Global Convergence 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]

July 5, 1993

Abstract. We consider the global convergence properties of a class of augmented Lagrangian methods for solving nonlinear programming problems. In the proposed method, linear constraints are treated separately from more general constraints. Thus only the latter are combined with the objective function in an augmented Lagrangian. The subproblem then consists of (approximately) minimizing this augmented Lagrangian subject to the linear constraints. In this paper, we prove the global convergence of the sequence of iterates generated by this technique to a first-order stationary point of the original problem. We consider various stopping rules for the iterative solution of the subproblem, including practical tests used in several existing packages for linearly constrained optimization. We also extend our results to the case where the augmented Lagrangian's definition involves several distinct penalty parameters.

[1] IBM T.J. Watson Research Center
Yorktown Heights, USA
[2] CERFACS
Toulouse, France
[3] Department of Mathematics,
Facultés Universitaires ND de la Paix,
Namur, Belgium

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

Global Convergence 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]

Report 93/01 July 5, 1993

Abstract. We consider the global convergence properties of a class of augmented Lagrangian methods for solving nonlinear programming problems. In the proposed method, linear constraints are treated separately from more general constraints. Thus only the latter are combined with the objective function in an augmented Lagrangian. The subproblem then consists of (approximately) minimizing this augmented Lagrangian subject to the linear constraints. In this paper, we prove the global convergence of the sequence of iterates generated by this technique to a first-order stationary point of the original problem. We consider various stopping rules for the iterative solution of the subproblem, including practical tests used in several existing packages for linearly constrained optimization. We also extend our results to the case where the augmented Lagrangian's definition involves several distinct penalty parameters.

[1] IBM T.J. Watson Research Center
Yorktown Heights, USA
[2] CERFACS
Toulouse, France
[3] Department of Mathematics,
Facultés Universitaires ND de la Paix,
Namur, Belgium

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

1 Introduction

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

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

where x is required to satisfy the general 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 $b \in \Re^p$.

A well known 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

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

where the components λ_i of the vector λ are known as *Lagrange multiplier estimates* and where μ is known as the *penalty parameter* (see, for instance, Hestenes (1969), Powell (1969) and Bertsekas (1982)) . The question then arises how to deal with the additional linear inequality constraints (1.3). The case where A is the identity matrix (that is when (1.3) specifies bounds on the variables) has been considered by Conn *et al.* (1991). They propose keeping these constraints explicitly outside the augmented Lagrangian formulation, 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 simple bounds, is attempted. In this approach all linear inequalities more general than bound constraints are incorporated in the augmented Lagrangian function. This strategy has been implemented and successfully

[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.

applied within the LANCELOT package for large-scale nonlinear optimization (see Conn *et al.* (1992)). However, such a method may be inefficient when linear constraints are present as there are a number of effective techniques specifically designed to handle such constraints directly (see Arioli *et al.* (1993), Forsgren and Murray (1993) or Lustig *et al.* (1989), for instance). This is especially noticeable for large-scale problems. The purpose of the present paper is therefore to define and analyze an algorithm where the constraints (1.3) are kept outside the augmented Lagrangian and handled at the level of the subproblem minimization.

Our proposal extends the method of Conn *et al.* (1991) in that not only bounds but general linear inequalities are treated separately. Fletcher (1987, page 295) remarks on the potential advantages of this strategy. We believe this is important because it provides considerable additional algorithmic flexibility to account for non-uniform changes in the behaviour of the general constraints. This can be helpful for instance when the degree of nonlinearity of (sets of) these constraints is substantially different.

Note that our formulation (1.3) allows us to specify linear equality constraints as two inequalities of opposite sign. Alternately, one can apply the theory presented below in the affine subspace of points feasible for the linear equality constraints, provided all iterates satisfy these equations. As a consequence, the augmented Lagrangian function need not contain any purely linear constraint, which means that the sparsity of its Hessian matrix is unaffected by these constraints.

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, while their global convergence analysis is developed in Section 4. Section 5 proposes alternatives to the criticality measure used in this analysis. We extend, in Section 6, the framework analyzed in Section 4 to cover the case where different penalty parameters are used for disjoint general equality constraints subsets. Finally, some conclusions and perspectives are outlined in Section 7.

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.

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

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

Thus

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

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 with respect to the c_i constraints only. If one defines *first-order Lagrange multiplier estimates*

$$\bar\lambda(x,\lambda,\mu) \stackrel{\rm def}{=} \lambda + c(x)/\mu, \tag{2.4}$$

one can verify the identity

$$\begin{array}{rl} \nabla_x\Phi(x,\lambda,\mu) = & \nabla_x f(x) + \sum_{i=1}^m \lambda_i\nabla_x c_i(x) + \frac{1}{\mu}\sum_{i=1}^m c_i(x)\nabla_x c_i(x) \\ = & g^\ell(x,\bar\lambda(x,\lambda,\mu)). \end{array} \tag{2.5}$$

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

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

where we have written

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

Let x satisfy (1.3). We then define the criticality measure

CM1:

$$\sigma(x,\lambda,\mu) \stackrel{\rm def}{=} \begin{array}{ll} |\min_{d\in\Re^n} & \nabla_x\Phi(x,\lambda,\mu)^T d|, \\ \text{subject to} & A(x+d)-b \geq 0, \\ & \|d\| \leq 1, \end{array} \tag{2.8}$$

representing the magnitude of the maximum decrease of the linearized model of the augmented Lagrangian function achievable on the intersection of the domain defined by the linear inequality constraints (1.3) with a ball of radius one centered at x. Here and below, $\|.\|$ denotes the ℓ_2-norm or the induced operator norm. The algorithms we are about to develop construct iterates which force $\sigma_k = \sigma(x_k,\lambda_k,\mu_k)$ to zero as k increases. The measure (2.8) was previously introduced by Conn *et al.* (1993a).

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

$$A_* Z_* = 0 \text{ and } Z_*^T Z_* = I. \tag{2.9}$$

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

$$\lambda(x) \stackrel{\rm def}{=} -((J(x)Z_*)^+)^T Z_*^T g(x) \tag{2.10}$$

at all points where the right generalized inverse

$$(J(x)Z_*)^+ \stackrel{\text{def}}{=} Z_*^T J(x)^T (J(x)Z_*Z_*^T J(x)^T)^{-1} \tag{2.11}$$

of $J(x)Z_*$ is well defined. We note that $\lambda(x)$ is differentiable; for completeness the derivative is given in the following lemma.

Lemma 2.1 *Suppose that AS2 holds. If $J(x)Z_*Z_*^T J(x)^T$ is nonsingular, $\lambda(x)$ is differentiable and its derivative is given by*

$$\nabla_x \lambda(x) = -((J(x)Z_*)^+)^T Z_*^T H^\ell(x, \lambda(x)) - (J(x)Z_*Z_*^T J(x)^T)^{-1} R(x) \tag{2.12}$$

where the ith row of $R(x)$ is $(Z_*^T g(x) + Z_*^T J(x)^T \lambda(x))^T Z_*^T H_i(x)$.

Proof. The result follows by observing that (2.10) may be rewritten as

$$r(x) - Z_*^T J(x)^T \lambda(x) = Z_*^T g(x) \;\text{ and }\; J(x)Z_* r(x) = 0 \tag{2.13}$$

for some vector $r(x)$. Differentiating (2.13) and eliminating the derivative of $r(x)$ from the resulting equations gives the required result. □

We stress that, as stated, the Lagrange multiplier estimate (2.10) is not directly calculable as it requires an a priori knowledge of x_*. It is merely introduced as an analytical device.

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

3 Statement of the algorithms

In order to solve the problem (1.1)–(1.3), we consider the following algorithmic models. Both models depend on an infinite sequence of positive numbers $\{\omega_k\}_{k=0}^\infty$ tending to zero.

Algorithm 3.1

Step 0 [Initialization]. An initial vector of Lagrange multiplier estimates λ_0 is given. The positive constants $\mu_s < 1$, $\tau < 1$, $\omega_* \ll 1$, $\eta_* \ll 1$, α_η, and β_η are specified. Set $\mu_0 = \mu_s$, $\eta_0 = \mu_0^{\alpha_\eta}$, and $k = 0$.

Step 1 [Inner iteration]. Find $x_k \in \mathcal{B}$ such that

$$\sigma_k \le \omega_k. \tag{3.1}$$

If

$$\|c(x_k)\| \le \eta_k, \tag{3.2}$$

execute Step 2. Otherwise, execute Step 3.

Step 2 [Test for convergence and update Lagrange multiplier estimates]. If $\sigma_k \le \omega_*$ and $\|c(x_k)\| \le \eta_*$, stop. Otherwise, set

$$\begin{aligned} \lambda_{k+1} &= \bar{\lambda}(x_k, \lambda_k, \mu_k), \\ \mu_{k+1} &= \mu_k, \\ \eta_{k+1} &= \eta_k \mu_{k+1}^{\beta_\eta}, \end{aligned} \tag{3.3}$$

increment k by one and go to Step 1.

Step 3 [Reduce the penalty parameter]. Set

$$\begin{aligned} \lambda_{k+1} &= \lambda_k, \\ \mu_{k+1} &= \tau \mu_k, \\ \eta_{k+1} &= \mu_{k+1}^{\alpha_\eta}, \end{aligned} \tag{3.4}$$

increment k by one and go to Step 1.

Algorithm 3.2

Step 0 [Initialization]. An initial vector of Lagrange multiplier estimates, λ_0, is given. The nonnegative constant α_η and the positive constants $\mu_s < 1$, $\tau < 1$, $\gamma < 1$, $\omega_* \ll 1$, $\eta_* \ll 1$, and β_η are specified. Set $\mu_0 = \mu_s$, $\eta_0 = \mu_0^{\alpha_\eta}$, and $k = 0$.

Step 1 [Inner iteration]. Find $x_k \in \mathcal{B}$ such that

$$\sigma_k \le \omega_k. \tag{3.5}$$

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 $\sigma_k \le \omega_*$ and $\|c(x_k)\| \le \eta_*$, stop. Otherwise, set

$$\begin{aligned} \mu_{k+1} &= \mu_k, \\ \lambda_{k+1} &= \begin{cases} \hat{\lambda}_{k+1} & \text{if } \|\hat{\lambda}_{k+1}\| \le \mu_{k+1}^{-\gamma}, \\ \lambda_k & \text{otherwise,} \end{cases} \\ \eta_{k+1} &= \eta_k \mu_{k+1}^{\beta_\eta}, \end{aligned} \tag{3.7}$$

increment k by one and go to Step 1.

Step 3 [Reduce the penalty parameter and update Lagrange multiplier estimates]. Set

$$\begin{aligned} \mu_{k+1} &= \tau \mu_k, \\ \lambda_{k+1} &= \begin{cases} \hat{\lambda}_{k+1} & \text{if } \|\hat{\lambda}_{k+1}\| \le \mu_{k+1}^{-\gamma}, \\ \lambda_k & \text{otherwise,} \end{cases} \\ \eta_{k+1} &= \mu_{k+1}^{\alpha_\eta}, \end{aligned} \tag{3.8}$$

increment k by one and go to Step 1.

As ω_k tends to zero, the inner iteration of Step 1 may thus be interpreted as the (increasingly accurate) minimization of the augmented Lagrangian function (1.4) subject to the linear inequality constraints (1.3). It is important to note that the algorithms considered here and those of Conn *et al.* (1991) are very similar: they essentially differ in the choice of the criticality test in (3.1) and (3.5) and in the manner in which ω_k tends to zero.

The motivation of both algorithms is quite simple and similar to that presented by Conn *et al.* (1991). In the worst case, global convergence is ensured by driving the penalty parameter to zero, in which case the algorithms essentially reduce to the quadratic penalty function method (see, for example, Gould (1989)). The tests (3.2) and (3.6) are designed to allow the multiplier updates of Step 2 to take over in the neighbourhood of a stationary point.

The algorithms differ, as in Conn *et al.* (1991), 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 (2.10).

The restriction $\mu_s < 1$ is imposed in order to simplify the exposition. In a more practical setting, it may be ignored provided one defines a quantity α_k by

$$\alpha_k = \min(\gamma_s, \mu_k), \tag{3.9}$$

for all k and for some constant $\gamma_s \in (0,1)$. This quantity then plays the role of the penalty parameter in the update to η_k. Algorithms 3.1 and 3.2 may be extended in other ways. For instance, one may replace the definition of η_0 and the third equation in (3.4)/(3.8) by

$$\eta_0 = \eta_s \alpha_0^{\alpha_\eta} \text{ and } \eta_{k+1} = \eta_s \alpha_{k+1}^{\alpha_\eta}, \tag{3.10}$$

for some $\eta_s > 0$, where we have used the quantity α_k introduced above. Finally, the acceptance test for λ_{k+1} in (3.7) and (3.8) may be replaced by

$$\|\hat{\lambda}_{k+1}\| \leq \nu \mu_{k+1}^{-\gamma} \tag{3.11}$$

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

As noted above, the definition of σ_k is identical to that of the criticality measure used by Conn *et al.* (1993a) in their algorithm for the solution of nonlinear optimization problems with convex constraints. As the iterates of this algorithm always stay feasible, this measure is well defined. Hence we can use this latter algorithm to solve the subproblem of Step 1 in both Algorithms 3.1 and 3.2.

The proposed algorithms use a number of parameters. The values used in the LANCELOT package in a similar context are $\alpha_\eta = \mu_s = \tau = \gamma_s = 0.1$, $\beta_\eta = 0.9$ (relation (3.10) is also used with $\eta_s = 0.12589$, ensuring that $\eta_0 = 0.01$). The values $\gamma = 0.9$ and $\nu = 1$ also seem suitable. The parameters ω_* and η_* specify the final accuracy requested by the user.

4 Global convergence analysis

We now proceed to show that both algorithms are globally convergent under the following assumptions.

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

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.

Assumption AS4 guarantees that the dimension of the nullspace of A is large enough to provide the number of degrees of freedom that are necessary for satisfying the nonlinear constraints, whose gradients (projected onto this nullspace) are assumed to be linearly independent at every limit point of the sequence of iterates. This assumption is the direct generalization of AS3 used by Conn *et al.* (1991).

We shall analyse the convergence of the algorithms of Section 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. These results show that the Lagrange multiplier estimates generated by either algorithm cannot behave too badly.

Lemma 4.1 [Conn *et al.* (1991), Lemma 4.1] *Suppose that μ_k converges to zero as k increases when Algorithm 3.1 is executed. Then the product $\mu_k\|\lambda_k\|$ converges to zero.*

Lemma 4.2 [Conn *et al.* (1991), Lemma 4.2] *Suppose that μ_k converges to zero as k increases when Algorithm 3.2 is executed. Then the product $\mu_k\|\lambda_k\|$ converges to zero.*

Both these lemmas are valid in our context because their proofs do not involve the differences between the algorithms of Conn *et al.* (1991) and those described above, namely the inner iteration termination criteria, (3.1) and (3.5), and the exact manner in which ω_k tends to zero.

We also need the following lemma, proving that the reduced gradient of the augmented Lagrangian may not be arbitrarily large when σ_k is not.

Lemma 4.3 *Let $\{x_k\}, k \in \mathcal{K}$, be a sequence which converges to the point x_* and suppose that $\sigma_k \le \omega_k$, where the ω_k are positive scalar parameters. Then, there is a positive constant κ_1 and an integer k_0 such that*

$$\|Z_*^T \nabla_x \Phi_k\| \le \kappa_1 \omega_k \tag{4.1}$$

for all $k \ge k_0, (k \in \mathcal{K})$.

Proof. The result obviously holds if $\|Z_*^T \nabla_x \Phi_k\| = 0$. Otherwise, we define the set

$$\mathcal{I}(x) \stackrel{\text{def}}{=} \{i \in \{1, \ldots, p\} \mid a_i^T x - b_i > 0\} \tag{4.2}$$

where a_i^T defines the i-th row of the matrix A and b_i the i-th component of the vector b. Considering the continuous function $\beta(\cdot)$ defined by

$$\beta(x) \stackrel{\text{def}}{=} \min_{i \in \mathcal{I}(x_*)} \{a_i^T x - b_i\}, \tag{4.3}$$

we have that $\beta_* \stackrel{\text{def}}{=} \beta(x_*)$ is strictly positive. This implies the inclusion

$$\mathcal{I}(x_*) \subseteq \mathcal{I}(x_k) \tag{4.4}$$

for all $k \in \mathcal{K}$ large enough. Indeed, this inclusion immediately follows if $\mathcal{I}(x_*)$ is empty. Otherwise, suppose by contradiction that there exists $i \in \mathcal{I}(x_*)$ such that $i \notin \mathcal{I}(x_k)$. We then have the following inequalities

$$\begin{array}{rcl} 0 < \beta_* & \leq & a_i^T x_* - b_i \\ & = & a_i^T x_* - b_i - (a_i^T x_k - b_i) \\ & = & a_i^T (x_* - x_k) \\ & \leq & \|a_i\| \|x_* - x_k\| \\ & \leq & \delta \|x_* - x_k\|, \end{array}$$

where

$$\delta \stackrel{\text{def}}{=} \max_{i \in \mathcal{I}(x_*)} \|a_i\| > 0, \tag{4.5}$$

which contradicts the convergence of the sequence $\{x_k\}$ to x_*. From (4.4), we deduce that

$$\beta_k \stackrel{\text{def}}{=} \beta(x_k) > 0 \tag{4.6}$$

for all $k \in \mathcal{K}$ sufficiently large.

We now define the vector

$$p_k \stackrel{\text{def}}{=} \epsilon_k \frac{Z_* Z_*^T \nabla_x \Phi_k}{\|Z_* Z_*^T \nabla_x \Phi_k\|}, \tag{4.7}$$

where

$$0 < \epsilon_k = \min\{1, \beta_k / \delta\}, \tag{4.8}$$

which is well defined because of (4.5). We then show that $-p_k$ is feasible for the minimization problem associated with σ_k in (2.8), that is,

$$a_i^T (x_k - p_k) - b_i \geq 0 \quad i = 1, \ldots, p \tag{4.9}$$

and

$$\|p_k\| \leq 1. \tag{4.10}$$

The last inequality directly follows from the definitions (4.7) and (4.8). We thus prove (4.9) only. Assume first that $i \notin \mathcal{I}(x_*)$. Then the definition of the matrix Z_* yields the desired inequality, since

$$\begin{array}{rcl} a_i^T (x_k - p_k) - b_i & = & a_i^T x_k - \epsilon_k a_i^T \frac{Z_* Z_*^T \nabla_x \Phi_k}{\|Z_* Z_*^T \nabla_x \Phi_k\|} - b_i \\ & = & a_i^T x_k - b_i \\ & \geq & 0. \end{array}$$

If, on the other hand, $i \in \mathcal{I}(x_*)$, we then obtain from (4.5), (4.7) and (4.8) that

$$a_i^T p_k \leq \|a_i\| \|p_k\| = \|a_i\| \epsilon_k \leq \beta_k. \tag{4.11}$$

Since $i \in \mathcal{I}(x_*)$, we also have that

$$\beta_k \leq a_i^T x_k - b_i. \tag{4.12}$$

Inequality (4.9) then follows from (4.11) and (4.12). Since $-p_k$ is a feasible direction for problem (2.8), using the assumption $\sigma_k \leq \omega_k$, we deduce that

$$|\nabla_x \Phi_k^T p_k| \leq \sigma_k \leq \omega_k. \tag{4.13}$$

By definition of p_k and since the matrix Z_* is orthogonal, we also have that

$$|\nabla_x \Phi_k^T p_k| = \epsilon_k \|Z_*^T \nabla_x \Phi_k\|. \tag{4.14}$$

Inequality (4.1) then follows directly from (4.13) and (4.14) with $\kappa_1 = 1$ when $\epsilon_k = 1$. If $\epsilon_k = \beta_k/\delta$, using the continuity of the function $\beta(\cdot)$, we obtain from (4.13) and (4.14) that

$$\|Z_*^T \nabla_x \Phi_k\| \leq \frac{\delta}{\beta_k}\omega_k \leq \frac{2\delta}{\beta_*}\omega_k$$

for $k \in \mathcal{K}$ sufficiently large. Inequality (4.1) thus follows for

$$\kappa_1 \stackrel{\text{def}}{=} \max\{1, 2\delta/\beta_*\}. \tag{4.15}$$

□

We now examine the behaviour of the sequence $\{\nabla_x \Phi_k\}$. We first recall a result extracted from the classical perturbation theory of convex optimization problems. This result is well known and can be found, for instance, on pp.14–17 of Fiacco (1983).

Lemma 4.4 *Assume that D is a continuous point-to-set mapping from $S \subseteq \Re^\ell$ into $\Re^n$ such that the set $D(\theta)$ is convex and non-empty for each $\theta \in S$. Assume that the real-valued function $F(y,\theta)$ is defined and continuous on the space $\Re^n \times S$ and convex in y for each fixed θ. Then, the real-valued function F_* defined by*

$$F_*(\theta) \stackrel{\text{def}}{=} \inf_{y \in D(\theta)} F(y,\theta) \tag{4.16}$$

is continuous on S.

We now show that, if it converges, the sequence $\{\nabla_x \Phi_k\}$ tends to a vector which is a linear combination of the rows of A_* with non-negative coefficients.

Lemma 4.5 *Let $\{x_k\}, k \in \mathcal{K}$, be a sequence which converges to the point x_* and suppose that the gradients $\nabla_x \Phi_k$, $k \in \mathcal{K}$, converge to some limit $\nabla_x \Phi_*$. Assume furthermore that σ_k approaches zero as $k \in \mathcal{K}$ increases. Then,*

$$\nabla_x \Phi_* = A_*^T \pi_* \tag{4.17}$$

for some vector $\pi_* \geq 0$, where A_* is the matrix whose rows are those of A corresponding to active constraints at x_*.

Proof. Consider the minimization problem

$$\begin{array}{ll} \min_{d \in \Re^n} & \nabla_x \Phi_*^T d, \\ \text{subject to} & A(x_* + d) - b \geq 0, \\ & \|d\| \leq 1. \end{array} \tag{4.18}$$

Since the sequences $\{\nabla_x \Phi_k\}$ and $\{x_k\}$ converge to $\nabla_x \Phi_*$ and x_* respectively, we deduce from Lemma 4.4 applied to the optimization problem (2.8) (with the choices $\theta^T = (\nabla_x \Phi^T, x^T)$, $D(\theta) = \{d \mid A(x+d) - b \geq 0,\ \|d\| \leq 1\}$, $y = d$, $F(y,\theta) = \nabla_x \Phi^T d$), and the convergence of the sequence σ_k to zero that the optimal value for problem (4.18) is zero. The vector $d = 0$ is thus a solution for problem (4.18) and satisfies

$$\nabla_x \Phi_* = A_*^T \pi_* - 2\zeta d = A_*^T \pi_* \tag{4.19}$$

for some vector $\pi_* \geq 0$, which ends the proof. □

The important part of our convergence analysis is the next lemma. It allows us to apply the theory developed in Conn *et al.* (1991) to deduce the desired convergence properties.

Lemma 4.6 *Suppose that AS1 and AS2 hold. Let $\{x_k\} \in \mathcal{B}, k \in \mathcal{K}$, be a sequence satisfying AS3 which converges to the point x_* for which AS4 holds and let $\lambda_* = \lambda(x_*)$, where λ satisfies (2.10). Assume that $\{\lambda_k\}$, $k \in \mathcal{K}$, is any sequence of vectors and that $\{\mu_k\}$, $k \in \mathcal{K}$, form a nonincreasing sequence of positive scalars. Suppose further that*

$$\sigma_k \leq \omega_k \tag{4.20}$$

where the ω_k are positive scalar parameters which converge to zero as $k \in \mathcal{K}$ increases. Then

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

$$\|\bar{\lambda}(x_k, \lambda_k, \mu_k) - \lambda_*\| \leq \kappa_2 \omega_k + \kappa_3 \|x_k - x_*\|, \tag{4.21}$$

$$\|\lambda(x_k) - \lambda_*\| \leq \kappa_3 \|x_k - x_*\|, \tag{4.22}$$

and

$$\|c(x_k)\| \leq \kappa_2 \omega_k \mu_k + \mu_k \|\lambda_k - \lambda_*\| + \kappa_3 \mu_k \|x_k - x_*\| \tag{4.23}$$

for all $k \geq k_1, (k \in \mathcal{K})$.

Suppose, in addition, that $c(x_*) = 0$. Then

(ii) 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 $\{\bar{\lambda}(x_k, \lambda_k, \mu_k)\}$ and $\{\lambda(x_k)\}$ converge to λ_* for $k \in \mathcal{K}$;

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

Proof. As a consequence of AS2–AS4, we have that for $k \in \mathcal{K}$ sufficiently large, $(J(x_k)Z_*)^+$ exists, is bounded and converges to $(J(x_*)Z_*)^+$. Thus, we may write

$$\|((J(x_k)Z_*)^+)^T\| \leq \kappa_0 \tag{4.24}$$

for some constant $\kappa_0 > 0$. Equations (2.6) and (2.7), the inner iteration termination criterion (4.20) and Lemma 4.3 give that

$$\|Z_*^T (g(x_k) + J(x_k)^T \bar{\lambda}_k)\| \leq \kappa_1 \omega_k \tag{4.25}$$

for all $k \in \mathcal{K}$ large enough, where κ_1 is defined in (4.15). By assumptions AS2, AS3, AS4 and (2.10), $\lambda(x)$ is bounded for all x in a neighbourhood of x_*. Thus we may deduce from (2.10), (4.24) and (4.25) that

$$\begin{aligned} \|\bar{\lambda}_k - \lambda(x_k)\| &= \|((J(x_k)Z_*)^+)^T Z_*^T g(x_k)\bar{\lambda}_k\| \\ &= \|((J(x_k)Z_*)^+)^T (Z_*^T g(x_k) + (J(x_k)Z_*)^T \bar{\lambda}_k)\| \\ &\leq \|((J(x_k)Z_*)^+)^T\| \kappa_1 \omega_k \\ &\leq \kappa_2 \omega_k, \end{aligned} \tag{4.26}$$

where $\kappa_2 = \kappa_1 \kappa_0$ is a positive constant. Moreover, from the integral mean value theorem and Lemma 2.1 we have that

$$\lambda(x_k) - \lambda(x_*) = \int_0^1 \nabla_x \lambda(x(s)) ds \cdot (x_k - x_*), \tag{4.27}$$

where $\nabla_x \lambda(x)$ is given by equation (2.12), 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 (4.27) gives

$$\|\lambda(x_k) - \lambda_*\| \leq \kappa_3 \|x_k - x_*\| \tag{4.28}$$

for all $k \in \mathcal{K}$ sufficiently large and for some constant $\kappa_3 > 0$, which is just the inequality (4.22). We then have that $\lambda(x_k)$ converges to λ_*. Combining (4.26) and (4.28) we obtain

$$\|\bar{\lambda}_k - \lambda_*\| \leq \|\bar{\lambda}_k - \lambda(x_k)\| + \|\lambda(x_k) - \lambda_*\| \leq \kappa_2 \omega_k + \kappa_3 \|x_k - x_*\|, \tag{4.29}$$

the required inequality (4.21). Then, since by construction ω_k tends to zero as k increases, (4.21) implies that $\bar{\lambda}_k$ converges to λ_* and therefore, from the identity (2.6), $\nabla_x \Phi_k$ converges to $g^\ell(x_*, \lambda_*)$. Furthermore, multiplying (2.4) by μ_k, we obtain

$$c(x_k) = \mu_k((\bar{\lambda}_k - \lambda_*) + (\lambda_* - \lambda_k)). \tag{4.30}$$

Taking norms of (4.30) and using (4.29) we derive (4.23).

Now suppose that

$$c(x_*) = 0. \tag{4.31}$$

Lemma 4.5 and the convergence of $\nabla_x \Phi_k$ to $g^\ell(x_*, \lambda_*)$ give that

$$g(x_*) + J(x_*)^T \lambda_* = A_*^T \pi_* \tag{4.32}$$

for some vector $\pi_* \geq 0$. This last equation and (4.31) show that x_* is a Kuhn-Tucker point and λ_* is the corresponding set of Lagrange multipliers. Moreover (4.21) and (4.22) ensure the convergence of the sequences $\{\bar{\lambda}(x_k, \lambda_k, \mu_k)\}$ and $\{\lambda(x_k)\}$ to λ_* for $k \in \mathcal{K}$. Hence the lemma is proved. $\square$

Now that we have the counterpart of Lemma 4.3 in Conn *et al.* (1991), we can derive the desired global convergence property of Algorithms 3.1 and 3.2.

Theorem 4.7 [Conn *et al.* (1991), Theorem 4.4] *Assume that AS1 and AS2 hold. Let x_* be any limit point of the sequence $\{x_k\}$ generated by Algorithm 3.1 or by Algorithm 3.2 of Section 3 for which AS3 and AS4 hold and let $\mathcal{K}$ be the set of indices of an infinite subsequence of the x_k whose limit is x_*. Finally, let $\lambda_* = \lambda(x_*)$. Then conclusions (i), (ii) and (iii) of Lemma 4.6 hold.*

The proof of this theorem is identical to that provided in Conn *et al.* (1991), and uses Lemmas 4.1 and 4.2.

5 Alternative criticality measures

In Algorithms 3.1 and 3.2, we used the criticality measure σ_k in order to define the stopping criterion of the inner iteration (see (3.1) and (3.5)), partly because it is general and partly to preserve coherence with the framework presented by Conn *et al.* (1993a). However, this quantity might not be easily computed in the course of the numerical method used to calculate x_k, especially when the dimension of the problem is high. It is therefore of interest to examine other criticality measures that might be easier to calculate. It is the purpose of this section to analyze such alternative proposals.

The first proposal is based on the identification of the linear constraints that are "dominant" at x_k (even though they might not be active) and on a measure of criticality for the part of the problem where those dominant constraints are irrelevant. We first define, for a vector x_k satisfying (1.3), the set of *dominant constraints at* x_k as the constraints whose index is in the set

$$D_k = \{i \in \{1, \ldots, p\} \mid a_i^T x_k - b_i \leq \kappa_4 \omega_k\}, \tag{5.1}$$

for some $\kappa_4 > 0$, and where $a_i^T \in \Re^n$ is the i-th row of the matrix A and b_i the corresponding component of the right-hand-side vector b. Denoting by A_{D_k} the submatrix of A consisting of the row(s) whose index is in D_k, we also define

$$N_k = \{A_{D_k}^T \xi \mid \xi \in \Re^{|D_k|} \text{ and } \xi_i \leq 0 \quad (i = 1, \ldots, |D_k|)\}, \tag{5.2}$$

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

$$T_k = N_k^0 = \text{cl}\{\nu d \mid \nu \geq 0 \text{ and } d^T v \leq 0 \text{ for all } v \in N_k\}, \tag{5.3}$$

where cl$\{S\}$ denotes the closure of the set S. The cone T_k is the tangent cone with respect to the dominant constraints at x_k. Note that D_k might be empty, in which case A_{D_k} is assumed to be zero, N_k reduces to the origin and T_k is the full space.

Given the tangent cone T_k, we consider the condition

CM2:

$$\|P_{T_k}(-\nabla_x \Phi_k)\| \leq \kappa_5 \omega_k, \tag{5.4}$$

for some $\kappa_5 \geq 0$, where $P_S(\cdot)$ is the projection onto the convex set S. For future reference, we also define Z_k and Y_k to be matrices whose columns form orthonormal bases for $\text{null}(A_{D_k})$ and $\text{span}\{a_i\}_{i \in D_k}$ respectively.

We first show that (5.4) measures criticality in a suitable way, which is compatible with the theory presented above. This requires the following simple geometric result.

Lemma 5.1 *Assume that all constraints in D_k (as defined by (5.1)) are active at some point x_*. Then*

$$\|Y_k^T(x_k - x_*)\| \leq \kappa_6 \omega_k \tag{5.5}$$

where κ_6 is some non-negative constant independent of x_k.

Proof. If D_k is empty, then Y_k is the zero matrix and (5.5) immediately follows. Assume therefore that $D_k \neq \emptyset$. We first select a submatrix $\hat{A}_{D_k}$ of A_{D_k} that is of full row-rank and note that the orthogonal projection onto the subspace spanned by the $\{a_i\}_{i \in D_k}$ is nothing but

$$Y_k Y_k^T = \hat{A}_{D_k}^T [\hat{A}_{D_k} \hat{A}_{D_k}^T]^{-1} \hat{A}_{D_k}. \tag{5.6}$$

Hence we obtain from the orthogonality of Y_k, the bound $|D_k| \leq p$, (5.1) and the fact that all constraints in D_k are active at x_*, that

$$\begin{aligned} \|Y_k^T(x_k - x_*)\| &\leq \|\hat{A}_{D_k}^T [\hat{A}_{D_k} \hat{A}_{D_k}^T]^{-1}\|.\|\hat{A}_{D_k}(x_k - x_*)\| \\ &\leq \|\hat{A}_{D_k}^T [\hat{A}_{D_k} \hat{A}_{D_k}^T]^{-1}\| p \kappa_4 \omega_k. \end{aligned} \tag{5.7}$$

But there are only a finite number of nonempty sets D_k for all possible choices of x_k, and we may thus deduce (5.5) from (5.7) by defining

$$\kappa_6 = p\kappa_4 \min \|\hat{A}_{D_k}^T [\hat{A}_{D_k} \hat{A}_{D_k}^T]^{-1}\|, \tag{5.8}$$

where the minimum is taken on all possible choices of D_k and $\hat{A}_{D_k}$. □

We use this result in the next theorem.

Theorem 5.2 *Assume that x_k satisfies the condition (5.4) (with D_k defined by (5.1)) and that the sequence $\{x_k\}$, $k \in \mathcal{K}$, belongs to $\mathcal{B}$ and converges to the limit point x_*. Then we have that*

$$\|Z_*^T \nabla_x \Phi_k\| \leq \kappa_5 \omega_k \tag{5.9}$$

for all $k \in \mathcal{K}$ sufficiently large. Furthermore, if the sequence $\{\|\nabla_x \Phi_k\|\}$ is bounded, then

$$\sigma_k \leq \kappa_7 \omega_k. \tag{5.10}$$

for some constant $\kappa_7 \geq 0$ and all $k \in \mathcal{K}$ sufficiently large.

Proof. Observe that, for $k \in \mathcal{K}$ sufficiently large , ω_k is sufficiently small and x_k sufficiently close to x_* to ensure that all the constraints in D_k are active at x_*. This implies that the subspace orthogonal to the normals of the dominant constraints at x_k, $\mathcal{V}_k$ say, contains the subspace orthogonal to the normals of the constraints active at x_*. Hence, we deduce that

$$\|Z_*^T \nabla_x \Phi_k\| \le \|Z_k^T \nabla_x \Phi_k\| \le \|P_{T_k}(-\nabla_x \Phi_k)\| \le \kappa_5 \omega_k, \tag{5.11}$$

where we have used the fact that $\mathcal{V}_k \subseteq T_k$ to obtain the second inequality and (5.4) to deduce the third. This proves (5.9). We now prove (5.10). We obtain from the definition of σ_k in (2.8), the Moreau decomposition (see Moreau (1962)) of $\nabla_x \Phi_k$ and the Cauchy-Schwarz inequality, that

$$\begin{aligned} \sigma_k &= \max_{\substack{A(x_k+d)-b\ge 0 \\ \|d\|\le 1}} \left(-\nabla_x {\Phi_k}^T d\right) \\ &\le \max_{\substack{A(x_k+d)-b\ge 0 \\ \|d\|\le 1}} P_{T_k}(-\nabla_x \Phi_k)^T d + \max_{\substack{A(x_k+d)-b\ge 0 \\ \|d\|\le 1}} P_{N_k}(-\nabla_x \Phi_k)^T d \\ &\le \|P_{T_k}(-\nabla_x \Phi_k)\| + \max_{d\in B_k} P_{N_k}(-\nabla_x \Phi_k)^T d, \end{aligned} \tag{5.12}$$

where $B_k \stackrel{\rm def}{=} \{d \in \Re^n \mid a_i^T(x_k+d) - b \ge 0 \;\; (i \in D_k) \;\; \text{and} \;\; \|d\| \le 1\}$. Defining Y_k as above, we obtain that

$$P_{N_k}(-\nabla_x \Phi_k)^T d = [Y_k Y_k^T P_{N_k}(-\nabla_x \Phi_k)]^T d = P_{N_k}(-\nabla_x \Phi_k)^T Y_k Y_k^T d. \tag{5.13}$$

As, for x_k sufficiently close to x_*, all the constraints in D_k must be active at x_*, we have that N_k is included in the normal cone at x_* and therefore the vector $P_{N_k}(-\nabla_x \Phi_k)$ belongs to this normal cone. Moreover, since the maximization problem of the last right-hand-side of (5.12) is a convex program, since x_* is feasible for this problem, and since $\|x_* - x_k\| \le 1$ for $k \in \mathcal{K}$ large enough, we thus deduce that $d = x_* - x_k$ is a global solution of this program, which yields, using the Cauchy-Schwarz inequality, that

$$\max_{d\in B_k} P_{N_k}(-\nabla_x \Phi_k)^T Y_k Y_k^T d \le \|P_{N_k}(-\nabla_x \Phi_k)\|.\|Y_k^T(x_k - x_*)\|. \tag{5.14}$$

We may then apply Lemma 5.1 and the contractive character of the projection onto a convex set to deduce from (5.13) and (5.14) that

$$\max_{d\in B_k} P_{N_k}(-\nabla_x \Phi_k)^T d \le \|P_{N_k}(-\nabla_x \Phi_k)\|.\|Y_k^T(x_k - x_*)\| \le \kappa_6 \omega_k \|\nabla_x \Phi_k\|. \tag{5.15}$$

Then substituting (5.4) and this last inequality in the last right-hand-side of (5.12) and finally using our boundedness assumption on $\|\nabla_x \Phi_k\|$, we obtain (5.10) for

$$\kappa_7 = \kappa_5 + \kappa_6 \max\{\|\nabla_x \Phi_k\|\}. \tag{5.16}$$

□

We now claim that the test (5.4) can be used in Algorithms 3.1 and 3.2 without modifying the conclusions of the convergence theory developed above.

Theorem 5.3 *Assume that AS1 and AS2 hold. Let x_* be any limit point of the sequence $\{x_k\}$ generated by Algorithm 3.1 or by Algorithm 3.2 of Section 3, where the tests (3.1) and (3.5) are replaced by (5.4). Assume furthermore that AS3 and AS4 hold and let $\mathcal{K}$ be the set of indices of an infinite subsequence of the x_k whose limit is x_*. Finally, let $\lambda_* = \lambda(x_*)$. Then conclusions (i), (ii) and (iii) of Lemma 4.6 hold.*

Proof. Consider Algorithm 3.1 first. To prove our result, we successively revisit the two occurrences of the criticality measure and test (3.1) in the theory presented above. The first occurrence is in Lemma 4.3, where (3.1) is used to ensure (4.1). But this latter condition is identical to (5.9), which is obtained when using (5.4), as proved in Theorem 5.2. The second occurrence is in Lemma 4.5. However, the assumptions of this lemma imply that the sequence $\{\|\nabla_x \Phi_k\|\}$ is bounded, and we may thus apply the second part of Theorem 5.2 to deduce (5.10), which is all we need to prove Lemma 4.5, given that ω_k tends to zero. The test (4.20) is only used in Lemma 4.6 in order to apply Lemmas 4.3 and 4.5. Hence we obtain the desired conclusion for the modified version of Algorithm 3.1. The same arguments hold for the replacement of (3.5) by (5.4) in Algorithm 3.2. □

Having shown that (5.4) can be used in our algorithms instead of (3.1) or (3.5), we now turn to a specialization of this rule. Given D_k, N_k, and A_{D_k} as above, we claim that (3.1) (and (3.5)) can be replaced by the requirement that there exists a set of non-positive "dominant multipliers" $\{\xi_{ik}\}_{i\in M_k}$ ($M_k \subseteq D_k$, $\xi_{ik} \leq 0$) such that

CM3:

$$\|\nabla_x \Phi_k + A_{D_k}^T \xi_k\| \leq \kappa_5 \omega_k, \tag{5.17}$$

where ξ_k is the $|D_k|$-dimensional vector whose i-th component is ξ_{ik} if $i \in M_k$ or zero otherwise, and where κ_5 is a non-negative constant. We prove this claim.

Lemma 5.4 *Assume that there exists a non-positive ξ_k such that (5.17) holds at x_k (with D_k defined by (5.1)). Then (5.4) also holds at x_k.*

Proof. Since the vector $A_{D_k}^T \xi_k$ belongs, by construction, to the cone N_k defined in (5.2), we can immediately deduce from (5.17) that

$$\|P_{T_k}(-\nabla_x \Phi_k)\| = \| - \nabla_x \Phi_k - P_{N_k}(-\nabla_x \Phi_k)\| \leq \| - \nabla_x \Phi_k - A_{D_k}^T \xi_k\| \leq \kappa_5 \omega_k, \tag{5.18}$$

which is the desired inequality. □

Condition (5.17) is interesting for two reasons. Firstly, a set of (possibly approximate) multipliers is available in many numerical procedures that could be used for performing the inner iteration and computing a suitable x_k; one can then select those corresponding to the dominant constraints, further restrict this choice to the non-positive ones and finally check (5.17). Alternatively, suitable multipliers can be computed, for instance by (approximately) solving the least-squares problem

$$\min_{\xi} \|\nabla_x \Phi_k + A_{D_k}^T \xi\| \tag{5.19}$$

and selecting the non-positive components of the resulting vector ξ, or by (approximately) solving the constrained least-squares problem

$$\min_{\xi \leq 0} \|\nabla_x \Phi_k + A_{D_k}^T \xi\|. \tag{5.20}$$

The second interest of (5.17) is to provide in a single condition both a stopping condition on the inner iteration and a measure of the tolerated "inexactness" in solving the associated least-squares problem, if this is the procedure chosen to obtain the dominant multipliers.

We may therefore deduce global convergence for our algorithms whenever (5.17) is used instead of (3.1)/(3.5), as stated in the next theorem.

Theorem 5.5 *Assume that AS1 and AS2 hold. Let x_* be any limit point of the sequence $\{x_k\}$ generated by Algorithm 3.1 or by Algorithm 3.2 of Section 3, where the tests (3.1) and (3.5) are replaced by (5.17). Assume furthermore that AS3 and AS4 hold and let $\mathcal{K}$ be the set of indices of an infinite subsequence of the x_k whose limit is x_*. Finally, let $\lambda_* = \lambda(x_*)$. Then conclusions (i), (ii) and (iii) of Lemma 4.6 hold.*

Condition (5.17) can be further specialized. For instance, one might choose to impose the familiar "reduced gradient" criterion

CM4:

$$\|Z(x_k)^T \nabla_x \Phi_k\| \leq \kappa_5 \omega_k, \tag{5.21}$$

where $Z(x_k)$ is an orthogonal matrix whose columns span the nullspace of the constraints active at x_k, provided that the multipliers associated with these linear constraints are all non-positive. In this case, we have that

$$\|P_{T_k}(-\nabla_x \Phi_k)\| \leq \|P_{T(x_k)}(-\nabla_x \Phi_k)\| = \|Z(x_k)^T \nabla_x \Phi_k\| \leq \kappa_5 \omega_k, \tag{5.22}$$

because $T(x_k)$, the tangent cone to the set determined by the linear inequality constraints active at x_k, contains T_k. As a consequence, global convergence is still obtained when this criterion, implemented by several subroutines for minimizing a (nonlinear) objective function subject to linear constraints (LSNNO by Toint and Tuyttens (1992), or `VE09`, `VE14` and `VE19` of the Harwell Subroutine Library, for instance), is used as an inner iteration stopping criterion in Algorithm 3.1 or 3.2.

Theorem 5.6 *Assume that AS1 and AS2 hold. Let x_* be any limit point of the sequence $\{x_k\}$ generated by Algorithm 3.1 or by Algorithm 3.2 of Section 3, where the tests (3.1) and (3.5) are replaced by (5.21) and the condition that the multipliers associated with the linear constraints active at x_k are all non-positive. Assume furthermore that AS3 and AS4 hold and let $\mathcal{K}$ be the set of indices of an infinite subsequence of the x_k whose limit is x_*. Finally, let $\lambda_* = \lambda(x_*)$. Then conclusions (i), (ii) and (iii) of Lemma 4.6 hold.*

We conclude this section on alternative criticality measures by showing that a strengthened version of criterion (3.1) (or (3.5)) implies (5.4) for a convergent algorithm.

Theorem 5.7 *Assume that $\{x_k\}$, $k \in \mathcal{K}$, is a convergent subsequence of vectors of $\mathcal{B}$ such that*

$$\sigma_k \leq {\omega_k}^2, \tag{5.23}$$

for each $k \in \mathcal{K}$, where the ω_k converge to zero as k increases in $\mathcal{K}$. Then the inequality (5.4) also holds for each $k \in \mathcal{K}$ sufficiently large.

Proof. We first consider the simple case where $p = 0$, that is when no linear inequality is present. In this case, it is easy to check from (2.8) that $\sigma_k = \|\nabla_x \Phi_k\|$. But we must have that $D_k = \emptyset$. Thus $\sigma_k = \|P_{T_k}(-\nabla_x \Phi_k)\|$. We therefore obtain that (5.4) holds with $\kappa_5 = 1$ and k large enough to ensure that $\omega_k \leq 1$.

Assume now that $p > 0$. The Moreau decomposition of $-\nabla_x \Phi_k$ (see Moreau (1962)) is given by

$$-\nabla_x \Phi_k = P_{T_k}(-\nabla_x \Phi_k) + P_{N_k}(-\nabla_x \Phi_k). \tag{5.24}$$

If $P_{T_k}(-\nabla_x \Phi_k)$ is zero, then (5.4) obviously holds for any choice of $\kappa_5 \geq 0$. Assume therefore that $P_{T_k}(-\nabla_x \Phi_k)$ is nonzero. We now show that $x_k + d_k \in \mathcal{B}$, where we define

$$d_k \stackrel{\text{def}}{=} \epsilon_k \frac{P_{T_k}(-\nabla_x \Phi_k)}{\|P_{T_k}(-\nabla_x \Phi_k)\|}, \quad \text{with} \quad \epsilon_k \stackrel{\text{def}}{=} \min\left[1, \frac{\kappa_4 \omega_k}{\|A\|_\infty}\right]. \tag{5.25}$$

Assume first that $i \in D_k$. Then $-a_i \in N_k$ and $a_i^T d_k \geq 0$ because of the polarity of N_k and T_k. Since $x_k \in \mathcal{B}$, we obtain that

$$a_i^T(x_k + d_k) - b_i = (a_i^T x_k - b_i) + a_i^T d_k \geq 0. \tag{5.26}$$

On the other hand, if $i \notin D_k$, we have that $a_i^T x_k - b_i > \kappa_4 \omega_k$ and hence

$$(a_i^T x_k - b_i) + a_i^T d_k > \kappa_4 \omega_k - \|a_i\|.\|d_k\| = \kappa_4 \omega_k - \epsilon_k \|a_i\| \geq \kappa_4 \omega_k - \kappa_4 \omega_k = 0. \tag{5.27}$$

Gathering (5.26) and (5.27), we obtain that $x_k + d_k \in \mathcal{B}$, as desired. Furthermore, since $\|d_k\| \leq 1$ by definition, we have verified that d_k is feasible for the minimization problem (2.8) associated with the definition of σ_k. Hence,

$$\begin{aligned} \sigma_k &\geq -\nabla_x \Phi_k{}^T d_k \\ &= P_{T_k}(-\nabla_x \Phi_k)^T d_k + P_{N_k}(-\nabla_x \Phi_k)^T d_k \\ &= \|P_{T_k}(-\nabla_x \Phi_k)^T\| \, \|d_k\| \\ &= \epsilon_k \|P_{T_k}(-\nabla_x \Phi_k)^T\|, \end{aligned} \tag{5.28}$$

where we have used successively the Moreau decomposition of $-\nabla_x \Phi_k$, the definition of d_k and the orthogonality of the terms in the Moreau decomposition. If $\epsilon_k = 1$, then (5.23) and (5.28) imply that

$$\|P_{T_k}(-\nabla_x \Phi_k)^T\| \leq \omega_k{}^2 \leq \omega_k \tag{5.29}$$

for $k \in \mathcal{K}$ sufficiently large. Otherwise, we deduce from (5.28), (5.23) and (5.25) that

$$\|P_{T_k}(-\nabla_x \Phi_k)^T\| \leq \frac{\|A\|_\infty}{\kappa_4} \omega_k. \tag{5.30}$$

As a consequence of (5.29) and (5.30), we therefore obtain that (5.4) holds with

$$\kappa_5 = \max\left[1, \frac{\|A\|_\infty}{\kappa_4}\right]. \tag{5.31}$$

Combining all cases, we conclude that (5.4) holds with this last value of κ_5. □

Of course, condition (5.23) implies (3.1) for k large enough, because ω_k tends to zero. Theorems 5.2 and 5.7 thus show the strong relationships between (5.9), (5.10) (5.23) and (5.4).

We finally note that Lemmas 5.1 and 5.4 as well as Theorems 5.2 and 5.7 do not depend on the actual form of the augmented Lagrangian (1.4), but are valid independently of the function minimized in the inner iteration. This observation can be useful if alternative techniques for augmenting the Lagrangian are considered for a merit function.

6 Partitioning the set of general equality constraints

We conclude this paper by a section devoted to an extension of our framework, with the intention of providing additional algorithmic flexibility. This will enable us to account for non-uniform changes, perhaps as a result of poor scaling, in the behaviour of (sets of) general constraints. One could of course scale the equality constraints a priori, or include suitable scaling matrices in the algorithm, as in Conn *et al.* (1991). However, the first of these approaches only provides static scaling, while the scaling matrices in the second are often difficult to determine, especially in a dynamic manner. In this section, we instead consider the possibility of assigning different penalty parameters to different groups of constraints. More precisely, we partition the set of constraints (1.2) into q disjoint subsets $\{\mathcal{Q}_j\}_{j=1}^q$, and redefine the augmented Lagrangian as

$$\Phi(x,\lambda,\mu) = f(x) + \sum_{j=1}^{q}\sum_{i\in\mathcal{Q}_j}\left[\lambda_i c_i(x) + \frac{1}{2\mu_j}c_i(x)^2\right], \tag{6.1}$$

where μ is now a q-dimensional vector, whose j-th component is $\mu_j > 0$, the penalty parameter associated with subset $\mathcal{Q}_j$. The idea is then to update the Lagrange multipliers or the penalty parameter associated with a subset of the constraints according to the constraint violation *within that subset.* Of course, we need to redefine our first-order multiplier update componentwise as

$$\bar{\lambda}(x,\lambda_{[\mathcal{Q}_j]},\mu_j)_{[\mathcal{Q}_j]} = \lambda_{[\mathcal{Q}_j]} + c(x)_{[\mathcal{Q}_j]}/\mu_j \quad (j=1,\ldots,q), \tag{6.2}$$

where $w_{[\mathcal{S}]}$ denotes the $|\mathcal{S}|$-dimensional subvector of w whose entries are indexed by the set $\mathcal{S}$. Note that the fundamental relation $\nabla_x\Phi(x,\lambda,\mu) = g^\ell(x,\bar{\lambda}(x,\lambda,\mu))$ is still valid in our new context.

The possibility of using multiple penalty parameters has been considered by many authors, including Fletcher (1987, page 292), Powell (1969) and Bertsekas (1982, page 124). Powell's approach, as described by Fletcher, increases the penalties corresponding to the constraints that are becoming too slowly feasible, based on the ℓ_∞-norm. Thus it is only when they have changed sufficiently so that they are all within the constraint violation tolerance that the Lagrange multiplier update is performed. By contrast, we would like to update the multipliers of the good constraints (assuming they correspond to a particular partition — which is likely since that is, partly at least, why the partitions exist) independently of more badly behaved constraints. In addition, by virtue of using the ℓ_2-norm, we do not give quite the same emphasis to the most violated constraint.

We now reformulate Algorithms 3.1 and 3.2 to distinguish amongst the constraint subsets. In this reformulation, 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 $\mathcal{S}$. As above, the algorithmic models depend on a infinite sequence of positive tolerances $\{\omega_k\}_{k=0}^{\infty}$ converging to zero.

Algorithm 6.1

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

$$\mu_{0,j} < 1 \quad (j = 1,\ldots,q). \tag{6.3}$$

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

Step 1 [Inner iteration]. Find $x_k \in \mathcal{B}$ such that

$$\sigma_k \leq \omega_k. \tag{6.4}$$

Step 2 [Test for convergence]. If $\sigma_k \leq \omega_*$ and $\|c(x_k)\| \leq \eta_*$, stop.

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

$$\|c(x_k)_{[\mathcal{Q}_j]}\| \leq \eta_k, \tag{6.5}$$

or Step 3b otherwise.

Step 3a [Update Lagrange multiplier estimates]. Set

$$\begin{aligned} \lambda_{k+1,[\mathcal{Q}_j]} &= \bar{\lambda}(x_k, \lambda_{k,[\mathcal{Q}_j]}, \mu_{k,j})_{[\mathcal{Q}_j]}, \\ \mu_{k+1,j} &= \mu_{k,j}. \end{aligned} \tag{6.6}$$

Step 3b [Reduce the penalty parameter]. Set

$$\begin{aligned} \lambda_{k+1,[\mathcal{Q}_j]} &= \lambda_{k,[\mathcal{Q}_j]}, \\ \mu_{k+1,j} &= \tau_{k,j}\mu_{k,j}, \end{aligned} \tag{6.7}$$

where

$$0 < \tau_{k,j} \leq \tau. \tag{6.8}$$

Step 4 [Aggregated updates]. Define

$$\alpha_{k+1} = \max_{j=1,\ldots,q} \mu_{k+1,j}, \tag{6.9}$$

If

$$\alpha_{k+1} < \alpha_k \tag{6.10}$$

then set

$$\eta_{k+1} = \alpha_{k+1}^{\alpha_\eta}, \tag{6.11}$$

otherwise set

$$\eta_{k+1} = \eta_k \alpha_{k+1}^{\beta_\eta}. \tag{6.12}$$

Increment k by one and go to Step 1.

Algorithm 6.2

The definition of this algorithm is identical to that of Algorithm 6.1, except that Step 3 is replaced by the following, for some $\gamma \in (0,1)$.

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

$$\|c(x_k)_{[\mathcal{Q}_j]}\| \le \eta_k, \tag{6.13}$$

or Step 3b otherwise.

Step 3a [Update Lagrange multiplier estimates]. Set

$$\begin{aligned} \lambda_{k+1,[\mathcal{Q}_j]} &= \begin{cases} \hat{\lambda}_{k+1,[\mathcal{Q}_j]} & \text{if } \|\hat{\lambda}_{k+1,[\mathcal{Q}_j]}\| \le \mu_{k+1,j}^{-\gamma}, \\ \lambda_{k,[\mathcal{Q}_j]} & \text{otherwise,} \end{cases} \\ \mu_{k+1,j} &= \mu_{k,j}. \end{aligned} \tag{6.14}$$

Step 3b [Reduce the penalty parameter]. Set

$$\begin{aligned} \lambda_{k+1,[\mathcal{Q}_j]} &= \begin{cases} \hat{\lambda}_{k+1,[\mathcal{Q}_j]} & \text{if } \|\hat{\lambda}_{k+1,[\mathcal{Q}_j]}\| \le \mu_{k+1,j}^{-\gamma}, \\ \lambda_{k,[\mathcal{Q}_j]} & \text{otherwise,} \end{cases} \\ \mu_{k+1,j} &= \tau_{k,j}\mu_{k,j}, \end{aligned} \tag{6.15}$$

where $\tau_{k,j}$ satisfies (6.8).

Note that the new Lagrange multiplier estimate $\hat{\lambda}_{k+1}$ in Algorithm 6.2 is computed as a single vector, but it could also be computed subset by subset. Observe that the dependence of $\tau_{k,j}$ on k and j introduces additional freedom to the algorithms.

Also note that the quantity α_k represents the maximum penalty parameter across all constraint subsets at iteration k. As for Algorithms 3.1 and 3.2, the restriction (6.3) can be removed by replacing the definition of α_0 and (6.9) by

$$\alpha_0 = \min\left(\gamma_s, \max_{j=1,\ldots,q} \mu_{0,j}\right) \quad \text{and} \quad \alpha_{k+1} = \min\left(\gamma_s, \max_{j=1,\ldots,q} \mu_{k+1,j}\right), \tag{6.16}$$

respectively, for some constant $\gamma_s \in (0,1)$. The other extensions mentioned for Algorithms 3.1 and 3.2 are also applicable to Algorithms 6.1 and 6.2, provided that (6.10) is replaced by

$$\max_{j=1,\ldots,q} \mu_{k+1,j} < \max_{j=1,\ldots,q} \mu_{k,j}. \tag{6.17}$$

We now indicate why the convergence results obtained in Section 4 can be extended to our more general framework.

1. Lemmas 4.1 and 4.2 are reformulated to ensure that, for each $j = 1, \ldots, q$, the product $\mu_{k,j} \|\lambda_{k,[\mathcal{Q}_j]}\|$ converges to zero whenever α_k tends to zero. The proof of the second lemma presented by Conn *et al.* (1991) still holds for each j. However, the proof of the first lemma must be extended to cover our current needs. We therefore state and prove the following result.

Lemma 6.1 *Suppose that α_k converges to zero as k increases when Algorithm 6.1 is executed. Then the product $\mu_{k,j} \|\lambda_{k,[\mathcal{Q}_j]}\|$ converges to zero for each $1 \le j \le q$.*

Proof. We shall prove the result for the t-th subset ($1 \le t \le q$).

As α_k converges to zero, so does $\mu_{k,t}$. Hence Step 3b must be executed infinitely often for the t-th subset. Let $\mathcal{K}_t = \{k_0, k_1, k_2, ...\}$ be the set of the indices of the iterations in which Step 3b is executed.

We consider how the t-th subset of Lagrange multiplier estimates changes between two successive iterations indexed in the set $\mathcal{K}_t$. Firstly note that $\lambda_{k_v+1,[\mathcal{Q}_t]} = \lambda_{k_v,[\mathcal{Q}_t]}$. At iteration $k_v + j$, for $k_v < k_v + j \le k_{v+1}$, we have

$$\lambda_{k_v+j,[\mathcal{Q}_t]} = \lambda_{k_v,[\mathcal{Q}_t]} + \sum_{l=1}^{j-1} \frac{c(x_{k_v+l})_{[\mathcal{Q}_t]}}{\mu_{k_v+l,t}} \tag{6.18}$$

where the summation is null if $j = 1$, and

$$\mu_{k_{v+1},t} = \mu_{k_v+j,t} = \mu_{k_v+1,t} = \tau_{k_v,t}\mu_{k_v,t}. \tag{6.19}$$

Substituting (6.19) into (6.18), multiplying both sides by $\mu_{k_v+j,t}$, taking norms and using (6.8), yields

$$\mu_{k_v+j,t}\|\lambda_{k_v+j,[\mathcal{Q}_t]}\| \le \tau\mu_{k_v,t}\|\lambda_{k_v,[\mathcal{Q}_t]}\| + \sum_{l=1}^{j-1} \|c(x_{k_v+l})_{[\mathcal{Q}_t]}\| \tag{6.20}$$

and hence

$$\mu_{k_v+j,t}\|\lambda_{k_v+j,[\mathcal{Q}_t]}\| \le \tau\mu_{k_v,t}\|\lambda_{k_v,[\mathcal{Q}_t]}\| + \sum_{l=1}^{k_{v+1}-k_v-1} \|c(x_{k_v+l})_{[\mathcal{Q}_t]}\|. \tag{6.21}$$

Using the fact that (6.5) holds for $k_v + 1 \le k_v + l \le k_{v+1} - 1$, we deduce that

$$\begin{aligned} \mu_{k_v+j,t}\|\lambda_{k_v+j,[\mathcal{Q}_t]}\| &\le \tau\mu_{k_v,t}\|\lambda_{k_v,[\mathcal{Q}_t]}\| + \sum_{l=1}^{k_{v+1}-k_v-1} \eta_{k_v+l} \\ &\le \tau\mu_{k_v,t}\|\lambda_{k_v,[\mathcal{Q}_t]}\| + \sum_{l=1}^{\infty} \eta_{k_v+l}. \end{aligned} \tag{6.22}$$

Now defining

$$\delta_v \stackrel{\text{def}}{=} \mu_{k_v,t}\|\lambda_{k_v,[\mathcal{Q}_t]}\| \text{ and } \rho_v \stackrel{\text{def}}{=} \sum_{l=1}^{\infty} \eta_{k_v+l}, \tag{6.23}$$

we obtain that

$$\mu_{k_v+j,t}\|\lambda_{k_v+j,[\mathcal{Q}_t]}\| \le \tau\delta_v + \rho_v \tag{6.24}$$

for all j such that $k_v < k_v + j \le k_{v+1}$, and, in particular,

$$\delta_{v+1} \le \tau \delta_v + \rho_v. \tag{6.25}$$

We now see from (6.25) and the inequality $\tau < 1$ that it is impossible for ρ_v to converge to zero while δ_v does not. Thus δ_v and hence, from (6.24), $\mu_{k_v+j,t}\|\lambda_{k_v+j,[\mathcal{Q}_t]}\|$ both converge to zero if ρ_v does. To complete the proof it therefore suffices to show that ρ_v converges to zero as v tends to infinity.

Using (6.8), the maximum penalty parameter must be decreased by at least a factor τ every q updates (6.11) and thus we have that each η_{k_v+l} must be bounded by a quantity of the form $(\tau^i \alpha_{k_v})^{\alpha_\eta + j\beta_\eta}$ for some indices i and j. Furthermore, at most q such terms can involve any particular i and j. Therefore, since $\tau \alpha_{k_v} < 1$, we obtain that

$$\begin{aligned} \rho_v &\le q \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (\tau^i \alpha_{k_v})^{\alpha_\eta + j\beta_\eta} \\ &= q \sum_{i=0}^{\infty} \frac{(\tau^i \alpha_{k_v})^{\alpha_\eta}}{1 - (\tau^i \alpha_{k_v})^{\beta_\eta}} \\ &\le q \sum_{i=0}^{\infty} \frac{(\tau^i \alpha_{k_v})^{\alpha_\eta}}{1 - (\tau \alpha_{k_v})^{\beta_\eta}} \\ &= q \frac{\alpha_{k_v}^{\alpha_\eta}}{(1 - \tau^{\alpha_\eta})(1 - (\tau \alpha_{k_v})^{\beta_\eta})}. \end{aligned} \tag{6.26}$$

Thus we see that as α_{k_v} converges to zero, so does ρ_v, completing the proof. □

2. Lemmas 2.1, 4.3, 4.4 and 4.5 need no modification.

3. The conclusions of Lemma 4.6 are essentially unmodified. Only (4.21),(4.22) and (4.23) require some reformulation. We now obtain the disaggregate estimates

$$\|(\bar{\lambda}(x_k, \lambda_k, \mu_k) - \lambda_*)_{[\mathcal{Q}_j]}\| \le \kappa_2 \omega_k + \kappa_3 \|x_k - x_*\|, \tag{6.27}$$

$$\|(\lambda(x_k) - \lambda_*)_{[\mathcal{Q}_j]}\| \le \kappa_3 \|x_k - x_*\|, \tag{6.28}$$

and, by using a technique of proof identical to that of Lemma 4.6 for each constraint subset,

$$\|c(x_k)_{[\mathcal{Q}_j]}\| \le \kappa_2 \omega_k \mu_{k,j} + \mu_{k,j} \|(\lambda_k - \lambda_*)_{[\mathcal{Q}_j]}\| + \kappa_3 \mu_{k,j} \|x_k - x_*\|, \tag{6.29}$$

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

4. Finally, Theorem 4.7 still holds, but the proof of Conn *et al.* (1991) must be adapted slightly: the only difficulty in the proof being to ensure that $c(x_*) = 0$, we must now repeat the argument for each $j = 1, \ldots, q$ (using Lemma 4.6 with (6.29)) to obtain that $c(x_*)_{[\mathcal{Q}_j]} = 0$ for each j. This relies on the fact that, if any $\mu_{k,t}$ is bounded away from zero, so is α_k. Furthermore, Lemma 6.1 is used instead of Lemma 4.1.

To summarize, we have proved the following global convergence result.

Theorem 6.2 *Assume that AS1 and AS2 hold. Let x_* be any limit point of the sequence $\{x_k\}$ generated by Algorithm 6.1 or by Algorithm 6.2 for which AS3 and AS4 hold and let $\mathcal{K}$ be the set of indices of an infinite subsequence of the x_k whose limit is x_*. Let $\lambda_* = \lambda(x_*)$. Then*

(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 (6.27), (6.28) and (6.29) hold for all $k \geq k_1, (k \in \mathcal{K})$;

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

We therefore conclude that the "disaggregate" updating technique described above does not alter the global convergence properties of our algorithmic framework.

7 Conclusions and perspectives

We have shown that the philosophy of keeping linear constraints as explicit constraints in the augmented Lagrangian minimization subproblem (hence avoiding their inclusion in the augmented Lagrangian function) does not modify the global convergence properties of the overall minimization algorithm. This strategy allows for explicit handling of linear constraints for improved efficiency. In particular, it avoids the fill-in caused by linear constraints in the Hessian of the augmented Lagrangian function. We also analyzed alternative stopping criteria for the subproblem, including commonly used practical rules. We finally observed that global convergence is still guaranteed if the set of general equality constraints is partitioned into several disjoint groups for which the Lagrange multiplier estimates and penalty parameter can be updated independently.

Our analysis was formally restricted to the case where the norm used in the last constraint of (2.8) is the classical ℓ_2-norm. This could be considered as a restriction, since other norms might be more appropriate to measure criticality of the subproblem, especially when poor scaling and/or ill-conditioning are present. This restriction is however only made for simplicity of exposition. The reader may indeed verify that we may choose, at each iteration k, a suitable norm $\|\cdot\|_{(k)}$ in (2.8) and still derive our global convergence results, provided all the considered norms are uniformly equivalent.

The next step in our development is to consider the asymptotic convergence properties of the algorithms studied above, including the rate of convergence of their iterates to a solution and the asymptotic boundedness of the penalty parameter. This step is the object of the companion paper Conn *et al.* (1993b).

8 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.

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

[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 a class of trust region algorithms for optimization using inexact projections on convex constraints. *SIAM Journal on Optimization*, 3(1):164–221, 1993.

[Conn *et al.*, 1993b] A. R. Conn, Nick Gould, A. Sartenaer, and Ph. L. Toint. Local convergence properties of two augmented Lagrangian algorithms for optimization with a combination of general equality and linear constraints. Technical Report 93/20, Department of Mathematics, FUNDP, Namur, Belgium, 1993.

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

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

[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.

[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.