Local Convergence Properties of two Augmented Lagrangian Algorithms for Optimization with a Combination of General Equality and Linear Constraints

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

CERFACS Report TR/PA/93/27

July 27, 1993

Abstract. We consider the local convergence properties of the class of augmented Lagrangian methods for solving nonlinear programming problems whose global convergence properties are analyzed by Conn *et al.* (1993a). In these methods, linear constraints are treated separately from more general constraints. These latter constraints are combined with the objective function in an augmented Lagrangian while the subproblem then consists of (approximately) minimizing this augmented Lagrangian subject to the linear constraints. The stopping rule that we consider for the inner iteration covers practical tests used in several existing packages for linearly constrained optimization. Our algorithmic class allows several distinct penalty parameters to be associated with different subsets of general equality constraints. In this paper, we analyze the local convergence of the sequence of iterates generated by this technique and prove fast linear convergence and boundedness of the potentially troublesome penalty parameters.

[1] IBM T.J. Watson Research Center, P.O.Box 218, Yorktown Heights, NY 10598, USA
Email : arconn@watson.ibm.com
[2] CERFACS, 42 Avenue Gustave Coriolis, 31057 Toulouse Cedex, France, EC
Email : gould@cerfacs.fr or nimg@letterbox.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 : as@math.fundp.ac.be or pht@math.fundp.ac.be
Current reports available by anonymous ftp from the directory
"reports" on thales.math.fundp.ac.be (internet 138.48.4.14)

Keywords: Constrained optimization, augmented Lagrangian methods, linear constraints, convergence theory.

Mathematics Subject Classifications : 65K05, 90C30

Local Convergence Properties of two Augmented Lagrangian Algorithms for Optimization with a Combination of General Equality and Linear Constraints

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

July 27, 1993

Abstract

We consider the local convergence properties of the class of augmented Lagrangian methods for solving nonlinear programming problems whose global convergence properties are analyzed by Conn *et al.* (1993a). In these methods, linear constraints are treated separately from more general constraints. These latter constraints are combined with the objective function in an augmented Lagrangian while the subproblem then consists of (approximately) minimizing this augmented Lagrangian subject to the linear constraints. The stopping rule that we consider for the inner iteration covers practical tests used in several existing packages for linearly constrained optimization. Our algorithmic class allows several distinct penalty parameters to be associated with different subsets of general equality constraints. In this paper, we analyze the local convergence of the sequence of iterates generated by this technique and prove fast linear convergence and boundedness of the potentially troublesome penalty parameters.

1 Introduction

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

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

where x is required to satisfy the general equality constraints

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

and the linear inequality constraints

$$Ax - b \ge 0. \tag{1.3}$$

Here f and c_i map $\Re^n$ into $\Re$, A is a p-by-n matrix and $b \in \Re^p$. We assume that $A \ne 0$ whenever $p > 0$.

A classical technique for solving problem (1.1)–(1.3) is to minimize a suitable sequence of *augmented Lagrangian functions*. If we only consider the problem (1.1)–(1.2), these functions are defined by

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

[0]This research 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.

where the components λ_i of the vector λ are known as *Lagrange multiplier estimates* and μ is known as the *penalty parameter* (see, for instance, Hestenes (1969), Powell (1969) and Bertsekas (1982)). The question of how to deal with the additional linear inequality constraints (1.3) then arises. This has been considered by Conn *et al.* (1993a): they suggest keeping these constraints explicitly outside the augmented Lagrangian formulation and handling them directly at the level of the augmented Lagrangian minimization. That is, a sequence of optimization problems, in which (1.4) is approximately minimized within the region defined by the linear constraints, is attempted. This proposal has the advantage that it can fully exploit a number of effective techniques specifically designed to handle linear constraints directly (see Arioli *et al.* (1993), Forsgren and Murray (1993) or Lustig *et al.* (1989), for instance). Such an approach is especially worthwhile for large-scale problems. This strategy has been implemented and successfully applied within the LANCELOT package for large-scale nonlinear optimization (see Conn *et al.* (1992)) in the more restrictive case where the linear constraints are simple bounds. Preliminary experiments with general linear constraints have also shown encouraging results and we intend to report on more exhaustive tests once we have implemented a robust general algorithm for linearly constrained optimization. The purpose of the present paper is therefore to examine the local convergence properties of this promising class of algorithms, completing the global convergence analysis developed in the companion paper by Conn *et al.* (1993a) and closing the gap between the theory for general linear constraints and that developed for simple bounds by Conn *et al.* (1991).

Furthermore, it is often worthwhile from the practical point of view to associate different penalty parameters to subsets of the general constraints (1.2) to reflect different degrees of nonlinearity. In this case, the formulation of the augmented Lagrangian (1.4) can be refined: we partition the set of constraints (1.2) into q disjoint *subsets* $\{\mathcal{Q}_j\}_{j=1}^q$, and redefine the augmented Lagrangian as

$$\Phi(x,\lambda,\mu) = f(x) + \sum_{j=1}^{q} \sum_{i\in\mathcal{Q}_j} \left[\lambda_i c_i(x) + \frac{1}{2\mu_j} c_i(x)^2 \right], \tag{1.5}$$

where μ is now a q-dimensional vector, whose j-th component is $\mu_j > 0$, the penalty parameter associated with subset $\mathcal{Q}_j$. Notice that this reformulation is covered by the global convergence theory developed in Conn *et al.* (1993a). Because of its potential usefulness, this refined formulation will be adopted in the present paper.

Since the theory presented below handles the linear inequality constraints in a purely geometric way, the same theory applies without modifications if linear equality constraints are also imposed and all the iterates are assumed to stay feasible with respect to these new constraints. It is indeed enough to apply the theory in the affine subspace corresponding to this feasible set. As a consequence, linear constraints need not be included in the augmented Lagrangian and have no impact on the structure of its Hessian matrix, a very desirable property.

The paper is organized as follows. In Section 2, we introduce our basic assumptions on the problem and the necessary terminology. Section 3 presents the proposed algorithms and the definition of suitable stopping criteria for the subproblem. The local convergence analysis is developed in Section 4 while second order conditions are investigated in Section 5. Finally, some conclusions and perspectives are outlined in Section 6.

2 The problem and related terminology

We consider the problem stated in (1.1)–(1.3) and make the following assumptions.

AS1: The region $\mathcal{B} = \{x \mid Ax - b \geq 0\}$ is nonempty.

AS2: The functions $f(x)$ and $c_i(x)$ are twice continuously differentiable for all $x \in \mathcal{B}$.

Assumption AS1 is clearly necessary for the problem to make sense. We note that it does not prevent $\mathcal{B}$ from being unbounded.

We now introduce the notation that will be used throughout the paper. It is identical to that introduced by Conn *et al.* (1993a), but is restated here for completeness.

Let $g(x)$ denote the gradient $\nabla_x f(x)$ of $f(x)$ and $H(x)$ denote its Hessian matrix $\nabla_{xx} f(x)$. We also define $J(x)$ to be the m-by-n Jacobian of $c(x)$, where

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

Hence

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

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, $\nabla_x \ell(x, \lambda)$, and Hessian matrix, $\nabla_{xx} \ell(x, \lambda)$, of the Lagrangian function

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

We note that $\ell(x, \lambda)$ is the Lagrangian solely with respect to the c_i constraints. If we define *first-order Lagrange multiplier estimates* componentwise as

$$\bar{\lambda}(x, \lambda_{[\mathcal{Q}_j]}, \mu_j)_{[\mathcal{Q}_j]} = \lambda_{[\mathcal{Q}_j]} + c(x)_{[\mathcal{Q}_j]}/\mu_j \quad (j = 1, \ldots, q), \tag{2.4}$$

where $w_{[\mathcal{S}]}$ denotes the $|\mathcal{S}|$-dimensional subvector of w whose entries are indexed by the set $\mathcal{S}$, we shall use the identity

$$\begin{aligned} \nabla_x \Phi(x, \lambda, \mu) = & \ \nabla_x f(x) + \textstyle\sum_{j=1}^{q} \sum_{i \in \mathcal{Q}_j} \left[\lambda_i \nabla_x c_i(x) + \frac{1}{\mu_j} c_i(x) \nabla_x c_i(x) \right] \\ = & \ g^\ell(x, \bar{\lambda}(x, \lambda, \mu)). \end{aligned} \tag{2.5}$$

Now suppose that $\{x_k \in \mathcal{B}\}$, $\{\lambda_k\}$ and $\{\mu_k\}$ are infinite sequences of n-vectors, m-vectors and positive q-vectors, respectively. For any function F, we shall use the notation that F_k denotes F evaluated with arguments x_k, λ_k and/or μ_k as appropriate. So, for instance, using the identity (2.5), we have that

$$\nabla_x \Phi_k = \nabla_x \Phi(x_k, \lambda_k, \mu_k) = g^\ell(x_k, \bar{\lambda}_k), \tag{2.6}$$

where we have written (2.4) in the compact form

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

We denote the vector w at iteration k by w_k and its i-th component by $w_{k,i}$. We also use $w_{k,[\mathcal{S}]}$ to denote the $|\mathcal{S}|$-dimensional subvector of w_k whose entries are indexed by $\mathcal{S}$.

Now let $\{x_k\}, k \in \mathcal{K}$, be a convergent subsequence with limit point x_*. Then we denote by A_* the matrix whose rows are those of A corresponding to active constraints at x_*, that is constraints that are satisfied as equalities at x_*. Furthermore, we choose Z_* to be a matrix whose columns form an orthonormal basis of the nullspace of A_*, that is

$$A_* Z_* = 0. \tag{2.8}$$

We define the *least-squares Lagrange multiplier estimates* (corresponding to A_*)

$$\lambda(x) \stackrel{\text{def}}{=} -((J(x) Z_*)^+)^T Z_*^T g(x) \tag{2.9}$$

at all points where the right generalized inverse

$$(J(x)Z_*)^+ \stackrel{\text{def}}{=} Z_*^T J(x)^T (J(x)Z_*Z_*^T J(x)^T)^{-1} \tag{2.10}$$

of $J(x)Z_*$ is well defined. We note that $\lambda(x)$ is differentiable and its derivative is given in Lemma 2.1 of Conn *et al.* (1993a).

We stress that, as stated, the Lagrange multiplier estimate (2.9) is not directly calculable as it requires a priori knowledge of x_*. It is merely introduced as an analytical device.

Finally, the symbol $\|\cdot\|$ will denote the ℓ_2-norm or the induced operator norm. We are now in position to describe more precisely the algorithms that we propose to use.

3 Statement of the algorithms

We consider the two algorithmic models we wish to use in order to solve the problem (1.1)–(1.3). Both models proceed at iteration k by approximately solving the subproblem

$$\min_{x\in\mathcal{B}} \Phi(x, \lambda_k, \mu_k), \tag{3.1}$$

where the values of the Lagrange multipliers λ_k and penalty parameters μ_k are fixed for the subproblem. Subsequently we update the Lagrange multipliers and/or decrease the penalty parameters, depending on how much the constraint violation for (1.2) has been reduced within each subset of the constraints. The motivation is simply to ensure global convergence by driving, in the worst case, the penalty parameters to zero, in which case the algorithms essentially reduce to the quadratic penalty function method (see, for example, Gould (1989)). The tests on the size of the general constraint violation are designed to allow the multiplier updates to take over in the neighbourhood of a stationary point.

The approximate minimization for problem (3.1) is performed in an *inner iteration* which is stopped as soon as its current iterate is "sufficiently critical". We propose to base this decision on the identification of the linear constraints that are "dominant" at x (even though they might not be active) and on a measure of criticality for the part of the problem where those constraints are irrelevant. Given $\omega \geq 0$, a criticality tolerance for the subproblem, we define, for a vector $x \in \mathcal{B}$, the set of *dominant constraints at* x as the constraints whose index is in the set

$$D(x,\omega) \stackrel{\text{def}}{=} \{i \in \{1,\ldots,p\} \mid a_i^T x - b_i \leq \kappa_1 \omega\}, \tag{3.2}$$

for some $\kappa_1 > 0$. Here $a_i^T \in \Re^n$ is the i-th row of the matrix A and b_i is the corresponding component of the right-hand-side vector b. Denoting by $A_{D(x,\omega)}$ the submatrix of A consisting of the row(s) whose index is in $D(x,\omega)$, we also define

$$N(x,\omega) = \{A_{D(x,\omega)}^T \xi \mid \xi \in \Re^{|D(x,\omega)|} \text{ and } \xi_i \leq 0, \quad (i = 1,\ldots,|D(x,\omega)|)\}, \tag{3.3}$$

the cone spanned by the outwards normals of the dominant constraints. The associated polar cone is then

$$T(x,\omega) = N(x,\omega)^0 = \text{cl}\{\nu d \mid \nu \geq 0 \text{ and } d^T v \leq 0 \text{ for all } v \in N(x,\omega)\}, \tag{3.4}$$

where $\text{cl}(V)$ denotes the closure of the set V. The cone $T(x,\omega)$ is the tangent cone with respect to the dominant constraints at x for the tolerance ω. Note that $D(x,\omega)$ might be empty, in which case $A_{D(x,\omega)}$ is assumed to be zero, $N(x,\omega)$ reduces to the origin and $T(x,\omega)$ is the full space.

We then formulate our "sufficient criticality" criterion for the subproblem as follows: we require that

$$\|P_{T(x_k,\omega_k)}(-\nabla_x \Phi_k)\| \leq \omega_k, \tag{3.5}$$

where $P_V(\cdot)$ is the projection onto the convex set V and ω_k is a suitable tolerance at iteration k.

It is important to note that this stopping rule covers a number of more specific choices, including the rule used in many existing software for linearly constrained optimization (such as LSNNO by Toint and Tuyttens (1992) or VE09, VE14 and VE19 from the Harwell Subroutine Library). The reader is referred to Section 5 of Conn *et al.* (1993a) for further details.

We are now in position to describe our algorithmic models more precisely.

Algorithm 3.1

Step 0 [Initialization]. A partition of the set $\{1,\ldots,m\}$ into q disjoint subsets $\{\mathcal{Q}_j\}_{j=1}^q$ is given, as well as initial vectors of Lagrange multiplier estimates λ_0 and positive penalty parameters μ_0 such that

$$\mu_{0,j} < 1, \quad (j = 1,\ldots,q). \tag{3.6}$$

The strictly positive constants $\omega_* \ll 1$, $\eta_* \ll 1$, $\tau < 1$, $\alpha_\eta < 1$, and $\beta_\eta < 1$ are specified. Set $\alpha_0 = \max_{j=1,\ldots,q} \mu_{0,j}$, $\omega_0 = \alpha_0$, $\eta_0 = \alpha_0^{\alpha_\eta}$ and $k = 0$.

Step 1 [Inner iteration]. Find $x_k \in \mathcal{B}$ that approximately solves (3.1), i.e. such that (3.5) holds.

Step 2 [Test for convergence]. If $\|P_{T(x_k,\omega_k)}(-\nabla_x \Phi_k)\| \leq \omega_*$ and $\|c(x_k)\| \leq \eta_*$, stop.

Step 3 [Disaggregated updates]. For $j = 1,\ldots,q$, execute Step 3a if

$$\|c(x_k)_{[\mathcal{Q}_j]}\| \leq \eta_k, \tag{3.7}$$

or Step 3b otherwise.

Step 3a [Update Lagrange multiplier estimates]. Set

$$\begin{array}{rcl} \lambda_{k+1,[\mathcal{Q}_j]} & = & \bar{\lambda}(x_k, \lambda_{k,[\mathcal{Q}_j]}, \mu_{k,j})_{[\mathcal{Q}_j]}, \\ \mu_{k+1,j} & = & \mu_{k,j}. \end{array} \tag{3.8}$$

Step 3b [Reduce the penalty parameter]. Set

$$\begin{array}{rcl} \lambda_{k+1,[\mathcal{Q}_j]} & = & \lambda_{k,[\mathcal{Q}_j]}, \\ \mu_{k+1,j} & = & \tau_{k,j}\mu_{k,j}, \end{array} \tag{3.9}$$

where

$$\tau_{k,j} = \begin{cases} \tau & \text{if } \mu_{k,j} = \alpha_k, \\ \min(\tau, \alpha_k) & \text{otherwise.} \end{cases} \tag{3.10}$$

Step 4 [Aggregated updates]. Define

$$\alpha_{k+1} = \max_{j,\ldots,q} \mu_{k+1,j}. \tag{3.11}$$

If

$$\alpha_{k+1} < \alpha_k, \tag{3.12}$$

then set

$$\begin{aligned} \omega_{k+1} &= \alpha_{k+1}, \\ \eta_{k+1} &= \alpha_{k+1}^{\alpha_\eta}, \end{aligned} \tag{3.13}$$

otherwise set

$$\begin{aligned} \omega_{k+1} &= \omega_k \alpha_{k+1}, \\ \eta_{k+1} &= \eta_k \alpha_{k+1}^{\beta_\eta}. \end{aligned} \tag{3.14}$$

Increment k by one and go to Step 1.

Algorithm 3.2

This algorithm is identical to Algorithm 3.1, except that Step 3 is replaced by the following, where γ is a constant in $(0,1)$.

Step 3 [Disaggregated updates]. Compute a new vector of Lagrange multiplier estimates $\hat{\lambda}_{k+1}$. For $j = 1, \ldots, q$, execute Step 3a if

$$\|c(x_k)_{[\mathcal{Q}_j]}\| \leq \eta_k, \tag{3.15}$$

or Step 3b otherwise.

Step 3a [Update Lagrange multiplier estimates]. Set

$$\begin{aligned} \lambda_{k+1,[\mathcal{Q}_j]} &= \begin{cases} \hat{\lambda}_{k+1,[\mathcal{Q}_j]} & \text{if } \|\hat{\lambda}_{k+1,[\mathcal{Q}_j]}\| \leq \mu_{k+1,j}^{-\gamma}, \\ \lambda_{k,[\mathcal{Q}_j]} & \text{otherwise}, \end{cases} \\ \mu_{k+1,j} &= \mu_{k,j}. \end{aligned} \tag{3.16}$$

Step 3b [Reduce the penalty parameter]. Set

$$\begin{aligned} \lambda_{k+1,[\mathcal{Q}_j]} &= \begin{cases} \hat{\lambda}_{k+1,[\mathcal{Q}_j]} & \text{if } \|\hat{\lambda}_{k+1,[\mathcal{Q}_j]}\| \leq \mu_{k+1,j}^{-\gamma}, \\ \lambda_{k,[\mathcal{Q}_j]} & \text{otherwise}, \end{cases} \\ \mu_{k+1,j} &= \tau_{k,j}\mu_{k,j}, \end{aligned} \tag{3.17}$$

where $\tau_{k,j}$ is defined by (3.10).

Except for the inner iteration stopping rule (3.5) and the more precise way in which ω_k and the $\tau_{k,j}$ are reduced, Algorithms 3.1 and 3.2 are identical to those analyzed in Section 6 of Conn *et al.* (1993a). The algorithms differ, here, as there, by their use of multiplier updates. Algorithm 3.1 is specifically designed for the first-order estimate (2.4), a formula with potential advantages for large-scale computations. Algorithm 3.2 allows a more flexible choice of the multipliers, but requires that some control is enforced to prevent their growth at an unacceptably fast rate. It covers, among others, the choice of the least-squares estimates $\lambda(x)$ as defined in (2.9).

The restriction (3.6) is imposed in order to simplify the exposition. In a more practical setting, it may be ignored provided the definition of α_0 and (3.11) are replaced by

$$\alpha_0 = \min\left(\gamma_s, \max_{j=1,\ldots,q} \mu_{0,j}\right) \text{ and } \alpha_{k+1} = \min\left(\gamma_s, \max_{j=1,\ldots,q} \mu_{k+1,j}\right), \tag{3.18}$$

respectively, for some constant $\gamma_s \in (0,1)$, and that (3.12) is replaced by

$$\max_{j=1,\ldots,q} \mu_{k+1,j} < \max_{j=1,\ldots,q} \mu_{k,j}. \tag{3.19}$$

Algorithms 3.1 and 3.2 may be extended in other ways. For instance, one may replace the definition of ω_0, the first equation in (3.13) and the first equation of (3.14) by

$$\omega_0 = \omega_s \alpha_0^{\alpha_\omega}, \quad \omega_{k+1} = \omega_s \alpha_{k+1}^{\alpha_\omega} \text{ and } \omega_{k+1} = \omega_s \alpha_{k+1}^{\beta_\omega}, \tag{3.20}$$

for some $\omega_s > 0$, $\alpha_\omega > \alpha_\eta$ and $\beta_\omega > \beta_\eta$. The definition of η_0 and the second equation in (3.13) may then be replaced by

$$\eta_0 = \eta_s \alpha_0^{\beta_\eta} \text{ and } \eta_{k+1} = \eta_s \alpha_{k+1}^{\alpha_\eta}, \tag{3.21}$$

for some $\eta_s > 0$. In the same spirit, it is also possible to replace (3.10) by

$$\tau_{k,j} = \begin{cases} \tau & \text{if } \mu_{k,j} = \max_{j=1,\ldots,q} \mu_{k,j}, \\ \min(\tau, \alpha_k^{\beta_\tau}) & \text{otherwise.} \end{cases} \tag{3.22}$$

for some $\beta_\tau > \beta_\eta$. Finally, the acceptance test for λ_{k+1} in (3.16) and (3.17) may be replaced by

$$\|\hat{\lambda}_{k+1,[\mathcal{Q}_j]}\| \leq \nu \mu_{k+1,j}^{-\gamma} \tag{3.23}$$

for some $\nu > 0$. None of these extensions alter the results of the convergence theory developed below.

The proposed algorithms use a number of parameters. The values used in the LANCELOT package in a similar context are $\alpha_\eta = \tau = \gamma_s = 0.1$, and $\beta_\eta = 0.9$ (relation (3.21) is also used with $\eta_s = 0.12589$, ensuring that $\eta_0 = 0.01$). The values $\gamma = 0.9$, $\nu = \omega_s = \beta_\tau = \alpha_\omega = \beta_\omega = 1$ and $\mu_{0,j} = 0.1$ $(j = 1, \ldots, q)$ also seem suitable. The parameters ω_* and η_* specify the final accuracy requested by the user.

We now re-examine the inner iteration stopping rule (3.5). This criterion was proposed by Conn *et al.* (1993a) as a computationally attractive alternative to the rule

$$\sigma_k = \sigma(x_k, \lambda_k, \mu_k) \leq \omega_k, \tag{3.24}$$

where the quantity

$$\sigma(x, \lambda, \mu) \stackrel{\text{def}}{=} \begin{array}{ll} |\min_{d \in \Re^n} & \nabla_x \Phi(x, \lambda, \mu)^T d|, \\ \text{subject to} & A(x+d) - b \geq 0, \\ & \|d\| \leq 1 \end{array} \tag{3.25}$$

represents the magnitude of the maximum decrease in the linearized augmented Lagrangian achievable on the intersection of $\mathcal{B}$ with a ball of radius one centered at the current point. This criterion was first considered by Conn *et al.* (1993b) and subsequently by Conn *et al.* (1993a), in both cases to prove global convergence of algorithms for convex constraints. Because we are concerned with stronger asymptotic convergence properties, it is not surprising that (3.24) needs some strengthening. A variant of (3.24), also discussed in the latter reference, is indeed appropriate for our present purposes. Thus we may decide to stop the inner iteration if

$$\sigma(x_k, \lambda_k, \mu_k) \leq {\omega_k}^2. \tag{3.26}$$

This criterion is clearly stronger than (3.24) since ω_k is driven to zero by our algorithms. Because Theorem 5.7 of Conn *et al.* (1993a) shows that the rule (3.26) implies (3.5) whenever the algorithm is convergent, it will be sufficient to develop our theory for this last choice to cover the rule (3.26) as well.

Finally, the purpose of the update (3.10) is to put more emphasis on the feasibility of the constraints whose violation is proportionally higher, in order to achieve a "balance" amongst all constraint violations. This balance then allows the true asymptotic regime of the algorithm to be reached. The advantage of (3.10) is that this balancing effect is obtained gradually, and not enforced at every major iteration, as is the case in Powell (1969).

It is important to note that Algorithms 3.1 and 3.2 were proved by Conn *et al.* (1993a) to be globally convergent with either tests (3.5) or (3.24) under the following additional assumptions.

AS3: The iterates $\{x_k\}$ considered lie within a closed, bounded domain.

AS4: The matrix $J(x_*)Z_*$ has column rank no smaller than m at any limit point, x_*, of the sequence $\{x_k\}$ considered in this paper.

Suppose that the gradients of the nonlinear constraints projected onto the nullspace of A are assumed to be linearly independent at every limit point of the sequence of iterates. The assumption AS4 guarantees that the dimension of this nullspace is large enough to provide the number of degrees of freedom that are necessary to satisfy the nonlinear constraints.

We now recall the global convergence result.

Theorem 3.1 [Conn *et al.* (1993a), Theorem 6.2] *Assume that AS1 and AS2 hold. Let $\{x_k\}$ be generated by Algorithm 3.1 or by Algorithm 3.2, where, at each iteration, the test (3.5) can be replaced by either (3.24) or (3.26). Assume that AS3 holds for this sequence. Let $\mathcal{K}$ be the set of indices of an infinite subsequence of the x_k whose limit is x_*, for which AS4 is satisfied. Let $\lambda_* = \lambda(x_*)$. Then the following conclusions hold.*

(i) *x_* is a Kuhn-Tucker point (first-order stationary point) for the problem (1.1)–(1.3), λ_* is the corresponding vector of Lagrange multipliers, and the sequences $\{\bar{\lambda}(x_k, \lambda_{k,[\mathcal{Q}_j]}, \mu_{k,j})_{[\mathcal{Q}_j]}\}$ and $\{\lambda(x_k)_{[\mathcal{Q}_j]}\}$ converge to $\lambda_{*,[\mathcal{Q}_j]}$ for $k \in \mathcal{K}$ and for all $j = 1, \ldots, q$;*

(ii) *There are positive constants κ_2, κ_3, and an integer k_1 such that*

$$\|(\bar{\lambda}(x_k, \lambda_k, \mu_k) - \lambda_*)_{[\mathcal{Q}_j]}\| \leq \kappa_2 \omega_k + \kappa_3 \|x_k - x_*\|, \tag{3.27}$$

$$\|(\lambda(x_k) - \lambda_*)_{[\mathcal{Q}_j]}\| \leq \kappa_3 \|x_k - x_*\|, \tag{3.28}$$

and

$$\|c(x_k)_{[\mathcal{Q}_j]}\| \leq \kappa_2 \omega_k \mu_{k,j} + \mu_{k,j} \|(\lambda_k - \lambda_*)_{[\mathcal{Q}_j]}\| + \kappa_3 \mu_{k,j} \|x_k - x_*\|, \tag{3.29}$$

for all $j = 1, \ldots, q$ and all $k \geq k_1, (k \in \mathcal{K})$.

(iii) *The gradients $\nabla_x \Phi_k$ converge to $g^\ell(x_*, \lambda_*)$ for $k \in \mathcal{K}$.*

4 Asymptotic convergence analysis

We shall now analyze the local convergence of the algorithms of Section 3 in the case where the convergence tolerances ω_* and η_* are both zero.

We first need to introduce some additional notation. We define $D_k \stackrel{\text{def}}{=} D(x_k, \omega_k)$ the set of dominant constraints at x_k, and its complement, the set of *floating* constraints at the same point. We also denote by A_{D_k} the submatrix whose rows consist of the rows of A indexed by D_k. As in (3.3) and (3.4), we define

$$N_k \stackrel{\text{def}}{=} \{A_{D_k}^T \xi \mid \xi \in \Re^{|D_k|} \text{ and } \xi_i \leq 0 \quad (i \in D_k)\} \tag{4.1}$$

and $T_k = N_k^0$, the outward normal cone and tangent cone with respect to the dominant constraints at x_k. For future reference, we define Z_k to be a matrix whose columns form an orthonormal basis of $\mathcal{V}_k$, the nullspace of A_{D_k}, and Y_k to be a matrix whose columns form an orthonormal basis of $\mathcal{W}_k = \mathcal{V}_k^\perp$. As above, we have that T_k is the full space and N_k reduces to the origin when D_k is empty. We note that, in this case, $Z_k = P_{T_k} = I$, the identity operator, and $Y_k = P_{N_k} = 0$. We also note that $\mathcal{V}_k \subseteq T_k$, and hence that

$$\|Z_k^T \nabla_x \Phi_k\| = \|Z_k Z_k^T \nabla_x \Phi_k\| \leq \|P_{T_k}(-\nabla_x \Phi_k)\|, \tag{4.2}$$

since $Z_k Z_k^T$ is the orthogonal projection onto $\mathcal{V}_k$.

We next state a useful consequence of the partition of the linear inequality constraints into dominant and floating.

Lemma 4.1 *Assume that $\{x_k\}$ is a convergent sequence of points whose limit point is x_* and such that (3.5) holds for a sequence $\{\omega_k\}$ of positive numbers tending to zero. Then*

$$\|Z_k^T \nabla_x \Phi_k\| \leq \omega_k \;\; \textit{and} \;\; \|Y_k^T(x_k - x_*)\| \leq \kappa_4 \omega_k \tag{4.3}$$

for some $\kappa_4 \geq 0$ and all k sufficiently large.

Proof. As $T_k = T(x_k, \omega_k)$, we deduce the first part of (4.3) directly from (4.2) and (3.5). Because $\{x_k\}$ converges to x_* and since ω_k tends to zero, we have that, for all k sufficiently large, any constraint in D_k is also active at x_*. Hence the second inequality of (4.3) results from Lemma 5.1 in Conn *et al.* (1993a). □

The distinction between floating and dominant linear inequality constraints also has some implications in terms of the identification of those constraints that are active at a limit point of the sequence of iterates generated by either algorithm. Given such a point x_* we know from Theorem 3.1 that it is critical, i.e. that $-g^\ell(x_*, \lambda_*) \in N_* = N(x_*, 0)$ for the corresponding Lagrange multipliers λ_*. If we now consider a linear constraint with index $i \in \{1, \ldots, p\}$ that is active at x_*, we may define the normal cone $N_*^{[i]}$ to be the cone spanned by the outwards normals to all linear inequality constraints active at x_*, except the i-th one. We then say that the i-th linear inequality constraint is *strongly active* at x_* if $-g^\ell(x_*, \lambda_*) \notin N_*^{[i]}$. In other words, the i-th constraint is strongly active at a critical point if this point ceases to be critical when this constraint is ignored. Let us denote by $S(x_*)$ the set of strongly active constraints at x_*. All non-strongly active constraints at x_* are called *weakly active* at x_*. We next prove the interesting result that all strongly active constraints at a limit point x_* are dominant for k large enough.

Theorem 4.2 *Assume that AS1–AS3 hold. Let $\{x_k\}$, $k \in \mathcal{K}$, be a convergent subsequence of iterates produced by Algorithm 3.1 or 3.2, whose limit point is x_* with corresponding Lagrange multipliers λ_*. Assume furthermore that AS4 holds at x_*. Then*

$$S(x_*) \subseteq D_k \tag{4.4}$$

for all k sufficiently large.

Proof. Consider a linear inequality constraint $i \in S(x_*)$. Then, by definition of this latter set, we have that $-g^\ell(x_*, \lambda_*) \notin N_*^{[i]}$. Since $\nabla_x \Phi_k$ converges to $g^\ell(x_*, \lambda_*)$, as guaranteed by Theorem 3.1, and because $N_*^{[i]}$ is closed, we have that $-\nabla_x \Phi_k \notin N_*^{[i]}$ for $k \in \mathcal{K}$ large enough. Therefore, one obtains from the Moreau decomposition (see Moreau (1962)) of $-\nabla_x \Phi_k$ that

$$\|P_{T_*^{[i]}}(-\nabla_x \Phi_k)\| \geq \epsilon \tag{4.5}$$

for some $\epsilon > 0$ and for all sufficiently large $k \in \mathcal{K}$, where $T_*^{[i]} = \left[N_*^{[i]}\right]^0$. We have also from (3.5) that $\|P_{T_k}(-\nabla_x \Phi_k)\|$ is arbitrarily small, because ω_k tends to zero. Assume now that, for some arbitrarily large $k \in \mathcal{K}$, we have that $i \notin D_k$. This implies that $T_*^{[i]} \subseteq T_k$, and hence that (4.5) is impossible. We therefore deduce that i must belong to D_k, which proves the theorem. □

This result is important and is the direct generalization of Theorem 5.4 by [Conn *et al.*, 1991]. It can also be interpreted as a means of active constraint identification, as is clear from the following easy corollary.

Corollary 4.3 *Suppose that the conditions of Theorem 4.2 hold. Assume furthermore that all linear inequality constraints active at x_* have linearly independent normals and are non-degenerate, in the sense that*

$$-g^\ell(x_*, \lambda_*) \in \mathrm{ri}[N_*], \tag{4.6}$$

where ri$[V]$ *denotes the relative interior of a convex set V. Then D_k is identical to the set of active linear inequality constraints at x_* for all $k \in \mathcal{K}$ sufficiently large.*

Proof. The non-degeneracy assumption and the linear independence of the active constraints' normals imply that λ_* is unique and only has strictly negative components. Therefore each of the active linear inequality constraint at x_* is strongly active at x_*, and the desired conclusion follows from Theorem 4.2. □

We note here that the non-degeneracy assumption corresponds to strict complementarity slackness in our context (see, for instance, Dunn (1987) or Burke *et al.* (1990)).

We now make some additional assumptions before pursuing our local convergence analysis. We first intend to show that all penalty parameters are bounded away from zero.

AS5: The second derivatives of the functions $f(x)$ and $c_i(x)$ $(1 \le i \le m)$ are Lipschitz continuous at any limit point x_* of the sequence of iterates $\{x_k\}$.

AS6: Suppose that (x_*, λ_*) is a Kuhn-Tucker point for problem (1.1)–(1.3) and define $\mathcal{I}$ to be a subset of the linear inequality constraints active at x_* containing all strongly active constraints at x_* $(S(x_*) \subseteq \mathcal{I})$ plus an arbitrary subset of weakly active constraints at x_*. Then, if the columns of the matrix Z form an orthonormal basis of the subspace orthogonal to the normals of the constraints in $\mathcal{I}$, we assume that the matrix

$$\begin{pmatrix} Z^T H^\ell(x_*, \lambda_*) Z & Z^T J(x_*)^T \\ J(x_*) Z & 0 \end{pmatrix}$$

is nonsingular for all possible choices of the weakly active constraints in the set $\mathcal{I}$.

We note that AS6 implies AS4 and seems reasonable in that the definition of strongly and weakly active constraints may vary with small perturbations in the problem, for instance when $g^\ell(x_*, \lambda_*)$ lies in one of the extreme faces of the cone N_*. Our assumption might be seen as a safeguard against the possible effect of all such perturbations.

We now make the distinction between the subsets for which the penalty parameter converges to zero and those for which it stays bounded away from it. We define

$$\mathcal{Z} \stackrel{\text{def}}{=} \{j \in \{1, \ldots, q\} \mid \lim_{k \to \infty} \mu_{k,j} = 0\} \ \text{ and } \ \mathcal{P} \stackrel{\text{def}}{=} \{1, \ldots, q\} \setminus \mathcal{Z}. \tag{4.7}$$

We also denote

$$\mu_{k,\mathcal{Z}} \stackrel{\text{def}}{=} \max_{j \in \mathcal{Z}} \mu_{k,j} \tag{4.8}$$

and

$$\rho_k \stackrel{\text{def}}{=} \sum_{j\in\mathcal{Z}} \mu_{k,j} \|(\lambda_k - \lambda_*)_{[\mathcal{Q}_j]}\|. \tag{4.9}$$

We now prove an analog to Lemma 5.1 by Conn *et al.* (1991) which is suitable for our more general framework.

Lemma 4.4 *Assume that AS1–AS3 hold. Let* $\{x_k\}$, $k \in \mathcal{K}$, *be a convergent subsequence of iterates produced by Algorithm 3.1 or 3.2, whose limit point is* x_* *with corresponding Lagrange multipliers* λ_*. *Assume that AS5 and AS6 hold at* x_*. *Assume furthermore that* $\mathcal{Z} \neq \emptyset$.

(i) If $\mathcal{P} = \emptyset$, *there are positive constants* $\bar\alpha < 1$, κ_5, κ_6, κ_7, κ_8 *and an integer* k_2 *such that, if* $\alpha_{k_2} \le \bar\alpha$, *then*

$$\|x_k - x_*\| \le \kappa_5\omega_k + \kappa_6\alpha_k\|\lambda_k - \lambda_*\|, \tag{4.10}$$

$$\|\bar\lambda(x_k,\lambda_k,\mu_k) - \lambda_*\| \le \kappa_7\omega_k + \kappa_8\alpha_k\|\lambda_k - \lambda_*\|, \tag{4.11}$$

and

$$\|c(x_k)\| \le q\kappa_7\omega_k\alpha_k + q\alpha_k(1+\kappa_8\alpha_k)\|\lambda_k - \lambda_*\|, \tag{4.12}$$

for all $k \ge k_2$, $k \in \mathcal{K}$.

(ii) If, on the other hand, $\mathcal{P} \neq \emptyset$, *there are positive constants* $\bar\alpha < 1$, κ_5, κ_6, κ_7, κ_8 *and an integer* k_2 *such that, if* $\mu_{k_2,\mathcal{Z}} \le \bar\alpha$, *then*

$$\|(\bar\lambda(x_k,\lambda_k,\mu_k) - \lambda_*)_{[\mathcal{Q}_j]}\| \le \kappa_7\eta_k + \kappa_8\rho_k, \tag{4.13}$$

and

$$\|c(x_k)_{[\mathcal{Q}_j]}\| \le \kappa_7\eta_k\mu_{k,\mathcal{Z}} + (1+\kappa_8\mu_{k,\mathcal{Z}})\rho_k, \tag{4.14}$$

for all $k \ge k_2$, $k \in \mathcal{K}$, *and all* $j \in \mathcal{Z}$.

Proof. We will denote the gradient and Hessian of the Lagrangian function, taken with respect to x, at the limit point (x_*, λ_*) by g_*^ℓ and H_*^ℓ, respectively. Similarly, J_* will denote $J(x_*)$. We also define $\delta_k = x_k - x_*$. We observe that the assumptions of the lemma guarantee that Theorem 3.1 can be used.

We first note that there is only a finite number of possible D_k, and we may thus consider subsequences of $\mathcal{K}$ such that D_k is constant in each subsequence. We also note that each $k \in \mathcal{K}$ belongs to a unique such subsequence. In order to prove our result, it is thus sufficient to consider an arbitrary infinite subsequence $\bar{\mathcal{K}}$ such that, for $k \in \bar{\mathcal{K}}$, D_k is independent of k. This "constant" index set will be denoted by D. As a consequence, the cones N_k and T_k, the subspaces $\mathcal{V}_k$ and $\mathcal{W}_k$ and the orthogonal matrices Z_k and Y_k are also independent of k; they are denoted by N, T, $\mathcal{V}$, $\mathcal{W}$, Z and Y, respectively.

Using (2.6) and Taylor's expansion around x_*, we obtain that

$$\begin{aligned}\nabla_x\Phi_k &= g_k + J_k^T\bar\lambda_k \\ &= g(x_*) + H(x_*)\delta_k + J_*^T\bar\lambda_k + \textstyle\sum_{i=1}^m \bar\lambda_{k,i}H_i(x_*)\delta_k + r_1(x_k,x_*,\bar\lambda_k) \\ &= g_*^\ell + H_*^\ell\delta_k + J_*^T(\bar\lambda_k - \lambda_*) + r_1(x_k,x_*,\bar\lambda_k) + r_2(x_k,x_*,\bar\lambda_k,\lambda_*),\end{aligned} \tag{4.15}$$

where

$$r_1(x_k,x_*,\bar\lambda_k) \stackrel{\text{def}}{=} \int_0^1 [H^\ell(x_*+s\delta_k,\bar\lambda_k) - H^\ell(x_*,\bar\lambda_k)]\delta_k\,ds \tag{4.16}$$

and

$$r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*) \stackrel{\text{def}}{=} \sum_{i=1}^{m} (\bar{\lambda}_{k,i} - \lambda_{*,i}) H_i(x_*) \delta_k. \tag{4.17}$$

The boundedness and Lipschitz continuity of the Hessian matrices of f and c_i in a neighbourhood of x_*, together with the convergence of $\bar{\lambda}_k$ to λ_* then imply that

$$\|r_1(x_k, x_*, \bar{\lambda}_k)\| \leq \kappa_9 \|\delta_k\|^2, \tag{4.18}$$

and

$$\|r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*)\| \leq \kappa_{10} \|\delta_k\| \, \|\bar{\lambda}_k - \lambda_*\| \tag{4.19}$$

for some positive constants κ_9 and κ_{10}. Moreover, using Taylor's expansion again, along with the fact that Theorem 3.1 ensures the equality $c(x_*) = 0$, we obtain that

$$c(x_k) = J_* \delta_k + r_3(x_k, x_*), \tag{4.20}$$

where

$$[r_3(x_k, x_*)]_i = \int_0^1 s \int_0^1 \delta_k^T H_i(x_* + ts\delta_k) \delta_k \, dt \, ds \tag{4.21}$$

(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

$$\|r_3(x_k, x_*)\| \leq \kappa_{11} \|\delta_k\|^2 \tag{4.22}$$

for some positive constant κ_{11}. Combining (4.15) and (4.20), we obtain

$$\begin{pmatrix} H_*^\ell & J_*^T \\ J_* & 0 \end{pmatrix} \begin{pmatrix} \delta_k \\ \bar{\lambda}_k - \lambda_* \end{pmatrix} = \begin{pmatrix} \nabla_x \Phi_k - g_*^\ell \\ c(x_k) \end{pmatrix} - \begin{pmatrix} r_1 + r_2 \\ r_3 \end{pmatrix}, \tag{4.23}$$

where we have suppressed the arguments of the residuals r_1, r_2 and r_3 for brevity. Using the orthogonal decomposition of $\Re^n$ into $\mathcal{V} \oplus \mathcal{W}$, we may rewrite this last equation, premultiplied by $(Z^T \; Y^T \; I)$, as

$$\begin{pmatrix} Z^T H_*^\ell Z & Z^T H_*^\ell Y & Z^T J_*^T \\ Y^T H_*^\ell Z & Y^T H_*^\ell Y & Y^T J_*^T \\ J_* Z & J_* Y & 0 \end{pmatrix} \begin{pmatrix} Z^T \delta_k \\ Y^T \delta_k \\ \bar{\lambda}_k - \lambda_* \end{pmatrix} = \begin{pmatrix} Z^T(\nabla_x \Phi_k - g_*^\ell) \\ Y^T(\nabla_x \Phi_k - g_*^\ell) \\ c(x_k) \end{pmatrix} - \begin{pmatrix} Z^T r_4 \\ Y^T r_4 \\ r_3 \end{pmatrix}, \tag{4.24}$$

where $r_4 \stackrel{\text{def}}{=} r_1 + r_2$. We now observe that (3.5), the inclusion $\mathcal{V} \subseteq T$ and the fact that ω_k tends to zero imply that

$$Z^T g_*^\ell = 0. \tag{4.25}$$

Substituting (4.25) in (4.24), removing the middle horizontal block and rearranging the terms of this latter equation then yields that

$$\begin{pmatrix} Z^T H_*^\ell Z & Z^T J_*^T \\ J_* Z & 0 \end{pmatrix} \begin{pmatrix} Z^T \delta_k \\ \bar{\lambda}_k - \lambda_* \end{pmatrix} = \begin{pmatrix} Z^T(\nabla_x \Phi_k - H_*^\ell Y Y^T \delta_k) \\ c(x_k) - J_* Y Y^T \delta_k \end{pmatrix} - \begin{pmatrix} Z^T r_4 \\ r_3 \end{pmatrix}. \tag{4.26}$$

Roughly speaking, we now proceed by showing that the right-hand side of this relation is of the order of $\theta_k + \rho_k$, where

$$\theta_k \stackrel{\text{def}}{=} \begin{cases} \omega_k & \text{if } \mathcal{P} = \emptyset, \\ \eta_k & \text{if } \mathcal{P} \neq \emptyset. \end{cases} \tag{4.27}$$

We will then ensure that the vector on the left-hand side is of the same size, which is essentially the result we aim to prove. We first observe that

$$\|\delta_k\| = \|ZZ^T\delta_k + YY^T\delta_k\| \le \|Z^T\delta_k\| + \kappa_4\omega_k \tag{4.28}$$

from (4.3). We then obtain from (3.27) and (4.28) that

$$\|\bar{\lambda}_k - \lambda_*\| \le \sum_{j=1}^{q} \|(\bar{\lambda}_k - \lambda_*)_{[\mathcal{Q}_j]}\| \le \kappa_{12}\omega_k + q\kappa_3\|Z^T\delta_k\|, \tag{4.29}$$

where $\kappa_{12} = q(\kappa_2 + \kappa_3\kappa_4)$. Furthermore, from (4.18), (4.19), (4.22), (4.28) and (4.29),

$$\left\| \begin{pmatrix} Z^T r_4 \\ r_3 \end{pmatrix} \right\| \le \kappa_{13}\|Z^T\delta_k\|^2 + \kappa_{14}\|Z^T\delta_k\|\omega_k + \kappa_{15}{\omega_k}^2, \tag{4.30}$$

where $\kappa_{13} = \kappa_9 + q\kappa_3\kappa_{10} + \kappa_{11}$, $\kappa_{14} = 2\kappa_4(\kappa_9 + \kappa_{11}) + \kappa_{10}(\kappa_{12} + q\kappa_3\kappa_4)$, and $\kappa_{15} = \kappa_4^2(\kappa_9 + \kappa_{11}) + \kappa_4\kappa_{10}\kappa_{12}$. We now bound $c(x_k)$ by distinguishing components from $\mathcal{Z}$ and $\mathcal{P}$. We first note that, since the penalty parameters for each subset in $\mathcal{P}$ are bounded away from zero, the test (3.7)/(3.15) is satisfied for all k sufficiently large. Moreover, the remaining components of $c(x_k)$ satisfy the bound

$$\|c(x_k)_{[\mathcal{Q}_j]}\| \le \kappa_2\omega_k\mu_{k,j} + \mu_{k,j}\|(\lambda_k - \lambda_*)_{[\mathcal{Q}_j]}\| + \kappa_3\mu_{k,j}\|x_k - x_*\|, \tag{4.31}$$

for all $j \in \mathcal{Z}$ and all k sufficiently large, using (3.29). Hence, using (4.8), (3.7)/(3.15) and (4.31), we deduce that

$$\begin{aligned} \|c(x_k)\| &\le \sum_{j\in\mathcal{P}} \|c(x_k)_{[\mathcal{Q}_j]}\| + \sum_{j\in\mathcal{Z}} \|c(x_k)_{[\mathcal{Q}_j]}\| \\ &\le q\eta_k + q\kappa_2\omega_k\mu_{k,\mathcal{Z}} + \rho_k + q\kappa_3\mu_{k,\mathcal{Z}}\|\delta_k\|. \end{aligned} \tag{4.32}$$

Note that the first term of the last right-hand side only appears if $\mathcal{P}$ is not empty. Since both algorithms ensure that

$$\omega_k \le \eta_k \tag{4.33}$$

because $\alpha_\eta < 1$ and $\beta_\eta < 1$, we may obtain from (4.3), (4.32) and (4.28) that

$$\left\| \begin{pmatrix} Z^T(\nabla_x\Phi_k - H_*^\ell YY^T\delta_k) \\ c(x_k) - J_*YY^T\delta_k \end{pmatrix} \right\| \le \kappa_{16}\theta_k + \rho_k + \kappa_{12}\mu_{k,\mathcal{Z}}\theta_k + q\kappa_3\mu_{k,\mathcal{Z}}\|Z^T\delta_k\|, \tag{4.34}$$

where $\kappa_{16} = q + \left(1 + \kappa_4(\|Z^TH_*^\ell Y\| + \|J_*Y\|)\right)$. By assumption AS6, the coefficient matrix on the left-hand side of (4.26) is nonsingular. Let M be the norm of its inverse. Multiplying both sides of the equation by this inverse and taking norms, we obtain from (4.27), (4.30), (4.33) and (4.34) that

$$\begin{aligned} \left\| \begin{pmatrix} Z^T\delta_k \\ \bar{\lambda}_k - \lambda_* \end{pmatrix} \right\| \le\ & M[\kappa_{13}\|Z^T\delta_k\|^2 + \kappa_{14}\|Z^T\delta_k\|\theta_k + \kappa_{15}\theta_k^2 + \kappa_{16}\theta_k \\ & +\rho_k + \kappa_{12}\mu_{k,\mathcal{Z}}\theta_k + q\kappa_3\mu_{k,\mathcal{Z}}\|Z^T\delta_k\|]. \end{aligned} \tag{4.35}$$

Suppose now that k is sufficiently large to ensure that

$$\theta_k \le \frac{1}{4M\kappa_{14}} \tag{4.36}$$

and let

$$\bar{\alpha} \stackrel{\text{def}}{=} \min\left[\alpha_0, \frac{1}{4Mq\kappa_3}\right]. \tag{4.37}$$

Recall that α_0 and hence $\bar{\alpha} < 1$. Then, if $\mu_{k,\mathcal{Z}} \leq \bar{\alpha}$, the relations (4.35)–(4.37) give

$$\|Z^T \delta_k\| \leq \frac{1}{2}\|Z^T \delta_k\| + M[\kappa_{17}\theta_k + \rho_k + \kappa_{13}\|Z^T \delta_k\|^2], \tag{4.38}$$

where $\kappa_{17} = \kappa_{12} + \kappa_{15} + \kappa_{16}$. As δ_k, and hence $\|Z^T \delta_k\|$ converge to zero, we have that

$$\|Z^T \delta_k\| \leq \frac{1}{4M\kappa_{13}} \tag{4.39}$$

for k large enough. Hence inequalities (4.38) and (4.39) yield that

$$\|Z^T \delta_k\| \leq 4M(\kappa_{17}\theta_k + \rho_k). \tag{4.40}$$

If $\mathcal{P}$ is empty, we use (4.28), (4.40) and (4.27), the fact that $\mu_{k,\mathcal{Z}} = \alpha_k$ and the inequality

$$\rho_k \leq q\alpha_k \|\lambda_k - \lambda_*\| \tag{4.41}$$

to deduce (4.10), where $\kappa_5 \stackrel{\text{def}}{=} 4M\kappa_{17} + \kappa_4$ and $\kappa_6 = 4Mq$. Defining now $\kappa_7 \stackrel{\text{def}}{=} q(\kappa_2 + \kappa_3\kappa_5)$ and $\kappa_8 \stackrel{\text{def}}{=} q\kappa_3\kappa_6$, we deduce (4.11) from (3.27) and (4.10). Now, using (2.4),

$$\|c(x_k)\| \leq \sum_{j=1}^{q} \|c(x_k)_{[\mathcal{Q}_j]}\| = \sum_{j=1}^{q} \mu_{k,j}\|(\bar{\lambda}_k - \lambda_k)_{[\mathcal{Q}_j]}\| \leq q\alpha_k(\|\bar{\lambda}_k - \lambda_*\| + \|\lambda_k - \lambda_*\|) \tag{4.42}$$

and (4.12) then follows from (4.42) and (4.11).

If, on the other hand, $\mathcal{P}$ is not empty, (4.13) results from (3.27), (4.28), (4.40) with $\theta_k = \eta_k$ and (4.33), with $\kappa_7 \stackrel{\text{def}}{=} 4M\kappa_3\kappa_{17} + \kappa_2 + \kappa_3\kappa_4$ and $\kappa_8 \stackrel{\text{def}}{=} 4M\kappa_3$. Finally, (4.14) results from (2.4) and (4.13). □

We then deduce the following simple consequence of this lemma.

Corollary 4.5 *Suppose that the conditions of Lemma 4.4 hold and that $\hat{\lambda}_{k+1}$ is any Lagrange multiplier estimate for which*

$$\|\hat{\lambda}_{k+1} - \lambda_*\| \leq \kappa_{18}\|x_k - x_*\| + \kappa_{19}\omega_k, \tag{4.43}$$

for some positive constants κ_{18} and κ_{19} and all $k \in \mathcal{K}$ sufficiently large.

(i) If $\mathcal{P} = \emptyset$, then there are positive constants $\bar{\alpha} < 1$, κ_5, κ_6, κ_7, κ_8, and an integer k_2 such that, if $\alpha_{k_2} \leq \bar{\alpha}$, then (4.10),

$$\|\hat{\lambda}_{k+1} - \lambda_*\| \leq \kappa_7\omega_k + \kappa_8\alpha_k\|\lambda_k - \lambda_*\|, \tag{4.44}$$

and (4.12) hold for all $k \geq k_2$, $k \in \mathcal{K}$.

(ii) If, on the other hand, $\mathcal{P} \neq \emptyset$, then there are positive constants $\bar{\alpha} < 1$, κ_5, κ_6, κ_7, κ_8, and an integer k_2 such that, if $\mu_{k_2,\mathcal{Z}} \leq \bar{\alpha}$, then

$$\|(\hat{\lambda}_{k+1} - \lambda_*)_{[\mathcal{Q}_j]}\| \leq \kappa_7\eta_k + \kappa_8\rho_k, \tag{4.45}$$

and (4.14) hold for all $k \geq k_2$, $k \in \mathcal{K}$, and all $j \in \mathcal{Z}$.

Proof. Inequality (4.44) immediately results from (4.43) and (4.10), with $\kappa_7 = \kappa_5\kappa_{18} + \kappa_{19}$ and $\kappa_8 = \kappa_6\kappa_{18}$. To obtain (4.45), we deduce from (4.43), (4.28), (4.40), (4.27) and the assumption that $\mathcal{P} \neq \emptyset$, that

$$\begin{aligned} \|(\hat{\lambda}_{k+1} - \lambda_*)_{[\mathcal{Q}_j]}\| &\leq \kappa_{18}\|Z^T\delta_k\| + (\kappa_4\kappa_{18} + \kappa_{19})\omega_k \\ &\leq 4M\kappa_{18}(\kappa_{17}\theta_k + \rho_k) + (\kappa_4\kappa_{18} + \kappa_{19})\omega_k \\ &= 4M\kappa_{17}\kappa_{18}\eta_k + (\kappa_4\kappa_{18} + \kappa_{19})\omega_k + 4M\kappa_{18}\rho_k. \end{aligned} \tag{4.46}$$

Using (4.33), this yields the desired bound with $\kappa_7 \stackrel{\text{def}}{=} 4M\kappa_{17}\kappa_{18} + (\kappa_4\kappa_{18} + \kappa_{19})$ and $\kappa_8 \stackrel{\text{def}}{=} 4M\kappa_{18}$. □

We now show that, if the maximum penalty parameter α_k converges to zero, then the Lagrange multiplier estimates λ_k converge to their true values λ_*.

Lemma 4.6 *Assume AS1 and AS2 hold. Assume that $\{x_k\}$, the sequence of iterates generated by Algorithm 3.1 or 3.2, converges to the single limit point x_* at which AS6 holds, and with corresponding Lagrange multipliers λ_*.*

(i) If Algorithm 3.1 is used and if α_k tends to zero, the sequence λ_k converges to λ_*.

(ii) If, on the other hand, Algorithm 3.2 is used, if (4.43) holds for sufficiently large k, if $\mathcal{Z} \neq \emptyset$ and $\mu_{k,\mathcal{Z}}$ tends to zero, then the multiplier updates are accepted by this algorithm for all k sufficiently large and $\lambda_{k,[\mathcal{Q}_j]}$ converges to $\lambda_{*,[\mathcal{Q}_j]}$ for all $j \in \mathcal{Z}$.

Proof. Note first that AS6 implies AS4 and therefore that our assumptions are sufficient to apply Theorem 3.1.

Consider Algorithm 3.1 first. We observe that the desired convergence holds if $\lambda_{k,[\mathcal{Q}_j]}$ converges to $\lambda_{*,[\mathcal{Q}_j]}$, for all $j = 1, \ldots, q$. It is thus sufficient to show this latter result for an arbitrary j between 1 and q. This is obvious if Step 3a is executed infinitely often for the j-th subset. Indeed, each time this step is executed, $\lambda_{k+1,[\mathcal{Q}_j]} = \bar{\lambda}_{k,[\mathcal{Q}_j]}$ and the inequality (3.27) guarantees that $\bar{\lambda}_{k,[\mathcal{Q}_j]}$ converges to $\lambda_{*,[\mathcal{Q}_j]}$. Suppose therefore that Step 3a is not executed infinitely often for this subset. Then $\|(\lambda_k - \lambda_*)_{[\mathcal{Q}_j]}\|$ will remain fixed for all $k \geq k_3$, for some $k_3 > 0$, as Step 3b is executed for each remaining iteration. But (3.29) then implies that $\|c(x_k)_{[\mathcal{Q}_j]}\| \leq \kappa_{20}\mu_{k,j}$, for some constant $\kappa_{20} > 0$ and for all $k \geq k_4 \geq k_3$. As α_k tends to zero and $\alpha_\eta < 1$, $\kappa_{20}\mu_{k,j} \leq \kappa_{20}\alpha_k \leq \alpha_k^{\alpha_\eta} = \eta_k$ for all k sufficiently large for which α_k strictly decreases. But then inequality (3.7) must be satisfied for some $k \geq k_4$, which is impossible, as this would imply that Step 3a is again executed for the j-th subset. Hence Step 3a must be executed infinitely often.

Now consider Algorithm 3.2. Inequality (4.43) gives that, for $j \in \mathcal{Z}$,

$$\|\hat{\lambda}_{k+1,[\mathcal{Q}_j]}\| \leq \|\hat{\lambda}_{k+1}\| \leq \|\lambda_*\| + \kappa_{18}\|x_k - x_*\| + \kappa_{19}\omega_k \leq \mu_{k+1,j}^{-\gamma} \tag{4.47}$$

for all k sufficiently large, because ω_k is bounded for all k, $\{x_k\}$ converges to x_* and $\mu_{k+1,j}$ tends to zero. Hence, the multiplier update $\lambda_{k+1,[\mathcal{Q}_j]} = \hat{\lambda}_{k+1,[\mathcal{Q}_j]}$ will always be performed, for k sufficiently large and for each $j \in \mathcal{Z}$. The convergence of $\lambda_{k,[\mathcal{Q}_j]}$ to $\lambda_{*,[\mathcal{Q}_j]}$ now results from (4.43). □

We now consider the behaviour of the maximum penalty parameter α_k and show the important result that, under stated assumptions, it is bounded away from zero for both Algorithms 3.1 and 3.2. The proof of this result is inspired by the technique developed by Conn *et al.* (1991). When the single penalty parameter definition of the augmented Lagrangian (1.4) is used (or, equivalently, when $q = 1$), one then avoids a steadily increasing ill-conditioning of the Hessian of the augmented Lagrangian. Note that this ill-conditioning is also avoided when $q > 1$, as we show below in Theorem 4.8.

Theorem 4.7 *Assume AS1 and AS2 hold and suppose that the sequence of iterates* $\{x_k\}$ *of Algorithm 3.1 or 3.2 converges to a single limit point* x_* *with corresponding Lagrange multipliers* λ_**, at which AS5 and AS6 hold. Assume furthermore that (4.43) holds for sufficiently large* k *when Algorithm 3.2 is used. Then there is a constant* $\alpha_{\min} \in (0,1)$ *such that* $\alpha_k \geq \alpha_{\min}$ *for all* k*.*

Proof. Suppose otherwise that α_k tends to zero (that is $\mathcal{P} = \emptyset$), and hence that $\mu_{k,t}$ tends to zero for each t between 1 and q. Then Step 3b of either algorithm must be executed infinitely often for each subset. We aim to obtain a contradiction to this statement by showing that Step 3a is always executed for each subset for sufficiently large k. We note that our assumptions are sufficient to apply Theorem 3.1. Furthermore, we may apply Lemma 4.4 (or Corollary 4.5) to the complete sequence of iterates generated by either algorithm.

First recall that

$$\alpha_k \leq \bar{\alpha} < 1 \tag{4.48}$$

for all $k \geq k_2$, where $\bar{\alpha}$ and k_2 are those of Lemma 4.4 and Corollary 4.5. Note that

$$\omega_k \leq \alpha_k \tag{4.49}$$

for all $k \geq k_2$. This follows by definition if (3.13) is executed. Otherwise it is a consequence of the fact that α_k is unchanged while ω_k is reduced, when (3.14) occurs. Let k_5 be the smallest integer such that

$$\alpha_k^{1-\alpha_\eta} \leq \frac{1}{q(2+\kappa_7)}, \tag{4.50}$$

$$\alpha_k^{1-\beta_\eta} \leq \min\left[\frac{1}{\kappa_{21}}, \frac{1}{q(2\kappa_{21}+\kappa_7)}\right], \tag{4.51}$$

where $\kappa_{21} = \max(1, \kappa_7 + \kappa_8)$, and, if Algorithm 3.2 is used,

$$\lambda_{k+1,[\mathcal{Q}_t]} = \hat{\lambda}_{k+1,[\mathcal{Q}_t]} \tag{4.52}$$

for all $t = 1, \ldots, q$, which is possible because of Lemma 4.6. Note that (4.48) and (4.51) imply that

$$\alpha_k \leq \alpha_k^{1-\beta_\eta} \leq \frac{1}{\kappa_{21}} \leq \frac{1}{\kappa_8} \tag{4.53}$$

for all $k \geq \max(k_2, k_5)$. Furthermore, let k_6 be such that

$$\|\lambda_k - \lambda_*\| \leq 1 \tag{4.54}$$

for all $k \geq k_6$, which is possible because of Lemma 4.6 for either algorithm. Now define $k_7 = \max(k_2, k_5, k_6)$, let Γ be the set $\{k \mid (3.13) \text{ is executed at iteration } k-1 \text{ and } k \geq k_7\}$ and let k_0 be the smallest element of Γ. By the assumption that α_k tends to zero, Γ has an infinite number of elements.

By definition of Γ, for iteration k_0, $\omega_{k_0} = \alpha_{k_0}$ and $\eta_{k_0} = \alpha_{k_0}^{\alpha_\eta}$. Then inequality (4.12) gives that, for each t,

$$\begin{array}{rcll} \|c(x_{k_0})_{[\mathcal{Q}_t]}\| & \leq & \|c(x_{k_0})\| & \\ & \leq & q(\alpha_{k_0} + \kappa_8 \alpha_{k_0}^2)\|\lambda_{k_0} - \lambda_*\| + q\kappa_7 \omega_{k_0} \alpha_{k_0} & \\ & \leq & 2q\alpha_{k_0}\|\lambda_{k_0} - \lambda_*\| + q\kappa_7\omega_{k_0}\alpha_{k_0} & \text{(from (4.53))} \\ & \leq & q\alpha_{k_0}(2 + \kappa_7\alpha_{k_0}) & \text{(from (4.54))} \\ & \leq & q(2+\kappa_7)\alpha_{k_0} & \text{(from (4.48))} \\ & \leq & \alpha_{k_0}^{\alpha_\eta} = \eta_{k_0} & \text{(from (4.50)).} \end{array} \tag{4.55}$$

As a consequence of this inequality, Step 3a of Algorithm 3.1 or Algorithm 3.2 will be executed for each t with $\lambda_{k_0+1,[\mathcal{Q}_t]} = \bar{\lambda}(x_{k_0}, \lambda_{k_0,[\mathcal{Q}_t]}, \mu_{k_0,t})_{[\mathcal{Q}_t]}$ or $\lambda_{k_0+1,[\mathcal{Q}_t]} = \hat{\lambda}_{k_0+1,[\mathcal{Q}_t]}$ respectively. Inequality (4.11)/(4.44) together with (4.54) guarantee that

$$\|\lambda_{k_0+1} - \lambda_*\| \le \kappa_7 \omega_{k_0} + \kappa_8 \alpha_{k_0} \|\lambda_{k_0} - \lambda_*\| \le \kappa_{21} \alpha_{k_0}. \tag{4.56}$$

We shall now make use of an inductive proof. Assume that, for each t, Step 3a of either algorithm is executed for iterations $k_0 + i$, $(0 \le i \le j)$, and that

$$\|\lambda_{k_0+i+1} - \lambda_*\| \le \kappa_{21} \alpha_{k_0}^{1+\beta_\eta i}. \tag{4.57}$$

Inequalities (4.55) and (4.56) show that this is true for $j = 0$. We aim to show that the same is true for $i = j+1$. Our assumption that Step 3a is executed gives that, for iteration $k_0 + j + 1$, $\alpha_{k_0+j+1} = \alpha_{k_0}$, $\omega_{k_0+j+1} = \alpha_{k_0}^{j+2}$, and $\eta_{k_0+j+1} = \alpha_{k_0}^{\beta_\eta(j+1)+\alpha_\eta}$. Then, inequality (4.12) yields that, for each t,

$$\begin{aligned}
\|c(x_{k_0+j+1})_{[\mathcal{Q}_t]}\| &\le \|c(x_{k_0+j+1})\| \\
&\le q(\alpha_{k_0+j+1} + \kappa_8 \alpha_{k_0+j+1}^2)\|\lambda_{k_0+j+1} - \lambda_*\| \\
&\quad + q\kappa_7 \omega_{k_0+j+1}\alpha_{k_0+j+1} \\
&\le 2q\alpha_{k_0+j+1}\|\lambda_{k_0+j+1} - \lambda_*\| + q\kappa_7\omega_{k_0+j+1}\alpha_{k_0+j+1} && \text{(from (4.53))} \\
&\le 2q\kappa_{21}\alpha_{k_0}\alpha_{k_0}^{1+\beta_\eta j} + q\kappa_7 \alpha_{k_0}^{j+3} && \text{(from (4.57))} \\
&\le 2q\kappa_{21}\alpha_{k_0}\alpha_{k_0}^{\alpha_\eta+\beta_\eta j} + q\kappa_7 \alpha_{k_0}^{\alpha_\eta+\beta_\eta(j+1)+1} \\
&\le q(2\kappa_{21} + \kappa_7)\alpha_{k_0}^{1-\beta_\eta}\alpha_{k_0}^{\beta_\eta(j+1)+\alpha_\eta} && \text{(from (4.53))} \\
&\le \alpha_{k_0}^{\beta_\eta(j+1)+\alpha_\eta} = \eta_{k_0+j+1} && \text{(from (4.51))}.
\end{aligned} \tag{4.58}$$

Hence Step 3a of either algorithm will again be executed for each t with

$$\lambda_{k_0+j+2,[\mathcal{Q}_t]} = \bar{\lambda}(x_{k_0+j+1}, \lambda_{k_0+j+1,[\mathcal{Q}_t]}, \mu_{k_0+j+1,t})_{[\mathcal{Q}_t]} \text{ or } \lambda_{k_0+j+2,[\mathcal{Q}_t]} = \hat{\lambda}_{k_0+j+2,[\mathcal{Q}_t]},$$

respectively. Inequality (4.11)/(4.44) then implies that

$$\begin{aligned}
\|\lambda_{k_0+j+2} - \lambda_*\| &\le \kappa_7\omega_{k_0+j+1} + \kappa_8\alpha_{k_0+j+1}\|\lambda_{k_0+j+1} - \lambda_*\| \\
&\le \kappa_7\alpha_{k_0}^{j+2} + \kappa_8\kappa_{21}\alpha_{k_0}\alpha_{k_0}^{1+\beta_\eta j} && \text{(from (4.57))} \\
&\le \kappa_7\alpha_{k_0}^{1+\beta_\eta(j+1)} + \kappa_8\kappa_{21}\alpha_{k_0}\alpha_{k_0}^{1+\beta_\eta j} \\
&= (\kappa_7 + \kappa_8\kappa_{21}\alpha_{k_0}^{1-\beta_\eta})\alpha_{k_0}^{1+\beta_\eta(j+1)} \\
&\le (\kappa_7 + \kappa_8)\alpha_{k_0}^{1+\beta_\eta(j+1)} && \text{(from (4.51))} \\
&\le \kappa_{21}\alpha_{k_0}^{1+\beta_\eta(j+1)}
\end{aligned} \tag{4.59}$$

which establishes (4.57) for $i = j + 1$. Thus Step 3a of the appropriate algorithm is executed for each $t = 1, \ldots, q$ for all iterations $k \ge k_0$. But this implies that Γ is finite, which contradicts the assumption that Step 3b is executed infinitely often for each subset. Hence the theorem is proved. □

We now prove the stronger result that *all* penalty parameters stay bounded away from zero.

Theorem 4.8 *Assume AS1 and AS2 hold and suppose that the sequence of iterates $\{x_k\}$ of Algorithm 3.1 or 3.2 converges to a single limit point x_* with corresponding Lagrange multipliers λ_*, at which AS5 and AS6 hold. Assume furthermore that (4.43) holds for sufficiently large k when Algorithm 3.2 is used. Then there is a constant $\mu > 0$ such that $\mu_{k,j} \ge \mu$ for all k and all $j = 1, \ldots, q$.*

Proof. Assume otherwise that $\mathcal{Z}$ is not empty, and hence that $\mu_{k,\mathcal{Z}}$ converges to zero. Then Step 3b of either algorithm must be executed infinitely often for $t \in \mathcal{Z}$. We aim to obtain a contradiction to this statement by showing that, for any $t \in \mathcal{Z}$, Step 3a is always executed for sufficiently large k. We may deduce from Theorem 4.7 that α_k attains its minimum value $\alpha_{\min} \in (0,1)$ at iteration $k_{\max}$, say. Hence, $\mathcal{P} \neq \emptyset$. Furthermore, we may apply Lemma 4.4 (or Corollary 4.5) to the complete sequence of iterates generated by either algorithm. Let $k_8 \geq k_{\max}$ be the smallest integer for which

$$\mu_{k,\mathcal{Z}} \leq \min\left[\bar{\alpha}, \frac{1}{2\kappa_7+\kappa_8}, \frac{\alpha_{\min}^{\beta_\eta+\epsilon}-\alpha_{\min}}{q(2\kappa_7+\kappa_8)}\right] \tag{4.60}$$

for all $k \geq k_8 \geq k_2$, where $\bar{\alpha}$ and k_2 are those of Lemma 4.4 and Corollary 4.5, and where $\epsilon = \frac{1}{2}(1-\beta_\eta)$. Note that $\alpha_{\min}^{\beta_\eta+\epsilon} > \alpha_{\min}$ as $\beta_\eta < 1$.

Assume first that Algorithm 3.1 is used and consider the t-th subset, for some $t \in \mathcal{Z}$. At iteration $k \geq k_8$, this algorithm ensures that

$$\mu_{k+1,t}\|(\lambda_{k+1}-\lambda_*)_{[Q_t]}\| \leq \alpha_{\min}\mu_{k,t}\|(\lambda_k-\lambda_*)_{[Q_t]}\| \tag{4.61}$$

if Step 3b is executed for the t-th subset, while (4.13) ensures that

$$\mu_{k+1,t}\|(\lambda_{k+1}-\lambda_*)_{[Q_t]}\| \leq \mu_{k,t}(\kappa_7\eta_k+\kappa_8\rho_k) \tag{4.62}$$

if Step 3a is executed for the same subset. Summing on all $t \in \mathcal{Z}$, and defining

$$\begin{aligned}\mathcal{Z}_{k,a} &= \{t \in \mathcal{Z} \mid \text{Step 3a is executed for the } t\text{-th subset at iteration } k\}\\ \mathcal{Z}_{k,b} &= \{t \in \mathcal{Z} \mid \text{Step 3b is executed for the } t\text{-th subset at iteration } k\},\end{aligned} \tag{4.63}$$

we obtain that

$$\begin{aligned}\rho_{k+1} &\leq \alpha_{\min}\sum_{t\in\mathcal{Z}_{k,b}}\mu_{k,t}\|(\lambda_k-\lambda_*)_{[Q_t]}\| + \sum_{t\in\mathcal{Z}_{k,a}}\mu_{k,t}(\kappa_7\eta_k+\kappa_8\rho_k)\\ &\leq (\alpha_{\min}+\kappa_8 q\mu_{k,\mathcal{Z}})\rho_k + \kappa_7 q\mu_{k,\mathcal{Z}}\eta_k.\end{aligned} \tag{4.64}$$

For the purpose of obtaining a contradiction, assume now that

$$\rho_k \geq \frac{1}{2}\eta_k \tag{4.65}$$

for all $k \geq k_8$. Then (4.64) gives that, for all $k \geq k_8$,

$$\frac{\rho_{k+1}}{\rho_k} \leq \alpha_{\min} + \kappa_8 q\mu_{k,\mathcal{Z}} + 2\kappa_7 q\mu_{k,\mathcal{Z}} \leq \alpha_{\min}^{\beta_\eta+\epsilon} < 1 \tag{4.66}$$

because of (4.60). Hence we obtain from (4.66) that

$$\rho_{k+1} \leq \rho_{k_8}\alpha_{\min}^{(k-k_8+1)(\beta_\eta+\epsilon)}. \tag{4.67}$$

Therefore, since $\rho_{k_8}\alpha_{\min}^{(k-k_8+1)\epsilon}$ tends to zero, we obtain that

$$\rho_{k+1} < \frac{1}{2}\alpha_{\min}^{\alpha_\eta+(k_8-k_{\max})\beta_\eta}\alpha_{\min}^{(k-k_8+1)\beta_\eta} = \frac{1}{2}\alpha_{\min}^{\alpha_\eta+(k-k_{\max}+1)\beta_\eta} = \frac{1}{2}\eta_{k+1} \tag{4.68}$$

for all sufficiently large k, where the last equality results from the definition of $k_{\max}$. But this contradicts (4.65), which implies that (4.65) does not hold for all k sufficiently large. As a consequence, there exists a subsequence $\mathcal{K}$ such that

$$\rho_k < \frac{1}{2}\eta_k \tag{4.69}$$

for all $k \in \mathcal{K}$. Consider such a k. Then, using (4.64) and (4.69), we deduce that

$$\rho_{k+1} < \frac{1}{2}\eta_k(\alpha_{\min} + q\kappa_8\mu_{k,\mathcal{Z}} + 2q\kappa_7\mu_{k,\mathcal{Z}}) \le \frac{1}{2}\alpha_{\min}^{\beta_\eta+\epsilon}\eta_k \le \frac{1}{2}\eta_{k+1}, \tag{4.70}$$

where we have used (4.60) to obtain the second inequality. As a consequence, $k+1 \in \mathcal{K}$ and (4.69) holds for all k sufficiently large. Returning to subset $t \in \mathcal{Z}$, we now obtain from (4.14) and (4.69) that

$$\|c(x_k)_{[\mathcal{Q}_t]}\| \le \eta_k(\kappa_7\mu_{k,\mathcal{Z}} + \frac{1}{2}(1 + \kappa_8\mu_{k,\mathcal{Z}})) \le \eta_k, \tag{4.71}$$

for all k sufficiently large, because of (4.60). Hence Step 3a of either algorithm is executed for the subset t and for all sufficiently large k, which implies that t does not belong to $\mathcal{Z}$. Therefore $\mathcal{Z}$ is empty and the proof of the theorem is completed as far as Algorithm 3.1 is concerned.

The reasoning is similar for Algorithm 3.2. If we consider the t-th subset, for $t \in \mathcal{Z}$, we see that, because of (4.45) and Lemma 4.6, one has that

$$\mu_{k+1,t}\|(\lambda_{k+1} - \lambda_*)_{[\mathcal{Q}_t]}\| \le \mu_{k,t}(\kappa_7\eta_k + \kappa_8\rho_k) \tag{4.72}$$

for all k sufficiently large. Summing over all the subsets in $\mathcal{Z}$, dividing by ρ_k and assuming that (4.65) holds, one then obtains, for all $k \ge k_8$, that

$$\frac{\rho_{k+1}}{\rho_k} \le \mu_{k,\mathcal{Z}}(2q\kappa_7 + q\kappa_8) \le \alpha_{\min}^{\beta_\eta+\epsilon} < 1 \tag{4.73}$$

where the second inequality holds because of (4.60). As above we may therefore deduce a contradiction with (4.65) and obtain (4.69) for a subsequence $\mathcal{K}$. If we now consider a $k \in \mathcal{K}$, we obtain, analogously to (4.70), that

$$\rho_{k+1} \le \mu_{k,\mathcal{Z}}(q\kappa_7\eta_k + q\kappa_8\rho_k) < \frac{1}{2}\eta_k\alpha_{\min}^{\beta_\eta+\epsilon} \le \frac{1}{2}\eta_{k+1} \tag{4.74}$$

where we used (4.72), the definitions of ρ_{k+1} and $\mu_{k,\mathcal{Z}}$, (4.69) and (4.60). Hence $k+1 \in \mathcal{K}$ and the rest of the proof follows as for Algorithm 3.1 □

Note, in particular, that if Algorithm 3.2 is used with $\hat{\lambda}_{k+1}$ chosen as either the first-order or least-squares multiplier estimates, the penalty parameters $\mu_{k,j}$ will stay bounded away from zero. This follows directly from Theorem 4.8 because each of the inequalities (3.27) and (3.28) imply (4.43).

We also note that Corollary 4.3 ensures that the least-squares multiplier estimates (2.9) are implementable when the non-degeneracy condition (4.6) holds. By this we mean that the estimates

$$\hat{\lambda}_k = -((J(x_k)Z_k)^+)^T Z_k^T g(x_k) \tag{4.75}$$

are identical to those defined in (2.9) for all k sufficiently large, and, unlike (2.9), are well defined when x_* is unknown.

As in Conn *et al.* (1991), we examine the rate of convergence of our algorithms.

Theorem 4.9 *Under the assumptions of Theorem 4.8, the iterates x_k, the Lagrange multipliers $\bar{\lambda}_k$ of Algorithm 3.1 and any $\hat{\lambda}_k$ satisfying (4.43) for Algorithm 3.2 are at least R-linearly convergent with R-factor at most $\alpha_{\min}^{\beta_\eta}$, where $\alpha_{\min}$ is the smallest value of the maximum penalty parameter generated by the algorithm in question.*

Proof. The proof parallels that of Lemma 4.4. First, Theorem 4.7 shows that the maximum penalty parameter α_k stays bounded away from zero, and thus remains fixed at some value $\alpha_{\min} > 0$, for $k \geq k_{\max}$. For all subsequent iterations,

$$\omega_{k+1} = \alpha_{\min}\omega_k \text{ and } \eta_{k+1} = \alpha_{\min}^{\beta_\eta}\eta_k \tag{4.76}$$

hold. Moreover, Theorem 4.8 implies that, for all $j = 1, \ldots, q$, (3.7)/(3.15) hold for all $k \geq k_9 \geq k_{\max}$, say. Hence and because of (4.3), the bound on the right-hand side of (4.34) may be replaced by $\kappa_{16}\omega_k + q\eta_k$, and thus

$$\|Z^T\delta_k\| \leq M[\kappa_{16}\omega_k + q\eta_k + \kappa_{13}\|Z^T\delta_k\|^2 + \kappa_{14}\|Z^T\delta_k\|\omega_k + \kappa_{15}\omega_k^2]. \tag{4.77}$$

Therefore, if k is sufficiently large that

$$\omega_k \leq \frac{1}{2M\kappa_{14}} \tag{4.78}$$

and

$$\|Z^T\delta_k\| \leq \frac{1}{4M\kappa_{13}}, \tag{4.79}$$

inequalities (4.77)–(4.79) can be rearranged to yield

$$\|Z^T\delta_k\| \leq 4M(\kappa_{22}\omega_k + q\eta_k), \tag{4.80}$$

where $\kappa_{22} = \kappa_{15} + \kappa_{16}$. But then (4.28) gives that

$$\|\delta_k\| \leq \kappa_{23}\omega_k + \kappa_{24}\eta_k \tag{4.81}$$

where $\kappa_{23} = \kappa_4 + 4M\kappa_{22}$ and $\kappa_{24} = 4Mq$. As $\beta_\eta < 1$ and $\alpha_{\min} < 1$, (4.76) and (4.81) show that x_k converges to x_* at least R-linearly, with R-factor $\alpha_{\min}^{\beta_\eta}$. Inequalities (3.27)/(4.43) and (4.81) then guarantee the same property for $\bar{\lambda}_k$ and $\hat{\lambda}_k$. $\square$

To conclude this section, we note that the conclusions of Theorems 4.7, 4.8 and 4.9 require that the complete sequence of iterates converges to a unique limit point. This assumption cannot be relaxed. The counterexample presented by Conn *et al.* (1991) (where the linear inequality constraints are simple bound constraints on the problem's variables) shows that the sequence of penalty parameters may indeed converge to zero, if there is more than a single limit point.

5 Second order conditions

If we further strengthen the stopping test for the inner iteration beyond (3.5) to include second-order conditions, we can then guarantee that our algorithms converge to an isolated local solution. More specifically, we require the following additional assumption.

AS7: Suppose that x_k satisfies (3.5), converges to x_* for $k \in \mathcal{K}$, such that Z_* has a rank strictly greater than m. Then, if Z is defined as in AS6, we assume that $Z^T\nabla_{xx}\Phi_k Z$ is uniformly positive definite (that is, its smallest eigenvalue is uniformly bounded away from zero) for all $k \in \mathcal{K}$ sufficiently large.

We can then prove the following result.

Theorem 5.1 *Under assumptions AS1–AS3, AS5–AS7, the iterates x_k, $k \in \mathcal{K}$, generated by either Algorithm 3.1 or 3.2 converge to an isolated local solution of (1.1)–(1.3).*

Proof. By definition of Φ,

$$\nabla_{xx}\Phi_k = H^\ell(x_k, \bar{\lambda}_k) + \sum_{j=1}^{q} \frac{1}{\mu_{k,j}} J_{\mathcal{Q}_j}(x_k)^T J_{\mathcal{Q}_j}(x_k), \tag{5.1}$$

where $J_{\mathcal{Q}_j}(x)$ is the Jacobian of $c(x)_{[\mathcal{Q}_j]}$. Note that the rank of Z is at least that of Z_*. AS7 then implies that there exists a nonzero vector s such that

$$J(x_k)Zs = 0 \tag{5.2}$$

and hence

$$J_{\mathcal{Q}_j}(x_k)Zs = 0 \tag{5.3}$$

for each j. For any such vector, AS7 further implies that

$$s^T Z^T \nabla_{xx}\Phi_k Zs \geq \kappa_{25}\|s\|^2 \tag{5.4}$$

for some $\kappa_{25} > 0$, which in turn gives that

$$s^T Z^T H^\ell(x_k, \bar{\lambda}_k)Zs \geq \kappa_{25}\|s\|^2, \tag{5.5}$$

because of (5.1) and (5.3). By continuity of H^ℓ as x_k and $\bar{\lambda}_k$ approach their limits, this ensures that

$$s^T Z^T H^\ell(x_*, \lambda_*)Zs \geq \kappa_{25}\|s\|^2 \tag{5.6}$$

for all nonzero s satisfying

$$J(x_*)Zs = 0, \tag{5.7}$$

which implies that x_* is an isolated local solution of (1.1)–(1.3) (see, for instance, Avriel (1976), Thm. 3.11). □

If we assume that the inner iteration stopping test is tightened so that $\nabla_{xx}\Phi_k$ is required to be uniformly positive definite in the nullspace of the dominant constraints, and if we assume that the non-degeneracy condition (4.6) holds, then Corollary 4.3 ensures that $Z_k = Z = Z_*$ for sufficiently large k and Theorem 5.1 holds. A weaker version of this result also holds, where only positive semi-definiteness of the augmented Lagrangian's Hessian is required, yielding then that x_* is a (possibly not isolated) minimizer of the problem.

6 Conclusion

We have considered two augmented Lagrangian algorithms for constrained nonlinear optimization, where the linear constraints present in the problem are handled directly and where multiple penalty parameters are allowed. These algorithms have the advantage that efficient techniques for handling linear constraints may be used at the inner iteration level, and also that the sparsity pattern of the Hessian of the augmented Lagrangian is independent of that of the linear constraints. The local convergence results available for the specific case where linear constraints reduce to simple bounds have been extended to the more general and useful context where any form of linear constraint is permitted.

We finally note that the theory presented is directly relevant to practical computation, as the inner iteration stopping rule (3.5) covers the type of optimality tests used in available packages for linearly constrained problems. This means that these packages can therefore be applied for obtaining an (approximate) solution of the subproblem, which constitues a realistic and attractive algorithmic development.

7 Acknowledgements

The authors wish to acknowledge funding provided by a NATO travel grant.

References

[Arioli *et al.*, 1993] M. Arioli, T.F. Chan, I.S. Duff, Nick Gould, and J.K. Reid. Computing a search direction for large-scale linearly constrained nonlinear optimization calculations. Technical Report (in preparation), CERFACS, Toulouse, France, 1993.

[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 Multipliers Methods.* Academic Press, London, 1982.

[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.

[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.*, 1992] 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.*, 1993a] A. R. Conn, Nick Gould, A. Sartenaer, and Ph. L. Toint. Global convergence of two augmented Lagrangian algorithms for optimization with a combination of general equality and linear constraints. Technical Report TR/PA/93/26, CERFACS, Toulouse, France, 1993.

[Conn *et al.*, 1993b] A. R. Conn, Nick Gould, A. Sartenaer, and Ph. L. Toint. Global convergence of a class of trust region algorithms for optimization using inexact projections on convex constraints. *SIAM Journal on Optimization*, 3(1):164–221, 1993.

[Dunn, 1987] J. C. Dunn. On the convergence of projected gradient processes to singular critical points. *Journal of Optimization Theory and Applications*, 55:203–216, 1987.

[Forsgren and Murray, 1993] A. L. Forsgren and W. Murray. Newton methods for large-scale linear equality-constrained minimization. *SIAM Journal on Matrix Analysis and Applications*, 14(2):560–587, 1993.

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

[Gruver and Sachs, 1980] W. A. Gruver and E. Sachs. *Algorithmic methods in optimal control.* Pitman, Boston, USA, 1980.

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

[Lustig *et al.*, 1989] 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 Applications*, 152:191–222, 1989.

[Moreau, 1962] J. J. Moreau. Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires. *Comptes-Rendus de l'Académie des Sciences (Paris)*, 255:238–240, 1962.

[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.

[Toint and Tuyttens, 1992] Ph. L. Toint and D. Tuyttens. LSNNO: a Fortran subroutine for solving large scale nonlinear network optimization problems. *ACM Transactions on Mathematical Software*, 18(3):308–328, 1992.