# SANDIA REPORT

SAND2006-5315
Unlimited Release
Printed August 2006

# A generating set direct search augmented Lagrangian algorithm for optimization with a combination of general and linear constraints

T. G. Kolda, R. M. Lewis, and V. Torczon

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.



Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.



Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

| | |
|---|---|
| Telephone: | (865) 576-8401 |
| Facsimile: | (865) 576-5728 |
| E-Mail: | reports@adonis.osti.gov |
| Online ordering: | http://www.osti.gov/bridge |

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

| | |
|---|---|
| Telephone: | (800) 553-6847 |
| Facsimile: | (703) 605-6900 |
| E-Mail: | orders@ntis.fedworld.gov |
| Online ordering: | http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online |

SAND2006-5315
Unlimited Release
Printed August 2006

# A generating set direct search augmented Lagrangian algorithm for optimization with a combination of general and linear constraints

Tamara G. Kolda
Computational Science and Mathematics Research Department
Sandia National Laboratories
Livermore, CA 94550-9159
tgkolda@sandia.gov

Robert Michael Lewis
Department of Mathematics
College of William & Mary
Williamsburg, Virginia, 23187-8795
buckaroo@math.wm.edu

Virginia Torczon
Department of Computer Science
College of William & Mary
Williamsburg, Virginia, 23187-8795
va@cs.wm.edu

**Abstract**

We consider the solution of nonlinear programs in the case where derivatives of the objective function and nonlinear constraints are unavailable. To solve such problems, we propose an adaptation of a method due to Conn, Gould, Sartenaer, and Toint that proceeds by approximately minimizing a succession of linearly constrained augmented Lagrangians. Our modification is to use a

derivative-free generating set direct search algorithm to solve the linearly constrained subproblems. The stopping criterion proposed by Conn, Gould, Sartenaer and Toint for the approximate solution of the subproblems requires explicit knowledge of derivatives. Such information is presumed absent in the generating set search method we employ. Instead, we show that stationarity results for linearly constrained generating set search methods provide a derivative-free stopping criterion, based on a step-length control parameter, that is sufficient to preserve the convergence properties of the original augmented Lagrangian algorithm.

# Acknowledgments

We thank Rakesh Kumar of The MathWorks for helpful discussions regarding the performance of generating set search algorithms. We are grateful to him both for pointing out to us the desirability of relaxing the update for the stopping tolerance in the augmented Lagrangian subproblems (discussed in Section 6) [13] and for being the first to confirm that using a GSS algorithm to directly handle explicit linear constraints, rather than just bounds, within the context of an augmented Lagrangian algorithm can have an appreciable effect on performance [14].

This page intentionally left blank.

# Contents



## Appendix

This page intentionally left blank.

# 1 Introduction

In this paper, the problems of interest are general nonlinear optimization problems of the following form:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c(x) = 0 \\ & Ax \geq b. \end{array} \tag{1}$$

Here the objective function is $f : \mathbb{R}^n \rightarrow \mathbb{R}$. The constraints are a mixture of explicit linear constraints and general equality constraints. The matrix $A \in \mathbb{R}^{p \times n}$ defines the explicit linear constraints, including linear equality constraints and bounds on the variables; $a_i^T$ denotes the $i$th row of $A$. The general equality constraints are $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$; $c_i(x)$ denotes the $i$th equality constraint. General inequality constraints are assumed to be converted to equalities by introducing nonnegative slack variables; see Section 7.

The motivation for the work reported here is the situation in which the derivatives of both $f$ and $c$ are either unavailable or unreliable. The algorithm we present for solving (1) is an adaptation of an augmented Lagrangian method due to Conn, Gould, Sartenaer, and Toint [5] (related work may be found in [3, 4, 6]). In their approach the linear constraints are dealt with directly, but derivatives of $f$ and $c$ are presumed to be available. Our adaptation of their algorithm makes use of generating set search (GSS) methods [11], which neither require nor explicitly approximate these derivatives, and yet possess standard first-order convergence properties. Specifically, we use a derivative-free GSS variant for linearly constrained problems that is known to possess good convergence behavior in both theory [12] and practice [15, 9].

In the augmented Lagrangian method due to Conn, Gould, Sartenaer, and Toint [5], only the general nonlinear equality constraints are included in the augmented Lagrangian $\Phi$:

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

where the components $\lambda_i$ of the vector $\lambda$ are the Lagrange multiplier estimates and $\mu$ is the penalty parameter. Their method then involves successive linearly constrained minimization of a more general version of (2). The basic form of the algorithm is:

As noted in [5], an attractive feature of this framework is that the linear constraints are kept outside the augmented Lagrangian and are handled at the level of the subproblem minimization. This reduces the number of Lagrange multipliers that must be estimated. It also allows the use of algorithms that ensure that the iterates produced remain feasible with respect to the linear constraints.

**Initialization.** Choose $\lambda_0$, $\mu_0$, and $x_0$ satisfying $Ax_0 \geq b$ as well as various parameters for stopping tolerances in the inner and outer iterations.

**Outer iteration.** For $k = 0, 1, 2, \ldots$

**Inner iteration.** Find a solution $x_{k+1}$ that approximately minimizes $\Phi(x; \lambda_k, \mu_k)$ subject to $Ax_{k+1} \geq b$, according to an appropriate stopping criterion.

**Test for convergence.** If the final convergence tests are satisfied, stop with the solution $x_{k+1}$.

**Updates.** Else

- update the Lagrange multipliers to obtain $\lambda_{k+1}$,
- update the penalty parameter to obtain $\mu_{k+1}$, and
- update assorted parameters, including the stopping tolerances for the inner iteration.

**Algorithm 1.1:** A basic framework for the augmented Lagrangian approach that leaves the linear constraints explicit.

We adapt the above framework, making use of a GSS method for solving linearly constrained problems [12] to solve the inner iteration while preserving the convergence properties of the augmented Lagrangian algorithm in [5]. The catch for us is that the "appropriate stopping criterion" for the inner iteration, as originally defined in [5], involves the explicit gradient of the augmented Lagrangian. The GSS adaptation substitutes a suitable derivative-free stopping criterion. We show that our stopping criterion for the linearly constrained subproblem can be substituted for the one used in [5] without sacrificing the convergence properties of the original approach. Thus we, too, are able to proceed by successive, inexact minimization of the augmented Lagrangian via GSS methods, even though we do not know directly how inexact the minimization is.

Dealing with general nonlinear constraints in the absence of derivatives is challenging and has received considerable attention over the years. Summaries of early work appear in [7] and [1, Section 13.1]. Approaches to handling general nonlinear constraints can be partitioned into three basic alternatives. If the derivatives of $c$ are available, or reliable estimates can be obtained, then it is possible to make explicit use of these derivatives to compute feasible directions at the boundary of the feasible region. See [11, Section 8.3.1] for a summary of work involving this approach. The second alternative is an augmented Lagrangian approach. The work we present here may be viewed as an extension of the work in [16]; see also [11, Section 8.3.2]. For both the feasible directions [21, 17] and augmented Lagrangian approaches [16], under standard assumptions it is possible to prove convergence to Karush–Kuhn–Tucker (KKT) points of problem (1), as we will do here. The third alternative involves approaches such as inexact penalization, exact penalization, barrier methods, and a variety of heuristics-based approaches. See [11, Section 8.3.3] for a summary of algorithmic developments along these lines.

The paper proceeds as follows. Section 2 lays out the augmented Lagrangian algorithm from [5] and reviews the relevant notation. Section 3 summarizes the GSS algorithm from [12] for handling problems with linear constraints and recalls a critical stationarity result. In Section 4 we show how to incorporate the GSS algorithm to solve the subproblems in the augmented Lagrangian algorithm by introducing a derivative-free stopping condition. Section 5 summarizes the convergence results from [5] that the GSS adaptation possesses. In Section 6 we discuss a way to relax the subproblem stopping criterion update introduced in [16] so as to improve computational efficiency. Section 7 discusses the conversion of inequality constraints to equality constraints through the introduction of slack variables. We close with some final observations in Section 8.

This page intentionally left blank.

# 2 The augmented Lagrangian algorithm of Conn, Gould, Sartenaer, and Toint

Our augmented Lagrangian GSS approach is based on Algorithm 3.1 of [5], which we review in this section.

## 2.1 A comment on notation

To facilitate comparison of our approach with the original algorithm, we adhere to the notation of [5] throughout. Subscripts may denote either a component of a vector or an iteration index. Thus $w_i$ (or $w_j$) denotes the $i$th (or $j$th) component of the vector $w$ while $w_k$ denotes the vector $w$ from the $k$th iteration of the algorithm. When combined, $w_{k,j}$ denotes that $j$th component of the vector $w_k$. A vector can also be subscripted by a set, i.e., $w_{[\mathcal{S}]}$ denotes the $|\mathcal{S}|$-dimensional subvector of $w$ whose entries are indexed by the set $\mathcal{S}$. Moreover, subset indexing may be combined with the iteration index as $w_{k,[\mathcal{S}]}$.

## 2.2 The augmented Lagrangian

For the formulation of the augmented Lagrangian, the constraints $c(x)$ are assumed to be partitioned into $q$ disjoint subsets $\{\mathcal{Q}_j\}_{j=1}^q$ such that $\bigcup_{j=1}^q \mathcal{Q}_j = \{1, \ldots, m\}$. The partitioning of $c(x)$ enables the algorithm to place greater emphasis on achieving feasibility for subsets of constraints that are, at any particular iteration, proportionally more violated than the others.

The basic augmented Lagrangian given in (2) is replaced by

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

The vector $\lambda = (\lambda_1, \ldots, \lambda_m)^T$ is the Lagrange multiplier estimate for the equality constraints and the vector $\mu = (\mu_1, \ldots, \mu_q)^T$ contains the penalty parameters associated with each partition of $c(x)$.

## 2.3 The linear constraints and cones

Mechanisms are required for handling the set of linear constraints that are nearly binding at $x$. Let $\kappa_0 > 0$ be fixed. Define $\mathcal{B} = \{\, x \mid Ax \geq b \,\}$. For $x \in \mathcal{B}$, let

$$D(x,\omega) = \left\{\, i \in \{1,\ldots,p\} \mid a_i^T x - b_i \leq \kappa_0 \omega \,\right\} \tag{3}$$

denote the indices of the linear constraints that, with respect to $\omega$, are considered nearly binding at $x$. From this, define the $\omega$-normal and $\omega$-tangent cones

$$N(x,\omega) = \left\{ v \mid v = \sum_{i \in D(x,\omega)} \xi_i a_i,\ \xi_i \leq 0 \right\} \quad \text{and} \quad T(x,\omega) = N^\circ(x,\omega).$$

The cone $N(x,\omega)$ is the cone generated by the outward pointing normals to the nearly binding linear constraints and $T(x,\omega)$ is its polar. If $D(x,\omega) = \emptyset$, then $N(x,\omega) = \{\mathbf{0}\}$ so that $T(x,\omega) = \mathbb{R}^n$. The projection of the vector $v$ onto $T(x,\omega)$ is denoted $P_{T(x,\omega)}(v)$.

## 2.4 The subproblem

At the $k$th outer iteration of the augmented Lagrangian method, an inexact solution to the following subproblem is required:

$$\begin{array}{ll} \text{minimize} & \Phi_k(x) \\ \text{subject to} & Ax \geq b, \end{array} \tag{4}$$

where

$$\Phi_k(x) \equiv \Phi(x; \lambda_k, \mu_k) \tag{5}$$

and the vectors $\lambda_k$ and $\mu_k$ are updated each outer iteration.

The solution $x_k$ of (4) must satisfy:

$$\| \, P_{T(x_k,\omega_k)}(-\nabla_x \Phi_k) \, \| \leq \omega_k, \tag{6}$$

where

$$\nabla_x \Phi_k \equiv \nabla_x \Phi(x; \lambda_k, \mu_k)$$

and the scalar $\omega_k > 0$ is a suitable tolerance that is updated at each outer iteration in a way that ensures $\omega_k \to 0$ as $k \to \infty$.

## 2.5 The full algorithm

We reproduce Algorithm 3.1 in [5] as Algorithm 2.1. The parameters $\omega_k$ and $\eta_k$ represent stationarity and feasibility tolerances at iteration $k$, respectively. Updates for these two parameters, as well as for $\lambda_k$ and $\mu_k$ are specified. The Lagrange multiplier estimates $\lambda_k$ are updated according to the first-order Hestenes–Powell [10, 19] update rule:

$$\bar{\lambda}(x, \lambda_{[\mathcal{Q}_j]}, \mu_j)_{[\mathcal{Q}_j]} = \lambda_{[\mathcal{Q}_j]} + c(x)_{[\mathcal{Q}_j]}/\mu_j \quad (j = 1, \ldots, q). \tag{7}$$

Since the Hestenes–Powell multiplier update and its variants do not require information about derivatives of $f$ and $c$, unlike other update formulas (see, for instance, [2, 20]), they are appropriate for derivative-free methods.

The following assumptions are made in [5] for the purposes of their convergence analysis, which we review in Section 5. We also make similar assumptions for the results in Sections 3 and 4 so that the hypotheses are consistent.

**AS1** *[5, p. 676] The set $\mathcal{B}$ is nonempty.*

**AS2** *[5, p. 676] The functions $f(x)$ and $c(x)$ are twice continuously differentiable for all $x \in \mathcal{B}$.*

**AS3** *[5, p. 681] The iterates $\{x_k\}$ lie within a closed, bounded domain $\Omega$.*

Note that we place smoothness assumption on $f$ and $c$ for the purposes of analysis only. Neither the first nor second derivatives of $f$ and $c$ are required or used in the algorithms that follow.

**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 $\lambda_0$ and positive penalty parameters $\mu_0$ such that

$$\mu_{0,j} < 1, \; j = 1,\ldots,q.$$

The strictly positive constants $\kappa_0$, $\omega_* \ll 1$, $\eta_* \ll 1$, $\tau < 1$, $\alpha_\eta < 1$, 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 (4), i.e., such that (6) holds.

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

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

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

or Step 3b otherwise.

- **Step 3a. [Update Lagrange multiplier estimates].** Set

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

- **Step 3b. [Reduce the penalty parameter].** Set

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

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

**Step 4.** **[Aggregated updates].** Define

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

If $\alpha_{k+1} < \alpha_k$, then set

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

otherwise set

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

Increment k by one and go to Step 1.

**Algorithm 2.1:** Augmented Lagrangian algorithm of Conn, Gould, Sartenaer, and Toint.

# 3 Generating set search for linearly constrained problems

To adapt the augmented Lagrangian framework in the absence of derivatives for $f$ and $c$, we solve subproblem (4) using a GSS algorithm for linearly constrained optimization [12]. In this section, we review salient details of the algorithm.

## 3.1 A comment on notation

We use a "hat" (as in $\hat{f}$, $\hat{D}$, $\hat{N}$, $\hat{T}$, $\hat{\alpha}$, and $\hat{\kappa}$) to distinguish between variables discussed in [12] and those discussed in Section 2 since the notation is similar but the quantities are not necessarily equivalent. Once again subscripts may denote either a component of a vector or an iteration index. Thus $b_i$ denotes the $i$th component of the vector $b$ while $x_\ell$ denotes the vector $x$ from the $\ell$th iteration of the GSS algorithm for linearly constrained problems.

## 3.2 The linearly constrained problem

GSS for linearly constrained optimization solves problems of the form:

$$\begin{array}{ll} \text{minimize} & \hat{f}(x) \\ \text{subject to} & Ax \geq b. \end{array} \tag{8}$$

Here $\hat{f} : \mathbb{R}^n \to \mathbb{R}$ and $A$ is the same as in (1). The set $\mathcal{B} = \{\ x \ \mid \ Ax \geq b\ \}$ denotes the feasible region for problem (8).

## 3.3 The linear constraints and cones

The GSS methods in [12] also use the cone generated by the working set of nearby constraints and the polar of this cone, but the definitions are not identical to those in Section 2.3. The definition of $D(x, \omega)$ in (3) defines the working set by looking at the values of the constraints. In the GSS case, the definition of the working set uses distance to the constraints. We use $\varepsilon$ rather than $\omega$ and define the $\varepsilon$-binding constraints (the working set) for $x \in \mathcal{B}$ as

$$\hat{D}(x, \varepsilon) = \left\{\ i \in \{1, \ldots, p\} \ \mid \ \frac{a_i^T x - b_i}{\|\ a_i\ \|} \leq \varepsilon\ \right\}. \tag{9}$$

From this, we define the $\varepsilon$-normal and $\varepsilon$-tangent cones as before:

$$\hat{N}(x,\varepsilon) = \left\{ v \mid v = \sum_{i \in \hat{D}(x,\varepsilon)} \xi_i a_i,\ \xi_i \le 0 \right\} \quad \text{and} \quad \hat{T}(x,\varepsilon) = \hat{N}^\circ(x,\varepsilon).$$

## 3.4 The GSS algorithm

In Algorithm 3.1, we present Algorithm 5.1 from [12], stated in the notation and style of presentation adopted here.

Iteration $\ell$ of a GSS algorithm proceeds as follows. Compute a set of search directions $\mathcal{G}_\ell$ that conforms to the boundary defined by the $\varepsilon$-binding constraints $\hat{D}(x_\ell, \varepsilon_\ell)$. Generate feasible trial points by taking a step, whose length is determined by the step-length control parameter $\Delta_\ell$, along each search direction. If any of the trial points yields sufficient decrease, specifically

$$\hat{f}(x_\ell + \tilde{\Delta}_\ell\, d_\ell) < \hat{f}(x_\ell) - \hat{\alpha}\Delta_\ell^2, \tag{10}$$

where $\tilde{\Delta}_\ell = \max\{\, \Delta \in [0, \Delta_\ell] \mid x_\ell + \Delta d_\ell \in \mathcal{B} \,\}$, then the iteration is deemed *successful* and that trial point becomes $x_{\ell+1}$. Otherwise no trial point admits sufficient decrease and the iteration is deemed *unsuccessful*, in which case the step-length control parameter $\Delta_\ell$ is reduced. The set of unsuccessful iterations is denoted by $\mathcal{U}$ and plays an important role in the analysis of GSS methods.

Here, for convenience, we leave $\Delta_\ell$ unchanged after a successful iteration and halve $\Delta_\ell$ after an unsuccessful iteration. These updates could be altered, subject to the conditions given in [12], without detriment to the analysis presented here.

The following two conditions on the search directions in Step 1 play a critical role in the theory that follows, so we assume that both hold whenever we reference Algorithm 3.1. We start with the following definition from [12]. For any finite set of vectors $\mathcal{G}$, we define

$$\kappa(\mathcal{G}) = \inf_{\substack{v \in \mathbb{R}^n \\ P_K(v) \neq 0}} \max_{d \in \mathcal{G}} \frac{v^T d}{\| P_K(v) \| \, \| d \|}, \quad \text{where } K \text{ is the cone generated by } \mathcal{G}$$

and $P_K(v)$ is the projection of the vector $v$ onto the cone $K$. This is a generalization of the quantity given in [11, (3.10)], where $\mathcal{G}$ generates $\mathbb{R}^n$.

**Step 0.** **[Initialization].** Let $x_0 \in \mathcal{B}$ be the initial guess. Let $\Delta_{\text{tol}} > 0$ be the tolerance used to test for convergence. Let $\Delta_0 > \Delta_{\text{tol}}$ be the initial value of the step-length control parameter. Let $\varepsilon_{\max} > \beta_{\max}\Delta_{\text{tol}}$ be the maximum distance used to identify nearby constraints ($\varepsilon_{\max} = +\infty$ is permissible). Let $\hat{\alpha} > 0$.

**Step 1.** **[Choose search directions].** Let $\varepsilon_\ell = \min\{\varepsilon_{\max}, \beta_{\max}\Delta_\ell\}$. Choose a set of search directions $\mathcal{D}_\ell = \mathcal{G}_\ell \cup \mathcal{H}_\ell$ satisfying Conditions 3.1 and 3.2.

**Step 2.** **[Successful iteration].** If there exists $d_\ell \in \mathcal{D}_\ell$ and a corresponding

$$\tilde{\Delta}_\ell = \max\{\ \Delta \in [0, \Delta_\ell]\ \mid\ x_\ell + \Delta d_\ell \in \mathcal{B}\ \}$$

such that

$$\hat{f}(x_\ell + \tilde{\Delta}_\ell d_\ell) < \hat{f}(x_\ell) - \hat{\alpha}\Delta_\ell^2,$$

then:

- Set $x_{\ell+1} = x_\ell + \tilde{\Delta}_\ell d_\ell$.
- Set $\Delta_{\ell+1} = \Delta_\ell$ (no change).

**Step 3.** **[Unsuccessful iteration].** Otherwise,

- Set $x_{\ell+1} = x_\ell$ (no change).
- Set $\Delta_{\ell+1} = \frac{1}{2}\Delta_\ell$.

If $\Delta_{\ell+1} < \Delta_{\text{tol}}$, then terminate.

**Step 4.** **[Advance].** Increment $\ell$ by one and go to Step 1.

**Algorithm 3.1:** Linearly constrained generating set search of Kolda, Lewis, and Torczon.

**Condition 3.1** *There exists a constant $\hat{\kappa}_{\min} > 0$, independent of $\ell$, such that for every $\ell$ for which $\hat{T}(x_\ell, \varepsilon_\ell) \neq \{\mathbf{0}\}$, the set $\mathcal{G}_\ell$ generates $\hat{T}(x_\ell, \varepsilon_\ell)$ and satisfies $\hat{\kappa}(\mathcal{G}_\ell) \geq \hat{\kappa}_{\min}$.*

**Condition 3.2** *There exist $\beta_{\max} \geq \beta_{\min} > 0$, independent of $\ell$, such that for every $\ell$ for which $\hat{T}(x_\ell, \varepsilon_\ell) \neq \{\mathbf{0}\}$, the following holds:*

$$\beta_{\min} \leq \|d\| \leq \beta_{\max} \qquad \textit{for all } d \in \mathcal{G}_\ell.$$

### 3.5 The critical stationarity result

The following restatement of a result from [12] is central to showing that we can recognize when the derivative-free GSS method has solved the augmented Lagrangian subproblem (4) to the accuracy (6) required by Algorithm 2.1. Theorem 3.3 gives a bound on the size of the projection onto $\hat{T}(x_\ell, \varepsilon_\ell)$ of $\nabla \hat{f}(x_\ell)$, which is not available to us in the derivative-free context, in terms of the explicitly known step-length control parameter $\Delta_\ell$.

**Theorem 3.3** *Suppose that the set $\mathcal{B}$ is nonempty, that the function $\hat{f}(x)$ is twice continuously differentiable for all $x \in \mathcal{B}$, and that the iterates produced by Algorithm 3.1 lie within a closed, bounded domain $\Omega$. Let $M$ be a Lipschitz constant for $\nabla \hat{f}$ on $\Omega$. If $\ell \in \mathcal{U}$ and and $\varepsilon_\ell$ satisfies $\varepsilon_\ell = \beta_{\max}\Delta_\ell$, then*

$$\| \, P_{\hat{T}(x_\ell, \varepsilon_\ell)}(-\nabla \hat{f}(x_\ell)) \, \| \leq \frac{1}{\hat{\kappa}_{\min}} \left( M\beta_{\max} + \frac{\hat{\alpha}}{\beta_{\min}} \right) \Delta_\ell. \tag{11}$$

*Here, $\hat{\alpha}$ is from the step acceptance criterion (10), $\hat{\kappa}_{\min}$ is from Condition 3.1, and $\beta_{\max}$ and $\beta_{\min}$ are from Condition 3.2.*

Theorem 3.3 is a variant of Theorem 6.3 from [12] using the specific step acceptance criterion (10). Theorem 3.3 assumes the iterates remain in a compact set $\Omega$ in order to be assured of the existence of $M$, while Theorem 6.3 from [12] accomplishes the same thing by assuming that the set $\{ \, x \in \mathcal{B} \mid \hat{f}(x) \leq \hat{f}(x_0) \, \}$ is compact. Furthermore, Theorem 6.3 from [12] assumes only that $\nabla \hat{f}$ is Lipschitz continuous. Looking ahead to the results in Section 5, here we assume the stronger condition that $\hat{f}(x)$ is $C^2$. The proof of Theorem 3.3 is the same as that of Theorem 6.3.

Finally, we need the following immediate consequence of the step acceptance criterion (10).

**Theorem 3.4** *Suppose that the function $\hat{f}(x)$ is continuous on $\mathcal{B}$ and that the iterates produced by Algorithm 3.1 lie within a closed, bounded domain $\Omega$. Then* $\liminf_{\ell \to \infty} \Delta_\ell = 0$.

# 4 A GSS adaption of the augmented Lagrangian algorithm

We now provide a modified version of the augmented Lagrangian algorithm of Conn, Gould, Sartenaer, and Toint (Algorithm 2.1) that uses our linearly constrained GSS algorithm (Algorithm 3.1) to solve the subproblem (4). In Algorithm 4.1 we have highlighted the differences between the new algorithm and its progenitor, Algorithm 2.1, which lie in the choice of stopping criterion for the inner iteration, the update for the associated stopping tolerance, and the test for convergence of the outer iteration.

## 4.1 The derivative-free stopping criterion

The substantive change to be addressed is that of finding a suitable stopping criterion for the solution of (4). As noted earlier, we do not assume access to the derivatives of $f$ and $c$ and thus cannot compute $\nabla_x \Phi_k$ as required for the original stopping condition (6). Instead, we make use of the conclusion of Theorem 3.3 to craft an appropriate termination test for the subproblem. Specifically, we stop the inner iteration at the first $u \in \mathcal{U}$ (the subsequence of unsuccessful GSS iterations) for which

$$\Delta_{k,u} \leq \delta_k, \tag{12}$$

where $\delta_k \to 0$ is a sequence of stopping tolerances for the inner iteration that is updated at each outer iteration $k$. Theorem 3.4 assures us that this stopping criterion eventually will be satisfied and thus that the inner iteration will terminate.

This change in the termination test for the inner iteration also affects the test for convergence for the outer iteration. The original outer iteration convergence criteria are that for some $\omega_* \ll 1$ and $\eta_* \ll 1$,

$$\| \, P_{T_k}(-\nabla_x \Phi_k) \, \| \leq \omega_* \quad \text{and} \quad \| \, c(x_k) \, \| \leq \eta_*.$$

These become

$$\delta_k \leq \delta_* \quad \text{and} \quad \| \, c(x_k) \, \| \leq \eta_*$$

for some $\delta_* \ll 1$. The test for feasibility is unchanged, but the test for constrained stationarity is necessarily altered.

Clearly, if the convergence analysis from [5] is to hold, then the new sequence of stopping tolerances $\delta_k$ needs to be tied to the original sequence of stopping tolerances $\omega_k$. We also need to ensure a stable relationship between the stopping criterion (12)

**Step 0.** **[Initialization].** A partition of the set $\{1,\dots,m\}$ into $q$ disjoint subsets $\{\mathcal{Q}_j\}_{j=1}^{q}$ is given, as well as initial vectors of Lagrange multiplier estimates $\lambda_0$ and positive penalty parameters $\mu_0$ such that $\mu_{0,j} < 1$, $j = 1,\dots,q$. **Set** $\kappa_0 = \min_{1,\dots,p}\{\|a_i\|\}$**.** The strictly positive constants $\omega_\star \ll 1$, $\eta_\star \ll 1$, $\tau < 1$, $\alpha_\eta < 1$, and $\beta_\eta < 1$ are specified. Set $\alpha_0 = \max_{j=1,\dots,q}\mu_{0,j}$, $\omega_0 = \alpha_0$, $\eta_0 = \alpha_0^{\alpha_\eta}$, and $k = 0$. **In addition, specify strictly positive constants** $\delta_\star \ll 1$ **and** $\theta_{\text{tol}} \gg 1$**. Set** $\delta_0 = \omega_0/\left(\beta_{\max}\,\theta(\lambda_0,\mu_0)\right)$.

**Step 1.** **[Inner iteration].** Find $x_\ell \in \mathcal{B}$ that approximately solves (4), i.e., such that **(12)** holds.

**Step 2.** **[Test for convergence].** If $\delta_k \le \delta_\star$ and $\|\, c(x_k)\,\| \le \eta_\star$, stop.

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

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

or Step 3b otherwise.

- **Step 3a. [Update Lagrange multiplier estimates].** Set

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

- **Step 3b. [Reduce the penalty parameter].** Set

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

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

**Step 4.** **[Aggregated updates].** Define

$$\alpha_{k+1} = \max_{j=1,\dots,q}\mu_{k+1,j}.$$

If $\alpha_{k+1} < \alpha_k$, then set

$$\begin{aligned}\omega_{k+1} &= \alpha_{k+1},\\ \eta_{k+1} &= \alpha_{k+1}^{\alpha_\eta},\\ \delta_{\mathbf{k+1}} &= \omega_{\mathbf{k+1}}/\theta(\lambda_{\mathbf{k+1}},\mu_{\mathbf{k+1}});\end{aligned}$$

otherwise set

$$\begin{aligned}\omega_{k+1} &= \omega_k\alpha_{k+1},\\ \eta_{k+1} &= \eta_k\alpha_{k+1}^{\beta_\eta},\\ \delta_{\mathbf{k+1}} &= \omega_{\mathbf{k+1}}/\theta(\lambda_{\mathbf{k+1}},\mu_{\mathbf{k+1}}).\end{aligned}$$

Increment k by one and go to Step 1.

**Algorithm 4.1:** A generating set search augmented Lagrangian algorithm.

in the GSS solution of the subproblems and the stationarity condition (6). To do so, let $\theta_{\text{tol}} \gg 1$ be given and define

$$\theta(\lambda, \mu) = \max\left\{1, \ \left(1 + \| \lambda \| + \sum_{j=1}^{q} 1/\mu_j\right) / \theta_{\text{tol}}\right\}. \tag{13}$$

Any function $\theta(\lambda, \mu)$ such that $(\| \lambda \| + \sum_{j=1}^{q} 1/\mu_j) = O(\theta(\lambda, \mu))$ as $(\| \lambda \| + \sum_{j=1}^{q} 1/\mu_j) \to \infty$ suffices for the purposes of establishing global convergence properties. We discuss the role of $\theta$ further in Section 6.

Finally, we need to take into account the fact that in Algorithm 3.1

$$\varepsilon = \min\{\varepsilon_{\max}, \beta_{\max}\Delta\}. \tag{14}$$

This leads to the following update for $\delta_k$:

$$\delta_{k+1} = \omega_{k+1} / \left(\beta_{\max}\, \theta(\lambda_{k+1}, \mu_{k+1})\right). \tag{15}$$

Observe that the the initialization of $\delta_0$ in Algorithm 4.1 and the definition of $\theta$ in (13), together with the update rule (15), ensure that for all $k \geq 0$,

$$\delta_k \leq \min\left\{1, \theta_{\text{tol}} / \left(1 + \| \lambda_k \| + \sum_{j=1}^{q} 1/\mu_{k,j}\right)\right\} \frac{\omega_k}{\beta_{\max}} \leq \frac{\omega_k}{\beta_{\max}}. \tag{16}$$

As a practical matter, in the implementation of Algorithm 3.1 discussed in [15] the directions in $\mathcal{G}$ are normalized, so $\beta_{\max} = 1$, which simplifies both (15) and (16).

## 4.2 The linear constraints and cones

A technical matter to be addressed is that in (3) the condition for inclusion in the set $D(x, \omega)$ for the augmented Lagrangian is

$$a_i^T x - b_i \leq \kappa_0 \omega,$$

whereas in (9) the condition for inclusion in the set $\hat{D}(x, \varepsilon)$ for GSS is

$$a_i^T x - b_i \leq \|a_i\| \varepsilon. \tag{17}$$

Since this affects the definitions of the cones $T(x, \omega)$ and $\hat{T}(x, \varepsilon)$, which we wish to relate in the results that follow in the next section, we reconcile this difference within Algorithm 4.1 by setting

$$\kappa_0 = \max_{i=1,\ldots,p} \|a_i\|.$$

This particular choice of $\kappa_0$ simplifies the upcoming proof of Proposition 4.1, but is not essential. Observe that with this choice of $\kappa_0$, if (17) holds for all $i \in \{1, \ldots, p\}$ and $\varepsilon \leq \omega$, then we have $a_i^T x - b_i \leq \|a_i\|\varepsilon \leq \|a_i\|\omega \leq \max_{i=1,\ldots,p} \|a_i\|\omega$ for all $i \in \{1, \ldots, p\}$. Thus $D(x, \omega) \supseteq \hat{D}(x, \varepsilon)$, so $N(x, \omega) \supseteq \hat{N}(x, \varepsilon)$ and

$$T(x, \omega) \subseteq \hat{T}(x, \varepsilon), \tag{18}$$

a fact we use shortly.

## 4.3 The relationship between $\Delta_k$ and stationarity

We begin by relating the GSS stopping criterion (12) to the original stopping criterion (6).

**Proposition 4.1** *Suppose that AS1–AS3 hold, and let $M$ be a Lipschitz constant for $\nabla f(x)$, $\nabla c(x)$, and $\nabla c(x)\, c(x)$ on $\Omega$. Then the following bound holds at outer iteration $k$ of Algorithm 4.1:*

$$\| P_{T(x_k, \omega_k)}(-\nabla_x \Phi_k) \| \leq \frac{1}{\hat{\kappa}_{\min}} \left( M_k \delta_k \beta_{\max} + \frac{\hat{\alpha}}{\beta_{\min}} \delta_k \right), \tag{19}$$

*where*

$$M_k = M \left( 1 + \| \lambda_k \| + \sum_{j=1}^{q} 1/\mu_{k,j} \right). \tag{20}$$

**Proof.** We first bound the Lipschitz constant for $\nabla_x \Phi(x; \lambda_k, \mu_k)$ and then apply Theorem 3.3. We have

$$\nabla_x \Phi(x; \lambda_k, \mu_k) = \nabla_x f(x) + \sum_{j=1}^{q} \sum_{i \in \mathcal{Q}_j} \left[ \lambda_{k,i} \, \nabla_x c_i(x) + \frac{1}{\mu_{k,j}} c_i(x) \, \nabla_x c_i(x) \right],$$

so

$$\| \nabla_x \Phi(x; \lambda_k, \mu_k) - \nabla_x \Phi(y; \lambda_k, \mu_k) \| \leq M \left( 1 + \| \lambda_k \| + \sum_{j=1}^{q} 1/\mu_{k,j} \right) \| x - y \|.$$

Since the stopping criterion (12) is invoked at unsuccessful iterations of the GSS solution of (8), we may apply Theorem 3.3, from which we obtain

$$\| P_{\hat{T}(x_k, \varepsilon_k)}(-\nabla_x \Phi_k) \| \leq \frac{1}{\hat{\kappa}_{\min}} \left( M_k \delta_k \beta_{\max} + \frac{\hat{\alpha}}{\beta_{\min}} \delta_k \right), \tag{21}$$

where $\varepsilon_k$ is the value of $\varepsilon$ in the solution of (4) at the time the stopping criterion (12) is triggered.

From (12), (14), and (16) we know that $\varepsilon_k \leq \beta_{\max} \delta_k \leq \omega_k$, so (18) holds. Therefore,

$$\| P_{T(x_k, \omega_k)}(-\nabla_x \Phi_k) \| \leq \| P_{\hat{T}(x_k, \varepsilon_k)}(-\nabla_x \Phi_k) \|.$$

Combining this with (21) yields the result. □

The next proposition is central to our approach. It says that the asymptotic behavior of $\| P_{T_k(x_k, \omega_k)}(-\nabla_x \Phi_k) \|$ in Algorithm 4.1 is like its behavior in the original algorithm.

**Proposition 4.2** *Suppose that AS1–AS3 hold. Then there exists a constant $C > 0$, independent of $k$, such that the following holds at outer iteration $k$ of Algorithm 4.1:*

$$\| P_{T(x_k, \omega_k)}(-\nabla_x \Phi_k) \| \leq C \omega_k. \tag{22}$$

**Proof.** From Proposition 4.1 we have

$$\| P_{T(x_k, \omega_k)}(-\nabla_x \Phi_k) \| \leq \frac{1}{\hat{\kappa}_{\min}} \left( M \left( 1 + \| \lambda_k \| + \sum_{j=1}^{q} 1/\mu_{k,j} \right) \delta_k \beta_{\max} + \frac{\hat{\alpha}}{\beta_{\min}} \delta_k \right),$$

where $M$ is the Lipschitz constant appearing in Proposition 4.1. The upper bounds on $\delta_k$ from (16) tell us that $\delta_k \leq \omega_k / \beta_{\max}$ as well as

$$\left( 1 + \| \lambda_k \| + \sum_{j=1}^{q} 1/\mu_{k,j} \right) \beta_{\max} \, \delta_k \leq \theta_{\text{tol}} \, \omega_k.$$

Therefore

$$\| \, P_{T(x_k,\omega_k)}(-\nabla_x \Phi_k) \, \| \leq \frac{1}{\hat{\kappa}_{\min}} \left( M \, \theta_{\mathrm{tol}} \, \omega_k + \frac{\hat{\alpha}}{\beta_{\min}} \frac{\omega_k}{\beta_{\max}} \right)$$

and the result follows. $\square$

# 5 The applicable convergence results from Conn, Gould, Sartenaer, and Toint

The import of Proposition 4.2 is that the convergence analysis for the original algorithm can be applied and the original proofs of these results still hold with only minor changes involving the values of some constants that appear (see Appendix A for details). We now briefly review some of the convergence properties of the augmented Lagrangian algorithm from [5] that hold for our GSS adaptation. Before doing so, we recall a little more notation and one additional assumption.

Suppose $\{x_k\}_{k\in\mathcal{K}}$ is a subsequence that converges to $x_*$. We denote by $A_*$ the matrix whose rows are the linear constraints that are binding (i.e., hold as equalities) at $x_*$, and denote by $Z_*$ a matrix whose columns form an orthonormal basis for the nullspace of $A_*$. If $J(x)$ is the Jacobian of $c(x)$, then the least-squares multiplier estimate at $x_*$ corresponding to $A_*$ is defined to be

$$\lambda_* \equiv -((J(x_*)Z_*)^+)^T Z_*^T \nabla f(x_*).$$

**AS4** *[5, p. 681] The matrix $J(x_*)Z_*$ has column rank no smaller than $m$ at any limit point $x_*$ of the sequence of $\{x_k\}$.*

The fundamental convergence result for Algorithm 4.1 corresponds to Theorem 4.6 in [5, p. 686].

**Theorem 5.1** *Assume that AS1 and AS2 hold. Let $x_*$ be any limit point of the sequence $\{x_k\}$ generated by Algorithm 4.1 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*

*(i) there are positive constants $\kappa_2$ and $\kappa_3$ such that*

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

$$\|\lambda(x_k) - \lambda_*\| \le \kappa_3 \|x_k - x_*\|,$$

*and*

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

*for all $j = 1, \ldots, q$ and all $k \in \mathcal{K}$ sufficiently large.*

*(ii) $x_*$ is a Karush–Kuhn–Tucker point (first-order stationary point) for the problem (1), $\lambda_*$ is the corresponding vector of Lagrange multipliers, and the sequence $\{\bar{\lambda}(x_k, \lambda_k, \mu_k)\}$ converges to $\lambda_*$ for $k \in \mathcal{K}$.*

Stronger results follow under additional assumptions on the regularity of $f$ and $c$ and on the stability of the reduced KKT system under small perturbations of the problem. Lemma 5.3 of [5, p. 689] then relates the convergence of the iterates to the error in the multipliers, a relationship characteristic of augmented Lagrangian methods [2]. Finally, if $\{x_k\}$ has a single limit point $x_*$, then Theorem 5.6 of [5, p. 695] says that we may reasonably expect the penalty parameters $\mu_{k,j}$ to remain bounded away from zero and Theorem 5.7 of [5, p. 696] gives a rate of convergence result for the outer iteration.

# 6 Updating the subproblem stopping tolerance

The quantity $\theta$, defined in (13), figures in the update (15) of the stopping criterion $\delta_k$ for the augmented Lagrangian subproblems. It provides a mechanism to dealing with the nonlinearity of the augmented Lagrangian that can occur for very small values of the weights $\mu_{k,j}$ or very large values of the multiplier estimates $\lambda_k$.

As defined in (20), the Lipschitz constant $M_k$ for $\nabla_x \Phi(x; \lambda_k, \mu_k)$ depends in an essential way on $\lambda_k$ and $\mu_k$. This, in turn, affects the relationship between the stationarity condition (6) and $\delta$, as can be seen in the bound (22) in Proposition 4.1. The Lipschitz constant $M_k$ will increase as the penalty parameters $\mu_{k,j}$ decrease or the multipliers $\lambda_k$ increase in magnitude. To counter this effect we must tighten the stopping tolerance accordingly.

A device similar to $\theta$ was used in [16]. However, independent testing revealed that, for a few test problems, the update rule from [16] causes trouble because the stopping tolerance $\delta_k$ quickly becomes very small [13]. This leads to the subproblems being over-solved, with an attendant increase in the overall computational cost. Further experiments indicated there was rarely a disadvantage to omitting the rescaling by $\theta$ in the update of $\delta_k$, while there were some dramatic improvements in efficiency [13].

For this reason we define $\theta$ in (13) so that it becomes active only if $1 + \| \lambda \| + \sum_{j=1}^{q} 1/\mu_{k,j}$ exceeds the prescribed threshold $\theta_{\text{tol}}$. In this way, if $\theta_{\text{tol}}$ is sufficiently large, and the penalty parameters $\mu_{k,j}$ remain uniformly bounded away from zero (as is the case in Theorem 5.6 from [5], for instance), and the multipliers are converging to their correct values, then we have $\delta_k = \omega_k / \beta_{\max}$ for all $k$, and avoid a rapid decrease in $\delta_k$. This threshold trigger $\theta_{\text{tol}}$ was not present in the update rule for $\delta_k$ in [16].

The choice of $\theta$ in (13) is still sufficient to prove convergence of Algorithm 4.1 even if some of the penalty parameters tend to zero. As a practical matter, however, we do not expect a GSS algorithm to be efficient in this case. The augmented Lagrangian in (8) will become increasingly nonlinear and ill-conditioned as the penalty parameters become very small and GSS algorithms tend to converge slowly when confronted with badly scaled problems.

The difficulty here, the unbounded nonlinearity of the augmented Lagrangian if some of the penalty parameters tend to zero, also arises in the original Algorithm 2.1 if one solves the subproblem (4) using finite differences to estimate the Jacobian of the constraints. In this context, the nonlinearity surfaces in the truncation error of the finite difference estimates. If some of the penalty parameters tend to zero, then the finite difference perturbation used will need to decrease more quickly than the $\mu_{k,j}$ in order to control the truncation error and have assurance that if the finite difference approximation of $\nabla_x \Phi_k$ satisfies (6), then the exact gradient $\nabla_x \Phi_k$ does as well.

This page intentionally left blank.

# 7 Application to inequality constrained minimization

In the framework considered here, applying a standard approach to dealing with the nonlinearly inequality constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c(x) \geq 0 \\ & Ax \geq b \end{array}$$

leads to an augmented Lagrangian of the form

$$\Phi(x, z(x); \lambda, \mu) = f(x) + \sum_{j=1}^{q} \frac{\mu_j}{2} \sum_{i \in Q_j} \left( \left[ \max\left\{0,\ \lambda_i + \frac{1}{\mu_j} c_i(x) \right\} \right]^2 - \lambda_i^2 \right)$$

and the solution of successive subproblems of the form

$$\begin{array}{ll} \text{minimize} & \Phi(x, z(x); \lambda, \mu) \\ \text{subject to} & Ax \geq b. \end{array}$$

The multiplier update formula (7) is also modified:

$$\bar{\lambda}(x, \lambda_{[Q_j]}, \mu_j)_{[Q_j]} = \max\left\{0,\ \lambda_{[Q_j]} + c(x)_{[Q_j]}/\mu_j \right\},\ j = 1, \ldots, q.$$

See [2, Chapter 3] for details.

This page intentionally left blank.

# 8 Conclusion

We have crafted a derivative-free GSS augmented Lagrangian algorithm for optimization with a combination of general and linear constraints based on the augmented Lagrangian framework of Conn, Gould, Sartenaer, and Toint [5]. To do so we use a linearly constrained GSS method [12] to solve the subproblems (4) and replace the derivative-based stopping criterion (6) with the derivative-free stopping criterion (12). In Proposition 4.2 we have shown that this substitution still allows us to satisfy the optimality condition (6). As a consequence the derivative-free adaptation inherits the first-order convergence properties of the original augmented Lagrangian algorithm (Theorem 5.1), even in the absence of explicit knowledge of derivatives for $f$ and $c$. This extends the results in [16], which dealt directly with bound constraints only, just as [5] extends the results in [6] (upon which [16] is based).

In addition, we have improved upon the update rule introduced in [16] for the sequence of stopping tolerances for the subproblems. The new rule relaxes the stringency of the test for an approximate solution to subproblem (4) while still satisfying (6), even if some of the penalty parameters tend to zero.

This page intentionally left blank.

# References


[1] M. Avriel, *Nonlinear Programming: Analysis and Methods*, Prentice-Hall, Englewood Cliffs, New Jersey, 1976. Reissued by Dover Publications, Mineola, New York, 2003.

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

[3] A. R. Conn, N. Gould, A. Sartenaer, and P. L. Toint, *Global convergence of two augmented Lagrangian algorithms for optimization with a combination of general equality and linear constraints*, Tech. Report 93/01, Department of Mathematics, FUNDP, Namur, Belgium, July 1993.

[4] ——, *Local convergence properties of two augmented Lagrangian algorithms for optimization with a combination of general equality and linear constraints*, Tech. Report 93/20, Department of Mathematics, FUNDP, Namur, Belgium, July 1993.

[5] ——, *Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints*, SIAM Journal on Optimization, 6 (1996), pp. 674–703.

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

[7] D. Davies and W. H. Swann, *Review of constrained optimization*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 187–202.

[8] A. V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York, 1983.

[9] J. D. Griffin, T. G. Kolda, and R. M. Lewis, *Asynchronous parallel generating set search for linearly-constrained optimization*, Tech. Report SAND2006-4621, Sandia National Laboratories, Albuquerque, NM and Livermore, CA, July 2006.

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

[11] T. G. Kolda, R. M. Lewis, and V. Torczon, *Optimization by direct search: New perspectives on some classical and modern methods*, SIAM Review, 45 (2003), pp. 385–482.

[12] ——, *Stationarity results for generating set search for linearly constrained optimization*, Tech. Report SAND2003-8550, Sandia National Laboratories, Livermore, California, 2003. Revised version to appear in *SIAM Journal on Optimization.*

[13] R. KUMAR. Private communication, 2005.

[14] ——, *An augmented Lagrangian pattern search algorithm for linear and nonlinear constrained optimization.* Talk presented at the SIAM Annual Meeting, Boston, Massachusetts, July 11, 2006.

[15] R. M. LEWIS, A. SHEPHERD, AND V. TORCZON, *Implementing generating set search methods for linearly constrained minimization*, Tech. Report WM–CS–2005–01, Department of Computer Science, College of William and Mary, July 2005. Revised July 2006 for *SIAM Journal on Scientific Computing.*

[16] R. M. LEWIS AND V. TORCZON, *A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds*, SIAM Journal on Optimization, 12 (2002), pp. 1075–1089.

[17] S. LUCIDI, M. SCIANDRONE, AND P. TSENG, *Objective-derivative-free methods for constrained optimization*, Mathematical Programming, 92 (2002), pp. 37–59.

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

[19] M. J. D. POWELL, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.

[20] R. A. TAPIA, *Diagonalized multiplier methods and quasi-Newton methods for constrained optimization*, Journal of Optimization Theory and Applications, 22 (1977), pp. 135–194.

[21] W. YU AND Y. LI, *A direct search method by the local positive basis for the nonlinearly constrained optimization*, Chinese Annals of Mathematics, 2 (1981), pp. 269–280.

# A Global convergence analysis

Our goal in this section is to demonstrate that the key convergence results from [5] still hold when we substitute our stopping criterion (12) for the original stopping condition (6). This reduces to an exercise in chasing through the constant $C$ that appears on the right-hand side of (22) to ensure that it does not change the conclusions of the results we need from [5]. Most of the presentation that follows is reproduced verbatim from [5]. The changes required to accommodate the stopping criterion we have substituted are highlighted.

Let $g(x)$ denote the gradient $\nabla_x f(x)$ of $f(x)$, with $g_k = g(x_k)$. Let $g^\ell(x, \lambda)$ and $H^\ell(x, \lambda)$, respectively, denote the gradient, $\nabla_x \ell(x, \lambda)$, and the Hessian matrix, $\nabla_{xx} \ell(x, \lambda)$, of the Lagrangian function

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

We note that $\ell(x, \lambda)$ is the Lagrangian solely with respect to the $c_i$ constraints.

Recalling the definitions of $A_*$, $Z_*$, and $J(x)$ from Section 5, we define the *least-squares Lagrange multiplier estimates* (corresponding to $A_*$) as

$$\lambda(x) \stackrel{\text{def}}{=} -((J(x)Z_*)^+)^T Z_*^T g(x) \tag{23}$$

at all points where the right generalized inverse

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

of $J(x)Z_*$ is well defined. We note that whenever $J(x)Z_*$ has full rank, $\lambda(x)$ is differentiable and its derivative is given in the following lemma.

**Lemma A.1** *[5, Lemma 2.1, p. 677] 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*

$$\nabla_x \lambda(x) = -((J(x)Z_*)^+)^T Z_*^T H^\ell(x, \lambda(x)) - (J(x) Z_* Z_*^T J(x)^T)^{-1} R(x), \tag{24}$$

*where the ith row of $R(x)$ is $(Z_*^T g(x) + Z_*^T J(x)^T \lambda(x))^T Z_*^T \nabla_{xx} c_i(x)$.*

We rewrite (7) in the compact form

$$\bar{\lambda}_k = \bar{\lambda}(x_k, \lambda_k, \mu_k). \tag{25}$$

We shall use the identity

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

which we then write using the compact notation

$$\nabla_x \Phi_k = \nabla_x \Phi(x_k;\lambda_k,\mu_k) = g^{\ell}(x_k,\bar{\lambda}_k). \tag{26}$$

Once $x_k$ satisfying (6) has been determined by the inner iteration, we denote

$$D_k = D(x_k,\omega_k), \quad N_k = N(x_k,\omega_k), \quad T_k = T(x_k,\omega_k). \tag{27}$$

Denote by $A_{D_k}$ the submatrix of $A$ consisting of the row(s) whose index is in $D_k$. For future reference, we define $Z_k$ to be a matrix whose columns form an orthonormal basis of $\mathcal{V}_k$, the null space of $A_{D_k}$, and $Y_k$ to be a matrix whose columns form an orthonormal basis of $\mathcal{W}_k = \mathcal{V}_k^{\perp}$. We note that $\mathcal{V}_k \subseteq T_k$ and, hence, that

$$\| Z_k^T \nabla_x \Phi_k \| = \| Z_k Z_k^T \nabla_x \Phi_k \| \leq \| P_{T_k}(-\nabla_x \Phi_k) \|, \tag{28}$$

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

Recall that the stopping criterion (12) for Algorithm 4.1 implies that under assumptions AS1–AS3 the result (22) of Proposition 4.2 holds. Thus

$$\Delta_k \leq \delta_k \;\Rightarrow\; \| P_{T_k}(-\nabla_x \Phi_k) \| \leq C\omega_k. \tag{29}$$

We notice that AS3 implies that there exists at least a convergent subsequence of iterates but does not, of course, guarantee that this subsequence converges to a stationary point, i.e., that "the algorithm works." We also note that it is always satisfied in practice because the linear constraints in (1) include lower and upper bounds on the variables, either actual or implied by the finite precision of computer arithmetic.

We now proceed to show that Algorithm 4.1 is globally convergent under the additional assumption AS4, which guarantees that the dimension of the null space of $A_*$ is large enough to provide the number of degrees of freedom that are necessary to satisfy the nonlinear constraints, and we require that the gradients of these constraints (projected onto this null space) are linearly independent at every limit point of the sequence of iterates. This assumption is the direct generalization of AS3 used by Conn, Gould, and Toint [6].

We shall analyze the convergence of Algorithm 4.1 in the case where the convergence tolerances $\omega_*$, and $\eta_*$ are both zero. We first need the following lemma, proving that (29) prevents both the reduced gradient of the augmented Lagrangian and its orthogonal complement from being arbitrarily large when $\omega_k$ is small.

**Lemma A.2** *[5, Lemma 4.1, p. 681] Let $\{x_k\} \subset \mathcal{B}$, $k \in \mathcal{K}$, be a sequence that converges to the point $x_*$ and suppose that*

$$\| P_{T_k}(-\nabla_x \Phi_k) \| \leq C\omega_k,$$

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

$$\| Z_*^T \nabla_x \Phi_k \| \leq \| Z_k^T \nabla_x \Phi_k \| \leq C\omega_k \text{ and } \| Y_k^T(x_k - x_*) \| \leq \kappa_1 \omega_k \tag{30}$$

*for some $\kappa_1 > 0$ and for all $k \in \mathcal{K}$ sufficiently large.*

**Proof.** Observe that, for $k \in \mathcal{K}$ sufficiently large, $\omega_k$ 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 $x_k$, $\mathcal{V}_k$, contains the subspace orthogonal to the normals of the constraints active at $x_*$. Hence, we deduce that

$$\| Z_*^T \nabla_x \Phi_k \| \leq \| Z_k^T \nabla_x \Phi_k \| \leq \| P_{T_k}(-\nabla_x \Phi_k) \| \leq C\omega_k,$$

where we have used (28) to obtain the second inequality and (29) to deduce the third. This proves the first part of (30).

We now turn to the second. If $D_k$ is empty, then $Y_k$ is the zero matrix and the second part of (30) 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 maximal full row-rank and note that the orthogonal projection onto the subspace spanned by the $\{a_i\}_{i \in D_k}$ is nothing but

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

Hence we obtain from the orthogonality of $Y_k$, the bound $| D_k | \leq p$, (3) and (27), and the fact that all the constraints in $D_k$ are active at $x_*$ for $k$ sufficiently large, that

$$\begin{aligned} \| Y_k^T(x_k - x_*) \| &\leq \| \hat{A}_{D_k}^T [\hat{A}_{D_k} \hat{A}_{D_k}^T]^{-1} \| \cdot \| \hat{A}_{D_k}(x_k - x_*) \| \\ &\leq \| \hat{A}_{D_k}^T [\hat{A}_{D_k} \hat{A}_{D_k}^T]^{-1} \| \, p\kappa_0 \omega_k. \end{aligned} \tag{31}$$

But there are [sic] only a finite number of nonempty sets $D_k$ for all possible choices of $x_k$ and $\omega_k$, and we may thus deduce the second part of (30) from (31) by defining

$$\kappa_1 = p\kappa_0 \min \| \hat{A}_{D_k}^T [\hat{A}_{D_k} \hat{A}_{D_k}^T]^{-1} \|,$$

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

We now examine the behavior 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, in [8, pp. 14–17].

**Lemma A.3** *[5, Lemma 4.2, p. 682] Assume that $U$ is a continuous point-to-set mapping from $S \subset \mathbb{R}^\ell$ into the power set of $\mathbb{R}^n$ such that the set $U(\theta)$ is convex and nonempty for each $\theta \in S$. Assume that the real-valued function $F(y, \theta)$ is defined and continuous on the space $\mathbb{R}^n \times S$ and convex in $y$ for each fixed $\theta$. Then the real-valued function $F_*$ defined by*

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

*is continuous on $S$.*

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

**Lemma A.4** *[5, Lemma 4.3, p. 682] Let $\{x_k\} \subset \mathcal{B}$, $k \in \mathcal{K}$, be a sequence that 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 (29) holds for $k \in \mathcal{K}$ and that $\omega_k$ tends to zero as $k \in \mathcal{K}$ increases. Then*

$$\nabla_x \Phi_* = A_*^T \pi_*$$

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

**Proof.** We first define

$$\sigma_k \stackrel{\text{def}}{=} \max_{\substack{A(x_k+d)-b \geq 0 \\ \| d \| \leq 1}} (-\nabla_x \Phi_k^T d) \tag{32}$$

with the aim of showing that this quantity tends to zero when $k \in \mathcal{K}$ increases. We obtain from (32), the Moreau decomposition [18] of $\nabla_x \Phi_k$, and the Cauchy–Schwarz inequality that

$$\begin{aligned}\sigma_k &\leq \max_{\substack{A(x_k+d)-b\geq 0\\ \| d \|\leq 1}} P_{T_k}(-\nabla_x \Phi_k)^T d + \max_{\substack{A(x_k+d)-b\geq 0\\ \| d \|\leq 1}} P_{N_k}(-\nabla_x \Phi_k)^T d \\ &\leq \| P_{T_k}(-\nabla_x \Phi_k) \| + \max_{d\in B_k} P_{N_k}(-\nabla_x \Phi_k)^T d,\end{aligned} \tag{33}$$

where $B_k \stackrel{\text{def}}{=} \{ d \in \mathbb{R}^n \mid a_i^T(x_k+d) - b_i \geq 0 \ (i \in D_k) \text{ and } \| d \| \leq 1 \}$. Since, for $x_k$ sufficiently close to $x_*$ and $\omega_k$ sufficiently small, all the constraints in $D_k$ must be active at $x_*$, we have that $N_k$ is included in the normal cone $N(x_*, 0)$ 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 (33) is a concave program, since $x_*$ is feasible for (1) and since $\| x_* - x_k \| \leq 1$ for $k \in \mathcal{K}$ large enough, we thus deduce that $d = x_* - x_k$ is a global solution of this problem. Observing that

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

we obtain

$$\max_{d\in B_k} P_{N_k}(-\nabla_x \Phi_k)^T d = \max_{d\in B_k} P_{N_k}(-\nabla_x \Phi_k)^T Y_k Y_k^T d \leq \|P_{N_k}(-\nabla_x \Phi_k)\| \cdot \|Y_k^T (x_k - x_*)\|, \tag{34}$$

where we have used the Cauchy–Schwarz inequality to deduce the last inequality. We may now apply Lemma A.2 and deduce from the second part of (30), (34), and the contractive character of the projection onto a convex set containing the origin that

$$\max_{d\in B_k} P_{N_k}(-\nabla_x \Phi_k)^T d \leq \kappa_1 \omega_k \| \nabla_x \Phi_k \|,$$

and thus, from (33) and our assumptions, that

$$\sigma_k \leq C\omega_k + \kappa_1 \omega_k \| \nabla_x \Phi_k \|.$$

Our assumption on the $\omega_k$ sequence then implies that $\sigma_k$ converges to zero as $k$ increases in $\mathcal{K}$.

Consider now the minimization problem

$$\begin{array}{ll}\min\limits_{d\in\mathbb{R}^n} & \nabla_x \Phi_*^T d \\ \text{subject to} & A(x_* + d) - b \geq 0 \\ & \| d \| \leq 1.\end{array} \tag{35}$$

Since the sequences $\{\nabla_x \Phi_k\}$ and $\{x_k\}$ converge to $\nabla_x \Phi_*$ and $x_*$, respectively, we deduce from Lemma A.3 applied to the optimization problem (32) (with the choices $\theta^T = (\nabla_x \Phi^T, x^T)$, $U(\theta) = \{ d \mid A(x+d) - b \geq 0, \| d \| \leq 1 \}$, $y = d$, $F(y, \theta) = \nabla_x \Phi^T d$), and the convergence of the sequence $\sigma_k$ to zero that the optimal value for problem (35) is zero. The vector $d = 0$ is thus a solution for problem (35) and satisfies

$$\nabla_x \Phi_* = A_*^T \pi_* - 2\zeta d = A_*^T \pi_*$$

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

The important part of our convergence analysis is the next lemma.

**Lemma A.5** *[5, Lemma 4.4, p. 683] Suppose that AS1 and AS2 hold. Let $\{x_k\} \subset \mathcal{B}$, $k \in \mathcal{K}$, be a sequence satisfying AS3 that converges to the point $x_*$ for which AS4 holds and let $\lambda_* = \lambda(x_*)$, where $\lambda$ satisfies (23). 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 q-dimensional vectors. Suppose further that (29) holds where the $\omega_k$ are positive scalar parameters that converge to zero as $k \in \mathcal{K}$ increases. Then*

*(i) there are positive constants $\kappa_2$ and $\kappa_3$ such that*

$$\| \bar{\lambda}(x_k, \lambda_k, \mu_k) - \lambda_* \| \leq \kappa_2 \omega_k + \kappa_3 \| x_k - x_* \|, \tag{36}$$

$$\| \lambda(x_k) - \lambda_* \| \leq \kappa_3 \| x_k - x_* \|, \tag{37}$$

*and,*

$$\| c(x_k)_{[\mathcal{Q}_j]} \| \leq \kappa_2 \omega_k \mu_{k,j} + \mu_{k,j} \| (\lambda_k - \lambda_*)_{[\mathcal{Q}_j]} \| + \kappa_3 \mu_{k,j} \| x_k - x_* \| \tag{38}$$

*for all $j = 1, \ldots, q$ and all $k \in \mathcal{K}$ sufficiently large.*

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

*(ii) $x_*$ is a Karush-Kuhn-Tucker point (first-order stationary point) for the problem (1), $\lambda_*$ is the corresponding vector of Lagrange multipliers, and the sequences $\{\bar{\lambda}(x_k, \lambda_k, \mu_k)\}$ and $\{\lambda(x_k)\}$ converge to $\lambda_*$ 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, $(J_k Z_*)^+$ exists, is bounded, and converges to $(J(x_*)Z_*)^+$. Thus, we may write

$$\| ((J_k Z_*)^+)^T \| \leq \kappa_2' \tag{39}$$

for some constant $\kappa_2' > 0$. Equations (26) and (25), the inner iteration termination criterion (29), and Lemma A.2 give that

$$\| Z_*^T(g_k + J_k^T \bar{\lambda}_k) \| \leq C\omega_k \tag{40}$$

for all $k \in \mathcal{K}$ large enough. By assumptions AS2, AS3, AS4, and (23), $\lambda(x)$ is bounded for all $x$ in a neighborhood of $x_*$. Thus we may deduce from (23), (39), and (40) that

$$\begin{aligned} \| \bar{\lambda}_k - \lambda(x_k) \| &= \| ((J_k Z_*)^+)^T Z_*^T g_k + \bar{\lambda}_k \| \\ &= \| ((J_k Z_*)^+)^T (Z_*^T g_k + (J_k Z_*)^T \bar{\lambda}_k) \| \\ &\leq \| ((J_k Z_*)^+)^T \| \, C \, \omega_k \\ &\leq C \, \kappa_2' \, \omega_k. \end{aligned} \tag{41}$$

Moreover, from the integral mean value theorem and Lemma A.1 we have that

$$\lambda(x_k) - \lambda(x_*) = \int_0^1 \nabla_x \lambda(x(s)) ds \cdot (x_k - x_*), \tag{42}$$

where $\nabla_x \lambda(x)$ is given by equation (24) and $x(s) = x_k + s(x_* - x_k)$ [sic]. Now the terms within the integral sign are bounded for all $x$ sufficiently close to $x_*$ and hence (42) gives

$$\| \lambda(x_k) - \lambda_*) \| \leq \kappa_3 \| x_k - x_* \| \tag{43}$$

for all $k \in \mathcal{K}$ sufficiently large and for some constant $\kappa_3 > 0$, which implies inequality (37). We then have that $\lambda(x_k)$ converges to $\lambda_*$. Combining (41) and (43) we obtain

$$\| \bar{\lambda}_k - \lambda_* \| \leq \| \bar{\lambda}_k - \lambda(x_k) \| + \| \lambda(x_k) - \lambda_* \| \leq C\kappa_2' \omega_k + \kappa_3 \| x_k - x_* \|, \tag{44}$$

which gives the required inequality (36) with $\kappa_2 = C\kappa_2{}'$. Then, since by assumption $\omega_k$ tends to zero as $k$ increases, (44) implies that $\bar{\lambda}_k$ converges to $\lambda_*$ and therefore, from the identity (26), $\nabla_x \Phi_k$ converges to $g^\ell(x_*, \lambda_*)$. Furthermore, multiplying (7) by $\mu_{k,j}$ we obtain

$$c(x_k)_{[\mathcal{Q}_j]} = \mu_{k,j}((\bar{\lambda}_k - \lambda_*)_{[\mathcal{Q}_j]} + (\lambda_* - \lambda_k)_{[\mathcal{Q}_j]}). \tag{45}$$

Taking norms of (45) and using (44), we derive (38).

Now suppose that

$$c(x_*) = 0. \tag{46}$$

Lemma A.4 and the convergence of $\nabla_x \Phi_k$ to $g^\ell(x_*, \lambda_*)$ give that

$$g(x_*) + J(x_*)^T \lambda_* = A_*^T \pi_*$$

for some vector $\pi_* \geq 0$. This last equation and (46) show that $x_*$ is a Karush–Kuhn–Tucker point and $\lambda_*$ is the corresponding set of Lagrange multipliers. Moreover, (36) and (37) ensure the convergence of the sequences $\{\bar{\lambda}(x_k, \lambda_k, \mu_k)\}$ and $\{\lambda(x_k)\}$ to $\lambda_*$ for $k \in \mathcal{K}$. Hence the lemma is proved. $\square$

No further changes to the original analysis in [5] are needed beyond this point.

# DISTRIBUTION:

3 Prof. Robert Michael Lewis
Dept. of Mathematics
College of William & Mary
P.O. Box 8795
Williamsburg, VA 23187-8795

3 Prof. Virginia Torczon
Dept. of Computer Science
College of William & Mary
P.O. Box 8795
Williamsburg, VA 23187-8795

1 Rakesh Kumar
The MathWorks, Inc.
3 Apple Hill Drive
Natick, MA 01760-2098

1 MS 0370
Tim Trucano, 1411

1 MS 1318
Danny Dunlavy, 1411

1 MS 1320
Scott Collis, 1414

1 MS 1318
Mike Eldred, 1411

1 MS 1318
Bill Hart, 1415

1 MS 1318
Scott Mitchell, 1411

5 MS 9159
Tamara G. Kolda, 8962

1 MS 9159
Joshua D. Griffin, 8962

1 MS 9159
Heidi Ammerlahn, 8962

1 MS 9159
Genetha Gray, 8962

1 MS 9159
Patty Hough, 8962

1 MS 9159
Kevin Long, 8962

1 MS 9159
Pam Williams, 8962

2 MS 9018
Central Technical Files, 8944

2 MS 0899
Technical Library, 4536