Librería Británica, S.A.
Serapio Rendón 125, Col. San Rafael, 06470-México, D.F.

TEL. 7-05-05-85
CON 8 LINEAS

APDO. POSTAL: 30-502
CABLES: GREPEMEX

SEPTEMBER 18 TH, 1989.

UNIVERSITY OF WATERLOO
DEPARTMENT OF COMPUTER SCIENCE
WATERLOO, ONTARIO N2L 3G1
CANADA.

OCT - 5 1989

DEARS SIRS:

PLEASE PLACE THE AMOUNT OF THE ENCLOSED CHECK No.3300 M. BANK DALLAS FOR $ 8.00 (EIGHT DOLLARS 00/100 USCY), INVOICE No. 8839 IN FULL.

received Oct. 10/89

THANKING YOU FOR YOUR ATTENTION, WE REMAIN.

YOUR SINCERELY

LIBRERIA BRITANICA, S. A.

JOHN H. GREPE FIELD
GENERAL MANAGER

TeleFax (5)-5352009

University of Waterloo
Department of Computer Science
Waterloo, Ontario N2L 3G1

June 26, 1989

INVOICE

Libreria Britanica
Serapio Rendon 125
Col. San Rafael
06470
Mexico, D.F.
Order Number 4319

REPORT(S) ORDERED

CS-88-39, 88-38

TOTAL COST $8.00 (This price includes postage)

Would you please make your cheque or international bank draft payable to the **Computer Science Department, University of Waterloo** and forward to my attention.

Thanking you in advance.

Yours truly,

Susan DeAngelis

Susan DeAngelis (Mrs.)
Research Report Secretary
Computer Science Dept.

/sd

Encl.

Librería Británica S.A.

address -
SERAPIO RENDON 125, COL. SAN RAFAEL 06470 MEX. D.F.
☎ 566 69 89 535 23 42 535 80 48 535 82 68
CABLE: GREPEMEX

ORDER TO

UNIVERSITY OF WATERLOO,
DEPARTMENT OF COMPUTER SCIENCE,
WATERLOO, ONTARIO N2L 3G1
CANADA

ORDER NUMBER: Nº 4319 SR

DATE: 25 MAYO 1989.

PLEASE QUOTE OUR ORDER NUMBER

QUANTITY	TITLE & AUTOR	PRICE
1	CONN, A.R.- THE COMPUTATIONAL STRUCTURES AND CHARACTERIZATION OF NONLINEAR NONLINEAR DISCRETE CHEBYSHEV PROBLEMS (CS-88-39 RESEARCH REPORTS 1988 september to december)	
1	A GLOBALLY CONVERGENT AUGMENTED LAGRANGIAN ALGORITHM FOR OPTIMIZATION WITH GENERAI CONSTRAINTS AND SIMPLE BOUNDS.(CS-88-38 RESEARCH REPORTES 1988 september to december.	

IMPORTANT NOTICE

Please include a HANDSINGED copy of the invoice in every package.
Please follow shipping instructions stated at the bottom of this order. Thank you

IF YOU CANNOT SUPPLY WITHIN 3 MONTHS, DO NOT BACK ORDER
SEND PAPERBACK IF AVAILABLE, UNLESS OTHERWISE STATED

SHIP TO:

☐ CLARK WORLDWIDE TRANSPORTATION, INC.
(Librería Británica) México
C/O HIGHWAY FILM DELIVERY
80 S. James Stret Kansas City, KS 66118
SHIP VIA:
Prepaid Freight & Include Packing list

☐ MERCURY AIR FREIGHT INTL. LTD.
Unitair Center
Great South West Road East Bedfont
FELTHAM, MIDDLESEX, TW 14 8NJ
For separate consolidation to:
LIBRERIA BRITANICA, S.A. MEXICO

☐ TeleFax (5)- 5 3 5 2 0 0 9
AGENCIA ADUANAL GALVAN FLOGALVA S.C.
Comercial Forwarders of Laredo
2301 Moctezuma st. P.O. BOX 326
LAREDO TEXAS 78040
For consignment to:
LIBRERIA BRITANICA Mexico City, Mexico

- PLEASE BE SURE TO STATE ON ALL INVOICES (for customs purposes):
  Los Títulos de los libros amparados por la presente factura no pueden ser Traducidos al español por ser Títulos y ediciones originales en idioma distinto al español. El valor de la presente factura es de_____ (U.S; L.E; Fr. Fr; etc.)
- AIRFREIGHT AND SEAFREIGHT Two copies HANDSIGNED with shipment in package N° 1 or with B/Lading. Two copies by airmail to us.
- N.B. WE ARE A TRADE BOOKSELLER. PLEASE GIVE US YOUR BEST DISCOUNT SEND SHIPMENT AS INSTRUCTED. DO NOT ADD SHIPPING CHARGES AS WE ALREADY PAY HEAVY TRANSPORT CHANGES.

T 29 Rev. 8/72

THE UNIVERSITY OF TENNESSEE
SPECIAL REMITTANCE AND ORDER FORM

(TO BE SENT TO THE TREASURER'S OFFICE IN DUPLICATE IN LIEU OF REQUISITION ONLY WHEN REMITTANCE IN ADVANCE IS REQUIRED AND IT IS IMPOSSIBLE TO OBTAIN AN INVOICE. SUPPORT, SUCH AS ORDER FORM, SUBSCRIPTION NOTICE, ETC., SHOULD BE FURNISHED IN DUPLICATE WITH THIS FORM.)

Sequence Number: _____ (77-78)

To: Research Report Secretary (6-30)
University of Waterloo (6-30)
Dept. of Computer Science (6-30)
Waterloo, Ontario N2L 3G1 (6-30)
(NAME & ADDRESS OF VENDOR)

Date May 22, 1989 (1-5)

Amount $ 4.00 (Canadian $s) (63-72)

Enclosed is remittance for the following items to be delivered to the address shown below.

Quantity	Description	Unit Price	Total
	Technical Reports		
		Canddian Rates	
1	CS-88-35 - Errata: Theory of Computation Author: Derick Wood	2.00	2.00
1	CS-88-39 The Computational Structure and Characterization of Nonlinear Discrete Chebyshev Problems Authors: A. R. Conn, Y.Li	2.00	2.00
		Total Canadian Rate	4.00

Please Make Check Payable to:
Dept. of Computer Science

Sent June 7

Deliver to The University of Tennessee

Computer Science Dept.

107 Ayres Hall

Attn: Ms. Christine Sheetz

Knoxville, Tenn. 37916

Name of Account	Account Number	Obj. Code	$ Amount
Computer Science	E01-1030	334	4.00 (Canadian

(47-52) (53-5) (56-62) (54-5) (63-72)

APPROVAL

CS-88-48 - SWITCH-LEVEL TESTABILITY OF THE DYNAMITE CMOS PLA

AUTHORS:B.F. Cockburn, J.A. Brzozowski

ABSTRACT:

Functional testing, as opposed to parametric testing, plays an important role in testing VLSI integrated circuits. However, it appears that designs are not always carefully analysed a priori to determine precisely which faults are *clean*, i.e. testable by logic means alone. The programmable logic array (PLA) is a popular circuit form used to implement a system of Boolean functions over a set of input variables. This report considers the testability of the dynamic CMOS PLA with respect to an extended set of switch-level faults models, namely: node faults, transistor stuck-opens and stuck-ons, interconnect breaks, ohmic shorts, and crosspoint faults. Single occurrences of each fault model are classified as either clean, unclean, or clean subject to conditions on the logical products and output functions computed by the PLA. Finally, a modified dynamic CMOS PLA design is described and its increased switch-level testability properties are stated.

PRICE:$2.00

If you would like to order any reports please forward your order, along with a cheque or international bank draft payable to the **Department of Computer Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, to the Research Report Secretary.**

Please indicate your current mailing address. CS-88-38 1 copy

MAILING ADDRESS:

David M. Himmelblau
Dept. Chemical Engineering
University of Texas
Austin TX 78712

sent Mar. 22/89

Printing Requisition / Graphic Services

26180

1. **Please complete unshaded areas** on form as applicable.
2. **Distribute copies** as follows: White and Yellow to Graphic Services. **Retain Pink Copies for your records.**
3. On completion of order the **Yellow copy will be returned** with the printed material.
4. **Please direct enquiries, quoting requisition number and account number,** to extension 3451.

TITLE OR DESCRIPTION: **A Globally Convergent Augmented Lagrangian Algorithm For Optimization With General Constraints And Simple Bounds** CS-88-38

DATE REQUISITIONED	DATE REQUIRED	ACCOUNT NO.
Nov. 2/88	ASAP	1 2 8 6 0 1 7 4 1

REQUISITIONER – PRINT	PHONE	SIGNING AUTHORITY
A.R. Conn	4458	S. DeAngelis / Andy Conn

MAILING INFO – NAME	DEPT.	BLDG. & ROOM NO.	
Sue DeAngelis	C.S.	DC 2314	[X] DELIVER [ ] PICK-UP

Copyright: I hereby agree to assume all responsibility and liability for any infringement of copyrights and/or patent rights which may arise from the processing of, and reproduction of, any of the materials herein requested. I further agree to indemnify and hold blameless the University of Waterloo from any liability which may arise from said processing or reproducing. I also acknowledge that materials processed as a result of this requisition are for educational use only.

NUMBER OF PAGES: 30 NUMBER OF COPIES: 100

TYPE OF PAPER STOCK: [X] BOND [ ] NCR ____ PT. [X] COVER [ ] BRISTOL [X] SUPPLIED [ ] ____

PAPER SIZE: [X] 8½ x 11 [ ] 8½ x 14 [ ] 11 x 17 [ ] 100M

PAPER COLOUR: [X] WHITE [ ] ____ INK: [X] BLACK [ ] ____

PRINTING: [ ] 1 SIDE ____ PGS. [X] 2 SIDES ____ PGS. NUMBERING FROM ____ TO ____

BINDING/FINISHING: [X] COLLATING [X] STAPLING 3 down left side [ ] ____ DRILLED [ ] PUNCHED [ ] PLASTIC RING

FOLDING/PADDING: 3 STAPLES CUTTING SIZE

Special Instructions

Math fronts and backs enclosed.

COPY CENTRE — OPER. NO. | BLDG. | MACH. NO.

DESIGN & PASTE-UP — OPER. NO. | TIME | LABOUR CODE

				D01
				D01
				D01

TYPESETTING — QUANTITY

PAP000000				T01
PAP000000				T01
PAP000000				T01

PROOF

PRF				
PRF				
PRF				

NEGATIVES — QUANTITY | OPER. NO. | TIME | LABOUR CODE

FLM				C01
FLM				C01
FLM				C01
FLM				C01
FLM				C01

PMT

PMT				C01
PMT				C01
PMT				C01

PLATES

PLT				P01
PLT				P01
PLT				P01

STOCK

		15		001
				001
				001
				001

BINDERY

RNG		26		B01
RNG		26	30	B01
RNG				B01
MIS000000				B01

OUTSIDE SERVICES

$____ COST

TAXES – PROVINCIAL [ ] FEDERAL [ ]

GRAPHIC SERV. OCT. 85 482-2

A Globally Convergent Augmented Lagrangian Algorithm For Optimization With General Constraints And Simple Bounds

A.R. Conn, N.I.M. Gould, Ph.L. Toint

Research Report CS-88-38
November, 1988

A GLOBALLY CONVERGENT AUGMENTED LAGRANGIAN ALGORITHM FOR OPTIMIZATION WITH GENERAL CONSTRAINTS AND SIMPLE BOUNDS

by A.R. Conn†, N.I.M. Gould‡ and Ph.L. Toint§

Abstract. The paper extends an algorithm for optimization with simple bounds (Conn, Gould and Toint, Siam Journal of Numerical Analysis 25, 433-460, 1988) to handle general constraints. The extension is achieved using an augmented Lagrangian approach. Global convergence is proved and it is established that a potentially troublesome penalty parameter is bounded away from zero.

†Department of Combinatorics and Optimization
University of Waterloo
Waterloo, Canada

‡Harwell Laboratory,
Harwell, England

§Department of Mathematics
Facultés Universitaires ND de la Paix
B-5000 Namur, Belgium

The work of the first author was supported in part by NSERC grant A8639.
Computer Science Research Report CS-88-38
This report is being issued simultaneously by the Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada, the Computer Science and Systems Division, Harwell Laboratory, Oxfordshire, U.K. and the Department of Mathematics, Facultés Notre Dame de la Paix, University of Namur, Namur, Belgium

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 constraints

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

and the simple bounds

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

Here f and c_i map R^n into R and the inequalities (1.3) are considered component-wise; we shall assume that the region $B = \{x \mid l \le x \le u\}$ is non-empty and may be infinite. We further assume that

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

We assume that any general inequality constraints $c_i(x) \ge 0$ have already been converted into equations by the introduction of slack variables (see, e.g. Fletcher, 1981, p.8); we wish the combinatorial side of the minimization problem to be represented purely in terms of simple bound constraints. We shall attempt to solve our problem by means of a sequential minimization of the *augmented Lagrangian function*

$$\Phi(x, \lambda, S, \mu) = f(x) + \sum_{i=1}^{m} c_i(x)\,\lambda_i + (1/2\mu) \sum_{i=1}^{m} s_{ii}\, c_i(x)^2, \tag{1.4}$$

where the components, λ_i, of the vector λ are known as Lagrange multiplier estimates, where the entries s_{ii} of the diagonal matrix S are positive scaling factors and where μ is known as the penalty parameter. Notice that we *do not* include the simple bounds (1.3) in the augmented Lagrangian function; rather the intention is that the sequential minimization will automatically ensure that these constraints are always satisfied.

Our principal interest is in solving large-scale problems. With a few notable exceptions (see, for example, Murtagh and Saunders, 1980, Lasdon, 1982, Drud, 1985), there has been little progress in constructing algorithms for such problems; this is somewhat understandable in view of the lack of a consensus as to the "best" algorithm for solving small nonlinear programs. Nevertheless, there are many large-scale applications awaiting a suitable algorithm.

A similar situation existed for unconstrained optimization in the early 1970's. However, during the past ten years, this deficiency has been redressed primarily through the development of three important ideas. The first has been the recognition that large problems normally have considerable structure and that such structure usually manifests itself as sparsity or low rank of the relevant matrices. This has lead to suitable ways of storing and approximating problem data (function, gradient and Hessian approximations), see, for example, Griewank and Toint (1982). The second development has been the realization that, although Newton's method (or a good approximation to it) is necessary for rapid asymptotic convergence of an algorithm, in early iterations only very crude approximations to the solution of the Newton equations are needed to guarantee global convergence. In particular, the steepest descent method often makes very good initial progress towards a minimizer. This lead to a study of realistic conditions that suffice to guarantee global convergence of an algorithm and also of methods which satisfied such conditions, the truncated conjugate gradient method being a particularly successful example. This work is described, for example, by Toint (1981), Dembo, Eisenstat and Steihaug (1982)

and Steihaug (1983). Thirdly, the development of trust-region methods (see, e.g., Moré, 1983) has allowed a sensible handling of negative curvature in the objective function; for large-scale problems whose second derivatives are available (contrary to popular belief, an extremely common circumstance in many problem areas), this enables meaningful steps towards the solution to be made when the Hessian matrix is indefinite. Significantly, these ideas have had a important impact on the design of algorithms not only for large problems but also for small ones (see, Toint, 1988, Dixon, Dolan and Price, 1988).

One issue that is not present in unconstrained minimization, but is in evidence here, is the *combinatorial problem* of finding which of the variables lie at a bound at the solution. In active set algorithms, the intention is to predict these variables and to minimize the function with respect to the remaining variables. Obviously, an incorrect prediction is undesirable and it is then useful (indeed essential for large problems) to be able to make rapid changes in the active set to correct for wrong initial choices. Unfortunately, many existing algorithms for constrained optimization only allow very small changes in the active set at each iteration and consequently, for large problems, there is the possibility of requiring a large number of iterations to find the solution. Fortunately, for simple bound constraints, it is easy to allow for rapid changes in the active set in the design of algorithms (see, e.g., Bertsekas, 1982b, pp. 76-92, Conn, Gould and Toint, 1988a).

Our intention here is to develop a fairly general algorithm which may benefit from the above-mentioned advances. We have recently developed and tested (Conn, Gould, Lescrenier and Toint, 1987, Conn, Gould and Toint, 1988a,b) an algorithm for solving *bound constrained minimization problems* (problems of the form minimize (1.1) subject to (1.3)) which is appropriate in the large scale case. Our basic idea is now to use this algorithm within an augmented Lagrangian framework, that is to use the algorithm to find an *approximation* to a minimizer of the augmented Lagrangian function (1.4) *subject to the bounds (1.3)* for a sequence of different values of S, λ and μ. The novelty comes from being able to solve the augmented Lagrangian problems approximately and on being able to deal with the bounds in an efficient manner.

The augmented Lagrangian method was proposed independently by Hestenes (1969) and Powell (1969), partly as a reaction to the unfortunate side effects associated with ill-conditioning of the simpler differentiable penalty and barrier functions (Murray, 1971). Indeed, Powell showed, using a very simple device, how to ensure that the penalty parameter does not converge to zero and hence that the resulting ill-conditioning does not occur. A similar device is employed in the algorithms that we are concerned with in this paper with the same consequence. A concise statement of the salient features of augmented Lagrangian methods are given, for example, by Fletcher (1981). The most comprehensive references on augmented Lagrangians are the paper by Tapia (1977) and the book by Bertsekas (1982b). Interest in augmented Lagrangians declined with the introduction of successive quadratic programming (SQP) techniques but recently has gained in popularity – see for example the papers of Schittkowski (1981) and Gill, Murray, Saunders and Wright (1986) which combines SQP with an augmented Lagrangian merit function. Both these methods are not *pure* augmented Lagrangian techniques since they perform a line search on the augmented Lagrangian as a function of both the position, x, and the multipliers, λ, in contrast to the method described in this paper.

One strong disadvantage of SQP methods for large-scale problems is that, although there is a

theory of how to truncate the solution process in the early iterations (see, Dembo and Tulowitzki, 1984) – as used so successfully in the unconstrained case – it is not clear to us how to construct an efficient algorithm which conforms to this theory. We feel that solving a quadratic programming problem to completion at each iteration is probably too expensive a calculation for large-scale problems in the same way that solving the Newton equations exactly is considered too expensive in large-scale unconstrained minimization. We thus feel compelled to try to use an alternative to the SQP approach.

Bertsekas (1982a) and others, however, have remarked that augmented Lagrangians are particularly attractive for large problems, where active set strategies are inappropriate and we tend to agree with this sentiment. In this paper we explore some of the issues involved in using an augmented Lagrangian approach for large-scale problems. We have deliberately not included the results of numerical testing as, in our view, the construction of appropriate software is by no means trivial and we wish to make a thorough job of it. We will report on our numerical experience in due course.

Our exposition will be considerably simplified if we consider the special case where $l_i = 0$ and $u_i = \infty$ for all $1 \le i \le n$ in (1.3). Although straightforward, the modification required to handle more general constraints will be indicated at the end of the paper. Thus we consider the problem:

$$\text{minimize } f(x), \tag{1.5}$$

subject to the constraints

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

and the non-negativity restrictions

$$x \in B = \{x \in R^n \mid x \ge 0\}. \tag{1.7}$$

The paper is organised as follows. In §2 we introduce concepts and definitions and then state a pair of related algorithms for solving (1.5)–(1.7) in §3. Global convergence is established in §4, while issues of asymptotic convergence follow in §5. An example showing the importance of a certain assumption in §5 is given in §6, while in §7 the consequences of satisfying second order conditions are given. We conclude in §8 by indicating how this theory applies to the original problem (1.1)–(1.3).

2 Notation

In this section we introduce the notation that will be used throughout the paper. We will use the projection operator defined component-wise by

$$(P[x])_i = \begin{cases} 0 \text{ if } x_i \le 0 \\ x_i \text{ otherwise.} \end{cases} \tag{2.1}$$

This operator projects the point x onto the region B. Furthermore, we will make use of the "projection"

$$P(x, v) = x - P[x - v]. \tag{2.2}$$

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

$$c(x) = [c_1(x), ..., c_m(x)]^T, \tag{2.3}$$

and let $H_i(x)$ denote the Hessian matrix, $\nabla_{xx} c_i(x)$, of $c_i(x)$. Finally, let $g_L(x,\lambda)$ and $H_L(x,\lambda)$ denote the gradient and Hessian matrix (taken with respect to its first argument) of the Lagrangian function

$$L(x,\lambda)=f(x)+\sum_{i=1}^{m}\lambda_i c_i(x). \tag{2.4}$$

We note that $L(x,\lambda)$ is the Lagrangian function with respect to the c_i constraints only. If we define *first order Lagrange multiplier estimates*

$$\tilde{\lambda}(x,\lambda,S,\mu)=\lambda+S\,c(x)/\mu, \tag{2.5}$$

we shall make much use of the identity

$$\nabla_x \Phi(x,\lambda,S,\mu)=g_L(x,\tilde{\lambda}(x,\lambda,S,\mu)). \tag{2.6}$$

Now suppose that $\{x^{(k)}\geq 0\}$ and $\{\lambda^{(k)}\}$ are infinite sequences of n– and m–vectors respectively, that $\{S^{(k)}\}$ is an infinite sequence of positive definite diagonal matrices and that $\{\mu^{(k)}\}$ is an infinite sequence of positive scalars. We shall use the notation that $F^{(k)}$ denotes the generic function $F(\cdot)$ evaluated with arguments $x^{(k)},\lambda^{(k)},S^{(k)}$ or $\mu^{(k)}$ as appropriate. So, for instance, using the identity (2.6), we have

$$\nabla_x \Phi^{(k)}=\nabla_x \Phi(x^{(k)},\lambda^{(k)},S^{(k)},\mu^{(k)})=g_L(x^{(k)},\tilde{\lambda}^{(k)}), \tag{2.7}$$

where we have written

$$\tilde{\lambda}^{(k)}=\tilde{\lambda}(x^{(k)},\lambda^{(k)},S^{(k)},\mu^{(k)}). \tag{2.8}$$

For any $x^{(k)}$, we have two possibilities for each component $x_i^{(k)}$, namely

$$\begin{aligned} &\text{(i)}\quad 0\leq x_i^{(k)}\leq(\nabla_x \Phi^{(k)})_i \text{ or}\\ &\text{(ii)}\quad (\nabla_x \Phi^{(k)})_i < x_i^{(k)} \end{aligned} \tag{2.9}$$

In case (i) we then have

$$(P(x^{(k)},\nabla_x \Phi^{(k)}))_i=x_i^{(k)} \tag{2.10}$$

whereas in case (ii) we have

$$(P(x^{(k)},\nabla_x \Phi^{(k)}))_i=(\nabla_x \Phi^{(k)})_i. \tag{2.11}$$

We shall refer to an $x_i^{(k)}$ which satisfies (i) as a *dominated* variable; a variable which satisfies (ii) is known as a *floating* variable. The algorithms which we are about to develop construct iterates which force $P(x^{(k)},\nabla_x \Phi^{(k)})$ to zero. The dominated variables are thus pushed to zero, while the floating variables are allowed to find their own level.

If, in addition, there is a convergent subsequence $\{x^{(k)}\}, k\in K$, with limit point x^*, we wish to partition the set $N=\{1,2,....,n\}$ into the following four subsets which are related to the two possibilities (i) and (ii) above and to the corresponding x^*:

$$\begin{aligned} I_1&=\{i\,|\,x_i^{(k)} \text{ are floating for all } k\in K \text{ sufficiently large and } x_i^*>0\},\\ I_2&=\{i\,|\,x_i^{(k)} \text{ are dominated for all } k\in K \text{ sufficiently large}\},\\ I_3&=\{i\,|\,x_i^{(k)} \text{ are floating for all } k\in K \text{ sufficiently large but } x_i^*=0\} \text{ and}\\ I_4&=N\setminus(I_1\cup I_2\cup I_3). \end{aligned} \tag{2.12}$$

From time-to-time we will slightly abuse notation by saying that a variable x_i belongs to (for instance) I_1, when strictly we should say that the index of the variable belongs to I_1. We will also mention the components of a (given) vector in the set I_1 when strictly we mean the components of the vector whose indices lie in I_1.

If the iterates are chosen so that $P(x^{(k)}, \nabla_x \Phi^{(k)})$ approaches zero as k increases, we have the following result:

Lemma 2.1. *Suppose $x^{(k)}$, $\lambda^{(k)}$, $S^{(k)}$ and $\mu^{(k)}$ are as above and that $P(x^{(k)}, \nabla_x \Phi^{(k)})$ approaches zero as $k \in K$ increases. Then*

(i) *the variables in sets I_2, I_3 and I_4 all converge to their bounds;*

(ii) *the components of $(\nabla_x \Phi^{(k)})_i$ in the sets I_1 and I_3 converge to zero; and*

(iii) *if a component of $(\nabla_x \Phi^{(k)})_i$ in the set I_4 converges to a finite limit, the limit is zero.*

Proof. (i) The result is true for variables in I_2 from (2.10), for those in I_3 by definition and for those in I_4 as, again from (2.10), there must be a subsequence of the $k \in K$ for which $x_i^{(k)}$ is equal to a sequence converging to zero.

(ii) The result follows for i in I_1 and I_3 from (2.11).

(iii) This is true for i in I_4 as there must be a subsequence of the $k \in K$ for which, from (2.11), $(\nabla_x \Phi^{(k)})_i$ is equal to a sequence whose limit is zero. ■

It will sometimes be convenient to group the variables in sets I_3 and I_4 together and call the resulting set

$$I_5 = I_3 \cup I_4. \tag{2.1.1}$$

As we see from Lemma 2.1, I_5 gives variables which are zero at the solution and which may correspond to zero components of the gradient of the augmented Lagrangian function. These variables are potentially (dual) degenerate at the solution of the nonlinear programming problem.

We will let $\hat{g}(x)$ denote the components of $g(x)$ indexed by I_1. Similarily, $\hat{A}(x)$ denotes the corresponding columns of the Jacobian matrix; indeed any matrix $\hat{M}$ refers to the columns of the generic matrix M indexed by I_1. In addition we will define the *least-squares Lagrange multiplier estimates* (corresponding to the set I_1)

$$\lambda(x) = -(\hat{A}(x)^+)^T \hat{g}(x) \tag{2.2.1}$$

at all points where the right generalized inverse

$$\hat{A}(x)^+ = \hat{A}(x)^T (\hat{A}(x)\hat{A}(x)^T)^{-1} \tag{2.2.2}$$

of $\hat{A}(x)$ is well defined. We note that $\lambda(x)$ is differentiable; for completeness the derivative is given in the following lemma.

Lemma 2.2. *Suppose that (AS1) holds. If $\hat{A}(x)\hat{A}(x)^T$ is non-singular, $\lambda(x)$ is differentiable and its derivative is given by*

$$\nabla_x \lambda(x) = -(\hat{A}(x)^+)^T \hat{H}_L(x, \lambda(x)) - (\hat{A}(x)\hat{A}(x)^T)^{-1} R(x), \tag{2.2.3}$$

where the i–th row of $R(x)$ *is* $(\hat{g}(x)+\hat{A}(x)^T\lambda(x))^T\hat{H}_i(x)$.

Proof. The result follows by observing that (2.2.1) may be rewritten as

$$r(x)-\hat{A}(x)^T\lambda(x)=\hat{g}(x) \text{ and } \hat{A}(x)\,r(x)=0 \tag{2.2.4}$$

for some vector $r(x)$. Differentiating (2.2.4) and eliminating the derivative of $r(x)$ from the resulting equations gives the required result. ■

We stress that, as stated, the Lagrange multiplier estimate (2.2.1) is not a directly calculable quantity as it requires an a priori knowledge of x^*. It is merely introduced as an analytical device but we shall show in due course that a variant of this estimate may be calculated and used.

We are now in a position to describe more precisely the algorithms we propose to use.

3 Statement of the algorithms

In order to solve problem (1.5)–(1.7), we consider the following algorithmic models.

Algorithm 1:

step 0 : [Initialization] An initial vector of Lagrange multiplier estimates, $\lambda^{(0)}$, is given. The positive constants η_0, μ_0, $\tau<1$, $\omega_0<1$, $\gamma_1<1$, $\omega_* \ll 1$, $\eta_* \ll 1$, α_ω, β_ω, α_η and β_η are specified. The diagonal matrices S_1 and S_2, for which $0<S_1^{-1}\leq S_2<\infty$, are given (the inequalities are to be understood element-wise for the diagonal elements). Set $\mu^{(0)}=\mu_0$, $\alpha^{(0)}=\min(\mu^{(0)},\gamma_1)$, $\omega^{(0)}=\omega_0\,(\alpha^{(0)})^{\alpha_\omega}$, $\eta^{(0)}=\eta_0\,(\alpha^{(0)})^{\alpha_\eta}$ and $k=0$.

step 1 : [Inner iteration] Define a diagonal scaling matrix $S^{(k)}$ for which $S_1^{-1}\leq S^{(k)}\leq S_2$. Find $x^{(k)}\in B$ such that

$$\|P(x^{(k)},\nabla_x\Phi^{(k)})\|\leq\omega^{(k)}. \tag{3.1}$$

If

$$\|c(x^{(k)})\|\leq\eta^{(k)}, \tag{3.2}$$

execute step 2. Otherwise, execute step 3.

step 2 : [Test for convergence and update Lagrange multiplier estimates] If $\omega^{(k)}\leq\omega_*$ and $\eta^{(k)}\leq\eta_*$, stop. Otherwise, set

$$\begin{aligned}
\lambda^{(k+1)}&=\bar{\lambda}(x^{(k)},\lambda^{(k)},S^{(k)},\mu^{(k)}),\\
\mu^{(k+1)}&=\mu^{(k)},\\
\alpha^{(k+1)}&=\min(\mu^{(k+1)},\gamma_1),\\
\omega^{(k+1)}&=\omega^{(k)}\,(\alpha^{(k+1)})^{\beta_\omega},\\
\eta^{(k+1)}&=\eta^{(k)}\,(\alpha^{(k+1)})^{\beta_\eta},
\end{aligned} \tag{3.3}$$

increment k by one and go to step 1.

step 3 : [Reduce the penalty parameter] Set

$$
\begin{aligned}
\lambda^{(k+1)} &= \lambda^{(k)}, \\
\mu^{(k+1)} &= \tau\mu^{(k)}, \\
\alpha^{(k+1)} &= \min(\mu^{(k+1)}, \gamma_1), \\
\omega^{(k+1)} &= \omega_0 (\alpha^{(k+1)})^{\alpha_\omega}, \\
\eta^{(k+1)} &= \eta_0 (\alpha^{(k+1)})^{\alpha_\eta},
\end{aligned} \tag{3.4}
$$

increment k by one and go to step 1.

Algorithm 2:

step 0 : [Initialization] An initial vector of Lagrange multiplier estimates, $\lambda^{(0)}$, is given. The non-negative constant α_η and the positive constants η_0, μ_0, $\tau<1$, $\gamma<1$, $\omega_0<1$, $\gamma_1<1$, $\omega_* \ll 1$, $\eta_* \ll 1$, ν, α_ω, β_ω and β_η are specified. The diagonal matrices S_1 and S_2, for which $0<S_1^{-1}\le S_2<\infty$, are given. Set $\mu^{(0)}=\mu_0$, $\alpha^{(0)}=\min(\mu^{(0)},\gamma_1)$, $\omega^{(0)}=\omega_0(\alpha^{(0)})^{\alpha_\omega}$, $\eta^{(0)}=\eta_0(\alpha^{(0)})^{\alpha_\eta}$ and $k=0$.

step 1 : [Inner iteration] Define a diagonal scaling matrix $S^{(k)}$ for which $S_1^{-1}\le S^{(k)}\le S_2$. Find $x^{(k)}\in B$ such that

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

Compute a new vector of Lagrange multiplier estimates $\hat{\lambda}^{(k+1)}$. If

$$
\|c(x^{(k)})\| \le \eta^{(k)}, \tag{3.6}
$$

execute step 2. Otherwise, execute step 3.

step 2 : [Test for convergence and update Lagrange multiplier estimates] If $\omega^{(k)}\le\omega_*$ and $\eta^{(k)}\le\eta_*$, stop. Otherwise, set

$$
\begin{aligned}
\mu^{(k+1)} &= \mu^{(k)}, \\
\lambda^{(k+1)} &= \begin{cases} \hat{\lambda}^{(k+1)} & \text{if } \|\hat{\lambda}^{(k+1)}\| \le \nu(\mu^{(k+1)})^{-\gamma} \\ \lambda^{(k)} & \text{otherwise,} \end{cases} \\
\alpha^{(k+1)} &= \min(\mu^{(k+1)}, \gamma_1), \\
\omega^{(k+1)} &= \omega^{(k)} (\alpha^{(k+1)})^{\beta_\omega}, \\
\eta^{(k+1)} &= \eta^{(k)} (\alpha^{(k+1)})^{\beta_\eta},
\end{aligned} \tag{3.7}
$$

increment k by one and go to step 1.

step 3 : [Reduce the penalty parameter and update Lagrange multiplier estimates] Set

$$
\begin{aligned}
&\mu^{(k+1)} = \tau\mu^{(k)}, \\
&\lambda^{(k+1)} = \begin{cases} \hat{\lambda}^{(k+1)} & \text{if } \|\hat{\lambda}^{(k+1)}\| \le \nu(\mu^{(k+1)})^{-\gamma} \\ \lambda^{(k)} & \text{otherwise,} \end{cases} \\
&\alpha^{(k+1)} = \min(\mu^{(k+1)}, \gamma_1), \qquad (3.8) \\
&\omega^{(k+1)} = \omega_0 (\alpha^{(k+1)})^{\alpha_\omega}, \\
&\eta^{(k+1)} = \eta_0 (\alpha^{(k+1)})^{\alpha_\eta},
\end{aligned}
$$

increment k by one and go to step 1.

The motivation for both algorithms is quite straightforward. Traditional augmented Lagrangian methods are known to be locally convergent if the penalty parameter is sufficiently small and if the augmented Lagrangian is approximately minimized at each stage (see, for instance, Bersekas, 1982b, §2.5). In order to ensure that the method is globally convergent, as a last resort we must drive the penalty parameter to zero and ensure that the Lagrange multiplier estimates do not behave too badly. The convergence of such a scheme is guaranteed, since in this case, the iteration is essentially that used in the quadratic penalty function method (see, for example, Gould, 1987). We consider this further in §4. In order to try to allow the traditional multiplier iteration to take over, the test on the size of the constraints (3.2)/(3.6) is based upon the size that might be expected if the multiplier iteration is converging. This aspect is considered in §5.

The algorithms differ in their use of multiplier updates. Algorithm 1 is designed specifically for the first-order estimate (2.5); the multiplier estimates are encouraged to behave well as a consequence of the test (3.2). For large-scale computations, it is likely that first-order estimates will be used and thus algorithm 1 is directly applicable. Algorithm 2 allows any multiplier estimate to be used. This extra freedom means that tighter control must be maintained on the acceptance of the estimates to make sure that they do not grow unacceptably fast. In this algorithm, we have in mind using any of the well known Lagrange multiplier update formulae, including the first order update (2.5) (used in algorithm 1), the least-squares update (2.2.1) and other updates summarized, for instance, by Tapia (1977). We note, however, that some of these updates may require a significant amount of computation and this may prove prohibitively expensive for large-scale problems.

Both algorithms use a number of free parameters. To give the reader some feel for what might be typical values, we suggest that for well scaled problems $\alpha_\omega = \beta_\omega = \gamma = \nu = \eta_0 = \omega_0 = 1$, $\alpha_\eta = \mu_0 = \gamma_1 = 0.1$, $\beta_\eta = 0.9$ and $\tau = 0.01$ are appropriate.

4 Global convergence analysis

In this section we shall make use of the the following assumptions.

AS2: The iterates generated by algorithms 1 or 2 all lie within a closed, bounded domain Ω.

AS3: The matrix $\hat{A}(x^*)$ has column rank no smaller than m at any limit point, x^*, of the sequences $\{x^{(k)}\}$ generated by algorithms 1 or 2.

Notice that AS3 excludes the possibility that I_1 is empty unless there are no general constraints. In view of Lemma 2.1, this seems reasonable as otherwise we are allowing the possibility that all the constraints and bounds are satisfied as equations at x^*.

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

We shall analyse the convergence of the algorithms of §3 in the case where the convergence tolerances ω_* and η_* are both zero. We require the following pair of lemmas in the proof of global convergence of our algorithms. Essentially, the results show that the Lagrange multiplier estimates generated by either algorithm cannot behave too badly.

Lemma 4.1. *Suppose that $\mu^{(k)}$ converges to zero as k increases when algorithm 1 is executed. Then the product $\mu^{(k)}\|\lambda^{(k)}\|$ converges to zero.*

Proof. If $\mu^{(k)}$ converges to zero, step 3 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 3 of the algorithm is executed and for which

$$\mu^{(k)} \le \min(\tfrac{1}{2}^{1/\beta_\eta}, \gamma_1). \tag{4.1.1}$$

We consider how the Lagrange multiplier estimates change between two successive iterations indexed in the set K. At iteration k_i+j, for $k_i<k_i+j\le k_{i+1}$, we have

$$\lambda^{(k_i+j)} = \lambda^{(k_i)} + \sum_{l=1}^{j-1} S^{(k_i+l)}\, c(x^{(k_i+l)})/\mu^{(k_i+l)} \tag{4.1.2}$$

and

$$\mu^{(k_{i+1})} = \mu^{(k_i+j)} = \mu^{(k_i+1)} = \tau\mu^{(k_i)}, \tag{4.1.3}$$

where the summation in (4.1.2) is null if $j=1$. Now suppose that $j>1$. Then for the set of iterations k_i+l, $1\le l<j$, step 2 of the algorithm must have been executed and hence, from (3.2), (4.1.3) and the recursive definition of $\eta^{(k)}$, we must also have

$$\|c(x^{(k_i+l)})\| \le \eta_0(\mu^{(k_{i+1})})^{\beta_\eta(l-1)+\alpha_\eta} \tag{4.1.4}$$

Combining equations (4.1.1) to (4.1.4) and using the imposed upper bound on $S^{(k)}$, we obtain the bound

$$\begin{aligned}
\|\lambda^{(k_i+j)}\| &\le \|\lambda^{(k_i)}\| + \sum_{l=1}^{j-1}\|S^{(k_i+l)}\, c(x^{(k_i+l)})\|/\mu^{(k_i+l)} \\
&\le \|\lambda^{(k_i)}\| + s_2\,\eta_0(\mu^{(k_{i+1})})^{\alpha_\eta-1}\sum_{l=1}^{j-1}(\mu^{(k_{i+1})})^{\beta_\eta(l-1)} \\
&\le \|\lambda^{(k_i)}\| + s_2\,\eta_0(\mu^{(k_{i+1})})^{\alpha_\eta-1}/(1-(\mu^{(k_{i+1})})^{\beta_\eta}) \\
&\le \|\lambda^{(k_i)}\| + 2s_2\,\eta_0(\mu^{(k_{i+1})})^{\alpha_\eta-1},
\end{aligned} \tag{4.1.5}$$

where s_2 is the norm of S_2. Thus we obtain that

$$\mu^{(k_i+j)}\|\lambda^{(k_i+j)}\| \le \tau\mu^{(k_i)}\|\lambda^{(k_i)}\| + 2s_2\,\eta_0(\mu^{(k_{i+1})})^{\alpha_\eta}. \tag{4.1.6}$$

Equation (4.1.6) is also satisfied when $j=1$ as equations (3.4) and (4.1.3) give $\mu^{(k_i+1)}\|\lambda^{(k_i+1)}\| = \tau\mu^{(k_i)}\|\lambda^{(k_i)}\|$.

Hence from (4.1.6),

$$\mu^{(k_{i+1})}\|\lambda^{(k_{i+1})}\| \le \tau\mu^{(k_i)}\|\lambda^{(k_i)}\| + 2s_2\,\eta_0(\mu^{(k_{i+1})})^{\alpha_\eta}. \tag{4.1.7}$$

Equation (4.1.7) then gives that $\mu^{(k_i)}\|\lambda^{(k_i)}\|$ converges to zero as k increases. For, if we define

(4.1.8) $$\alpha_i = \mu^{(k_i)} \|\lambda^{(k_i)}\| \text{ and } \beta_i = 2s_2\, \eta_0 (\mu^{(k_i)})^{\alpha_\eta},$$

equations (4.1.3), (4.1.7) and (4.1.8) give that

(4.1.9) $$\alpha_{i+1} \le \tau\alpha_i + \tau^{\alpha_\eta}\beta_i \text{ and } \beta_{i+1} = \tau^{\alpha_\eta}\beta_i$$

and hence that

(4.1.10) $$0 \le \alpha_i \le \tau^i \alpha_0 + (\tau^{\alpha_\eta})^i \sum_{l=0}^{i-1} (\tau^{1-\alpha_\eta})^l \beta_0.$$

If $\alpha_\eta < 1$, the sum in (4.1.10) can be bounded to give

(4.1.11) $$0 \le \alpha_i \le \tau^i \alpha_0 + (\tau^{\alpha_\eta})^i \beta_0 / (1 - \tau^{1-\alpha_\eta}),$$

whereas if $\alpha_\eta > 1$ we obtain the alternative

(4.1.12) $$0 \le \alpha_i \le \tau^i (\alpha_0 + \tau^{\alpha_\eta - 1}\beta_0 / (1 - \tau^{\alpha_\eta - 1})),$$

and if $\alpha_\eta = 1$,

(4.1.13) $$0 \le \alpha_i \le \tau^i \alpha_0 + i\tau^i \beta_0.$$

But, both α_0 and β_0 are finite. Thus, as i increases, α_i converges to zero; equation (4.1.9) implies that β_i converges to zero. Therefore, as the right-hand-side of (4.1.6) converges to zero, the truth of the lemma is established. ■

We note that lemmas 4.1 may be proved under much weaker conditions on the sequence $\{\eta^{(k)}\}$ than those imposed in algorithm 1. All that is needed is that, in the proof just given, $\sum_{l=1}^{j-1} \|c(x^{(k_i+l)})\|$ in (4.1.5) should be bounded by some multiple of a positive power of $\mu^{(k_{i+1})}$.

Turning to algorithm 2, we have the following easier-to-establish result.

Lemma 4.2. *Suppose that $\mu^{(k)}$ converges to zero as k increases when algorithm 2 is executed. Then the product $\mu^{(k)}\|\lambda^{(k)}\|$ converges to zero.*

Proof. Let $K = \{k_0, k_1, k_2, ...\}$ be the iterations on which

(4.2.1) $$\|\hat{\lambda}^{(k+1)}\| \le \nu (\mu^{(k+1)})^{-\gamma}$$

and consequently on which $\lambda^{(k+1)} = \hat{\lambda}^{(k+1)}$. Then, from (4.2.1),

(4.2.2) $$\mu^{(k_i+1)} \|\lambda^{(k_i+1)}\| \le \nu (\mu^{(k_i+1)})^{1-\gamma}.$$

If K is finite, $\lambda^{(k)}$ will be fixed for all k sufficiently large and the result is immediate. If K is infinite, for any $k_i < k \le k_{i+1}$, $\lambda^{(k)} = \lambda^{(k_i+1)}$ and $\mu^{(k)} \le \mu^{(k_i+1)}$. Hence, from (4.2.2)

(4.2.3) $$\mu^{(k)} \|\lambda^{(k)}\| \le \nu (\mu^{(k_i+1)})^{1-\gamma}.$$

By hypothesis, the right-hand side of (4.2.3) can be made arbitrarily small by chosing k_i large enough, and so $\mu^{(k)}\|\lambda^{(k)}\|$ converges to zero. ■

We now establish that both algorithms 1 and 2 possess a powerful global convergence property under relatively weak conditions.

Theorem 4.3. *Assume that (AS1-AS3) hold. Let x^* be any limit point of the sequence $\{x^{(k)}\}$ generated by algorithm 1 or by algorithm 2 of §3 and let K be the set of indices of an infinite subsequence of the $x^{(k)}$ whose limit is x^*. Then*

(i) *x^* is a Kuhn-Tucker point (first-order stationary point) for the problem (1.5)–(1.7) and the sequences $\{\bar{\lambda}(x^{(k)}, \lambda^{(k)}, S^{(k)}, \mu^{(k)})\}$ and $\{\lambda(x^{(k)})\}$ converge to the corresponding vector of Lagrange multipliers λ^* for $k \in K$;*

(ii) *the gradients $\nabla_x \Phi^{(k)}$ converge to $g_L(x^*, \lambda^*)$ for $k \in K$;*

(iii) *there are positive constants a_1, a_2, s_1 and an integer k_0 such that*

$$\|\bar{\lambda}(x^{(k)}, \lambda^{(k)}, S^{(k)}, \mu^{(k)}) - \lambda^*\| \leq a_1 \omega^{(k)} + a_2 \|x^{(k)} - x^*\|, \tag{4.3.1}$$

$$\|\lambda(x^{(k)}) - \lambda^*\| \leq a_2 \|x^{(k)} - x^*\| \tag{4.3.2}$$

and

$$\|c(x^{(k)})\| \leq s_1 \left(a_1 \omega^{(k)} \mu^{(k)} + \mu^{(k)} \|\lambda^{(k)} - \lambda^*\| + a_2 \mu^{(k)} \|x^{(k)} - x^*\|\right) \tag{4.3.3}$$

for all $k \geq k0$, $(k \in K)$.

Proof. As a consequence of AS1 and AS3, we have that for $k \in K$ sufficiently large, $\hat{A}(x^{(k)})^+$ exists, is bounded and converges to $\hat{A}(x^*)^+$. Thus we may write

$$\|(\hat{A}(x^{(k)})^+)^T\| \leq a_1 \tag{4.3.4}$$

for some constant $a_1 > 0$. As the variables in the set I_1 are floating, equations (2.7), (2.8), (2.11) and the inner iteration termination criterion (step 1) give that

$$\|\hat{g}(x^{(k)}) + \hat{A}(x^{(k)})^T \bar{\lambda}^{(k)}\| \leq \omega^{(k)}. \tag{4.3.5}$$

Define $\lambda^* = \lambda(x^*)$. By assumption, $\lambda(x)$ is bounded for all x in a neighbourhood of x^*. Thus we may deduce from (2.2.1), (4.3.4) and (4.3.5) that

$$\begin{aligned} \|\bar{\lambda}^{(k)} - \lambda(x^{(k)})\| &= \|(\hat{A}(x^{(k)})^+)^T \hat{g}(x^{(k)}) + \bar{\lambda}^{(k)}\| \\ &= \|(\hat{A}(x^{(k)})^+)^T (\hat{g}(x^{(k)}) + \hat{A}(x^{(k)})^T \bar{\lambda}^{(k)})\| \\ &\leq \|(\hat{A}(x^{(k)})^+)^T\| \, \omega^{(k)} \leq a_1 \omega^{(k)} \end{aligned} \tag{4.3.6}$$

Moreover, from the integral mean value theorem and lemma 2.2 we have that

$$\lambda(x^{(k)}) - \lambda(x^*) = \int_0^1 \nabla_x \lambda(x(s)) ds \cdot (x^{(k)} - x^*), \tag{4.3.7}$$

where $\nabla_x \lambda(x)$ is given by equation (2.2.3) and where $x(s) = x^{(k)} + s(x^* - x^{(k)})$. Now the terms within the integral sign are bounded for all x sufficiently close to x^* and hence (4.3.7) gives

$$\|\lambda(x^{(k)}) - \lambda^*\| \leq a_2 \|x^{(k)} - x^*\| \tag{4.3.8}$$

for some constant $a_2 > 0$, which is just the inequality (4.3.2). We then have that $\lambda(x^{(k)})$ converges to λ^*. Combining (4.3.6) and (4.3.8) we obtain

$$\|\bar{\lambda}^{(k)} - \lambda^*\| \leq \|\bar{\lambda}^{(k)} - \lambda(x^{(k)})\| + \|\lambda(x^{(k)}) - \lambda^*\| \leq a_1 \, \omega^{(k)} + a_2 \|x^{(k)} - x^*\|, \tag{4.3.9}$$

the required inequality (4.3.1). Then, since by construction $\omega^{(k)}$ tends to zero as k increases, (4.3.1) implies that $\bar{\lambda}^{(k)}$ converges to λ^* and from (4.3.5) we have that

$$\hat{g}_L(x^*, \lambda^*) = \hat{g}(x^*) + \hat{A}(x^*)^T \lambda^* = 0. \tag{4.3.10}$$

Moreover, from the identity (2.6), $\nabla_x \Phi^{(k)}$ converges to $g_L(x^*, \lambda^*)$. Furthermore, multiplying (4.3.9) by $\mu^{(k)}$ and using (2.5), we obtain (4.3.3), where s_1 is the norm of S_1^{-1}. We also have that

$$c(x^*) = 0. \tag{4.3.11}$$

To see this, we consider two separate cases.

(i) If $\mu^{(k)}$ is bounded away from zero, step 2 must be executed every iteration for k sufficiently large. But this implies that (3.2) is always satisfied (k large enough) and $\eta^{(k)}$ converges to zero. Hence $c(x^{(k)})$ converges to zero.

(ii) If $\mu^{(k)}$ converges to zero, lemma 4.1 for algorithm 1 and lemma 4.2 for algorithm 2 show that $\mu^{(k)}\|\lambda^{(k)}-\lambda^*\|$ converges to zero. But then, inequality (4.3.3) gives the required result.

Finally, we consider the status of the variables in the sets I_1, I_2 and I_5. Lemma 2.1 and the convergence of $\nabla_x \Phi^{(k)}$ to $g_L(x^*,\lambda^*)$ show that the complementary slackness condition

$$g_L(x^*,\lambda^*)^T x^* = 0 \tag{4.3.12}$$

is satisfied. The variables in the set I_1 are, by definition, positive at x^*. The components of $g_L(x^*,\lambda^*)$ indexed by I_2 are all non-negative from (2.9) as their corresponding variables are dominated. This then gives the conditions

$$\begin{aligned} &x_i^* > 0 \text{ and } (g_L(x^*,\lambda^*))_i = 0 \quad \text{for } i \in I_1, \\ &x_i^* = 0 \text{ and } (g_L(x^*,\lambda^*))_i \geq 0 \quad \text{for } i \in I_2 \text{ and} \\ &x_i^* = 0 \text{ and } (g_L(x^*,\lambda^*))_i = 0 \quad \text{for } i \in I_5 . \end{aligned} \tag{4.3.13}$$

Equations (4.3.11) and (4.3.13) thus show that x^* is a Kuhn-Tucker point and λ^* are the corresponding set of Lagrange multipliers. Hence the theorem is proved. ■

Notice that theorem 4.3 would remain true regardless of the actual choice of $\{\omega^{(k)}\}$ provided that the sequence converges to zero.

5 Asymptotic convergence analysis

We now give our first rate-of-convergence result. It is inconvenient that the estimates (4.3.1)–(4.3.3) depend upon $\|x^{(k)}-x^*\|$. The next lemma removes this dependence and gives a result similar to the classical theory in which the errors in x are bounded by the errors in the multiplier estimates $\|\lambda^{(k)}-\lambda^*\|$, (see Bertsekas, 1982b, p.108); however, as an inexact minimization of the augmented Lagrangian function is made, a term reflecting this is also present in the bound. Before giving our result, we need to make an extra assumption.

We use the notation that, if J_1 and J_2 are any subsets of N, $H_L(x^*,\lambda^*)_{[J_1,J_2]}$ is the matrix formed by taking the *rows and columns* of $H_L(x^*,\lambda^*)$ indexed by J_1 and J_2 respectively and $A(x^*)_{[J_1]}$ is the matrix formed by taking the *columns* of $A(x^*)$ indexed by J_1. We use the following assumption

AS5: Suppose that (x^*,λ^*) is a Kuhn-Tucker point for problem (1.5)–(1.7) and that

$$\begin{aligned} J_1 &= \{i \mid (g_L(x^*,\lambda^*))_i = 0 \text{ and } x_i^* > 0\} \\ J_2 &= \{i \mid (g_L(x^*,\lambda^*))_i = 0 \text{ and } x_i^* = 0\} \end{aligned} \tag{5.1.1}$$

Then we assume that the matrix

$$\begin{bmatrix} H_L(x^*,\lambda^*)_{[J,J]} & A^T(x^*)_{[J]} \\ A(x^*)_{[J]} & 0 \end{bmatrix}$$

is non-singular for all sets J, where J is any set made up from the union of J_1 and any subset of J_2.

We note that assumption AS5 implies AS3. Furthermore, any point satisfying the well known second-order sufficiency condition for a minimizer of (1.5)–(1.7) (see, e.g., Fletcher, 1981, Theorem 9.3.2) automatically satisfies AS5 (see, e.g., Gould, 1985).

Lemma 5.1. *Suppose that (AS1) and (AS4) hold. Let* $\{x^{(k)}\}, k \in K$, *be a subsequence which converges to the Kuhn-Tucker point* x^* *for which (AS5) holds and let* λ^* *be the corresponding vector of Lagrange multipliers. Assume that* $\{\lambda^{(k)}\}, k \in K$, *is any sequence of vectors, that* $\{S^{(k)}\}, k \in K$, *is any sequence of diagonal matrices satisfying* $0 < S_1^{-1} \leq S^{(k)} \leq S_2 < \infty$, *and that* $\{\mu^{(k)}\}, k \in K$, *form a non-increasing sequence of positive scalars, so that the product* $\mu^{(k)} \|\lambda^{(k)} - \lambda^*\|$ *converges to zero as k increases. Now, suppose further that*

$$\|P(x^{(k)}, \nabla_x \Phi^{(k)})\| \leq \omega^{(k)} \tag{5.1.2}$$

where the $\omega^{(k)}$ *are positive scalar parameters which converge to zero as* $k \in K$ *increases. Then there are positive constants* $\bar{\mu}, a_3, a_4, a_5, a_6$ *and* s_1 *and an integer value* k_0 *so that if* $\mu^{(k_0)} \leq \bar{\mu}$ *then*

$$\|x^{(k)} - x^*\| \leq a_3 \omega^{(k)} + a_4 \mu^{(k)} \|\lambda^{(k)} - \lambda^*\|, \tag{5.1.3}$$

$$\|\bar{\lambda}(x^{(k)}, \lambda^{(k)}, S^{(k)}, \mu^{(k)}) - \lambda^*\| \leq a_5 \omega^{(k)} + a_6 \mu^{(k)} \|\lambda^{(k)} - \lambda^*\| \tag{5.1.4}$$

and

$$\|c(x^{(k)})\| \leq s_1 (a_5 \omega^{(k)} \mu^{(k)} + (\mu^{(k)} + a_6 (\mu^{(k)})^2) \|\lambda^{(k)} - \lambda^*\|) \tag{5.1.5}$$

for all $k \geq k_0$, $(k \in K)$.

Proof. We will denote the gradient and Hessian of the Lagrangian function at the limit point (x^*, λ^*) by g_L^* and H_L^* respectively.

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 I_1 and I_2, defined in (2.12), have been determined. Then, for $k \in K$, the remaining variables either float (variables in I_3) or oscillate between floating and being dominated (variables in I_4). Now pick an infinite subsequence, $\bar{K}$ of K such that:

(i) $I_5 = I_6 \cup I_7$ with $I_6 \cap I_7 = \emptyset$;

(ii) variables in I_6 are floating for all $k \in \bar{K}$; and

(iii) variables in I_7 are dominated for all $k \in \bar{K}$.

Notice that the set I_3 of (2.12) is contained within I_6. Note, also, that there are only a finite number ($\leq 2^{|I_5|}$) of such subsequences $\bar{K}$ and that for k sufficiently large, each $k \in K$ is in one such subsequence. It is thus sufficient to prove the lemma for $k \in \bar{K}$.

Now, for $k \in \bar{K}$, define

$$I_F = I_1 \cup I_6 \text{ and } I_D = I_2 \cup I_7 . \tag{5.1.6}$$

So, the variables in I_F are floating while those in I_D are dominated. We may deduce that

$$\|\bar{\lambda}^{(k)} - \lambda^*\| \leq a_1 \omega^{(k)} + a_2 \|x^{(k)} - x^*\| \tag{5.1.7}$$

and

$$\|c(x^{(k)})\| \leq s_1 (a_1 \omega^{(k)} \mu^{(k)} + \mu^{(k)} \|\lambda^{(k)} - \lambda^*\| + a_2 \mu^{(k)} \|x^{(k)} - x^*\|) \tag{5.1.8}$$

for all sufficiently large $k \in \bar{K}$ in exactly the same way as we established equations (4.3.1) and

(4.3.3) in the proof of theorem 4.3. Moreover, $\bar{\lambda}^{(k)}$ converges to λ^* and hence $\nabla_x \Phi^{(k)}$ converges to g_L^*. Therefore, from lemma 2.1,

$$x_i^* = 0 \text{ for all } i \in I_D \text{ and } (g_L^*)_i = 0 \text{ for all } i \in I_F. \tag{5.1.9}$$

Using Taylor's theorem,

$$\begin{aligned}
\nabla_x \Phi^{(k)} &= g^{(k)} + A^{(k)T} \bar{\lambda}^{(k)} \\
&= g(x^*) + H(x^*)(x^{(k)} - x^*) + A(x^*)^T \bar{\lambda}^{(k)} \\
&\quad + \sum_{j=1}^{m} \bar{\lambda}_j^{(k)} H_j(x^*)(x^{(k)} - x^*) + r_1(x^{(k)}, x^*, \bar{\lambda}^{(k)}) \qquad (5.1.10) \\
&= g_L(x^*, \lambda^*) + H_L(x^*, \lambda^*)(x^{(k)} - x^*) + A(x^*)^T(\bar{\lambda}^{(k)} - \lambda^*) \\
&\quad + r_1(x^{(k)}, x^*, \bar{\lambda}^{(k)}) + r_2(x^{(k)}, x^*, \bar{\lambda}^{(k)}, \lambda^*),
\end{aligned}$$

where

$$r_1(x^{(k)}, x^*, \bar{\lambda}^{(k)}) = \int_0^1 (H_L(x^{(k)} + s(x^* - x^{(k)}), \bar{\lambda}^{(k)}) - H_L(x^*, \bar{\lambda}^{(k)}))(x^{(k)} - x^*)\, ds \tag{5.1.11}$$

and

$$r_2(x^{(k)}, x^*, \bar{\lambda}^{(k)}, \lambda^*) = \sum_{j=1}^{m} (\bar{\lambda}_j^{(k)} - \lambda_j^*) H_j(x^*)(x^{(k)} - x^*). \tag{5.1.12}$$

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^* then give that

$$\begin{aligned}
&\|r_1(x^{(k)}, x^*, \bar{\lambda}^{(k)})\| \le a_7 \|x^{(k)} - x^*\|^2 \\
&\text{and} \qquad (5.1.13) \\
&\|r_2(x^{(k)}, x^*, \bar{\lambda}^{(k)}, \lambda^*)\| \le a_8 \|x^{(k)} - x^*\| \|\bar{\lambda}^{(k)} - \lambda^*\|
\end{aligned}$$

for some positive constants a_7 and a_8. In addition, again using Taylor's theorem and that $c(x^*) = 0$,

$$c(x^{(k)}) = A(x^*)(x^{(k)} - x^*) + r_3(x^{(k)}, x^*), \tag{5.1.14}$$

where

$$(r_3(x^{(k)}, x^*))_i = \int_0^1 \int_0^1 s\, (x^{(k)} - x^*)^T H_i(x^* + ts(x^{(k)} - x^*))(x^{(k)} - x^*)\, dt\, ds \tag{5.1.15}$$

(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

$$\|r_3(x^{(k)}, x^*)\| \le a_9 \|x^{(k)} - x^*\|^2 \tag{5.1.16}$$

for some constant $a_9 > 0$. Combining (5.1.10) and (5.1.14) we obtain

$$\begin{pmatrix} H_L(x^*, \lambda^*) & A^T(x^*) \\ A(x^*) & 0 \end{pmatrix} \begin{pmatrix} x^{(k)} - x^* \\ \bar{\lambda}^{(k)} - \lambda^* \end{pmatrix} = \begin{pmatrix} \nabla_x \Phi^{(k)} - g_L(x^*, \lambda^*) \\ c(x^{(k)}) \end{pmatrix} - \begin{pmatrix} r_1 + r_2 \\ r_3 \end{pmatrix}, \tag{5.1.17}$$

where we have suppressed the arguments of r_1, r_2 and r_3 for brevity. To proceed further, we introduce the notation that $y_{[J]}$ is the vector formed by taking the components of the vector y indexed by the set J. We may then rewrite (5.1.17) as

$$\begin{pmatrix} H_L(x^*,\lambda^*)_{[I_F,I_F]} & H_L(x^*,\lambda^*)_{[I_F,I_D]} & A^T(x^*)_{[I_F]} \\ H_L(x^*,\lambda^*)_{[I_D,I_F]} & H_L(x^*,\lambda^*)_{[I_D,I_D]} & A^T(x^*)_{[I_D]} \\ A(x^*)_{[I_F]} & A(x^*)_{[I_D]} & 0 \end{pmatrix} \begin{pmatrix} (x^{(k)}-x^*)_{[I_F]} \\ (x^{(k)})_{[I_D]} \\ \bar{\lambda}^{(k)}-\lambda^* \end{pmatrix}$$

(5.1.18)

$$= \begin{pmatrix} (\nabla_x \Phi^{(k)})_{[I_F]} \\ (\nabla_x \Phi^{(k)} - g_L(x^*,\lambda^*))_{[I_D]} \\ c(x^{(k)}) \end{pmatrix} - \begin{pmatrix} (r_1+r_2)_{[I_F]} \\ (r_1+r_2)_{[I_D]} \\ r_3 \end{pmatrix}$$

using (5.1.9). Then, rearranging (5.1.18) and removing the middle horizontal block we obtain

$$\begin{pmatrix} H_L(x^*,\lambda^*)_{[I_F,I_F]} & A^T(x^*)_{[I_F]} \\ A(x^*)_{[I_F]} & 0 \end{pmatrix} \begin{pmatrix} (x^{(k)}-x^*)_{[I_F]} \\ \bar{\lambda}^{(k)}-\lambda^* \end{pmatrix} =$$

(5.1.19)

$$\begin{pmatrix} (\nabla_x \Phi^{(k)})_{[I_F]} - H_L(x^*,\lambda^*)_{[I_F,I_D]}(x^{(k)})_{[I_D]} \\ c(x^{(k)}) - A(x^*)_{[I_D]}(x^{(k)})_{[I_D]} \end{pmatrix} - \begin{pmatrix} (r_1+r_2)_{[I_F]} \\ r_3 \end{pmatrix}.$$

Roughly, the rest of the proof proceeds by showing that the dominant term on the right-hand side of (5.1.19) is the first term and, moreover, that this term is $O(\omega^{(k)})+O(\mu^{(k)}\|\lambda^{(k)}-\lambda^*\|)$. 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

$$\|x^{(k)}_{[I_D]}\| \le \omega^{(k)}, \tag{5.1.20}$$

from (2.10) and (5.1.2) and

$$\|(\nabla_x \Phi^{(k)})_{[I_F]}\| \le \omega^{(k)}, \tag{5.1.21}$$

from (2.11). Consequently, again using (5.1.9),

$$\|x^{(k)}-x^*\| \le \|(x^{(k)}-x^*)_{[I_F]}\| + \omega^{(k)}. \tag{5.1.22}$$

Let $\Delta x^{(k)} = \|(x^{(k)}-x^*)_{[I_F]}\|$ and $\Delta\lambda^{(k)} = \|\bar{\lambda}^{(k)}-\lambda^*\|$. Combining (5.1.7) and (5.1.22), we obtain

$$\Delta\lambda^{(k)} \le a_{10}\omega^{(k)} + a_2 \Delta x^{(k)}, \tag{5.1.23}$$

where $a_{10} = a_1 + a_2$. Furthermore, from (5.1.13), (5.1.16), (5.1.22) and (5.1.23),

$$\left\| \begin{pmatrix} (r_1+r_2)_{[I_F]} \\ r_3 \end{pmatrix} \right\| \le a_{11}(\Delta x^{(k)})^2 + a_{12}\Delta x^{(k)}\omega^{(k)} + a_{13}(\omega^{(k)})^2, \tag{5.1.24}$$

where $a_{11} = a_7 + a_9 + a_8 a_2$, $a_{12} = 2(a_7+a_9) + a_8(a_{10}+a_2)$ and $a_{13} = a_7 + a_9 + a_8 a_{10}$. Moreover, from (5.1.8), (5.1.20), (5.1.21) and (5.1.22),

$$\left\| \begin{pmatrix} (\nabla_x \Phi^{(k)})_{[I_F]} - H_L(x^*,\lambda^*)_{[I_F,I_D]}(x^{(k)})_{[I_D]} \\ c(x^{(k)}) - A(x^*)_{[I_D]}(x^{(k)})_{[I_D]} \end{pmatrix} \right\| \le$$

(5.1.25)

$$a_{14}\omega^{(k)} + s_1(\mu^{(k)}\|\lambda^{(k)}-\lambda^*\| + a_{10}\omega^{(k)}\mu^{(k)} + a_2\mu^{(k)}\Delta x^{(k)}),$$

where

(5.1.26) $$a_{14} = 1 + \left\| \begin{pmatrix} H_L(x^*, \lambda^*)_{[I_F, I_D]} \\ A(x^*)_{[I_D]} \end{pmatrix} \right\|.$$

By assumption AS5, the coefficient matrix on the left-hand side of (5.1.19) is non-singular. Let its inverse have norm M. Multiplying both sides of the equation by this inverse and taking norms, we obtain

(5.1.27) $$\left\| \begin{pmatrix} (x^{(k)} - x^*)_{[I_F]} \\ \bar{\lambda}^{(k)} - \lambda^* \end{pmatrix} \right\| \le M[a_{14}\omega^{(k)} + s_1(\mu^{(k)}\|\lambda^{(k)} - \lambda^*\| + a_{10}\omega^{(k)}\mu^{(k)} + a_2\mu^{(k)}\Delta x^{(k)}) + a_{11}(\Delta x^{(k)})^2 + a_{12}\Delta x^{(k)}\omega^{(k)} + a_{13}(\omega^{(k)})^2].$$

Now, suppose that k is sufficiently large that

(5.1.28) $$\omega^{(k)} \le \min(1, 1/(4Ma_{12})).$$

Furthermore, let

(5.1.29) $$\bar{\mu} = \min(1, 1/(4Ma_2 s_1)).$$

Then, if $\mu^{(k)} \le \bar{\mu}$, (5.1.27), (5.1.28) and (5.1.29) give

(5.1.30) $$\Delta x^{(k)} \le \tfrac{1}{2}\Delta x^{(k)} + M(a_{15}\omega^{(k)} + s_1\mu^{(k)}\|\lambda^{(k)} - \lambda^*\| + a_{11}(\Delta x^{(k)})^2),$$

where $a_{15} = s_1 a_{10} + a_{13} + a_{14}$. Thus

(5.1.31) $$\phi(\Delta x^{(k)}) \equiv Ma_{11}(\Delta x^{(k)})^2 - \tfrac{1}{2}\Delta x^{(k)} + M(a_{15}\omega^{(k)} + s_1\mu^{(k)}\|\lambda^{(k)} - \lambda^*\|) \ge 0,$$

a quadratic inequality in $\Delta x^{(k)}$. Now consider $\phi(\Delta x^{(k)})$ in detail. For convenience, write $\phi(\alpha) = a\alpha^2 - b\alpha + c$. Then

(a) a and b are positive constants while c is positive and approaches zero as k increases. Therefore, $\phi(0) = c$ is also positive.

(b) $\phi(\alpha)$ is convex and attains its smallest value $c - (b^2/4a)$ at $\alpha_{\min} = b/2a$. This smallest value is negative for all k sufficiently large, and hence $\phi(\alpha)$ has two positive real roots for such k. Inequality (5.1.31), and the fact that $\Delta x^{(k)}$ approaches zero as k increases then imply that $\Delta x^{(k)}$ is no larger than the smaller of these two roots, α_1 say, for all $k \in \bar{K}$ sufficiently large.

(c) Let $\psi(\alpha)$ be the linear function which interpolates $\phi(\alpha)$ at the values $\alpha = 0$ and $\alpha = \alpha_{\min}$. Then $\psi(\alpha) = c - \tfrac{1}{2}b\alpha$. As ϕ is convex, $\psi(\alpha) \ge \phi(\alpha)$ for all α between 0 and $\alpha_{\min}$. Hence, α_1 is no larger than the root $\alpha = 2c/b$ of $\psi(\alpha) = 0$. Hence inequality (5.1.31) gives that

(5.1.32) $$\Delta x^{(k)} \le 2c/b = 4M(a_{15}\omega^{(k)} + s_1\mu^{(k)}\|\lambda^{(k)} - \lambda^*\|).$$

Writing $a_3 = 4Ma_{15} + 1$ and $a_4 = 4Ms_1$, we obtain the desired inequality (5.1.3) from (5.1.22) and (5.1.32). Now, using (5.1.3) and (5.1.7), we obtain (5.1.4), where $a_5 = a_1 + a_2 a_3$ and $a_6 = a_2 a_4$. Finally, (5.1.5) follows from (5.1.4) by substituting for $\bar{\lambda}^{(k)}$, using (2.5), and multiplying the inequality by $\mu^{(k)}$. ■

We can obtain the following simple corollary.

Corollary 5.2. *Suppose that the conditions of lemma 5.1 hold and that $\hat{\lambda}^{(k+1)}$ is any Lagrange multiplier estimate for which*

(5.2.1) $$\|\hat{\lambda}^{(k+1)} - \lambda^*\| \le a_{16}\|x^{(k)} - x^*\| + a_{17}\omega^{(k)},$$

for some positive constants a_{16} and a_{17} and all $k \in K$ sufficiently large. Then there are positive constants $\bar{\mu}$, a_3, a_4, a_5, a_6, s_1 and an integer value k_0 so that if $\mu^{(k_0)} \leq \bar{\mu}$ then (5.1.3),

$$\|\hat{\lambda}^{(k+1)} - \lambda^*\| \leq a_5 \omega^{(k)} + a_6 \mu^{(k)} \|\lambda^{(k)} - \lambda^*\| \tag{5.2.2}$$

and (5.1.5) hold for all $k \geq k_0$, $(k \in K)$.

Proof. Inequality (5.2.2) follows immediately from (5.2.1) and (5.1.3). ■

We now show that the penalty parameter will normally be bounded away from zero in both algorithms 1 and 2. This is important as many methods for solving the inner iteration subproblem will encounter difficulties if the parameter converges to zero since this causes the Hessian of the augmented Lagrangian to become increasingly ill conditioned.

Theorem 5.3. *Suppose that algorithm 1 or 2 of §3 converges to the single limit point x^*, that (AS1), (AS4) and (AS5) hold, that α_η and β_η satisfy*

$$\alpha_\eta < \alpha \equiv \min(1, \alpha_\omega) \tag{5.3.1}$$

$$\beta_\eta < \min(1, \beta_\omega) \tag{5.3.2}$$

and that (5.2.1) holds for all k sufficiently large when algorithm 2 is used. Then there is a constant $\underline{\mu} > 0$ such that $\mu^{(k)} \geq \underline{\mu}$ for all k.

Proof. Suppose, otherwise, that $\mu^{(k)}$ tends to zero. Then, step 3 of the algorithm must be executed infinitely often. We aim to obtain a contradiction to this statement by showing that step 2 is always executed for k sufficiently large.

Firstly, we show that the sequence of Lagrange multipliers $\{\lambda^{(k)}\}$ converge to λ^*.

Consider algorithm 1. The result is clear if step 2 is executed infinitely often as each time the step is executed, $\lambda^{(k+1)} = \bar{\lambda}^{(k)}$ and the inequality (4.3.1) guarantees that $\bar{\lambda}^{(k)}$ converges to λ^*. Suppose that step 2 is not executed infinitely often. Then $\|\lambda^{(k)} - \lambda^*\|$ will remain fixed for all $k \geq k_1$ for some k_1, as step 3 is executed for each remaining iteration. But then (4.3.3) implies that $\|c(x^{(k)})\| \leq a_{17} \mu^{(k)}$ for some constant a_{17} for all $k \geq k_2 \geq k_1$. As $\mu^{(k)}$ converges to zero as k increases, $a_{17} \mu^{(k)} \leq \eta_0 (\mu^{(k)})^{\alpha_\eta} = \eta^{(k)}$ for all k sufficiently large. But then inequality (3.2) must be satisfied for some $k \geq k_1$ which is impossible as this would imply that step 2 is again executed. Hence, step 2 must be executed infinitely often.

Now consider algorithm 2. The result is clear if the multipliers updates are accepted infinitely often, as each time the update is performed $\lambda^{(k+1)} = \hat{\lambda}^{(k+1)}$ and assumption (5.2.1) guarantees that $\hat{\lambda}^{(k+1)}$ converges to λ^*. Suppose that the update is not accepted infinitely often. Then for all k sufficiently large, $\|\hat{\lambda}^{(k+1)}\| > \nu (\mu^{(k+1)})^{-\gamma}$ which implies that $\|\hat{\lambda}^{(k+1)}\|$ diverges. But this contradicts assumption (5.2.1) and hence $\lambda^{(k)}$ converges to λ^*.

Therefore $\mu^{(k)} \|\lambda^{(k)} - \lambda^*\|$ tends to zero as k increases for both algorithms.

Let k_1 be the smallest integer for which

$$\mu^{(k)} \leq \gamma_1 < 1 \tag{5.3.3}$$

for all $k \geq k_1$. Now let $\omega^{(k)}$ be as generated by either algorithm. Notice that, by construction and inequality (5.3.3),

$$\omega^{(k)} \leq \omega_0 (\mu^{(k)})^{\alpha_\omega} \tag{5.3.4}$$

for all $k \geq k_1$. We shall apply lemma 5.1 or corollary 5.2 to the iterates generated by the

algorithm; we identify the set K with the complete set of integers larger than k_1 and the scalars $\mu^{(k)}$ with the set of penalty parameters computed in steps 2 and 3 of either algorithm. Therefore we can ensure that $\mu^{(k)}$ is sufficiently small so that lemma 5.1 applies to step 1 of algorithm 1 (or corollary 5.2 to step 1 of algorithm 2) and thus that there is an integer k_2 and constants a_5, a_6 and s_1 so that (5.1.4)/(5.2.2) and (5.1.5) hold for all $k \geq k_2$. Let k_3 be the smallest integer such that

$$(\mu^{(k)})^{1-\alpha_\eta} \leq \frac{\eta_0}{\omega_0 s_1 (a_5+2)}, \tag{5.3.5}$$

$$(\mu^{(k)})^{1-\beta_\eta} \leq \min\left(\frac{1}{a_{18}}, \frac{\eta_0}{\omega_0 s_1 (a_5+2a_{18})}\right), \tag{5.3.6}$$

and, if algorithm 2 is used,

$$(\mu^{(k)})^{\gamma} \leq \frac{\nu}{\|\lambda^*\|+\omega_0 a_{18}}, \tag{5.3.7}$$

where $a_{18} = a_5 + a_6$. Notice that (5.3.3) and (5.3.6) imply that

$$\mu^{(k)} \leq (\mu^{(k)})^{1-\beta_\eta} \leq \frac{1}{a_{18}} \leq \frac{1}{a_6}. \tag{5.3.8}$$

for all $k \geq k_3$. Furthermore, let k_4 be such that

$$\|\lambda^{(k)} - \lambda^*\| \leq \omega_0 \tag{5.3.9}$$

for all $k \geq k_4$. Now define $k_5 = \max(k_1, k_2, k_3, k_4)$, let Γ be the set $\{k|$ Step 3 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.

If algorithm 2 is used, inequality (5.2.2) gives that

$$\begin{aligned}
\|\hat{\lambda}^{(k+1)}\| &\leq \|\lambda^*\| + a_5 \omega^{(k)} + a_6 \mu^{(k)} \|\lambda^{(k)} - \lambda^*\| \\
&\leq \|\lambda^*\| + a_5 \omega_0 (\mu^{(k)})^{\alpha_\omega} + a_6 \mu^{(k)} \|\lambda^{(k)} - \lambda^*\| && \text{(from (5.3.4))} \\
&\leq \|\lambda^*\| + \omega_0 (a_5 (\mu^{(k)})^{\alpha_\omega} + a_6 \mu^{(k)}) && \text{(from (5.3.9))} \\
&\leq \|\lambda^*\| + \omega_0 a_{18} (\mu^{(k)})^{\alpha} && \text{(from (5.3.1))} \\
&\leq \|\lambda^*\| + \omega_0 a_{18} && \text{(from (5.3.3))} \\
&\leq \nu (\mu^{(k+1)})^{-\gamma}
\end{aligned} \tag{5.3.10}$$

for all $k > k_5$, the last inequality following from (5.3.7) and because $\mu^{(k+1)} \leq \mu^{(k)}$. Hence, the multiplier update $\lambda^{(k+1)} = \hat{\lambda}^{(k+1)}$ in algorithm 2 will always take place when $k \geq k_0$.

For iteration k_0, $\omega^{(k_0)} = \omega_0 (\mu^{(k_0)})^{\alpha_\omega}$ and $\eta^{(k_0)} = \eta_0 (\mu^{(k_0)})^{\alpha_\eta}$. Then (5.1.5) gives

$$
\begin{aligned}
\|c(x^{(k_0)})\| &\le s_1\,((\mu^{(k_0)} + a_6(\mu^{(k_0)})^2)\|\lambda^{(k_0)} - \lambda^*\| + a_5\,\omega^{(k_0)}\,\mu^{(k_0)}) \\
&\le s_1\,(2\mu^{(k_0)}\|\lambda^{(k_0)} - \lambda^*\| + a_5\,\omega^{(k_0)}\,\mu^{(k_0)}) && \text{(from (5.3.8))} \\
&\le s_1\,(2\omega_0\mu^{(k_0)} + a_5\,\omega_0(\mu^{(k_0)})^{1+\alpha_\omega}) && \text{(from (5.3.9))} \\
&\le \omega_0\,s_1\,(a_5+2)\mu^{(k_0)} && \text{(from (5.3.3))} \\
&\le \eta_0(\mu^{(k_0)})^{\alpha_\eta} = \eta^{(k_0)} && \text{(from (5.3.5)).}
\end{aligned} \tag{5.3.11}
$$

Thus, from (5.3.11), Step 2 of algorithm 1 or the same step of algorithm 2 will be executed with $\lambda^{(k_0+1)} = \bar{\lambda}(x^{(k_0)}, \lambda^{(k_0)}, S^{(k_0)}, \mu^{(k_0)})$ or $\lambda^{(k_0+1)} = \hat{\lambda}^{(k_0+1)}$ respectively. Inequality (5.1.4)/(5.2.2) in conjunction with (5.3.1), (5.3.4) and (5.3.9) guarantee that

$$
\|\lambda^{(k_0+1)} - \lambda^*\| \le a_5\,\omega^{(k_0)} + a_6\,\mu^{(k_0)}\|\lambda^{(k_0)} - \lambda^*\| \le \omega_0\,a_{18}(\mu^{(k_0)})^{\alpha}\,. \tag{5.3.12}
$$

We shall now suppose that step 2 is executed for iterations k_0+i, $(0 \le i \le j)$, and that

$$
\|\lambda^{(k_0+i+1)} - \lambda^*\| \le \omega_0\,a_{18}\,(\mu^{(k_0)})^{\alpha+\beta_\eta i}. \tag{5.3.13}
$$

Inequalities (5.3.11) and (5.3.12) 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 $\mu^{(k_0+j+1)} = \mu^{(k_0)}$, $\omega^{(k_0+j+1)} = \omega_0(\mu^{(k_0)})^{\beta_\omega(j+1)+\alpha_\omega}$ and $\eta^{(k_0+j+1)} = \eta_0(\mu^{(k_0)})^{\beta_\eta(j+1)+\alpha_\eta}$. Then (5.1.5) gives

$$
\begin{aligned}
\|c(x^{(k_0+j+1)})\| &\le s_1((\mu^{(k_0+j+1)} + a_6(\mu^{(k_0+j+1)})^2)\|\lambda^{(k_0+j+1)} - \lambda^*\| \\
&\qquad + a_5\,\omega^{(k_0+j+1)}\,\mu^{(k_0+j+1)}) \\
&\le s_1\,(2\mu^{(k_0+j+1)}\|\lambda^{(k_0+j+1)} - \lambda^*\| \\
&\qquad + a_5\,\omega^{(k_0+j+1)}\,\mu^{(k_0+j+1)}) && \text{(from (5.3.8))} \\
&\le s_1\,(2\omega_0 a_{18}\mu^{(k_0)}(\mu^{(k_0)})^{\alpha+\beta_\eta j} \\
&\qquad + a_5\,\omega_0(\mu^{(k_0)})^{\alpha_\omega+\beta_\omega(j+1)+1}) && \text{(from (5.3.13))} \\
&\le s_1\,(2\omega_0 a_{18}\mu^{(k_0)}(\mu^{(k_0)})^{\alpha_\eta+\beta_\eta j} && \text{(from (5.3.1), (5.3.2)} \\
&\qquad + a_5\,\omega_0(\mu^{(k_0)})^{\alpha_\eta+\beta_\eta(j+1)+1}) && \text{and (5.3.3))} \\
&\le \omega_0\,s_1\,(a_5+2a_{18})(\mu^{(k_0)})^{1-\beta_\eta}(\mu^{(k_0)})^{\beta_\eta(j+1)+\alpha_\eta} && \text{(from (5.3.3))} \\
&\le \eta_0(\mu^{(k_0)})^{\beta_\eta(j+1)+\alpha_\eta} = \eta^{(k_0+j+1)} && \text{(from (5.3.6)).}
\end{aligned} \tag{5.3.14}
$$

Thus, from (5.3.14), step 2 of algorithm 1 or the same step of algorithm 2 will be executed with $\lambda^{(k_0+j+2)} = \bar{\lambda}(x^{(k_0+j+1)}, \lambda^{(k_0+j+1)}, S^{(k_0+j+1)}, \mu^{(k_0+j+1)})$ or $\lambda^{(k_0+j+2)} = \hat{\lambda}^{(k_0+j+2)}$ respectively. Inequality (5.1.4)/(5.2.2) then guarantees that

$$
\begin{aligned}
\|\lambda^{(k_0+j+2)}-\lambda^*\| &\le a_5\omega^{(k_0+j+1)}+a_6\mu^{(k_0+j+1)}\|\lambda^{(k_0+j+1)}-\lambda^*\| \\
&\le \omega_0 a_5(\mu^{(k_0)})^{\alpha_\omega+\beta_\omega(j+1)} \\
&\qquad +\omega_0 a_6 a_{18}\mu^{(k_0)}(\mu^{(k_0)})^{\alpha+\beta_\eta j} \qquad \text{(from (5.3.13))} \\
&\le \omega_0 a_5(\mu^{(k_0)})^{\alpha+\beta_\eta(j+1)} \qquad \text{(from (5.3.1), (5.3.2)} \\
&\qquad +\omega_0 a_6 a_{18}\mu^{(k_0)}(\mu^{(k_0)})^{\alpha+\beta_\eta j} \qquad \text{and (5.3.3))} \\
&= \omega_0(a_5+a_6 a_{18}(\mu^{(k_0)})^{1-\beta_\eta})(\mu^{(k_0)})^{\alpha+\beta_\eta(j+1)} \\
&\le \omega_0(a_5+a_6)(\mu^{(k_0)})^{\alpha+\beta_\eta(j+1)} \qquad \text{(from (5.3.6))} \\
&= \omega_0 a_{18}(\mu^{(k_0)})^{\alpha+\beta_\eta(j+1)}
\end{aligned} \tag{5.3.15}
$$

which establishes (5.3.13) for $i=j+1$. Hence, step 2 of the appropriate algorithm is executed for all iterations $k \ge k_0$. But this implies that Γ is finite which contradicts the assumption that step 3 is executed infinitely often. Hence the theorem is proved. ■

Notice, in particular, that if algorithm 2 is used with $\hat{\lambda}^{(k+1)}$ chosen as either the first order or least-squares multiplier estimates, the penalty parameter $\mu^{(k)}$ will stay bounded away from zero. This follows directly from theorem 5.3 because of the inequalities (4.3.1) and (4.3.2).

Our definition of floating variables has a further desirable consequence if we make the following additional assumption.

AS6: (Strict complementary slackness condition) If the iterates $x^{(k)}$, $k \in K$, converge to the the limit point x^* with corresponding Lagrange multipliers λ^*, we assume that the set

$$J_2 = \{i \mid (g_L(x^*,\lambda^*))_i = 0 \text{ and } x_i^* = 0\} \tag{5.4.1}$$

is empty.

Notice that if inequality constraints $c_i(x) \ge 0$ have been converted to equations by the subtraction of slack variables (i.e, rewritten as $c_i(x) - x_{n+i} = 0, x_{n+i} \ge 0$), this statement of strict complementary slackness is equivalent to the more usual one which says that no inequality constraint shall be both active (the constraint function vanishing) and have a corresponding zero Lagrange parameter (see, e.g., Fletcher, 1981, p.51). For it is easy to show that the Lagrange parameter for such a constraint is precisely the corresponding component of the gradient of the Lagrangian function. A constraint being active and having a corresponding zero Lagrange parameter is thus the same as the slack variable having the value zero and its corresponding element in the gradient of the Lagrangian function vanishing so the latter is excluded under AS6.

Theorem 5.4 *Suppose that the iterates $x^{(k)}$, $k \in K$, converge to the the limit point x^* with corresponding Lagrange multipliers λ^*, that (AS1-AS3) and (AS6) hold. Then for k sufficiently large, the set of floating variables are precisely those which lie away from their bounds at x^*.*

Proof. From theorem 4.3, $\nabla_x \Phi^{(k)}$ converges to $g_L(x^*,\lambda^*)$ and from Lemma 2.1, the variables in the set I_5 then converge to zero and the corresponding components of $g_L(x^*,\lambda^*)$ are zero. Hence, under AS6, I_5 is null. Therefore, each variable ultimately remains tied to one of the sets I_1 or I_2 for all k sufficiently large; a variable in I_1 is, by definition, floating and converges to a value away from its bound. Conversely, a variable in I_2 is dominated and converges to its bound. ■

As a consequence of theorem 5.4, the least-squares multiplier estimates (2.2.1) are implementable. By this we mean that if $\bar{A}^{(k)}$ and $\bar{g}^{(k)}$ are the columns of $A(x^{(k)})$ and components of $g(x^{(k)})$ corresponding to the floating variables at $x^{(k)}$ respectively, the estimates

$$\hat{\lambda}^{(k)} = -(\bar{A}^{(k)+})^T \bar{g}^{(k)} \tag{5.4.2}$$

are identical to those given by (2.2.1) for all k sufficiently large. The estimates (5.4.2), unlike (2.2.1), are well defined when x^* is unknown.

We conclude the section by giving a rate-of-convergence result for our algorithms. For a comprehensive discussion of convergence, the reader is referred to Ortega and Rheinboldt (1970).

Theorem 5.5. *Under the assumptions of Theorem 5.3, the iterates $x^{(k)}$, the Lagrange multiplier estimates $\bar{\lambda}^{(k)}$ of algorithm 1 and any $\hat{\lambda}^{(k)}$ satisfying (5.2.1) for algorithm 2 are at least R-linearly convergent with R-factor at most $\hat{\mu}^{\min(\beta_\omega, \beta_\eta)}$, where $\hat{\mu} = \min[\gamma_1, \mu]$ and where μ is the smallest value of the penalty parameter generated by the algorithm in question.*

Proof. The proof parallels that of Lemma 5.1. Firstly, for k sufficiently large, Theorem 5.3 shows that the penalty parameter $\mu^{(k)}$ remains fixed at some value μ, say, and, for all subsequent iterations inequalities (3.2)/(3.6) and

$$\omega^{(k+1)} = (\hat{\mu})^{\beta_\omega} \omega^{(k)} \text{ and } \eta^{(k+1)} = (\hat{\mu})^{\beta_\eta} \eta^{(k)} \tag{5.5.1}$$

hold. Then, from (3.2)/(3.6), (5.1.20) and (5.1.21), the bound on the right-hand side of (5.1.25) may be replaced by $a_{14}\omega^{(k)} + \eta^{(k)}$ and consequently

$$\Delta x^{(k)} \leq M(a_{14}\omega^{(k)} + \eta^{(k)} + a_{11}(\Delta x^{(k)})^2 + a_{12}\Delta x^{(k)}\omega^{(k)} + a_{13}(\omega^{(k)})^2). \tag{5.5.2}$$

Hence, if k is sufficiently large that

$$\omega^{(k)} \leq \min(1, 1/(2Ma_{12})), \tag{5.5.3}$$

(5.5.2) can be rearranged to give

$$\psi(\Delta x^{(k)}) \equiv Ma_{11}(\Delta x^{(k)})^2 - \tfrac{1}{2}\Delta x^{(k)} + M(a_{19}\omega^{(k)} + \eta^{(k)}) \geq 0, \tag{5.5.4}$$

where $a_{19} = a_{13} + a_{14}$. This is of the same form as ϕ in (5.1.31) and implies that

$$\Delta x^{(k)} \leq 4M(a_{19}\omega^{(k)} + \eta^{(k)}) \tag{5.5.5}$$

in the same way that (5.1.31) gave (5.1.32). But then (5.1.22) and (5.5.5) give

$$\|x^{(k)} - x^*\| \leq a_{20}\omega^{(k)} + a_{21}\eta^{(k)}, \tag{5.5.6}$$

where $a_{20} = 1 + 4Ma_{19}$ and $a_{21} = 4M$. As, by assumption, $\beta_\eta < 1$, (5.5.1) and (5.5.6) show that $x^{(k)}$ converges at least R-linearly, with R-factor $\hat{\mu}^{\min(\beta_\omega, \beta_\eta)}$, to x^*. That the same is true for $\lambda^{(k)}$ and $\hat{\lambda}^{(k)}$ follows directly from (5.1.7)/(5.2.1) and (5.5.6). ■

6 An example

In Theorem 5.3, we showed that, if there is a unique limit point for the iterates generated by the algorithms, the penalty parameter $\mu^{(k)}$ is necessarily bounded away from zero. We now show that, if there is more than a single limit point, it is indeed possible for the penalty parameter to become arbitrarily close to zero.

We consider the problem

(6.1) $$\underset{x}{\text{minimize}}\ \sigma x$$

subject to the single constraint

(6.2) $$x^2 - 1 = 0,$$

for some $\sigma > 0$. This problem has two stationary points, namely

(6.3) $$(x_1^*, \lambda_1^*) = (-1, \frac{\sigma}{2}) \text{ and } (x_2^*, \lambda_2^*) = (1, -\frac{\sigma}{2}).$$

No bounds appear in the problem, and hence $P(x^{(k)}, \nabla_x \Phi^{(k)}) = \nabla_x \Phi^{(k)}$ for all k. (Of course, strictly we have not yet defined our algorithms for such a case – this case is covered in §8; however, we might think of (6.1)–(6.2) as resulting from a transformation of variables where the non-negativity constraint has been shifted so as to play no role here). For simplicity, we choose $S^{(k)} = I$ for all k, and it can be verified that

(6.4) $$\nabla_x \Phi(x, \lambda, I, \mu) = \frac{2}{\mu} x(x^2 - 1) + 2x\lambda + \sigma.$$

We wish to show that algorithm 1 can generate a sequence of points that oscillate between x_2^* and a neighbourhood of x_1^*, and such that the penalty parameter $\mu^{(k)}$ tends to zero. The idea is to consider an infinite sequence of iteration cycles, each of length $j+1$, where j is the smallest integer such that

(6.5) $$\eta_0 (\min[\mu_0, \gamma_1])^{\alpha_\eta + j\beta_\eta} < \frac{\sigma}{2} \min[\mu_0, \gamma_1].$$

For every iteration, $\lambda^{(k)}$ is equal to λ_2^*. For the first j iterations of each cycle, $x^{(k)}$ is equal to x_2^* and step 2 is executed; for the iteration that remains, $x^{(k)}$ has a value less than x_1^* and the penalty parameter is reduced as step 3 is executed. The process is initialized with $\lambda^{(0)} = \lambda_2^*$.

It remains to show that such a sequence can be constructed. It involves two types of iterations:

(i) iterations such that $x^{(k)} = x_2^*$ and step 2 is executed,

(ii) iterations such that $x^{(k-1)} = x_2^*$, $x^{(k)} < x_1^*$ and step 3 is executed.

Consider case (i) first. Since $\lambda^{(k)} = \lambda_2^*$, $x^{(k)} = x_2^*$ is a stationary point of the augmented Lagrangian function Φ. Therefore conditions (3.1) and (3.2) are trivially satisfied. Hence step 2 is executed and, as a consequence, both $\eta^{(k)}$ and $\omega^{(k)}$ are reduced. However, since $c(x^{(k)}) = 0$, the Lagrange multiplier estimate is not modified.

We now consider case (ii). We have to show that it is possible to have $x^{(k)} < x_1^*$ with $\|c(x^{(k)})\| > \eta^{(k)}$. Equivalently, we show that inequality (3.1) (but not (3.2)) of step 1 is satisfied for some $x^{(k)}$ of the form

(6.6) $$x^{(k)} < -1.$$

We note that the mechanism of the algorithm and (6.5) imply that

(6.7) $$\frac{\eta^{(k)}}{\mu^{(k)}} < \frac{\sigma}{2}.$$

Dropping the superscripts and remembering that $\lambda^{(k)} = \lambda_2^* = -\frac{1}{2}\sigma$, we thus require that the inequalities (6.6),

(6.8) $$|\psi(x)| \equiv \left| \frac{2}{\mu} x(x^2 - 1) - \sigma x + \sigma \right| \leq \omega$$

and

(6.9) $$|x^2-1|>\eta$$

are satisfied. Now observe that any $x\in(-\sqrt{1+\sigma\mu},-\sqrt{1+\eta})$ satisfies (6.9). At the end-points of the interval, one has that

(6.10) $$\psi(-\sqrt{1+\eta})=\sqrt{1+\eta}\left[\frac{-2\eta}{\mu}+\sigma+\frac{\sigma}{\sqrt{1+\eta}}\right]>0$$

and

(6.11) $$\psi(-\sqrt{1+\sigma\mu})=\sigma\sqrt{1+\sigma\mu}\left[-1+\frac{1}{\sqrt{1+\sigma\mu}}\right]<0.$$

The continuity of the function ψ along with (6.10) and (6.11) implies the existence of a root inside the interval. Any x sufficiently close to this root will therefore satisfy the required inequalities (6.6), (6.8) and (6.9) and we select such a point to define $x^{(k)}$. Because of (6.9), step 3 is executed and $\lambda^{(k)}$ remains equal to λ_2^*.

Furthermore, since the interval $(-\sqrt{1+\sigma\mu^{(k)}},-\sqrt{1+\eta^{(k)}})$ of case (ii) shrinks to the single point $x_1^*=-1$ as k tends to infinity, this point is the only possible limit point of the sequence of iterates besides x_2^*.

This completes our example for algorithm 1.

We now show that a slightly modified form of this example applies to algorithm 2. Given μ_0, pick σ sufficiently large such that

(6.12) $$\mu_0<\frac{\sigma}{2}.$$

Assume first that

(6.13) $$|\lambda_1^*|=|\lambda_2^*|\leq\nu\mu_0^{-\gamma}$$

and we construct an infinite sequence of iteration cycles, each of length j+2 with j defined as before to be the smallest integer such that inequality (6.5) is satisfied. The first j iterations are identical to those already described in case (i) above. Iteration j+1 is identical to case (ii) above, except that the Lagrange multiplier estimate is set to λ_1^*. For the remaining iteration, a point $x^{(k)}$ close to x_2^* is selected and the Lagrange multiplier estimate reset to λ_2^*. Notice that for all iterations, (5.2.1) is trivially satisfied since its left hand side is zero, and also that the new Lagrange multiplier estimate is acceptable because of (6.13).

It remains to show that we can construct a suitable iterate $x^{(k)}$ at the $(j+2)^{\text{nd}}$ iteration of each cycle. Dropping the superscripts again, we thus require that

(6.14) $$\|\nabla_x\Phi(x,\lambda_1^*,I,\mu)\|=\left|\frac{2}{\mu}x(x^2-1)+\sigma x+\sigma\right|\leq\omega,$$

which can be achieved by choosing x sufficiently close to the zero

(6.15) $$1+\frac{\sqrt{1-2\sigma\mu}}{2}$$

of that function. The Lagrange multiplier estimate is then reset to λ_2^*, which is allowed because of (6.13) and the fact that the root (6.15) converges to $x_2^*=1$ as μ tends to zero. Whether or not (3.6) holds is irrelevant, since its failure only causes a further reduction of the penalty parameter, which suits our purpose.

Assume now that (6.13) is not satisfied. Observe therefore that the Lagrange multiplier estimate will not be updated until

$$|\lambda_1^*| = |\lambda_2^*| \leq \nu(\mu^{(k)})^{-\gamma}. \tag{6.16}$$

We will ensure that this happens by using the iterates described for algorithm 1. These iterates are applicable because, as already observed, the Lagrange multiplier estimates of algorithm 1 are never updated. As soon as (6.16) holds, we revert to the sequence described above for algorithm 2, ensuring again that $\mu^{(k)}$ tends to zero.

7 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 $x^{(k)}$ satisfies the following second-order sufficiency condition:

> **AS7:** Suppose that $x^{(k)}$ satisfies (3.1)/(3.5), converges to x^* for $k \in K$ and that J_1 and J_2 are as defined by (5.1.1). Then we assume that $\nabla_{xx}\Phi^{(k)}_{[J,J]}$ is uniformly positive definite (that is, its smallest eigenvalue is uniformly bounded away from zero) for all $k \in K$ sufficiently large and all sets J, where J is any set made up from the union of J_1 and any subset of J_2.

With such a condition we have the following result.

Theorem 7.1. *Under (AS1-AS3) and (AS7), the iterates $x^{(k)}$, $k \in K$, generated by either algorithm 1 or 2 converge to an isolated local solution of (1.5)–(1.7).*

Proof. By definition of Φ,

$$\nabla_{xx}\Phi^{(k)} = H_L(x^{(k)}, \bar{\lambda}^{(k)}) + A^{(k)T} S^{(k)} A^{(k)} / \mu^{(k)}. \tag{7.1}$$

Let $s_{[J]}$ be any non-zero vector satisfying

$$A^{(k)}_{[J]} s_{[J]} = 0, \tag{7.2}$$

where J is any set as described in (AS7). Then for any such vector,

$$s^T_{[J]} \nabla_{xx}\Phi^{(k)}_{[J,J]} s_{[J]} \geq \varepsilon \tag{7.3}$$

for some $\varepsilon > 0$ under (AS7). It follows from (7.1)–(7.3) that

$$s^T_{[J]} H_L(x^{(k)}, \bar{\lambda}^{(k)})_{[J,J]} s_{[J]} \geq \varepsilon \tag{7.4}$$

By continuity of H_L as $x^{(k)}$ and $\bar{\lambda}^{(k)}$ approach their limits, this gives that

$$s^T_{[J]} H_L(x^*, \lambda^*)_{[J,J]} s_{[J]} \geq \varepsilon \tag{7.5}$$

for all non-zero $s_{[J]}$ satisfying (7.2) which implies that x^* is an isolated local solution to (1.5)–(1.7) (see, for example, Fletcher, 1981, Theorem 9.3.2). ■

There is a weaker version of this result, proved in the same way, that if the assumption of uniform positive definiteness in AS7 is replaced by an assumption of positive semi-definiteness, the limit point then satisfies second-order necessary conditions (Fletcher, 1981, Theorem 9.3.1) for being a minimizer.

8 Further comments

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 (2.1) by

$$(P[x])_i = \begin{cases} l_i \text{ if } x_i \le l_i \\ u_i \text{ if } x_i \ge u_i \\ x_i \text{ otherwise} \end{cases} \tag{8.1}$$

and let $B = \{x \mid l \le x \le u\}$, we may then use the algorithms of §3 without further significant modification. Our concept of floating and dominated variables stays essentially the same; for any iterate $x^{(k)}$ in B we have three mutually exclusive possibilities for each component $x_i^{(k)}$, namely

$$\begin{aligned} &\text{(i)} \quad 0 \le x_i^{(k)} - l_i \le (\nabla_x \Phi^{(k)})_i \\ &\text{(ii)} \quad (\nabla_x \Phi^{(k)})_i \le x_i^{(k)} - u_i \le 0 \\ &\text{(iii)} \quad x_i^{(k)} - u_i < (\nabla_x \Phi^{(k)})_i < x_i^{(k)} - l_i. \end{aligned} \tag{8.2}$$

In case (i) we then have

$$(P(x^{(k)}, \nabla_x \Phi^{(k)}))_i = x_i^{(k)} - l_i \tag{8.3}$$

whereas in case (ii) we have

$$(P(x^{(k)}, \nabla_x \Phi^{(k)}))_i = x_i^{(k)} - u_i \tag{8.4}$$

and in case (iii)

$$(P(x^{(k)}, \nabla_x \Phi^{(k)}))_i = (\nabla_x \Phi^{(k)})_i. \tag{8.5}$$

The $x_i^{(k)}$ which satisfies (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 (2.12) are straightforward to define. I_1 now contains variables which float for all $k \in K$ sufficiently large and converge to the interior of B. I_2 is now the union of the two sets I_{2l}, made up of variables which are dominated above for all $k \in K$ sufficiently large, and I_{2u}, made up of variables which are dominated below for all $k \in K$ sufficiently large. Likewise I_3 is the union of the two sets I_{3l}, made up of variables which are floating for all sufficiently large $k \in K$ but converge to their lower bounds, and I_{3u}, made up of variables which are floating for all sufficiently large $k \in K$ but converge to their upper bounds. With such definitions, we may reprove all of the results of sections 3 to 7, assumptions AS5 and AS6 being extended in the obvious way and Theorem 5.4 being strengthened to say that, for all $k \in K$ sufficiently large, I_{2l} and I_{2u} are precisely the variables which lie at their lower and upper bounds (respectively) at x^*.

We have not made any statement here about how the scaling matrices $S^{(k)}$ should be constructed, merely that they may be used. We consider that constraint scaling is essential for any realistic algorithm and beliieve that it is important that the scaling can be changed (albeit not too drastically) as the computation proceeds. We defer a discussion of the issues of how to choose such scalings until we have performed significant numerical testing of our algorithms. We also note that the results given here are unaltered if the convergence tolerance (3.1)/(3.5) is replaced by

$$\|D^{(k)} P(x^{(k)}, \nabla_x \Phi^{(k)})\| \le \omega^{(k)}. \tag{8.6}$$

for any sequence of positive diagonal matrices $\{D^{(k)}\}$ with uniformly bounded condition number. This is important as the method of Conn, Gould and Toint (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.

Finally, although the rules for how the convergence tolerances $\eta^{(k)}$ and $\omega^{(k)}$ 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 doing so here as the resulting conditions on the updates seemed rather complicated and are unlikely to provide more practical updates.

9 Acknowledgement

We would like to thank Annick Sartenaer for her careful reading of this paper.

10 References

D.P. Bertsekas, "Augmented Lagrangian and exact penalty methods", in *Nonlinear Optimization 1981* (M.J.D.. Powell, ed.), 223-234, Academic Press, London and New York, 1982a.

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

A.R. Conn, N.I.M. Gould, M. Lescrenier and Ph.L. Toint, "Performance of a multifrontal scheme for partially separable optimization", Research Report CS-88-04, Dept. of Computer Science, University of Waterloo, Waterloo, Ontario (1988).

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, 1988a.

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, 1988b.

R. S. Dembo, S. C. Eisenstat and T. Steihaug, "Inexact Newton methods", *SIAM Journal on Numerical Analysis* **19** 400-408, 1982.

R.S. Dembo and U. Tulowitzki, "Local convergence analysis for successive inexact quadratic programming methods", School of Organization and Management Working paper series B no. 78, Yale University, 1984.

L.C.W. Dixon, P. Dolan and R. Price, "Finite element optimization: the use of structured automatic differentiation", in *Simulation and Optimization of Large Systems* (A. Osiadacz, ed.), 117-141, Oxford University Press, 1988.

A. Drud, "CONOPT: a GRG code for large sparse dynamic nonlinear optimization problems", *Mathematical Programming*, **31** 153-191, 1985.

R. Fletcher, *Practical Methods of Optimization, Volume 2* , John Wiley, London and New York, 1981.

P.E. Gill, M.A. Saunders, W. Murray and M.H. Wright, "Some theoretical properties of an augmented Lagrangian merit function", report SOL 86-6, Dept. of Operations Research, Stanford University, 1986.

N. I. M. Gould, "On the convergence of a sequential penalty function method for constrained minimization", report CSS 207, Harwell Laboratory, England, 1987, (to appear *SIAM Journal on Numerical Analysis*, 1988).

N. I. M. Gould, "On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem", *Mathematical Programming*, **32** 90-99, 1985.

A. Griewank and Ph.L. Toint, "Partitioned variable metric updates for large structured optimization problems", *Numerische Mathematik*, **39** 119-137, 1982.

W.A. Gruver and E. Sachs, "Algorithmic methods in optimal control", Research notes in mathematics **47** Pitman, Boston, 1980.

M.R. Hestenes, "Multiplier and gradient methods", *Journal of Optimization Theory and Applications*, **4** 303-320, 1969.

L.S. Lasdon, "Reduced gradient methods", in *Nonlinear Optimization 1981* (M.J.D.. Powell, ed.), 235-242, Academic Press, London and New York, 1982.

J.J. Moré, "Recent developments in algorithms and software for trust region methods", in *Mathematical Programming: The State of the Art* (A. Bachem, M. Grötschel and B. Korte, eds.), 258-287, Springer Verlag, Berlin, 1983.

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.

B.A. Murtagh and M.A. Saunders, "Minos/Augmented user's manual", Report SOL 80-14, Dept. Of Operations Research, Stanford University, 1980.

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

M.J.D. Powell, "A method for nonlinear constraints in minimization problems" in *Optimization* (R. Fletcher, ed.), Academic Press, London and New York, 1969.

K. Schittkowski, "The nonlinear programming method of Wilson, Han and Powell with an augmented Lagrangian type line search function", *Numerische Mathematik*, **38** 83-114, 1981.

T. Steihaug, "The conjugate gradient method and trust regions in large scale optimization", *SIAM Journal on Numerical Analysis* **20** 626-637, 1983.

R.A.Tapia, "Diagonalized multiplier methods and quasi-Newton methods for constrained optimization", *Journal of Optimization Theory and Applications*, **22** 135-194, 1977.

Ph.L. Toint, "Towards an efficient sparsity exploiting Newton method for minimization", in *Sparse matrices and their uses* (I. S. Duff, ed.), Academic Press, London and New York, 1981.

Ph. L. Toint, "Nonlinear optimization in a large number of variables", in *Simulation and Optimization of Large Systems* (A. Osiadacz, ed.), 83-116, Oxford University Press, 1988.