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

ROBERT MICHAEL LEWIS † AND VIRGINIA TORCZON ‡

Abstract. We give a pattern search method for nonlinearly constrained optimization that is an adaption of a bound-constrained augmented Lagrangian method first proposed by Conn, Gould, and Toint. In the pattern search adaptation we solve the bound-constrained subproblem approximately using a pattern search method. The stopping criterion proposed by Conn, Gould, and Toint for the solution of the subproblem requires explicit knowledge of derivatives. Such information is presumed absent in pattern search methods; however, we show how we can replace this with a stopping criterion based on the pattern size in a way that preserves the convergence properties of the original algorithm. In this way we proceed by successive, inexact, bound constrained minimization without knowing exactly how inexact the minimization is. So far as we know, this is the first provably convergent direct search method for general nonlinear programming.

Key words. augmented Lagrangian, constrained optimization, direct search, nonlinear programming, pattern search

Subject classification. Applied and Numerical Mathematics

1. Introduction. In this paper we consider the extension of pattern search methods to nonlinearly constrained optimization problems of the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c(x) = 0 \\ & \ell \le x \le u, \end{array} \tag{1.1}$$

where $f : I\!R^n \to I\!R$ and $c(x) = (c_1(x), \cdots, c_m(x))$. We allow the possibility that some of the variables are unbounded either above or below by permitting $\ell_j, u_j = \pm\infty$, $j \in \{1, \cdots, n\}$. This formulation assumes that any general inequality constraints have been converted into equality constraints by the introduction of nonnegative slack variables, leaving bounds as the only explicit inequality constraints.

The pattern search method presented here is an adaptation of an augmented Lagrangian method due to Conn, Gould, and Toint [4], which is the basis for the subroutine `AUGLG` in the `LANCELOT` optimization package [5]. The method of Conn, Gould, and Toint involves successive bound constrained minimization of an augmented Lagrangian. Since the analysis of pattern search methods has recently been extended to bound constrained minimization [17, 18], an adaptation of the augmented Lagrangian method of Conn, Gould, and

*Revision: 3.0 Date: 2001/09/06 22:47:56

†Department of Mathematics, College of William & Mary, P. O. Box 8795, Williamsburg, Virginia 23187–8795, `buckaroo@math.wm.edu`. This research was supported by the National Aeronautics and Space Administration under NASA Contract No. NAS1-97046 while the author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, Virginia 23681–2199.

‡Department of Computer Science, College of William & Mary, P. O. Box 8795, Williamsburg, Virginia 23187–8795, `va@cs.wm.edu`. This research was supported by the National Science Foundation under Grant CCR-9734044 and by the National Aeronautics and Space Administration under NASA Contract No. NAS1-97046 while the author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, Virginia 23681–2199.

Toint to pattern search naturally suggests itself. Furthermore, the multiplier update of Algorithm 1 in [4] does not involve information about derivatives of the objective or constraints, so the augmented Lagrangian approach is consistent with the derivative-free nature of pattern search algorithms.

Since there exist broad classes of pattern search methods for unconstrained [16, 28] and bound constrained minimization [17, 18], it seems to us natural to first extend pattern search methods to nonlinearly constrained minimization via algorithms that proceed by successive unconstrained or bound constrained minimization, such as the augmented Lagrangian method we discuss here. In the absence of information about derivatives of the objective and constraints, it is difficult to design pattern search algorithms for general nonlinearly constrained minimization that produce only feasible directions or feasible iterates. This is due to the fact that a pattern in a pattern search algorithm would need to include a sufficiently rich set of search directions to capture any feasible improvement in the objective. When nonlinear constraints are present, it is not clear how to design such a pattern without first-order information.

We show that despite the absence of an explicit estimation of any derivatives (a characteristic of pattern search methods), our pattern search augmented Lagrangian approach exhibits all of the first-order convergence properties of the original algorithm of Conn, Gould, and Toint. This at first is surprising, since the original algorithm allows its subproblems to be solved approximately, and the stopping criterion for the solution of the subproblems is based on the magnitude of a measure of first-order stationarity for bound constrained minimization. This information is not explicitly available in a direct search method. However, as we discuss in Section 5.1, there is a correlation between the size of the pattern in bound constrained pattern search and the amount of local feasible descent. Using this correlation we are able to establish convergence to Karush–Kuhn–Tucker points of (1.1) even without explicit knowledge of derivatives. That is, we are able to proceed by successive, inexact minimization of the augmented Lagrangian via pattern search methods, even without knowing exactly how inexact the minimization is.

This is the main contribution of the work presented here. Otherwise, the extension of pattern search to constrained minimization by means of the augmented Lagrangian approach of Conn, Gould, and Toint is straightforward, due to the strength and generality of the convergence analysis presented in [4].

The question of treating general nonlinear constraints with direct search minimization algorithms has a long history, beginning with the original work on direct search methods. Rosenbrock, in [24], proposed treating constraints using his rotating directions method by redefining the objective near the boundary of the feasible region in a way that would tend to keep the iterates feasible, a form of penalization. Similar ideas for modifying the objective in the case of bound constraints are discussed by Spendley, Hext, and Himsworth [26] and Nelder and Mead [21] in connection with their simplex-based methods. In these approaches the objective is given a suitably large value (in the case of minimization) at all infeasible points.

More systematic approaches to penalization have also appeared. The treatment of inequality constraints via exact, nonsmooth penalization (though not by that name) appears as early as the work of Hooke and Jeeves [12]. More recently, Kearsley and Glowinski [10, 13] have applied pattern search methods with exact, nonsmooth penalization to equality constrained problems arising in control. Weisman's `MINIMAL` algorithm [11] applies the pattern search algorithm of Hooke and Jeeves to a nonsmooth quadratic penalty function and incorporates an element of random search. Davies and Swann [6], in connection with applying the pattern search method of Hooke and Jeeves to constrained optimization, recommend the use of the reciprocal barrier method of Carroll [3, 8].

A direct search method for constrained minimization that has proven popular in application is the Complex method of Box [2], which was originally developed to address difficulties encountered with Rosenbrock's method. In this algorithm, the objective is sampled at a broader set of points than in the simplex-based methods to try to avoid premature termination. There is also an element of random search involved. The `ACSIM` algorithm of Dixon [7] combines ideas from the simplex method of Nelder and Mead and the Complex method with elements of hem-stitching and quadratic modeling to accelerate convergence.

In the special case of bound constraints, Spendley also suggested the expedient of simply setting to the corresponding bound any variable that would otherwise become infeasible when applying the simplex algorithm of Nelder and Mead [25]. In [14], Keefer proposed a hybrid, feasible iterates algorithm for bound constrained minimization that uses the algorithm of Nelder and Mead for variables suitably far from their bounds and the method of Hooke and Jeeves for variables that are on or near one of their bounds, since the pattern in the algorithm of Hooke and Jeeves conforms in a natural way to the boundary of the feasible region. In the case of linear constraints there is the algorithm of May [19], which is an extension of Mifflin's derivative-free unconstrained minimization method in [20]. May's algorithm also takes into account the particular geometry of the feasible region. May's algorithm is notable because it is accompanied by convergence analysis results; however, it is not a direct search method insofar as it does rely on a model of the objective.

Others have proposed modifications of the method of Hooke and Jeeves along the lines of feasible directions algorithms. These methods involve a limited calculation of sensitivity information to compute feasible directions at the boundary of the feasible region if the algorithm appears to have stalled. Klingman and Himmelblau [15] give an algorithm with a simple construction of a suitable feasible direction. The method of Glass and Cooper [9] is more sophisticated, and computes a new search direction by solving a linear programming problem involving a linear approximation of the objective and constraints, just as one would in a derivative-based feasible directions algorithm.

Finally, we note the flexible tolerance method of Paviani and Himmelblau [11, 22]. This algorithm, based on the method of Nelder and Mead, alternatively attempts to reduce the objective and constraint violation, depending on the extent to which the iterates are infeasible.

These proposals for direct search algorithms for constrained minimization have often proven effective in practice, but have not been accompanied by any convergence analysis. In historical context, this is not surprising. The early development of direct search methods (particularly the work cited here) predates even the first global convergence analysis of practical unconstrained minimization algorithms using the Armijo–Goldstein–Wolfe conditions. Instead, in the 1960s the emphasis in optimization was the development of new computational methods, not on proving theoretical properties. And, in fact, some of the heuristics in the approaches discussed above do not always work in practice. For instance, see Box's comments on Rosenbrock's method in [2], and Keefer's comments on Box's method in [14].

Nevertheless, some of the heuristics proposed in this early research on direct search methods can be placed on firm theoretical grounds. For instance, Keefer's observation that the pattern search method of Hooke and Jeeves works particularly well for bound constrained problems can be explained analytically [17]. In this paper we apply analytical and algorithmic advances since the original development of direct search methods to construct a direct search method for general nonlinear programming with provable first-order global convergence properties.

2. The augmented Lagrangian method of Conn, Gould, and Toint. We base our augmented Lagrangian pattern search method on Algorithm 1 of [4]. To facilitate comparison of the pattern search approach with the original algorithm, we adhere to the notation of [4] throughout.

The augmented Lagrangian in [4] is

$$\Phi(x;\lambda,S,\mu) = f(x) + \sum_{i=1}^{m} \lambda_i c_i(x) + \frac{1}{2\mu}\sum_{i=1}^{m} s_{ii} c_i(x)^2. \tag{2.1}$$

The vector $\lambda = (\lambda_1, \cdots, \lambda_m)^T$ is the Lagrange multiplier estimate for the equality constraints, μ is the penalty parameter, and the entries s_{ii} of the diagonal matrix S are positive weights. The equality constraints of (1.1) are incorporated in the augmented Lagrangian Φ while the simple bounds are left explicit. For a particular choice of multiplier estimate $\lambda^{(k)}$, penalty parameter $\mu^{(k)}$, and scaling $S^{(k)}$, we define

$$\Phi^{(k)}(x) = \Phi(x; \lambda^{(k)}, S^{(k)}, \mu^{(k)}).$$

Following [4], unless otherwise indicated by an explicit argument, $\nabla_x \Phi^{(k)}$ denotes

$$\nabla_x \Phi^{(k)} \equiv \nabla_x \Phi^{(k)}(x^{(k)}) = \nabla_x \Phi(x^{(k)}; \lambda^{(k)}, S^{(k)}, \mu^{(k)})$$

for the iterate $x^{(k)}$.

Conn, Gould, and Toint define the first-order Lagrange multiplier update to be

$$\overline{\lambda}(x, \lambda, S, \mu) = \lambda + Sc(x)/\mu. \tag{2.2}$$

This is the Hestenes–Powell multiplier update for the augmented Lagrangian (2.1). For the purposes of a pattern search augmented Lagrangian approach, which assumes no explicit knowledge of derivative information, one appears to have no choice other than some variant of the Hestenes–Powell multiplier update, since other multiplier update formulae (such as those discussed in [1, 27]) require information about derivatives.

We denote by P the projection onto the set $B = \{\, x \mid \ell \le x \le u \,\}$; P is defined component-wise by

$$(P[x])_i = \begin{cases} \ell_i & \text{if } x_i \le \ell_i \\ u_i & \text{if } x_i \ge u_i \\ x_i & \text{otherwise.} \end{cases}$$

Given $x \in B$ and a vector v, we define

$$P(x, v) = x - P[x - v].$$

The geometrical meaning of $P(x,v)$ is illustrated in Figure 2.1. If x is interior to B, then $P(x,v) = 0$ if and only if $v = 0$, while if x is on the boundary of B, then $P(x,v) = 0$ if and only if v is normal to B (in the sense of convex analysis).

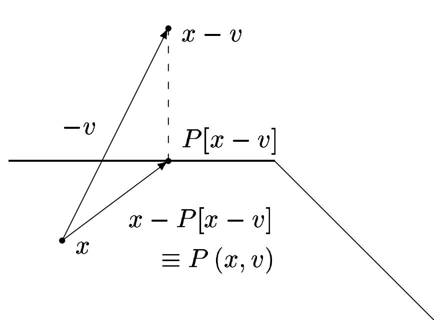

FIG. 2.1. *An example of* $P(\cdot, \cdot)$.

At iteration k of the original augmented Lagrangian algorithm described in [4], we approximately solve the subproblem

$$\begin{array}{ll} \text{minimize} & \Phi^{(k)}(x) \\ \text{subject to} & \ell \leq x \leq u. \end{array} \tag{2.3}$$

The degree to which this subproblem must be solved is given by

$$\| \, P(x^{(k)}, \nabla_x \Phi^{(k)}) \, \| \leq \omega^{(k)}, \tag{2.4}$$

where $\omega^{(k)}$ is updated at each iteration k. (Unless otherwise noted, we use $\| \cdot \|$ to denote the Euclidean vector norm or its induced matrix norm.)

We adapt Algorithm 1 in [4] to pattern search by solving the bound constrained subproblem (2.3) using a bound constrained pattern search method. However, pattern search methods do not have recourse to derivatives or explicit approximations thereof.

For that reason we replace (2.4) with a new stopping criterion that is based on the size of the pattern. As we discuss in Section 5, we retain the convergence properties of the original Conn, Gould and Toint algorithm because the size of the pattern and the stationarity condition (2.4) are correlated, even though we do not have explicit control of $\| \, P(x^{(k)}, \nabla_x \Phi^{(k)}) \, \|$.

3. Bound constrained pattern search algorithms. We next review relevant features of the general pattern search method for the bound constrained problem

$$\begin{array}{ll} \text{minimize} & F(x) \\ \text{subject to} & \ell \leq x \leq u. \end{array} \tag{3.1}$$

We concentrate only on features that we need for the results that follow. For a full discussion, see [17, 18].

3.1. The bound constrained pattern search method. Fig. 3.1 outlines the generalized pattern search method for minimization with bound constraints. To define a particular pattern search method, we must specify the pattern (a set of possible trial directions) $\Pi^{(j)}$, the bound constrained exploratory moves algorithm used to find a feasible step $s^{(j)}$, and the algorithms for updating $\Pi^{(j)}$ and $\Delta^{(j)}$. The options and conditions accompanying these choices are discussed in [17, 18].

Let $x^{(0)} \in B$ and $\Delta^{(0)} > 0$ be given.
For $j = 0, 1, \cdots,$
- a) Compute $F(x^{(j)})$.
- b) Determine a step $s^{(j)}$ using a bound constrained exploratory moves algorithm.
- c) If $F(x^{(j)} + s^{(j)}) < F(x^{(j)})$, then $x^{(j+1)} = x^{(j)} + s^{(j)}$. Otherwise $x^{(j+1)} = x^{(j)}$.
- d) Update $\Pi^{(j)}$ and $\Delta^{(j)}$.

FIG. 3.1. *The Generalized Pattern Search Method for Bound Constrained Problems.*

We make use of the following observations.

1. At iteration j, the step $s^{(j)}$ must be in the set $\Delta^{(j)}\Pi^{(j)}$ and $x^{(j)} + s^{(j)}$ must be feasible. We allow the possibility $s^{(j)} = 0$.
2. The pattern $\Pi^{(j)}$ contains a distinguished subset of trial directions known as the *core pattern*, which we denote by $\Gamma^{(j)}$. The core pattern is constructed to ensure that if $x^{(j)}$ is not a constrained stationary point of (3.1), then at least one element p in $\Gamma^{(j)}$ is a feasible direction of descent. The

elements of $\Gamma^{(j)}$ are required to be uniformly bounded in norm: there exists d^*, independent of j, such that $\| p \| \leq d^*$ for all $p \in \Gamma^{(j)}$.

3. We may accept any step $s^{(j)}$ that yields simple decrease in F.
4. If

$$\min\{ F(x^{(j)} + s) \mid s \in \Delta^{(j)}\Gamma^{(j)},\ x^{(j)} + s \in B \} < F(x^{(j)}), \tag{3.2}$$

then the step $s^{(j)}$ returned by the bound constrained exploratory moves algorithm must also produce simple decrease on $F(x^{(j)})$. (Note, though, that $s^{(j)}$ need not be an element of $\Delta^{(j)}\Gamma^{(j)}$.)

5. The update of $\Delta^{(j)}$ depends on whether or not the step $s^{(j)}$ satisfied the simple decrease criterion.

3.2. The update of $\Delta^{(j)}$. The conditions under which we allow $\Delta^{(j)}$ to be *reduced* are at the heart of the results that follow. The aim of the update of $\Delta^{(j)}$ is to force a strict reduction in F. An iteration with $F(x^{(j)} + s^{(j)}) < F(x^{(j)})$ is *successful*; otherwise, the iteration is *unsuccessful.* We cannot update $\Delta^{(j)}$ in an arbitrary manner, as discussed in [17, 18]. However, for the purposes of analyzing the augmented Lagrangian pattern search algorithm the update of $\Delta^{(j)}$ can be summarized as

$$\text{If } F(x^{(j)} + s^{(j)}) < F(x^{(j)}), \text{ then } \Delta^{(j+1)} \geq \Delta^{(j)}. \tag{3.3}$$

$$\text{If } F(x^{(j)} + s^{(j)}) \geq F(x^{(j)}), \text{ then } \Delta^{(j+1)} < \Delta^{(j)}. \tag{3.4}$$

If an iteration is successful it may be possible to increase the scale factor $\Delta^{(j)}$, but $\Delta^{(j)}$ is not allowed to decrease. If an iteration is unsuccessful, the scale factor $\Delta^{(j)}$ must be decreased.

4. The pattern search augmented Lagrangian method. We now state the augmented Lagrangian pattern search algorithm. At iteration k in the outermost loop of the algorithm, we denote by $\{x^{(k,j)}\}$ the sequence of iterates produced in the solution of (2.3) via a bound constrained pattern search algorithm. Thus, for a given value of k, we look for an approximate solution of the subproblem (2.3) starting from $x^{(k,0)} = x^{(k)}$ and proceed until we find j^* such that $x^{(k,j^*)}$ solves (2.3) to an acceptable degree. We modify the original algorithm by replacing the stopping criterion (2.4) for the solution of the subproblems with one that is suitable for pattern search while still allowing us to use the analysis from [4].

In order to relate the stopping criterion in the pattern search solution of the subproblems to the multiplier estimates and the penalty parameter, we introduce the function

$$\theta(\lambda, \mu) = (1 + \| \lambda \| + 1/\mu)^{-1}.$$

We note that any function $\theta(\lambda, \mu)$ such that $\theta(\lambda, \mu) = O((\| \lambda \| + 1/\mu)^{-1})$ as $(\| \lambda \| + 1/\mu) \to \infty$ suffices for the purposes of proving convergence.

Our algorithm closely resembles Algorithm 1 in [4]. We use boxes to highlight the elements that differ.

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

Step 1 [Inner iteration]. Define a scaling matrix $S^{(k)}$ for which $S_1^{-1} \leq S^{(k)} \leq S_2$.

Set $x^{(k,0)} = x^{(k)}$. Apply a bound constrained pattern search method to

$$\begin{array}{lll} (4.1) & \text{minimize} & \Phi^{(k)}(x) \\ & \text{subject to} & \ell \le x \le u \end{array}$$

to find the first iteration j^* for which the scale factor is sufficiently small; that is,

$$\Delta^{(k,j^*)} \le \delta^{(k)}. \tag{4.2}$$

Set $x^{(k)} = x^{(k,j^*)}$.

If

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

execute Step 2. Otherwise, execute Step 3.

Step 2 [Test for convergence and update Lagrange multiplier estimates]. If $\boxed{\delta^{(k)} \le \delta^*}$ and $\| c(x^{(k)}) \| \le \eta_*$, stop. Otherwise, set

$$\begin{aligned}
\lambda^{(k+1)} &= \overline{\lambda}(x^{(k)}, \lambda^{(k)}, S^{(k)}, \mu^{(k)}) \\
\mu^{(k+1)} &= \mu^{(k)} \\
\alpha^{(k+1)} &= \min(\mu^{(k+1)}, \gamma_1) \\
\omega^{(k+1)} &= \omega^{(k)}(\alpha^{(k+1)})^{\beta_\omega} \\
&\boxed{\delta^{(k+1)} = \theta(\lambda^{(k+1)}, \mu^{(k+1)})\ \omega^{(k+1)}} \\
\eta^{(k+1)} &= \eta^{(k)}(\alpha^{(k+1)})^{\beta_\eta},
\end{aligned}$$

increment k by one and go to Step 1.

Step 3 [Reduce the penalty parameter]. Set

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

increment k by one and go to Step 1.

We have replaced the stopping criterion (2.4) for the inner iteration of Algorithm 1 in [4] with (4.2), which is based on the scale factor Δ, because we do not assume explicit information about the derivatives. The remaining modifications to Algorithm 1 in [4] concern the management of the sequence $\{\delta^{(k)}\}$, which controls the stopping criterion we have introduced.

5. Convergence analysis. We now discuss the convergence properties of the augmented Lagrangian pattern search algorithm. As we shall see, altering the original algorithm by solving the bound constrained subproblem via pattern search leaves the convergence properties of the original algorithm almost entirely unchanged.

In [4], Conn, Gould, and Toint call a component of $x^{(k)}$ *floating* if

$$\ell_i < x_i^{(k)} - (\nabla_x \Phi^{(k)})_i < u_i.$$

For a convergent subsequence $\{x^{(k)}\}$, $k \in K$, with limit point x^* they define the index set

$$I_1 = \{\ i \ \mid\ x_i^{(k)} \text{ are floating for all } k \in K \text{ sufficiently large and } \ell_i < x_i^* < u_i\ \},$$

and let $\hat{A}(x)$ denote the corresponding columns of the Jacobian of $c(x)$, where $A(x)$ is the entire Jacobian of $c(x)$.

The following assumptions are made in [4].

AS1. *The functions $f(x)$ and $c(x)$ are twice continuously differentiable for all $x \in B$.*

AS2. *The iterates $\{x^{(k)}\}$ considered lie within a closed, bounded domain Ω.*

AS3. *The matrix $\hat{A}(x^*)$ has column rank no smaller than m at any limit point x^* of the sequences $\{x^{(k)}\}$ considered in this paper.*

In addition, in order to be assured that a bound constrained pattern search algorithm applied to the subproblem (4.1) will find an iterate satisfying (4.2), we assume the following.

PS1. *For a given k, the set $B \cap \{\ x \ \mid\ \Phi^{(k)}(x) \leq \Phi^{(k)}(x^{(k,0)})\ \}$ is compact.*

That is, we assume compactness of the set of $x \in B$ for which the augmented Lagrangian is no larger than the value of the augmented Lagrangian at the point at which we begin the solution of the subproblem. Under hypothesis (PS1), we are assured that in the inner iteration (the pattern search minimization of the bound constrained augmented Lagrangian),

$$\liminf_{j \to +\infty} \Delta^{(k,j)} = 0$$

(see [17, 18]). Thus the termination criterion (4.2) will eventually be satisfied, the pattern search solution of the augmented Lagrangian subproblem will halt, and the overall iteration of the pattern search augmented Lagrangian algorithm is well-defined.

We also assume the following uniform bound.

PS2. *There exists d^* such that for all k and j, we have $\|\ p\ \| \leq d^*$ for all $p \in \Gamma^{(k,j)}$.*

This uniformity in the pattern search algorithms used in the successive minimization of the augmented Lagrangian is not at all restrictive. For instance, one could simply choose for all (k, j) a single set Γ.

5.1. The relationship between the pattern size and stationarity. The following result is the key to analyzing the augmented Lagrangian pattern search method. The important observation in connection with the stopping criterion (4.2) is that at unsuccessful iterations of the pattern search solution of (4.1) there is a correlation between $\Delta^{(k,j)}$ and the stationarity of the augmented Lagrangian. The rules for updating $\Delta^{(k,j)}$, summarized in (3.3) and (3.4), mean that $\Delta^{(k,j)}$ can drop below $\delta^{(k)}$ only at an unsuccessful iteration of the pattern search. Thus (4.2) can only occur at an unsuccessful iteration of the solution of the subproblem. At an unsuccessful iteration we do *not* find an acceptable step in $\Delta^{(k,j)}\Gamma^{(k,j)}$; that is,

$$\Phi^{(k)}(x^{(k,j)} + s) \geq \Phi^{(k)}(x^{(k,j)}) \quad \text{for all } s \in \Delta^{(k,j)}\Gamma^{(k,j)} \text{ with } (x^{(k,j)} + s) \in B.$$

Now, the set of steps s for $s \in \Delta^{(k,j)}\Gamma^{(k,j)}$ includes a set of generators for the tangent cone of the bound constrained feasible region [17, 18]. The fact that none of the steps s yield a feasible trial point with a smaller value of $\Phi^{(k)}$ tells us something about the size of $\| P(x^{(k)}, \nabla_x \Phi^{(k)}) \|$. Proposition 5.1 makes this precise, and shows that the weaker condition (4.2) we have introduced guarantees that (2.4) will be satisfied.

PROPOSITION 5.1. *There exists $C_{5.1}$, independent of k, such that*

$$\| P(x^{(k)}, \nabla_x \Phi^{(k)}) \| \leq C_{5.1}\ \omega^{(k)}$$

for all iterations k of the pattern search augmented Lagrangian method.

Proof. Given k, we know that at the end of Step 1, the inner iteration, $x^{(k)} \equiv x^{(k,j^*)}$ for some j^*. For convenience, let

$$q^{(k,j^*)} = P(x^{(k)}, \nabla_x \Phi^{(k)}) \equiv P(x^{(k,j^*)}, \nabla_x \Phi^{(k)}(x^{(k,j^*)})).$$

First suppose

$$\Delta^{(k,j^*)} \geq \frac{\| q^{(k,j^*)} \|_\infty}{d^*}. \tag{5.1}$$

Then (5.1), (4.2), and the rule for updating $\delta^{(k)}$ in either Step 2 or Step 3 give us

$$\| q^{(k,j^*)} \|_\infty \ \leq\ d^* \Delta^{(k,j^*)} \ \leq\ d^* \delta^{(k)} \ \leq\ d^* \omega^{(k)},$$

and thus

$$\| q^{(k,j^*)} \| \leq n^{\frac{1}{2}} d^* \omega^{(k)}. \tag{5.2}$$

On the other hand, suppose

$$\Delta^{(k,j^*)} < \frac{\| q^{(k,j^*)} \|_\infty}{d^*}.$$

The proof of Proposition 5.2 in [17] shows that if $\Delta^{(k,j^*)} < \| q^{(k,j^*)} \|_\infty / d^*$, there is a step $s \in \Delta^{(k,j^*)}\Gamma^{(k,j^*)}$ such that $x^{(k,j^*)} + s \in B$ and

$$\nabla_x \Phi^{(k)}(x^{(k,j^*)})^T s < -n^{-\frac{1}{2}} \| q^{(k,j^*)} \| \| s \|. \tag{5.3}$$

Because $x^{(k,j^*)}$ is an unsuccessful iterate, we know from (3.2) that

$$0 \leq \Phi^{(k)}(x^{(k,j^*)} + s) - \Phi^{(k)}(x^{(k,j^*)}). \tag{5.4}$$

At the same time we have

$$\Phi^{(k)}(x^{(k,j^*)} + s) - \Phi^{(k)}(x^{(k,j^*)}) = \nabla_x \Phi^{(k)}(\xi)^T s \tag{5.5}$$

for some ξ in the line segment $(x^{(k,j^*)}, x^{(k,j^*)} + s)$ connecting $x^{(k,j^*)}$ and $x^{(k,j^*)} + s$. Thus from (5.4), (5.5), and (5.3) we obtain

$$\begin{aligned} 0 &\leq \Phi^{(k)}(x^{(k,j^*)} + s) - \Phi^{(k)}(x^{(k,j^*)}) \\ &= \nabla_x \Phi^{(k)}(x^{(k,j^*)})^T s + (\nabla_x \Phi^{(k)}(\xi) - \nabla_x \Phi^{(k)}(x^{(k,j^*)}))^T s \\ &\leq -n^{-\frac{1}{2}} \| q^{(k,j^*)} \| \| s \| + \| \nabla_x \Phi^{(k)}(\xi) - \nabla_x \Phi^{(k)}(x^{(k,j^*)}) \| \| s \|, \end{aligned}$$

which yields

$$\| q^{(k,j^*)} \| \le n^{\frac{1}{2}} \| \nabla_x \Phi^{(k)}(\xi) - \nabla_x \Phi^{(k)}(x^{(k,j^*)}) \|. \tag{5.6}$$

Applying the mean-value theorem again, for some $\zeta \in (x^{(k,j^*)}, \xi)$ we have

$$\nabla_x \Phi^{(k)}(\xi) - \nabla_x \Phi^{(k)}(x^{(k,j^*)}) = \nabla^2_{xx} \Phi^{(k)}(\zeta)(\xi - x^{(k,j^*)}),$$

so

$$\| \nabla_x \Phi^{(k)}(\xi) - \nabla_x \Phi^{(k)}(x^{(k,j^*)}) \| \ \le\ \| \nabla^2_{xx} \Phi^{(k)}(\zeta) \| \| \xi - x^{(k,j^*)} \| \ \le\ \| \nabla^2_{xx} \Phi^{(k)}(\zeta) \| \| s \|. \tag{5.7}$$

Now,

$$\nabla^2_{xx} \Phi^{(k)}(\zeta) = \nabla^2_{xx} f(\zeta) + \sum_{i=1}^{m} \lambda_i^{(k)} \nabla^2 c_i(\zeta) + \frac{1}{\mu^{(k)}} (\nabla c(\zeta) S \nabla c(\zeta)^T + \sum_{i=1}^{m} s_{ii} c_i(\zeta) \nabla^2 c_i(\zeta)).$$

By construction, $\omega^{(k)} \to 0$, so $\delta^{(k)} \to 0$, so by (AS2), ζ lies in a compact subset that is independent of k. Furthermore, the bound $S^{(k)} \le S_2$ is independent of k. Thus we can find M, independent of k, such that

$$\| \nabla^2_{xx} \Phi^{(k)}(\zeta) \| \ \le\ M + M \| \lambda^{(k)} \| + M \frac{1}{\mu^{(k)}} \ =\ M / \theta(\lambda^{(k)}, \mu^{(k)}).$$

Returning to (5.7) we have

$$\| \nabla_x \Phi^{(k)}(\xi) - \nabla_x \Phi^{(k)}(x^{(k,j^*)}) \| \le \left(M / \theta(\lambda^{(k)}, \mu^{(k)}) \right) \| s \|. \tag{5.8}$$

Thus from (5.6), (5.8), the fact that $s \in \Delta^{(k,j^*)} \Gamma^{(k,j^*)}$, and (4.2) we have

$$\begin{aligned}
\| q^{(k,j^*)} \| &\le n^{\frac{1}{2}} \| \nabla_x \Phi^{(k)}(\xi) - \nabla_x \Phi^{(k)}(x^{(k,j^*)}) \| \\
&\le n^{\frac{1}{2}} \left(M / \theta(\lambda^{(k)}, \mu^{(k)}) \right) \| s \| \\
&\le n^{\frac{1}{2}} d^* \left(M / \theta(\lambda^{(k)}, \mu^{(k)}) \right) \Delta^{(k,j^*)} \\
&\le n^{\frac{1}{2}} d^* \left(M / \theta(\lambda^{(k)}, \mu^{(k)}) \right) \delta^{(k)}.
\end{aligned}$$

Finally, the rule for updating $\delta^{(k)}$ in either Step 2 or Step 3 is $\delta^{(k)} = \theta(\lambda^{(k)}, \mu^{(k)}) \omega^{(k)}$, whence

$$\| q^{(k,j^*)} \| \le n^{\frac{1}{2}} d^* M \omega^{(k)}. \tag{5.9}$$

Combining (5.2) and (5.9) yields the proposition. □

5.2. Convergence results. Proposition 5.1 means that the asymptotic behavior of $\| P(x^{(k)}, \nabla_x \Phi^{(k)}) \|$ in the augmented Lagrangian pattern search algorithm is like that of the same quantity in the original algorithm. This, in turn, allows us to piggy-back the convergence analysis for the augmented Lagrangian pattern search algorithm on that for the original augmented Lagrangian algorithm in [4]. Because of Proposition 5.1 the original proofs of all these results still hold.

The first convergence result corresponds to Theorem 4.4 and Lemma 4.3 in [4]. Let

$$g_L(x; \lambda) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla c_i(x).$$

THEOREM 5.2. *Assume that (AS1) holds. Let x^* be any limit point of the sequence $\{x^{(k)}\}$ generated by the augmented Lagrangian pattern search algorithm for which (AS2) and (AS3) hold and let K be the set of indices of an infinite subsequence of the $x^{(k)}$ whose limit is x^*. Then*

(i) $c(x^*) = 0$.

(ii) x^* is a Karush-Kuhn-Tucker point (first-order stationary point) for the problem (1.1), λ^* is the corresponding vector of Lagrange multipliers, and the sequence $\{\overline{\lambda}(x^{(k)}, \lambda^{(k)}, S^{(k)}, \mu^{(k)})\}$ converges to λ^* for $k \in K$.

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

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

and

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

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

(iv) The gradients $\nabla_x \Phi^{(k)}$ converge to $g_L(x^*; \lambda^*)$ for $k \in K$.

As in [4], under additional assumptions we obtain stronger results. Following [4], if J_1 and J_2 are any index sets, and $H_L(x^*, \lambda^*)$ is the Hessian of the Lagrangian, then $H_L(x^*, \lambda^*)_{[J_1, J_2]}$ is the matrix formed by taking the rows and columns of $H_L(x^*, \lambda^*)$ indexed by J_1 and J_2, respectively, while $A(x^*)_{[J_1]}$ is the matrix formed by taking the columns of $A(x^*)$ indexed by J_1. We then make the following assumptions.

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

AS5. *Suppose that (x^*, λ^*) is a Karush-Kuhn-Tucker point for the problem (1.1) and that*

$$\begin{aligned} J_1 &= \{ \ i \ \mid \ (g_L(x^*; \lambda^*))_i = 0 \ \textit{and} \ \ell_i < x_i^* < u_i \ \} \\ J_2 &= \{ \ i \ \mid \ (g_L(x^*; \lambda^*))_i = 0 \ \textit{and} \ (x_i^* = \ell_i \ \textit{or} \ x_i^* = u_i) \ \}. \end{aligned}$$

Then we assume that the matrix

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

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

The next result from [4], which also holds for the augmented Lagrangian pattern search algorithm, is Lemma 5.1. This result relates the convergence of the iterates to the error in the multipliers, a relationship characteristic of augmented Lagrangian methods [1, 27]. Again, the proof in [4] holds for the pattern search variant because of Proposition 5.1.

LEMMA 5.3. *Suppose that (AS1) holds. Let $\{x^{(k)}\} \subset B$, $k \in K$, be a subsequence which converges to the Karush-Kuhn-Tucker point x^* for which (AS2), (AS4), and (AS5) hold, and let λ^* be the corresponding vector of Lagrange multipliers. Assume that $\{\lambda^{(k)}\}$, $k \in K$, is any sequence of vectors, that $\{S^{(k)}\}$, $k \in K$, is any sequence of diagonal matrices satisfying $0 < S_1^{-1} \leq S^{(k)} \leq S_2 < \infty$, and that $\{\mu^{(k)}\}$, $k \in K$, form a nonincreasing sequence of positive scalars, so that the product $\mu^{(k)} \| \lambda^{(k)} - \lambda^* \|$ converges to zero as k increases. Now, suppose further that*

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

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

$$\| x^{(k)} - x^* \| \leq a_3 \omega^{(k)} + a_4 \mu^{(k)} \| \lambda^{(k)} - \lambda^* \| \tag{5.10}$$

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

and

$$\| c(x^{(k)}) \| \leq s_1 (a_5 \omega^{(k)} \mu^{(k)} + (\mu^{(k)} + a_6 (\mu^{(k)})^2) \| \lambda^{(k)} - \lambda^* \|) \tag{5.11}$$

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

The following is Corollary 5.2 in [4].

COROLLARY 5.4. *Suppose that the conditions of Lemma 5.3 hold and that $\hat{\lambda}^{(k+1)}$ is any Lagrange multiplier estimate for which*

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

for some positive constants a_{16} and a_{17} and all $k \in K$ sufficiently large. Then there are positive constants $\overline{\mu}$, a_3, a_4, a_5, a_6, s_1 and an integer value k_0 so that if $\mu^{(k_0)} \leq \overline{\mu}$ then (5.10),

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

and (5.11) hold for all $k \geq k_0$, ($k \in K$).

We also inherit the following result indicating that we may generally expect the penalty parameter to remain bounded away from zero. This is Theorem 5.3 in [4]. Taken together with the convergence of the multiplier estimates, this means that the stopping tolerance for the inexact minimization of the augmented Lagrangian is decreasing at the same rate as in the original algorithm. However, in Section 6 of [4] the authors show that in the case of nonunique limit points one can have $\mu^{(k)} \to 0$, in which case the stopping tolerance δ^k decreases more like $(\mu^{(k)})^2$.

THEOREM 5.5. *Suppose that the iterates $\{x^{(k)}\}$ of the augmented Lagrangian pattern search algorithm converge to the single limit point x^*, that (AS1), (AS2), (AS4), and (AS5) hold, and that α_η and β_η satisfy $\alpha_\eta < \min(1, \alpha_\omega)$ and $\beta_\eta < \min(1, \beta_\omega)$. Then there is a constant $\mu > 0$ such that $\mu^{(k)} > \mu$ for all k.*

The proof of Theorem 5.5 makes use of the fact that $\| P(x^{(k)}, \nabla_x \Phi^{(k)}) \| = O(\omega^{(k)})$, whereas the proofs of the preceding convergence results require only that

$$\| P(x^{(k)}, \nabla_x \Phi^{(k)}) \| \to 0.$$

Finally, we have the following result on the rate of convergence of the outer iteration, corresponding to Theorem 5.5 in [4].

THEOREM 5.6. *Under the assumptions of Theorem 5.5, the iterates $x^{(k)}$ and the Lagrange multiplier estimates $\bar{\lambda}^{(k)}$ of the augmented Lagrangian pattern search algorithm are at least R-linearly convergent with R-factor at most $\hat{\mu}^{\min(\beta_\omega, \beta_\eta)}$, where $\hat{\mu} = \min[\gamma_1, \mu]$ and where μ is the smallest value of the penalty parameter generated by the algorithm in question.*

6. Application to inequality constrained minimization. Special consideration is due to the general problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \le 0 \\ & \ell \le x \le u, \end{array} \tag{6.1}$$

converted into the form (1.1) via the introduction of nonnegative slack variables:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) + z = 0 \\ & \ell \le x \le u \\ & z \ge 0. \end{array} \tag{6.2}$$

The augmented Lagrangian associated with (6.2) is

$$\Phi(x, z; \lambda, S, \mu) = f(x) + \lambda^T (g(x) + z) + \frac{1}{2\mu} \sum_{i=1}^{m} s_{ii} (g_i(x) + z_i)^2. \tag{6.3}$$

Explicit equality constraints may also be present in (6.1); we ignore them here for brevity.

The introduction of slacks increases the dimension of the bound constrained subproblem that we must solve at each outer iteration. Unfortunately, increases in dimension usually cause a degradation in performance for pattern search methods. We can avoid this increase in dimension because of the simple way in which the slacks z enter into (6.3). One approach [1, 23] is to note that given x, we can minimize $\Phi(x, z; \lambda, S, \mu)$ explicitly in z for $z \ge 0$. This leads to a subproblem in x alone:

$$\begin{array}{ll} \text{minimize} & \Phi(x, z(x); \lambda, S, \mu) \\ \text{subject to} & \ell \le x \le u, \end{array}$$

where

$$\Phi(x, z(x); \lambda, S, \mu) = f(x) + \frac{\mu}{2} \sum_{i=1}^{m} \frac{1}{s_{ii}} (\max(0, \lambda_i + \frac{s_{ii}}{\mu} g_i(x))^2 - \lambda_i^2).$$

The multiplier update formula (2.2) is also modified:

$$\overline{\lambda}_i(x, \lambda, S, \mu) = \max(0, \lambda_i + s_{ii} c_i(x)/\mu), \quad i = 1, \cdots, m.$$

See [1] for further discussion. The reduced augmented Lagrangian $\Phi(x, z(x); \lambda, S, \mu)$ has Lipschitz first derivatives. If one were using a quasi-Newton method for the minimization of the augmented Lagrangian one might be loath to eliminate z since the resulting problem is not C^2 and one loses any assurance of local superlinear convergence. However, pattern search methods do not have such favorable local convergence properties, so ostensibly nothing is lost, and much is gained by the reduction of dimension of the subproblems.

7. Conclusion. We have demonstrated that it is possible to construct a globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds. Extensive numerical tests of this algorithm remain to be done. We agree with the perspective of the authors in [4]:

> We have deliberately not included the results of numerical testing as, in our view, the construction of appropriate software is by no means trivial and we wish to make a thorough job of it. We will report on our numerical experience in due course.

This caution is particularly apt in view of the sort of problems to which pattern search is typically applied.

Acknowledgments. We wish to thank the associate editor and the two referees for their careful reading of the paper and their many helpful comments. The presentation is much clearer as a consequence.

REFERENCES

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

[2] M. J. Box, *A new method of constrained optimization and a comparison with other methods*, The Computer Journal, 8 (1965), pp. 42–52.

[3] C. W. Carroll, *The created response surface technique for optimizing nonlinear restrained systems*, Operations Research, 9 (1961).

[4] 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.

[5] ——, *LANCELOT: A Fortran package for large-scale nonlinear optimization (release A)*, vol. 17 of Springer Series in Computational Mathematics, Springer-Verlag, New York, 1992.

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

[7] L. C. W. Dixon, *ACSIM—An accelerated constrained simplex technique*, Computer Aided Design, 5 (1973), pp. 22–32.

[8] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Classics in Applied Mathematics, SIAM, Philadelphia, 1990.

[9] H. Glass and L. Cooper, *Sequential search: A method for solving constrained optimization problems*, Journal of the Association for Computing Machinery, 12 (1965), pp. 71–82.

[10] R. Glowinski and A. J. Kearsley, *On the simulation and control of some friction constrained problems*, SIAM Journal on Optimization, 5 (1995), pp. 681–694.

[11] D. M. Himmelblau, *Applied Nonlinear Programming*, McGraw-Hill, 1972.

[12] R. Hooke and T. A. Jeeves, *Direct search solution of numerical and statistical problems*, Journal of the Association for Computing Machinery, 8 (1961), pp. 212–229.

[13] A. J. Kearsley, *The use of optimization techniques in the solution of partial differential equations from science and engineering*, PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1996; available as Tech. Rep. 96–11, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005–1892.

[14] D. L. Keefer, *Simpat: Self-bounding direct search method for optimization*, Industrial and Engineering Chemistry Process Design and Development, 12 (1973), pp. 92–99.

[15] W. R. Klingman and D. M. Himmelblau, *Nonlinear programming with the aid of a multiple-gradient summation technique*, Journal of the Association for Computing Machinery, 11 (1964), pp. 400–415.

[16] R. M. Lewis and V. J. Torczon, *Rank ordering and positive bases in pattern search algorithms*, Tech. Report 96–71, Institute for Computer Applications in Science and Engineering, Mail Stop 132C, NASA Langley Research Center, Hampton, Virginia 23681–2199, 1996.

[17] ——, *Pattern search algorithms for bound constrained minimization*, SIAM Journal on Optimization, 9 (1999), pp. 1082–1099.

[18] ——, *Pattern search methods for linearly constrained minimization*, SIAM Journal on Optimization,

10 (2000), pp. 917–941.

[19] J. H. May, *Linearly Constrained Nonlinear Programming: A Solution Method That Does Not Require Analytic Derivatives*, PhD thesis, Yale University, December 1974.

[20] R. Mifflin, *A superlinearly convergent algorithm for minimization without evaluating derivatives*, Mathematical Programming, 9 (1975), pp. 100–117.

[21] J. A. Nelder and R. Mead, *A simplex method for function minimization*, The Computer Journal, 7 (1965), pp. 308–313.

[22] D. Paviani and D. M. Himmelblau, *Constrained nonlinear optimization by heuristic programming*, Operations Research, 17 (1969), pp. 872–882.

[23] R. T. Rockafellar, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Mathematical Programming, 5 (1973), pp. 354–373.

[24] H. H. Rosenbrock, *An automatic method for finding the greatest or least value of a function*, The Computer Journal, 3 (1960), pp. 175–184.

[25] W. Spendley, *Nonlinear least squares fitting using a modified simplex minimization method*, in Optimization, R. Fletcher, ed., Academic Press, 1969, pp. 259–270.

[26] W. Spendley, G. R. Hext, and F. R. Himsworth, *Sequential application of simplex designs in optimisation and evolutionary operation*, Technometrics, 4 (1962), pp. 441–461.

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

[28] V. Torczon, *On the convergence of pattern search algorithms*, SIAM Journal on Optimization, 7 (1997), pp. 1–25.