, we consider the following problem:

$$\underset{x\in\Re}{\text{minimize}} \quad \epsilon(x-1)^2 \quad \text{such that} \quad c(x) = x^2 - 4 \geq 0, \tag{C.2}$$

where ϵ is a (yet to be specified) positive constant. It is straightforward to show that the problem has two local solutions which occur at the Kuhn-Tucker points

$$(x_{*,1}, \lambda_{*,1}) = \left(-2, \frac{3\epsilon}{2}\right) \quad \text{and} \quad (x_{*,2}, \lambda_{*,2}) = \left(2, \frac{\epsilon}{2}\right). \tag{C.3}$$

and that the constraint is active at both local solutions. Moreover, there are no specific bounds on the variable in the problem and hence $P(x, \nabla_x \Psi(x, \lambda, s)) = \nabla_x \Psi(x, \lambda, s)$ for all x.

We intend to construct a cycle of iterates x_{k+i}, $i = 0, \ldots, j$, for some integer j, which are allowed by Algorithm A.1. The penalty parameter remains fixed throughout the cycle until it is reduced at the end of the final iteration. We start with $\lambda_0 = \lambda_{*,1}$. We also pick ϵ so that

$$\epsilon \leq \min\left(\frac{2}{3}, \frac{\omega_s}{(6 + \frac{1}{1-\mu_0})}, \frac{2\eta_s}{3\mu_0^{\frac{1}{6}}}\right). \tag{C.4}$$

We define j to be the smallest integer for which

$$\mu_0^{\frac{j-1}{6}} < \tfrac{1}{2}\epsilon/\eta_s. \tag{C.5}$$

We let μ denote the value of the penalty parameter at the start of the cycle.

i = 0: We have $\omega_k = \omega_s \mu$ and $\eta_k = \eta_s \mu^{\frac{5}{6}}$. We are given $\lambda_k = \lambda_{*,1}$. We pick x_k near $x_{*,1}$ so that $\bar{\lambda}_k = (1-\mu)\lambda_{*,1}$. We show that such a choice guarantees that the convergence and acceptance tests (A.5) and (A.7) are satisfied and thus Step 3 of the algorithm is executed.

i = 1, ... , j - 2: We have $\omega_{k+i} = \omega_s \mu^{1+i}$ and $\eta_{k+i} = \eta_s \mu^{\frac{5+i}{6}}$. We have $\lambda_{k+i} = (1-\mu_i)\lambda_{*,1}$. We pick x_{k+i} near $x_{*,1}$ so that $\bar{\lambda}_{k+i} = (1-\mu_{i+1})\lambda_{*,1}$. We again show that such a choice guarantees that the convergence and acceptance tests (A.5) and (A.7) are satisfied and thus Step 3 of the algorithm is executed.

i = j - 1: We have $\omega_{k+i} = \omega_s \mu^{1+i}$ and $\eta_{k+i} = \eta_s \mu^{\frac{5+i}{6}}$. We have $\lambda_{k+i} = (1-\mu_i)\lambda_{*,1}$. We pick x_{k+i} near $x_{*,1}$ so that $\bar{\lambda}_{k+i} = \lambda_{*,1}$. Once again, we show that such a choice guarantees that the convergence and acceptance tests (A.5) and (A.7) are satisfied and thus Step 3 of the algorithm is executed.

i = j: We have $\omega_{k+j} = \omega_s \mu^{1+j}$ and $\eta_{k+j} = \eta_s \mu^{\frac{5+j}{6}}$. We have $\lambda_{k+j} = \lambda_{*,1}$. We pick x_{k+i} as the local minimizer of the Lagrangian barrier function which is larger than $x_{*,2}$ which trivially ensures that the convergence test (A.5) is satisfied. We also show that the acceptance test (A.7) is violated at this point so that Step 4 of the algorithm will be executed and the penalty parameter reduced.

It is clear that if an infinite sequence of such cycles occur, the penalty parameter μ_k will converge to zero. We now show that this is possible.

If a is a real number, we will make entensive use of the trivial inequalities

$$1 \leq \sqrt{1+a} \leq 1+a \quad \text{whenever} \quad a \geq 0 \tag{C.6}$$

and

$$1-a \leq \sqrt{1-a} \leq 1-\tfrac{1}{2}a \quad \text{whenever} \quad 0 \leq a \leq 1. \tag{C.7}$$

We also remind the reader that

$$\mu \leq \mu_0 < 1. \tag{C.8}$$

1. Let

$$x_k = -2\sqrt{1+\tfrac{1}{4}\mu s_k/(1-\mu)}, \tag{C.9}$$

where the shift $s_k = \mu\sqrt{\frac{3}{2}\epsilon}$. Then it is easy to verify that $\bar{\lambda}_k = (1-\mu)\lambda_{*,1}$. Moreover

$$\begin{aligned} \nabla_x \Psi(x_k, \lambda_k, s_k) = &\ 2\epsilon(x_k - 1) - 3\epsilon(1-\mu)x_k = -\epsilon(2+(1-3\mu)x_k) \\ = &\ -2\epsilon\left(1-(1-3\mu)\sqrt{1+\mu s_k/(4(1-\mu))}\right). \end{aligned} \tag{C.10}$$

Taking norms of (C.10) and using (C.6) yields

$$\|P(x_k, \nabla_x \Psi(x_k, \lambda_k, s_k))\| \leq \begin{cases} 6\epsilon\mu & \text{if} \quad \mu \leq \frac{1}{3} \\ 2\epsilon\mu\left(3+\dfrac{(3\mu-1)s_k}{4(1-\mu)}\right) & \text{otherwise.} \end{cases} \tag{C.11}$$

Now (C.4) implies that $s_k \leq \mu < 1$ and thus we obtain the overall bound

$$\|P(x_k, \nabla_x \Psi(x_k, \lambda_k, s_k))\| \leq \epsilon\left(6+\frac{1}{1-\mu_0}\right)\mu \tag{C.12}$$

from (C.8) and (C.11). But then (C.12) and (C.4) give

$$\|P(x_k, \nabla_x \Psi(x_k, \lambda_k, s_k))\| \leq \omega_s \mu = \omega_k, \tag{C.13}$$

as $1 \leq \omega_s/(6+1/(1-\mu_0))\epsilon$. Furthermore, from (C.1) and (C.4),

$$\|c(x_k)\bar{\lambda}_k/\sqrt{\lambda_k}\| = \mu\|\lambda_{k,i} - \bar{\lambda}_{k,i}\| = \tfrac{3}{2}\mu^2\epsilon \leq \eta_s \mu^{\frac{5}{6}} = \eta_k. \tag{C.14}$$

as $\mu^{\frac{7}{6}} \leq \mu_0^{\frac{1}{6}} \leq 1 \leq 2\eta_s/3\epsilon$. Thus x_k satisfies (A.5) and (A.7) and hence Step 3 of the algorithm will be executed. Therefore, in particular, $\omega_{k+1} = \omega_s\mu^2$, $\eta_{k+1} = \eta_s\mu$ and $\lambda_{k+1} = (1-\mu)\lambda_{*,1}$.

2. For $i = 1, \cdots j-2$, let

$$x_{k+i} = -2\sqrt{1-\tfrac{1}{4}\mu^i(1-\mu)s_{k+i}/(1-\mu^{i+1})}, \tag{C.15}$$

where the shift $s_{k+i} = \mu\sqrt{\frac{3}{2}(1-\mu^i)\epsilon}$. Note that (C.15) is well defined as the second term within the square root is less than $\frac{1}{4}$ in magnitude because (C.4) and (C.8) imply that $s_k < \mu$ and $\mu^i(1-\mu)/(1-\mu^{i+1}) < 1$. It is then easy to verify that $\bar{\lambda}_{k+i} = (1-\mu^{i+1})\lambda_{*,1}$. Moreover,

$$\begin{aligned} \nabla_x \Psi(x_{k+i}, \lambda_{k+i}, s_{k+i}) = &\ 2\epsilon(x_{k+i}-1) - 3\epsilon(1-\mu^{i+1})x_{k+i} \\ = &\ -\epsilon(2+(1-3\mu^{i+1})x_{k+i}) \\ = &\ -2\epsilon\left(1-(1-3\mu^{i+1})\sqrt{1-\tfrac{\mu^i(1-\mu)s_{k+i}}{4(1-\mu^{i+1})}}\right). \end{aligned} \tag{C.16}$$

Now suppose $\mu^{i+1} \le \frac{1}{3}$. Then (C.16), (C.7), (C.8) and $s_k \le \mu$ yield

$$\begin{aligned} \|P(x_{k+i}, \nabla_x \Psi(x_{k+i}, \lambda_{k+i}, s_{k+i})\| \le & \; 2\epsilon \left(1 - (1 - 3\mu^{i+1})(1 - \tfrac{\mu^i(1-\mu)s_{k+i}}{8(1-\mu^{i+1})})\right) \\ = & \; 2\epsilon \left(3\mu^{i+1} + \tfrac{\mu^i(1-\mu)(1-3\mu^{i+1})s_{k+i}}{8(1-\mu^{i+1})}\right) \\ \le & \; 2\epsilon\mu^{i+1} \left(3 + \tfrac{(1-\mu)(1-3\mu^{i+1})}{8(1-\mu^{i+1})})\right) \\ \le & \; 2\epsilon\mu^{i+1} \left(3 + \tfrac{1}{8(1-\mu_0)}\right). \end{aligned} \tag{C.17}$$

If, on the other hand, $\mu^{i+1} > \frac{1}{3}$, the same relationships give

$$\begin{aligned} \|P(x_{k+i}, \nabla_x \Psi(x_{k+i}, \lambda_{k+i}, s_{k+i})\| \le & \; 2\epsilon \left(1 - (1 - 3\mu^{i+1})(1 - \tfrac{\mu^i(1-\mu)s_{k+i}}{4(1-\mu^{i+1})})\right) \\ = & \; 2\epsilon \left(3\mu^{i+1} + \tfrac{\mu^i(1-\mu)(1-3\mu^{i+1})s_{k+i}}{4(1-\mu^{i+1})}\right) \\ \le & \; 6\epsilon\mu^{i+1}. \end{aligned} \tag{C.18}$$

Thus, combining (C.17) and (C.18), we certainly have that

$$\|P(x_{k+i}, \nabla_x \Psi(x_{k+i}, \lambda_{k+i}, s_{k+i})\| \le \epsilon \left(6 + \frac{1}{1-\mu_0}\right) \mu^{i+1}. \tag{C.19}$$

But then (C.19) and (C.4) give

$$\|P(x_{k+i}, \nabla_x \Psi(x_{k+i}, \lambda_{k+i}, s_{k+i})\| \le \omega_s \mu^{1+i} = \omega_{k+i}, \tag{C.20}$$

as $1 \le \omega_s/((6 + 1/(1-\mu_0))\,\epsilon)$. Furthermore, from (C.1) and (C.4),

$$\begin{aligned} \|c(x_{k+i})\bar{\lambda}_{k+i}/\sqrt{\lambda_{k+i}}\| = & \; \mu\|\lambda_{k+i,i} - \bar{\lambda}_{k+i,i}\| = \tfrac{3}{2}\mu^{i+1}(1-\mu)\epsilon \\ \le & \; \tfrac{3}{2}\mu^{i+1}\epsilon \le \eta_s \mu^{\frac{5+i}{6}} = \eta_{k+i} \end{aligned} \tag{C.21}$$

as $\mu^{\frac{1+5i}{6}} \le \frac{2}{3}\eta_s/\epsilon$. Thus x_{k+i} satisfies (A.5) and (A.7) and hence Step 3 of the algorithm will be executed. Therefore, in particular, $\omega_{k+i+1} = \omega_s\mu^{i+2}$, $\eta_{k+i+1} = \eta_s\mu^{1+\frac{i}{6}}$ and $\lambda_{k+i+1} = (1 - \mu^{i+1})\lambda_{*,1}$.

3. Let

$$x_{k+j-1} = -2\sqrt{1 - \tfrac{1}{4}\mu^{j-1}s_{k+j-1}}, \tag{C.22}$$

where the shift $s_{k+j-1} = \mu\sqrt{\frac{3}{2}(1-\mu^{j-1})\epsilon}$. Once again, (C.4) and (C.8) imply that $s_{k+j-1} \le \mu$ and thus (C.22) is well defined. Furthermore it is easy to verify that $\bar{\lambda}_{k+j-1} = \lambda_{*,1}$. Moreover

$$\begin{aligned} \nabla_x \Psi(x_{k+j-1}, \lambda_{k+j-1}, s_{k+j-1}) = & \; 2\epsilon(x_{k+j-1} - 1) - 3\epsilon x_{k+j-1} \\ = & \; -\epsilon(2 + x_{k+j-1}) \\ = & \; -2\epsilon\left(1 - \sqrt{1 - \tfrac{1}{4}\mu^{j-1}s_{k+j-1}}\right). \end{aligned} \tag{C.23}$$

But then (C.7), (C.23) and the inequality $s_{k+j-1} \le \mu$ yield

$$\|P(x_{k+j-1}, \nabla_x \Psi(x_{k+j-1}, \lambda_{k+j-1}, s_{k+j-1})\| \le \tfrac{1}{2}\epsilon\mu^{j-1}s_{k+j-1} \le \tfrac{1}{2}\epsilon\mu^j. \tag{C.24}$$

Thus, (C.24) and (C.4) give

$$\|P(x_{k+j-1}, \nabla_x \Psi(x_{k+j-1}, \lambda_{k+j-1}, s_{k+j-1})\| \le \omega_s\mu^j = \omega_{k+j-1}, \tag{C.25}$$

as $1 \leq \omega_s/((6+1/(1-\mu_0))\epsilon) < 2\omega_s/\epsilon$. Furthermore, from (C.1) and (C.4),

$$\begin{aligned} \|c(x_{k+j-1})\bar{\lambda}_{k+j-1}/\sqrt{\lambda_{k+j-1}}\| = & \ \mu\|\lambda_{k+j-1,i} - \bar{\lambda}_{k+j-1,i}\| = \tfrac{3}{2}\mu^j\epsilon \\ \leq & \ \eta_s\mu^{\frac{4+j}{6}} = \eta_{k+j-1} \end{aligned} \tag{C.26}$$

as $\mu^{\frac{5j-4}{6}} \leq \frac{2}{3}\eta_s/\epsilon$. Thus x_{k+j-1} satisfies (A.5) and (A.7) and hence Step 3 of the algorithm will be executed. Therefore, in particular, $\omega_{k+j} = \omega_s\mu^{1+j}$, $\eta_{k+j} = \eta_s\mu^{\frac{5+j}{6}}$ and $\lambda_{k+j} =)\lambda_{*,1}$.

4. We pick x_{k+j} as the largest root of the nonlinear equation

$$\phi(x) \equiv 2(x-1) - \frac{3xs_{k+j}}{x^2 - 4 + s_{k+j}} = 0, \tag{C.27}$$

where $s_{k+j} = \mu\sqrt{\frac{3}{2}\epsilon}$, Equation (C.27) defines the stationary points of the Lagrangian barrier function for the problem (C.3). This choice ensures that (A.5) is trivially satisfied. As $\phi(2) = -4$ and $\phi(x)$ increases without bound as x tends to infinity, the largest root of (C.27) is greater than 2. The function $\bar{\lambda}$ given by (2.4) is a decreasing function of x as x grows beyond 2. Now let $\hat{x} = \sqrt{4 + \frac{1}{2}s_{k+j}}$. It is easy to show that $\bar{\lambda}(\hat{x}, \lambda_{*,1}, s_{k+j}) = \epsilon$. Moreover $\phi(\hat{x}) = 2(\hat{x}-1) - 2\hat{x} = -2$. Therefore, $x_{k+j} > \hat{x}$ and thus $\bar{\lambda}_{k+j} < \epsilon$. But then, using (C.1),

$$\begin{aligned} \|c(x_{k+j})\bar{\lambda}_{k+j}/\sqrt{\lambda_{k+j}}\| = & \ \mu(\lambda_{k+j} - \bar{\lambda}_{k+j}) \geq \mu(\tfrac{3}{2}\epsilon - \epsilon) = \tfrac{1}{2}\epsilon\mu \\ > & \ \eta_s\mu^{\frac{5+j}{6}} = \eta_{k+j} \end{aligned} \tag{C.28}$$

from (C.5). Thus the test (A.7) is violated and the penalty parameter subsequently reduced. This ensures that $\omega_{k+j+1} = \omega_s\mu$, $\eta_{k+j+1} = \eta_s\mu$ and $\lambda_{k+j+1} = \lambda_{*,1}$.

Hence, a cycle as described at the start of this section is possible and we conclude that, in the absence of AS6, the penalty parameter generated by Algorithm A.1 may indeed converge to zero. ∎

C.5 Proof of Theorem 5.5

Firstly, as (A.7) holds for all $k \geq k_0$, the penalty parameter μ_k remains fixed at some value μ_k, say, the convergence tolerances satisfy

$$\omega_{k+1} = \omega_k\mu_k \quad \text{and} \quad \eta_{k+1} = \eta_k\mu_k^{\frac{1}{6}} \tag{C.29}$$

and $\lambda_{k+1} = \bar{\lambda}_k$ for all $k > k_0$.

The Q-superlinear convergence of the Lagrange multiplier estimates for inactive constraints follows directly from Theorem 4.2, part (iv). Lemma 5.2, part (ii), the convergence of θ_k to zero and the relationships (4.1) and (4.8) then give that

$$\|\lambda_{k+1,\mathcal{I}^*}\| \leq \mu_k\|\lambda_{k,\mathcal{I}^*}\| \tag{C.30}$$

for all k sufficiently large.

The identities (2.5), (A.3) and the assumption that (A.7) holds for all $k \geq k_0$ gives

$$\begin{aligned} \|(\lambda_{k+1} - \lambda_k)_{\mathcal{A}^*}\| = & \ \mu_k^{-1}\left\|\left[c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}}\right]_{i\in\mathcal{A}^*}\right\| \\ \leq & \ \mu_k^{-1}\left\|\left[c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}}\right]_{i=1}^m\right\| \leq \mu_k^{-1}\eta_k \end{aligned} \tag{C.31}$$

for all such k. But then the triangle inequality and (C.31) implies that

$$\begin{aligned} \|(\lambda_{k+j}-\lambda_*)_{\mathcal{A}^*}\| \leq & \ \|(\lambda_{k+j+1}-\lambda_*)_{\mathcal{A}^*}\| + \|(\lambda_{k+j+1}-\lambda_{k+j})_{\mathcal{A}^*}\| \\ \leq & \ \|(\lambda_{k+j+1}-\lambda_*)_{\mathcal{A}^*}\| + \mu_k^{-1}\eta_{k+j} \end{aligned} \tag{C.32}$$

for all $k \geq k_0$. Thus summing (C.32) from $j=0$ to $j_{\max}-1$ and using the relationship (C.29) yields

$$\begin{aligned} \|(\lambda_k-\lambda_*)_{\mathcal{A}^*}\| \leq & \ \|(\lambda_{k+j_{\max}}-\lambda_*)_{\mathcal{A}^*}\| + \mu_k^{-1}\textstyle\sum_{i=0}^{j_{\max}-1}\eta_{k+j} \\ \leq & \ \|(\lambda_{k+j_{\max}}-\lambda_*)_{\mathcal{A}^*}\| + \mu_k^{-1}\eta_k(1-\mu_k^{\frac{1}{6}j_{\max}})/(1-\mu_k) \end{aligned} \tag{C.33}$$

Hence, letting $j_{\max}$ tend to infinity and recalling that λ_k converges to λ_*, (C.33) gives

$$\|(\lambda_k-\lambda_*)_{\mathcal{A}^*}\| \leq \frac{\mu_k^{-1}\eta_k}{1-\mu_k} \tag{C.34}$$

for all $k \geq k_0$. As η_k converges to zero R-linearly, with R-factor $\mu_k^{\frac{1}{6}}$, (C.34) gives the required result *(ii)*.

The remainder of the proof parallels that of Lemma 5.1. As (A.7) holds for all sufficiently large k, the definition (C.4) of $\mathcal{A}$ and the bound (C.8), ensure that

$$\begin{aligned} \|c_{k,\mathcal{A}}\| \leq & \ \left\|\left[\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i}\right]_{i\in\mathcal{A}}\right\| \left\|\left[c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}}\right]_{i\in\mathcal{A}}\right\| \\ \leq & \ a_{14}\mu_k^{\chi-1}\left\|\left[c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}}\right]_{i=1}^m\right\| \leq a_{14}\mu_k^{\chi-1}\eta_k. \end{aligned} \tag{C.35}$$

Thus combining (C.24) and (C.25), (C.30) and replacing (C.12) by (C.35), the bound on the right-hand side of (C.29) may be replaced by $a_{23}\omega_k + a_{24}\sigma_k\|\lambda_{k,\mathcal{I}^*}\| + a_{25}\mu_k^{1-\chi}\left\|\left[\sqrt{\lambda_{k,i}}\right]_{i\in\mathcal{A}_2^*}\right\| + a_{14}\mu_k^{\chi-1}\eta_k$ and consequently (C.31) replaced by

$$\begin{aligned} \Delta x_k \leq & \ M[a_{19}(\Delta x_k)^2 + a_{20}\Delta x_k\omega_k + a_{21}\omega_k^2 + \\ & \ a_{22}\sigma_k\|\lambda_{k,\mathcal{I}^*}\|(\omega_k+\Delta x_k) + a_{23}\omega_k + a_{24}\sigma_k\|\lambda_{k,\mathcal{I}^*}\| + \\ & \ a_{25}\mu_k^{1-\chi}\left\|\left[\sqrt{\lambda_{k,i}}\right]_{i\in\mathcal{A}_2^*}\right\| + a_{14}\mu_k^{\chi-1}\eta_k] \\ = & \ (Ma_{19}\Delta x_k + Ma_{20}\omega_k)\Delta x_k + (Ma_{21}\omega_k + Ma_{23})\omega_k + \\ & \ (Ma_{24} + Ma_{22}(\omega_k+\Delta x_k))\sigma_k\|\lambda_{k,\mathcal{I}^*}\| \\ & \ Ma_{25}\mu_k^{1-\chi}\left\|\left[\sqrt{\lambda_{k,i}}\right]_{i\in\mathcal{A}_2^*}\right\| + Ma_{14}\mu_k^{\chi-1}\eta_k. \end{aligned} \tag{C.36}$$

Hence, if k is sufficiently large that

$$\Delta x_k \leq 1/(4Ma_{19}), \quad \omega_k \leq \min(1, 1/(4Ma_{20})) \quad \text{and} \quad \sigma_k \leq 1, \tag{C.37}$$

(C.36) and (C.37) can be rearranged to give

$$\begin{aligned} \Delta x_k \leq & \ 2M[(a_{21}+a_{23})\omega_k + (a_{24}+2a_{22})\|\lambda_{k,\mathcal{I}^*}\| + \\ & \ a_{25}\mu_k^{1-\chi}\left\|\left[\sqrt{\lambda_{k,i}}\right]_{i\in\mathcal{A}_2^*}\right\| + a_{14}\mu_k^{\chi-1}\eta_k]. \end{aligned} \tag{C.38}$$

But then (C.26) and (C.38) give

$$\begin{aligned} \|x_k-x_*\| \leq & \ a_{32}\omega_k + a_{33}\|\lambda_{k,\mathcal{I}^*}\| \\ & \ +a_{34}\mu_k^{\chi-1}\eta_k + a_{35}\mu_k^{1-\chi}\left\|\left[\sqrt{\lambda_{k,i}}\right]_{i\in\mathcal{A}_2^*}\right\|, \end{aligned} \tag{C.39}$$

where $a_{32} \stackrel{\rm def}{=} 1 + 2M(a_{21} + a_{23})$, $a_{33} \stackrel{\rm def}{=} 2M(a_{24} + 2a_{22})$, $a_{34} \stackrel{\rm def}{=} 2Ma_{14}$ and $a_{35} \stackrel{\rm def}{=} 2Ma_{25}$. Each term on the right-hand-side of (C.39) converges at least R-linearly to zero; the R-factors (in order) being no larger than μ_k, μ_k, $\mu_k^{\frac{1}{6}}$ and $\mu_k^{\frac{1}{12}}$ respectively, following (C.29), (C.30) and (C.34). Hence (C.39) shows that x_k converges at least R-linearly with R-factor at most $\mu_k^{\frac{1}{12}}$.

C.6 Proof of Corollary 5.6

This follows directly from Theorem C.1 (Section C.3) as this ensures that (A.7) is satisfied for all k sufficiently large.

References

[Arrow and Solow, 1958] K. J. Arrow and R. M. Solow. Gradient methods for constrained maxima, with weakened assumptions. In K. J. Arrow, L. Hurwicz, and H. Uzawa, editors, *Studies in Linear and non-linear Programming*, pages 166–176, Stanford, USA, 1958. Stanford University Press.

[Avriel, 1976] M. Avriel. *Nonlinear Programming: Analysis and Methods.* Prentice-Hall, Englewood Cliffs, N.J., 1976.

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

[Breifeld and Shanno, 1993] M. G. Breifeld and D. F. Shanno. Computational experience with modified log-barrier methods for nonlinear programming. Research Report RRR 17-93, RUTCOR, Rutgers University, New Brunswick, NJ 08903, USA, 1993.

[Breitfeld and Shanno, 1993] M. G. Breitfeld and D. Shanno. Preliminary computational experience with modified log-barrier functions for large-scale nonlinear programming. Research Report RRR 08-93, Rutgers Center for Operations Research, New Brunswick, USA, 1993.

[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.*, 1991] 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(2):545–572, 1991.

[Conn *et al.*, 1992a] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. Technical Report 92/07, Department of Mathematics, FUNDP, Namur, Belgium, 1992.

[Conn *et al.*, 1992b] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. LANCELOT*: a Fortran package for large-scale nonlinear optimization (Release A).* Number 17 in Springer Series in Computational Mathematics. Springer Verlag, Heidelberg, Berlin, New York, 1992.

[Conn *et al.*, 1992c] 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*, pages 49–68. Longmans, 1992.

[Conn *et al.*, 1992d] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. On the number of inner iterations per outer iteration of a globally convergent algorithm for optimization with general nonlinear inequality constraints and simple bounds. Technical Report 92/14, Department of Mathematics, FUNDP, Namur, Belgium, 1992.

[Conn *et al.*, 1992e] A. R. Conn, Nick Gould, and Ph. L. Toint. A note on exploiting structure when using slack variables. LANCELOT B working note 1 92/18, Department of Mathematics, FUNDP, Namur, Belgium, 1992.

[Conn *et al.*, 1993] A. R. Conn, Nick Gould, and Ph. L. Toint. A note on using alternative second-order models for the subproblems arising in barrier function methods for minimization. Technical Report 93/16, Department of Mathematics, FUNDP, Namur, Belgium, 1993.

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

[Fiacco, 1983] A. V. Fiacco. *Introduction to sensitivity and stability analysis in nonlinear programming.* Academic Press, New York, 1983.

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

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

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

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

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

[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.

[Gruver and Sachs, 1980] W. A. Gruver and E. Sachs. *Algorithmic Methods in Optimal Control.* Pitman, Boston, USA, 1980.

[Güler *et al.*, 1993] O. Güler, D. den Hertog, C. Roos, T. Terlaky, and T. Tsuchiya. Degeneracy in interior point methods for linear programming. Technical report, Faculty of Technical Mathematics and Computer Science, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands, 1993.

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

[Hestenes, 1969] M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–320, 1969.

[Jarre and Saunders, 1991] F. Jarre and M. A. Saunders. Practical aspects of an interior point method for convex programming. Technical Report SOL 91-9, Operations Research Department, Stanford University, Stanford, USA, 1991.

[Jensen and Polyak, 1993] D. Jensen and R. Polyak. On the convergence of a modified barrier method for convex programming. Research Report RC 18570, IBM T. J. Watson Research Center, 1993.

[Jensen *et al.*, 1992] D. L. Jensen, R. Polyak, and R. Schneur. Numerical experience with modified barrier functions method for linear programming. Research Report RC 18415, IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, 1992.

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

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

[Kranich, 1991] E. Kranich. Interior point methods for mathematical programming: A bibliography. Discussion Paper 171, Institute of Economy and Operations Research, FernUniversität Hagen, P.O. Box1991. The bibliography can be accessed electronically by sending email to 'netlib@research.att.com' or 'netlib@ukc.ac.uk' with the message 'send intbib.bib from bib'.

[Lootsma, 1969] F. A. Lootsma. Hessian matrices of penalty functions for solving constrained optimization problems. *Philips Research Reports*, 24:322–331, 1969.

[Lustig *et al.*, 1991] I.J. Lustig, R.E. Marsten, and D.F. Shanno. Computational experience with a primal-dual interior point method for linear programming. *Linear Algebra and its Applications*, 152:191–222, 1991.

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

[Mifflin, 1975] R. Mifflin. Convergence bounds for nonlinear programming algorithms. *Mathematical Programming*, 8:251–271, 1975.

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

[Murray, 1971] W. Murray. Analytical expressions for eigenvalues and eigenvectors of the hessian matrices of barrier and penalty functions. *Journal of Optimization Theory and Applications*, 7:189–196, 1971.

[Murray, 1992] W. Murray. Ill-conditioning in barrier methods. In *Advances in numerical partial differential equations and optimization, Proceedings of the Sixth Mexico-United States Workshop*, Philadelphia, USA, 1992. SIAM.

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

[Nash *et al.*, 1993] S. G. Nash, R. Polyak, and A. Sofer. A numerical comparison of barrier and modified barrier methods for large-scale constrained optimization. Technical Report 93-02, Department of Operations Research, George Mason University, Fairfax, Virginia 22030, 1993.

[Nesterov and Nemirovsky, 1993] Y. Nesterov and A. Nemirovsky. *Self-concordant functions and polynomial-time methods in convex programming.* SIAM, Philadelphia, USA, 1993. To appear.

[Ortega and Rheinboldt, 1970] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables.* Academic Press, New York, 1970.

[Polyak, 1982] R. A. Polyak. Smooth optimization methods for solving nonlinear extremal and equilibrium problems with constraints. Paper presented at the XIth International Symposium on Mathematical Programming, Bonn, Federal Republic of Germany, August 1982.

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

[Ponceleón, 1990] D. B. Ponceleón. *Barrier methods for large-scale quadratic programming.* PhD thesis, Department of Computer Science, Stanford University, Stanford, California 94305, USA, 1990.

[Powell, 1969] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, London and New York, 1969. Academic Press.

[Powell, 1992a] M. J. D. Powell. Log barrier methods for semi-infinite programming calculations. Technical Report DAMTP NA11, Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge, UK, 1992.

[Powell, 1992b] M. J. D. Powell. Some convergence properties of the shifted log barrier method for linear programming. Technical Report DAMTP 1992/NA7, Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge, UK, 1992.

[Rendl *et al.*, 1993] F. Rendl, R. J. Vanderbei, and H. Wolkowicz. Primal-dual interior point algorithms for max-min eigenvalue problems. Technical Report CORR93-20, Faculty of Mathematics, University of Waterloo, 1993.

[Rockafellar, 1976] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.

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

[Wright, 1992] M. H. Wright. Interior methods for constrained optimization. volume 1 of *Acta Numerica*, pages 341–407. Cambridge University Press, New York, 1992.