Some Theoretical Properties of an Augmented Lagrangian Merit Function

Philip E. GILL* Walter MURRAY† Michael A. SAUNDERS†

Margaret H. WRIGHT‡

Technical Report SOL 86-6R§

September 1986

Abstract

Sequential quadratic programming (SQP) methods for nonlinearly constrained optimization typically use a *merit function* to enforce convergence from an arbitrary starting point. We define a *smooth* augmented Lagrangian merit function in which the Lagrange multiplier estimate is treated as a separate variable, and inequality constraints are handled by means of non-negative slack variables that are included in the linesearch. Global convergence is proved for an SQP algorithm that uses this merit function. We also prove that steps of unity are accepted in a neighborhood of the solution when this merit function is used in a suitable superlinearly convergent algorithm. Finally, some numerical results are presented to illustrate the performance of the associated SQP method.

Keywords: constrained optimization, sequential quadratic programming method, nonlinear programming method, quadratic programming.

1. Sequential Quadratic Programming Methods

Sequential quadratic programming (SQP) methods are widely considered to be effective general techniques for solving optimization problems with nonlinear constraints. (For a survey of results and references on SQP methods, see Powell [28].) One of the major issues of interest in recent research on SQP methods has been the choice of *merit function*—the measure of progress at each iteration. This paper describes some properties of a theoretical SQP algorithm (NPSQP) that uses a *smooth* augmented Lagrangian merit function. NPSQP is a simplified version of an SQP algorithm that has been implemented as the Fortran code NPSOL (Gill *et al.* [15]). (The main simplifications involve the form of the problem, use of a single penalty parameter, and strengthened assumptions.)

For ease of presentation, we assume that all the constraints are nonlinear *inequalities*. (The theory applies in a straightforward fashion to equality constraints.) The problem to be solved is thus:

$$\text{(NP)}\qquad \underset{x\in\mathbb{R}^n}{\text{minimize}}\quad f(x) \qquad \text{subject to}\quad c_i(x) \ge 0,\quad i = 1,\dots,m,$$

*Department of Mathematics, University of California San Diego, La Jolla, CA 92093, USA.

†Department of Management Science & Engineering, Stanford University, Stanford, CA 94305-4026, USA.

‡Department of Computer Science, Courant Institute, New York University, New York, NY 10012, USA.

§The material contained in this report is based upon research supported by the U.S. Department of Energy Contract DE-AA03-76SF00326, PA No. DE-AS03-76ER72018; National Science Foundation Grants DCR-8413211 and ECS-8312142; the Office of Naval Research Contract N00014-85-K-0343; and the U.S. Army Research Office Contract DAAG29-84-K-0156.

where f and $\{c_i\}$ are twice-continuously differentiable. Let $g(x)$ denote the gradient of $f(x)$, and $A(x)$ denote the Jacobian matrix of the constraint vector $c(x)$. A solution of NP will be denoted by x^*, and we assume that there are a finite number of solutions.

We assume that the second-order Kuhn-Tucker conditions hold (with strict complementarity) at x^*. Thus, there exists a Lagrange multiplier vector λ^* such that

$$g(x^*) = A(x^*)^T\lambda^*, \tag{1.1a}$$

$$c(x^*)^T\lambda^* = 0, \tag{1.1b}$$

$$\lambda_i^* > 0 \ \text{ if } \ c_i(x^*) = 0. \tag{1.1c}$$

(For a detailed discussion of optimality conditions, see, for example, Fiacco and McCormick [11], and Powell [26].) Conditions (1.1a) and (1.1b) may equivalently be stated as

$$Z(x^*)^T g(x^*) = 0,$$

where the columns of the matrix Z form a basis for the null space of constraints active at x^*.

At the k-th iteration, the new iterate x_{k+1} is defined as

$$x_{k+1} = x_k + \alpha_k p_k, \tag{1.2}$$

where x_k is the current iterate, p_k is an n-vector (the *search direction*), and α_k is a non-negative step length ($0 < \alpha_k \le 1$). For simplicity of notation, we henceforth suppress the subscript k, which will be implicit on unbarred quantities. A barred quantity denotes one evaluated at iteration $k+1$.

The central feature of an SQP method is that the search direction p in (1.2) is the solution of a quadratic programming subproblem whose objective function approximates the Lagrangian function and whose constraints are linearizations of the nonlinear constraints. The usual definition of the QP subproblem is the following:

$$\underset{p\in\mathbb{R}^n}{\text{minimize}} \quad g^Tp + \tfrac{1}{2}p^THp \tag{1.3a}$$

$$\text{subject to} \quad Ap \ge -c, \tag{1.3b}$$

where g, c and A denote the relevant quantities evaluated at x. The matrix H is a symmetric positive-definite quasi-Newton approximation to the Hessian of the Lagrangian function. The Lagrange multiplier vector μ of (1.3) satisfies

$$g + Hp = A^T\mu, \tag{1.4a}$$

$$\mu^T(Ap + c) = 0, \tag{1.4b}$$

$$\mu \ge 0. \tag{1.4c}$$

The remainder of this paper is organized as follows. Section 2 gives some background on the role of merit functions within SQP methods, and introduces the augmented Lagrangian merit function used in NPSQP. In Section 3 we state the assumptions about the problem and present the algorithm. Global convergence results are given in Section 4. Section 5 shows that the chosen merit function will not impede superlinear convergence. In Section 6 we present a selection of numerical results to indicate the robustness of the implementation NPSOL.

2. Background on Merit Functions

2.1. Introduction

Many authors have studied the choice of steplength in (1.2). Usually, α is chosen by a linesearch procedure to ensure a "sufficient decrease" (Ortega and Rheinboldt [23]) in a merit function that combines the objective and constraint functions in some way. A popular merit function for several years (Han [17, 18]; Powell [27] has been the ℓ_1 penalty function (Pietrzykowski [24]):

$$P(x, \rho) = f(x) + \rho \sum_{i=1}^{m} \max\big(0, -c_i(x)\big). \tag{2.1}$$

This merit function has the property that, for ρ sufficiently large, x^* is an unconstrained minimum of $P(x, \rho)$. In addition, ρ can always be chosen so that the SQP search direction p is a descent direction for $P(x, \rho)$. However, Maratos [21] observed that requiring a decrease in $P(x, \rho)$ at every iteration could lead to the inhibition of superlinear convergence. (See Chamberlain *et al.*, [7], for a procedure designed to avoid this difficulty.) Furthermore, $P(x, \rho)$ is not differentiable at the solution, and linesearch techniques based on smooth polynomial interpolation are therefore not applicable.

An alternative merit function that has recently received attention is the *augmented Lagrangian function*, whose development we now review. If all the constraints of (NP) are equalities, the associated augmented Lagrangian function is:

$$L(x, \lambda, \rho) \equiv f(x) - \lambda^T c(x) + \tfrac{1}{2}\rho c(x)^T c(x), \tag{2.2}$$

where λ is a multiplier estimate and ρ is a non-negative penalty parameter. Augmented Lagrangian functions were first introduced by Hestenes [19] and Powell [25] as a means of creating a sequence of *unconstrained subproblems* for the equality-constraint case. A property of (2.2) is that there exists a *finite* $\hat{\rho}$ such that for all $\rho \geq \hat{\rho}$, x^* is an unconstrained minimum of (2.2) when $\lambda = \lambda^*$(see, e.g., Fletcher [13]). The use of (2.2) as a merit function within an SQP method was suggested by Wright [35] and Schittkowski [32].

2.2. The Lagrange multiplier estimate

When (2.2) is used as a merit function, it is not obvious—even in the equality-constraint case— how the multiplier estimate λ should be defined at each iteration.

Most SQP methods (e.g., Han [17]; Powell [27]) define the approximate Hessian of the Lagrangian function using the QP multiplier μ (*cf.* (1.4)), which can be interpreted as the "latest" (and presumably "best") multiplier estimate, and requires no additional computation. However, using μ as the multiplier estimate in (2.2) has the effect of *redefining* the merit function at every iteration. Thus, since there is no monotonicity property with respect to a single function, difficulties may arise in proving global convergence of the algorithm.

Powell and Yuan [29] have recently studied an augmented Lagrangian merit function for the equality-constraint case in which λ in (2.2) is defined as the *least-squares multiplier estimate*, and hence is treated as a function of x rather than as a separate variable. (An augmented Lagrangian function of this type was first introduced and analyzed as an exact penalty function by Fletcher [12].) Powell and Yuan [29] prove several global and local convergence properties for this merit function.

Other smooth merit functions have been considered by Dixon [10], DiPillo and Grippo [9], Bartholomew-Biggs [3, 1, 2], and Boggs and Tolle [4, 5] (the latter only for the equality-constraint case).

An approach that makes an alternative use of the QP multiplier μ is to treat the elements of λ as *additional variables* (rather than to reset λ at every iteration). Thus, μ is used to define a "search direction" ξ for the multiplier estimate λ, and the linesearch is performed with respect to both x and λ. This idea was suggested by Tapia [34] in the context of an unconstrained subproblem and by Schittkowski [32] within an SQP method. This approach will also be taken in NPSQP.

2.3. Treatment of inequality constraints

When defining a merit function for a problem with *inequality* constraints, it is necessary to identify which constraints are "active". The ℓ_1 merit function (2.1) includes only the *violated* constraints. The original formulation of an augmented Lagrangian function for inequality constraints is due to Rockafellar [31]:

$$L(x, \lambda, \rho) = f(x) - \lambda^T c_+(x) + \tfrac{1}{2}\rho c_+(x)^T c_+(x), \tag{2.3}$$

where

$$(c_+)_i = \begin{cases} c_i & \text{if } \rho c_i \le \lambda_i; \\ \lambda_i/\rho & \text{otherwise.} \end{cases}$$

Some global results are given for this merit function in Schittkowski [32, 33].

When used as a merit function, (2.3) has two disadvantages: it is difficult to derive a convenient formulation for the choice of penalty parameter (see Lemma 4.3, below); and discontinuities in the second derivative may cause inefficiency in linesearch techniques based on polynomial interpolation. To avoid these difficulties, we augment the variables (x, λ) by a set of *slack variables* that are used *only in the linesearch*. At the k-th major iteration, a vector triple

$$y = \begin{pmatrix} p \\ \xi \\ q \end{pmatrix} \tag{2.4}$$

is computed to serve as a direction of search for the variables (x, λ, s). The new values are defined by

$$\begin{pmatrix} \bar{x} \\ \bar{\lambda} \\ \bar{s} \end{pmatrix} = \begin{pmatrix} x \\ \lambda \\ s \end{pmatrix} + \alpha \begin{pmatrix} p \\ \xi \\ q \end{pmatrix}, \tag{2.5}$$

and the vectors p, ξ and q are found from the QP subproblem (1.3), as described below.

In our algorithm (as in Schittkowski [32]), ξ is defined as

$$\xi \equiv \mu - \lambda, \tag{2.6}$$

so that if $\alpha = 1$, $\bar{\lambda} = \mu$. We take λ_0 as μ_0 (the QP multipliers at x_0), so that $\lambda_1 = \mu_0$ regardless of α_0.

The definition of s and q can be interpreted in terms of an idea originally given by Rockafellar [31] in the derivation of (2.3). In our algorithm, the augmented Lagrangian function includes a set of non-negative slack variables:

$$L(x, \lambda, s, \rho) = f(x) - \lambda^T \big(c(x) - s\big) + \tfrac{1}{2}\rho\big(c(x) - s\big)^T\big(c(x) - s\big), \quad \text{with } s \ge 0. \tag{2.7}$$

The vector s at the beginning of iteration k is taken as

$$s_i = \begin{cases} \max(0, c_i) & \text{if } \rho = 0; \\ \max(0, c_i - \lambda_i/\rho) & \text{otherwise,} \end{cases} \tag{2.8}$$

where ρ is the *initial* penalty parameter for that iteration (see Lemma 4.3, below). (When ρ is nonzero, the vector s defined by (2.8) yields the value of L minimized with respect to the slack variables alone, subject to the non-negativity restriction $s \ge 0$.) The vector q in (2.5) is then defined by

$$Ap - q = -(c - s), \tag{2.9a}$$

$$\text{so that } Ap + c = s + q. \tag{2.9b}$$

We see from (2.9) that $s+q$ is simply the residual of the inequality constraints from the QP (1.3). Therefore, it follows from (1.3) and (1.4) that

$$s+q \geq 0 \;\text{ and }\; \mu^T(s+q)=0. \tag{2.10}$$

2.4. Choice of the penalty parameter

Finally, we consider the definition of the penalty parameter in (2.7). Our numerical experiments have suggested strongly that efficient performance is linked to keeping the penalty parameter as small as possible, subject to satisfying the conditions needed for convergence. Hence, our strategy is to maintain a "current" ρ that is increased only when necessary to satisfy a condition that assures global convergence.

Several authors have stated that the need to choose a penalty parameter adds an arbitrary element to an SQP algorithm, or leads to difficulties in implementation. On the contrary, we shall see (Lemma 4.3) that in NPSQP *the penalty parameter is based directly on a condition needed for global convergence*, and hence should not be considered as arbitrary or heuristic.

3. Statement of the Algorithm

We make the following assumptions:

(i) x and $x+p$ lie in a closed, bounded region Ω of $\mathbb{R}^n$ for all k;

(ii) f, $\{c_i\}$, and their first and second derivatives are uniformly bounded in norm in Ω;

(iii) H is positive definite, with bounded condition number, and smallest eigenvalue uniformly bounded away from zero, i.e., there exists $\gamma > 0$ such that, for all k,

$$p^T Hp \geq \gamma \|p\|^2;$$

(iv) In every quadratic program (1.3), the active set is linearly independent, and strict complementarity holds.

(v) The quadratic program (1.3) always has a solution.

The following notation will be used in the remainder of this section. Given x, λ, s, p, ξ, q and ρ, we let $\phi(\alpha,\rho)$ denote $L(x+\alpha p, \lambda+\alpha\xi, s+\alpha q, \rho)$, i.e., the merit function as a function of the steplength, with the convention that $\phi_k(\rho)$ denotes evaluation of L at (x_k, λ_k, s_k). (The argument ρ may be omitted when its value is obvious.) The derivative of ϕ with respect to α will be denoted by ϕ'. Let v denote the "extended" iterate (x,λ,s), and let y denote the associated search direction (p,ξ,q). For brevity, we sometimes use the notation "(α)" to denote evaluation at $v+\alpha y$.

The steps of each iteration of algorithm NPSQP are:

1. Solve (1.3) for p. If $p=0$, set $\lambda=\mu$ and terminate. Otherwise, define $\xi=\mu-\lambda$.
2. Compute s from (2.8). Find ρ such that $\phi'(0) \leq -\frac{1}{2}p^T Hp$ (see Lemma 4.3, below).
3. Compute the steplength α, as follows. If

$$\phi(1)-\phi(0) \leq \sigma\phi'(0) \tag{3.1a}$$
$$\text{and }\; \phi'(1) \leq \eta\phi'(0) \;\text{ or }\; |\phi'(1)| \leq -\eta\phi'(0), \tag{3.1b}$$

where $0 < \sigma \le \eta < \frac{1}{2}$, set $\alpha = 1$. Otherwise, use safeguarded cubic interpolation (see, e.g., Gill and Murray [14], to find an $\alpha \in (0, 1)$ such that

$$\phi(\alpha) - \phi(0) \le \sigma\alpha\phi'(0) \tag{3.2a}$$
$$\text{and} \;\; |\phi'(\alpha)| \le -\eta\phi'(0). \tag{3.2b}$$

4. Update H so that (iii) is satisfied (e.g., using a suitably modified BFGS update; see Gill *et al.* [15], for the update used in NPSOL).

5. Update x and λ using (2.5).

4. Global Convergence Results

In order to prove global convergence, we first prove a set of lemmas that establish various properties of the algorithm.

Lemma 4.1. *When assumptions* (i)–(v) *are satisfied, the following properties hold for Algorithm* NPSQP*:*

(a) $\|p\| = 0$ *if and only if* x *is a Kuhn-Tucker point of NP;*

(b) *There exists* $\bar{\epsilon}$ *such that if* $\|p\| \le \bar{\epsilon}$*, the active set of the QP* (1.3) *is the same as the set of nonlinear constraints active at* x^**;*

(c) *There exists a constant* M_p*, independent of* k*, such that*

$$\|x^* - x_k\| \le M_p \|p_k\|. \tag{4.1}$$

Proof. The proofs of (a) and (b) are given in Robinson [30]. To show that (c) is true, let $\widehat{c}(\cdot)$ denote the vector of constraints active at x^*, $\widehat{A}$ the Jacobian of the active constraints, and Z an orthogonal basis for the null space of $\widehat{A}$. Expanding $\widehat{c}$ and $Z^T g$ about x^*, and noting that $\widehat{c}(x^*) = 0$ and $Z(x^*)^T g(x^*) = 0$, we obtain

$$\begin{pmatrix} \widehat{c}(x) \\ Z(x)^T g(x) \end{pmatrix} = \begin{pmatrix} \widehat{A}\big(x^* + \theta(x - x^*)\big) \\ V\big(x^* + \theta(x - x^*)\big) \end{pmatrix} (x - x^*) \equiv S\big(x^* + \theta(x - x^*)\big)(x - x^*), \tag{4.2}$$

where $0 < \theta < 1$, for an appropriate matrix function V. (See Goodman [16] for a discussion of the definition of V.)

For suitably small $\bar{\epsilon}$ in (b), $S\big(x^* + \theta(x_k - x^*)\big)$ is nonsingular, with smallest singular value uniformly bounded below (see, e.g., Robinson [30]). Because of assumption (i), the relation (4.1) is immediate if $\|p_k\| \ge \bar{\epsilon}$, and we henceforth consider only iterations k such that $\|p_k\| < \bar{\epsilon}$.

Taking $x = x_k$ in (4.2), and using the nonsingularity of S and norm inequalities, we obtain:

$$\|x_k - x^*\| \le \beta(\|\widehat{c}_k\| + \|Z_k^T g_k\|) \tag{4.3}$$

for some bounded β. We now seek an upper bound on the right-hand side of (4.3). Since the QP (1.3) identifies the correct active set, p_k satisfies the equations

$$\widehat{A}_k p_k = -\widehat{c}_k \;\; \text{and} \;\; Z_k^T H_k p_k = -Z_k^T g_k. \tag{4.4}$$

From (4.4), assumption (iii) and the property of S mentioned above, it follows that there must exist $\tilde{\beta} > 0$ such that

$$\tilde{\beta}(\|\widehat{c}_k\| + \|Z_k^T g_k\|) \le \|p_k\|. \tag{4.5}$$

Since β and $\tilde{\beta}$ are independent of k, combining (4.4) and (4.5) gives the desired result. ∎

Lemma 4.2. *For all $k \ge 1$,*

$$\|\lambda_k\| \le \max_{1\le j\le k-1} \|\mu_j\|,$$

and hence $\|\lambda_k\|$ is bounded for all k. In addition, $\|\xi_k\|$ is uniformly bounded for all k.

Proof. By definition,

$$\begin{aligned} \lambda_1 &= \mu_0; \\ \lambda_{k+1} &= \lambda_k + \alpha_k(\mu_k - \lambda_k), \quad k \ge 1. \end{aligned} \tag{4.6}$$

The proof is by induction. The result holds for λ_1 because of assumption (iv), which implies the boundedness of $\|\mu_k\|$ for all k. Assume that the lemma holds for λ_k. From (4.6) and norm inequalities, we have

$$\|\lambda_{k+1}\| \le \alpha_k\|\mu_k\| + (1-\alpha_k)\|\lambda_k\|.$$

Since $0 < \alpha \le 1$, applying the inductive hypothesis gives

$$\|\lambda_{k+1}\| \le \max \|\mu_j\|, \quad j = 1, \ldots, k,$$

which gives the first desired result.

The boundedness of $\|\xi_k\|$ follows immediately from its definition (2.6), assumption (iv), and the first result of this lemma. ∎

The next lemma establishes the existence of a non-negative penalty parameter such that the projected gradient of the merit function at each iterate satisfies a condition associated with global convergence.

Lemma 4.3. . *There exists $\hat\rho \ge 0$ such that*

$$\phi'(0, \rho) \le -\tfrac{1}{2} p^T H p \tag{4.7}$$

for all $\rho \ge \hat\rho$.

Proof. The gradient of L with respect to x, λ and s is given by

$$\nabla L(x, \lambda, s) \equiv \begin{pmatrix} g(x) - A(x)^T\lambda + \rho A(x)^T\big(c(x) - s\big) \\ -\big(c(x) - s\big) \\ \lambda - \rho\big(c(x) - s\big) \end{pmatrix}, \tag{4.8}$$

and it follows that $\phi'(0)$ is given by

$$\phi'(0) = p^T g - p^T A^T \lambda + \rho\, p^T A^T (c - s) - (c - s)^T \xi + \lambda^T q - \rho q^T (c - s), \tag{4.9}$$

where g, A, and c are evaluated at x.

Multiplying (1.4a) by p^T gives

$$g^T p = p^T A^T \mu - p^T H p. \tag{4.10}$$

Substituting (2.6), (2.9a) and (4.10) in (4.9), we obtain

$$\phi'(0) = -p^T H p + q^T \mu - 2(c - s)^T \xi - \rho (c - s)^T (c - s). \tag{4.11}$$

Substituting (4.11) in the desired inequality (4.7) and rearranging, we obtain as the condition to be satisfied:

$$q^T \mu - 2(c - s)^T \xi - \rho (c - s)^T (c - s) \le \tfrac{1}{2} p^T H p. \tag{4.12}$$

The complementarity conditions (1.4) and definition (2.9) imply that $q^T\mu \le 0$. Hence, if $\frac{1}{2}p^THp > -2(c-s)^T\xi$, then (4.12) holds for *all* non-negative ρ, and $\hat\rho$ may be taken as zero. (Note that this applies when $c-s$ is zero.) The determination of $\hat\rho$ is non-trivial only if

$$\tfrac{1}{2}p^THp \le -2(c-s)^T\xi. \tag{4.13}$$

When (4.13) holds, rearrangement of (4.12) shows that ρ must satisfy

$$\rho(c-s)^T(c-s) \ge q^T\mu - \tfrac{1}{2}p^THp - 2(c-s)^T\xi. \tag{4.14}$$

A value $\hat\rho$ such that (4.14) holds for all $\rho \ge \hat\rho$ is

$$\hat\rho = \frac{2\|\xi\|}{\|c-s\|}. \tag{4.15}$$

The value $\hat\rho$ is taken as (4.15) if (4.13) holds, and as zero otherwise. ∎

The penalty parameter in NPSQP is determined by retaining a "current" value, which is increased if necessary to satisfy (4.7). At iteration k, the penalty parameter ρ_k is thus defined by

$$\rho_k = \begin{cases} \rho_{k-1} & \text{if } \phi'(0,\rho_{k-1}) \le -\frac{1}{2}p_k^TH_kp_k; \\ \max(\hat\rho_k, 2\rho_{k-1}) & \text{otherwise.} \end{cases} \tag{4.16}$$

where $\rho_0 = 0$ and $\hat\rho_k$ is defined by Lemma 4.3.

Because of the definition of $\hat\rho$ and the strategy (4.16) for choosing ρ, there are two possible cases. When the value of ρ at every iteration is uniformly bounded, (4.16) implies that ρ_k eventually becomes fixed at some value, which is retained for all subsequent iterations. (This situation will be called the *bounded case.*)

Otherwise, when there is no upper bound on the penalty parameter, $\hat\rho$ (4.15) must also be tending to infinity. In this *unbounded case*, an infinite subsequence of iterations exists at which ρ is increased, with a finite number of iterations between each such increase. Let $\{k_l\}$, $l = 0, 1, \ldots$, denote the indices of the subsequence of iterations when the penalty parameter is increased. Thus, for any $l \ge 1$,

$$\rho_{k_l} = \max(\hat\rho_{k_l}, 2\rho_{k_{l-1}}), \tag{4.17}$$

and

$$\rho_{k_l} > \rho_{k_{l-1}}. \tag{4.18}$$

Note that (4.16) and the properties of $\hat\rho$ imply that

$$\rho_{k_l} \le 2\hat\rho_{k_l}. \tag{4.19}$$

In the proof of global convergence, it will be crucial to develop uniform bounds on various quantities multiplied by ρ. (Such bounds are immediate in the bounded case.) The next four lemmas provide the needed results.

Lemma 4.4. . *At any iteration k_l in which the penalty parameter is defined by* (4.17),

$$\rho_{k_l}\|p_{k_l}\|^2 \le N \tag{4.20}$$

$$\textit{and}\ \ \rho_{k_l}\|c_{k_l} - s_{k_l}\| \le N \tag{4.21}$$

for some uniformly bounded N.

Proof. Let unbarred quantities denote those associated with iteration k_l. When ρ is defined by (4.17), (4.13) must hold. From (4.13) and norm inequalities, we have

$$\tfrac{1}{2}p^T Hp \leq -2(c-s)^T\xi \leq 2\|c-s\|\|\xi\|, \tag{4.22}$$

which implies

$$\|c-s\| \geq \frac{\frac{1}{4}p^T Hp}{\|\xi\|}. \tag{4.23}$$

Let γ denote a lower bound on the smallest eigenvalue of H (*cf.* assumption (iii)). Combining (4.15) and (4.23) and noting that $p^T Hp \geq \gamma\|p\|^2$, we obtain

$$\hat{\rho} \leq \frac{2\|\xi\|^2}{\frac{1}{4}p^T Hp} \leq \frac{8\|\xi\|^2}{\gamma\|p\|^2}.$$

Using Lemma 4.2 and (4.19), (4.20) follows immediately. The relation (4.21) follows from (4.19) and the definition (4.15) of $\hat{\rho}$. ∎

In the unbounded case, the discussion following Lemma 4.3 shows that, given any $C > 0$, we can find $\bar{l}$ such that for all $l \geq \bar{l}$,

$$\rho_{k_l} > C.$$

Lemma 4.5 thus implies that, in the unbounded case, given any $\epsilon > 0$, we can find $\bar{l}$ such that, for $l \geq \bar{l}$,

$$\|p_{k_l}\| \leq \epsilon. \tag{4.24}$$

Hence, the norm of the search direction becomes arbitrarily small at a *subsequence* of iterations—namely, those at which the penalty parameter is increased. We now derive certain results for the intervening iterations during which the penalty parameter remains unchanged. We shall henceforth use "M" to denote a generic uniformly bounded constant (whose particular value can be deduced from the context).

Lemma 4.5. . *There exists a bounded constant M such that, for all l,*

$$\rho_{k_l}\big(\phi_{k_l}(\rho_{k_l}) - \phi_{k_{l+1}}(\rho_{k_l})\big) < M. \tag{4.25}$$

Proof. To simplify notation in this proof, we shall use the subscripts 0 and K to denote quantities associated with iterations k_l and k_{l+1} respectively. Thus, the penalty parameter is increased at x_0 and x_K in order to satisfy condition (4.12), and remains fixed at ρ_0 for iterations $1, \ldots, K-1$.

Consider the following identity:

$$\phi_0 - \phi_K = \sum_{k=0}^{K-1}(\phi_k - \phi_{k+1}), \tag{4.26}$$

where ϕ_k denotes $\phi_k(\rho_0)$. Because the merit function is decreased at iterations $0, 1, \ldots, K-1$, (4.26) implies

$$\phi_0 - \phi_K > 0. \tag{4.27}$$

Using definition (2.7), note that

$$\rho_0\phi = \rho_0 f - \rho_0\lambda^T(c-s) + \tfrac{1}{2}\rho_0^2(c-s)^T(c-s), \tag{4.28}$$

and recall from Lemma 4.4 that

$$\rho_0\|c_0 - s_0\| < M \;\text{ and }\; \rho_K\|c_K - s_K\| < M. \tag{4.29}$$

Since $\|\lambda\|$ is bounded (Lemma 4.2), the only term in (4.28) that might become unbounded is $\rho_0 f$. Because of the lower bound (4.27), the desired relation (4.25) will follow if an upper bound exists for $\rho_0(f_0 - f_K)$.

Expanding f about x^*, we have:

$$f_0 = f^* + (x_0 - x^*)^T g^* + O(\|x_0 - x^*\|^2) \tag{4.30a}$$
$$\text{and}\ \ f_K = f^* + (x_K - x^*)^T g^* + O(\|x_K - x^*\|^2). \tag{4.30b}$$

Subtraction of these two expressions gives:

$$f_0 - f_K = \big((x_0 - x^*) - (x_K - x^*)\big)^T g^* + O\big(\max(\|x_0 - x^*\|^2, \|x_K - x^*\|^2)\big). \tag{4.31}$$

Similarly, we expand c about x^*:

$$c_0 = c^* + A^*(x_0 - x^*) + O(\|x_0 - x^*\|^2) \tag{4.32a}$$
$$\text{and}\ \ c_K = c^* + A^*(x_K - x^*) + O(\|x_K - x^*\|^2). \tag{4.32b}$$

Substituting the expression $g^* = {A^*}^T \lambda^*$ and (4.32a)–(4.32b) in (4.31), we obtain

$$f_0 - f_K = (c_0 - c_K)^T \lambda^* + O\big(\max(\|x_0 - x^*\|^2, \|x_K - x^*\|^2)\big). \tag{4.33}$$

We thus seek to bound

$$\rho_0(f_0 - f_K) = \rho_0 c_0^T \lambda^* - \rho_0 c_K^T \lambda^* + \rho_0 O\big(\max(\|x_0 - x^*\|^2, \|x_K - x^*\|^2)\big). \tag{4.34}$$

To derive a bound for the first term on the right-hand side of (4.34), we first bound a similar term involving μ_0. Let ρ_0^- denote the penalty parameter at the *beginning* of iteration 0 that was "too small" to satisfy (4.7). Following the discussion in the proof of Lemma 4.3, it must hold that

$$\phi_0'(\rho_0^-) > -\tfrac{1}{2} p_0^T H_0 p_0. \tag{4.35}$$

Using (4.35) and expression (4.9), and noting that $\mu_0^T(s_0 + q_0) = 0$, we obtain the following inequality:

$$s_0^T \mu_0 < -\tfrac{1}{2} p_0^T H_0 p_0 - 2(c_0 - s_0)^T(\lambda_0 - \mu_0) - \rho_0^-(c_0 - s_0)^T(c_0 - s_0). \tag{4.36}$$

Since $s_0 \geq 0$ and $\mu_0 \geq 0$, the left-hand side of (4.36) is non-negative. The only term on the right-hand side of (4.36) that can be positive is the middle term, and so the following must hold:

$$\rho_0 s_0^T \mu_0 < -2\rho_0(c_0 - s_0)^T(\lambda_0 - \mu_0).$$

Because $\rho_0\|c_0 - s_0\|$, $\|\lambda_0\|$ and $\|\mu_0\|$ are bounded, we conclude that

$$\rho_0 s_0^T \mu_0 < M. \tag{4.37}$$

Because of Lemma 4.1 and strict complementarity, (4.37) implies that

$$\rho_0 s_0^T \lambda^* < M. \tag{4.38}$$

We know from Lemma 4.4 that $\rho_0\|c_0 - s_0\| < M$. Hence, the boundedness of $\|\lambda^*\|$ shows that

$$\rho_0 |(c_0 - s_0)^T \lambda^*| < M. \tag{4.39}$$

Combining (4.38) and (4.39) gives

$$\rho_0 c_0^T \lambda^* < M. \tag{4.40}$$

Now consider the second term on the right-hand side of (4.34). For sufficiently large $\bar{l}$, p_{K} will be small enough so that Lemma 4.1 implies that the inactive constraints at x^* are also inactive at x_K, i.e., $c_i(x_K) > 0$ for all i such that $c_i(x^*) > 0$. Since $\lambda_i^* = 0$ for an inactive constraint, it follows that

$$-\rho_0 c_K^T \lambda^* = -\rho_0 \widehat{c}_K^T \hat{\lambda}^*, \tag{4.41}$$

where $\widehat{c}$ and $\hat{\lambda}$ denote the components associated with the constraints active at x^*. The only terms in the sum on the right-hand side of (4.41) that affect the upper bound are those for which $\widehat{c}_i(x_K) < 0$. Let $\widetilde{c}$ denote the subvector of strictly negative components of $\widehat{c}$, so that

$$-\rho_0 c_K^T \lambda^* \le -\rho_0 \widetilde{c}_K^T \widetilde{\lambda}^*.$$

Since the associated slack variables are zero, the definition of $\widetilde{c}$ implies

$$\|\widetilde{c}_K\| = \|\widetilde{c}_K - \widetilde{s}_K\| \le \|c_K - s_K\|. \tag{4.42}$$

Applying Lemma 4.4 and the relation $\rho_0 < \rho_K$ to (4.42), it follows that $\rho_0\|\widetilde{c}(x_K)\|$ is bounded. Consequently, because $\|\lambda^*\|$ is bounded, we conclude that

$$-\rho_0 c_K^T \lambda^* < M. \tag{4.43}$$

Finally, consider the third term on the right-hand side of (4.34). Recall from Lemma 4.1 that

$$\|x_0 - x^*\| \le M_p\|p_0\| \quad\text{and}\quad \|x_K - x^*\| \le M_p\|p_K\|. \tag{4.44}$$

It follows from Lemma 4.4 and the relation $\rho_0 < \rho_K$ that

$$\rho_0\|p_0\|^2 < N \quad\text{and}\quad \rho_0\|p_K\|^2 < N,$$

and hence

$$\rho_0 O\big(\max(\|x_0 - x^*\|^2, \|x_K - x^*\|^2)\big) < M. \tag{4.45}$$

Combining (4.40), (4.43) and (4.45), we obtain the bound

$$\rho_0(f_0 - f_K) < M,$$

which implies the desired result. ∎

Lemma 4.6. *There exists a bounded constant M such that, for all l,*

$$\rho_{k_l} \sum_{k=k_l}^{k_{l+1}-1} \|\alpha_k p_k\|^2 < M. \tag{4.46}$$

Proof. As in the previous lemma, we shall use the subscripts 0 and K to denote quantities associated with iterations k_l and k_{l+1} respectively. Using the identity (4.26)

$$\phi_0 - \phi_K = \sum_{k=0}^{K-1} (\phi_k - \phi_{k+1}),$$

observe that properties (3.1) and (3.2) imposed by the choice of α_k imply that for $0 \le k \le K-1$,

$$\phi_k - \phi_{k+1} \ge -\sigma\alpha_k\phi_k', \tag{4.47}$$

where $0 < \sigma < 1$. Because the penalty parameter is not increased, (4.7) must hold for $k = 0$, $\ldots$, $K - 1$. It follows from (4.7) and assumption (iii) that

$$\phi'_k \le -\gamma \|p_k\|^2. \tag{4.48}$$

Since α_k, η and γ are positive, combining (4.26), (4.47) and (4.48) gives

$$\sum_{k=0}^{K-1} \gamma\sigma\alpha_k \|p_k\|^2 \le \phi_0 - \phi_K. \tag{4.49}$$

Rearranging and using the property that $0 < \alpha_k \le 1$, we obtain

$$\gamma\sigma \sum_{k=0}^{K-1} \|\alpha_k p_k\|^2 \le \phi_0 - \phi_K. \tag{4.50}$$

Since γ and σ are positive and bounded away from zero, the desired result follows by multiplying (4.50) by ρ_0 and using Lemma 4.5. ∎

Lemma 4.7. . *There exists a bounded constant M such that, for all k,*

$$\rho_k \|c_k - s_k\| \le M. \tag{4.51}$$

Proof. The result is nontrivial only for the unbounded case. Using the notation of the two previous lemmas, we note that (4.51) is immediate from Lemma 4.4 for $k = 0$ and $k = K$.

To verify a bound for $k = 1, \ldots, K-1$ (iterations at which the penalty parameter is not increased), we first consider x_1. Let unbarred and barred quantities denote evaluation at x_0 and x_1 respectively. If $\bar{c}_i \ge 0$, a bound on $\rho_0|\bar{c}_i - \bar{s}_i|$ follows from definition (2.8) and the boundedness of $\|\lambda\|$. Therefore, assume that $\bar{c}_i < 0$, and expand the i-th constraint function about x_0:

$$\bar{c}_i = c_i + \alpha_0 a_i^T p + O(\|\alpha_0 p_0\|^2). \tag{4.52}$$

We substitute from (2.9) into (4.52), noting that $\bar{s}_i = 0$, and obtain:

$$\begin{aligned} \bar{c}_i = \bar{c}_i - \bar{s}_i &= c_i + \alpha_0(-c_i + s_i + q_i) + O(\|\alpha_0 p_0\|^2) \\ &= (1 - \alpha_0)c_i + \alpha_0(s_i + q_i) + O(\|\alpha_0 p_0\|^2). \end{aligned} \tag{4.53}$$

Adding and subtracting $(1 - \alpha_0)s_i$ on the right-hand side of (4.53) gives

$$\bar{c}_i - \bar{s}_i = (1 - \alpha_0)(c_i - s_i) + (1 - \alpha_0)s_i + \alpha_0(s_i + q_i) + O(\|\alpha_0 p_0\|^2). \tag{4.54}$$

The properties of α_0, s_i and q_i imply that

$$(1 - \alpha_0)s_i + \alpha_0(s_i + q_i) \ge 0.$$

Since the quantity $(\bar{c}_i - \bar{s}_i)$ is strictly negative, (4.54) gives the following inequality that holds when $c_i < 0$:

$$\rho_0|\bar{c}_i - \bar{s}_i| \le \rho_0(1 - \alpha_0)|c_i - s_i| + \rho_0 O(\|\alpha_0 p_0\|^2). \tag{4.55}$$

There are two cases to consider in analyzing (4.55). First, when $c_i \ge 0$, the term $\rho|c_i - s_i|$ is bounded above, using (2.8). The second term on the right-hand side of (4.55) is bounded above, using Lemma 4.6. Thus, the desired bound

$$\rho_0|\bar{c}_i - \bar{s}_i| < M$$

follows if $c_i \ge 0$. Extending this reasoning over the sequence $k = 1, \ldots, K-1$, we see that the quantity $\rho_0 |c_i(x_k) - s_i(x_k)|$ is bounded whenever $c_i(x_k) \ge 0$ or $c_i(x_{k-1}) \ge 0$.

Consequently, the only remaining case involves components of c that are *negative* at two or more consecutive iterations. Let $\widetilde{c}$ denote the subvector of such components of c. Using the componentwise inequality (4.55) and the fact that $0 < \alpha \le 1$, we have

$$\rho_0 \|\widetilde{c}(x_1) - \widetilde{s}(x_1)\| \le \rho_0 \|\widetilde{c}(x_0) - \widetilde{s}(x_0)\| + \rho_0 O(\|\alpha_0 p_0\|^2). \tag{4.56}$$

Proceeding over the relevant sequence of iterations, the following inequality must hold for $k = 1, \ldots, K-1$:

$$\rho_0 \|\widetilde{c}(x_k) - \widetilde{s}(x_k)\| \le \rho_0 \|\widetilde{c}(x_0) - \widetilde{s}(x_0)\| + \rho_0 O\Big(\sum_{j=0}^{k-1} \|\alpha_j p_j\|^2\Big). \tag{4.57}$$

The desired result then follows by applying Lemmas 4.4 and 4.6 to (4.57). ∎

The next two lemmas establish the existence of a step bounded away from zero, *independent of k and the size of ρ*, for which a sufficient decrease condition is satisfied.

Lemma 4.8. *For $0 \le \theta \le \alpha$,*

$$\phi''(\theta) \le -\phi'(0) + q^T\mu + N\|p\|^2,$$

where N is bounded and independent of k.

Proof. Using (4.8), we have

$$\nabla^2 L = \begin{pmatrix} \nabla^2 f - \sum \big(\lambda_i + \rho(c_i - s_i)\big)\nabla^2 c_i + \rho A^T A & -A^T & -\rho A^T \\ -A & 0 & I \\ -\rho A & I & \rho I \end{pmatrix},$$

so that

$$\phi''(\theta) = y^T \nabla^2 L(v + \theta y) y = p^T W(\theta) p - \sum \rho\,\big(c_i(\theta) - s_i(\theta)\big) p^T \nabla^2 c_i(\theta) p$$
$$+ \rho\big(A(\theta)p - q\big)^T\big(A(\theta)p - q\big) - 2\xi^T\big(A(\theta)p - q\big), \tag{4.58}$$

where

$$W(\theta) = \nabla^2 f(\theta) - \sum(\lambda_i + \theta\xi_i)\nabla^2 c_i(\theta).$$

We now derive bounds on the first two terms on the right-hand side of (4.58). The first term is bounded in magnitude by a constant multiple of $\|p\|^2$ because of assumption (ii) and the boundedness of $\|\lambda\|$ (from Lemma 4.2). For the second term, we expand c_i in a Taylor series about x:

$$c_i(x + \theta p) = c_i(x) + \theta a_i(x)^T p + \tfrac{1}{2}\theta^2 p^T \nabla^2 c_i(x + \theta_i p) p, \tag{4.59}$$

where $0 < \theta_i < \theta$. Since $s_i(\theta) = s_i + \theta q_i$, using (2.9a) and multiplying by ρ, we have

$$\rho\,\big(c_i(x + \theta p) - (s_i + \theta q_i)\big) = \rho\,(1 - \theta)\big(c_i(x) - s_i\big) + \rho\,\tfrac{1}{2}\theta^2 p^T \nabla^2 c_i(x + \theta_i p) p. \tag{4.60}$$

We know from Lemma 4.7 that $\big|\rho\big(c_i(x) - s_i\big)\big|$ is bounded, and from Lemma 4.6 that $\rho\|\alpha p\|^2$ is bounded. Therefore,

$$\rho\big|\big(c_i(\theta) - s_i(\theta)\big)\big| \le J_i, \tag{4.61}$$

where J_i is bounded and independent of the iteration. Using (4.61), we obtain the overall bound

$$\sum |\rho\left(c_i(\theta) - s_i(\theta)\right)p^T\nabla^2 c_i(\theta)p| \le J\|p\|^2, \tag{4.62}$$

where J is bounded and independent of the iteration.

Now we examine the third term on the right-hand side of (4.58). Using Taylor series, we have

$$a_i(x+\theta p)^T p = a_i^T p + \theta p^T \nabla^2 c_i(\bar\theta_i)p, \tag{4.63}$$

where $0 < \bar\theta_i < \theta$. Using (2.9b) and Lemmas 4.6 and 4.7, we obtain

$$\rho\big(A(\theta)p - q\big)^T\big(A(\theta)p - q\big) \le \rho(c-s)^T(c-s) + L\|p\|^2, \tag{4.64}$$

where $|L|$ is bounded and independent of the iteration.

Using (4.63) and the boundedness of $\|\xi\|$, the final term on the right-hand side of (4.58) can be written as

$$-2\xi^T\big(A(\theta)p - q\big) \le 2\xi^T\big(c(x) - s\big) + M\|p\|^2, \tag{4.65}$$

where $|M|$ is bounded and independent of the iteration.

Combining (4.64) and (4.65), the last two terms on the right-hand side of (4.58) become

$$\begin{aligned}\rho\left(A(\theta)p - q\right)^T\left(A(\theta)p - q\right) - 2\xi^T\left(A(\theta)p - q\right) &\le \rho\,(c-s)^T(c-s) + 2\xi^T(c-s) + \tilde M\|p\|^2\\ &\le -\phi'(0) + q^T\mu + \bar M\|p\|^2,\end{aligned}$$

where $|\bar M|$ is bounded and independent of the iteration (using (4.11) and noting that the largest eigenvalue of H is bounded).

Combining all these bounds gives the required result. ∎

Lemma 4.9. *The line search in Step 3 of the algorithm defines a step length α ($0 < \alpha \le 1$) such that*

$$\phi(\alpha) - \phi(0) \le \sigma\alpha\phi'(0), \tag{4.66}$$

and $\alpha \ge \bar\alpha$, where $0 < \sigma < 1$, and $\bar\alpha > 0$ is bounded away from zero and independent of the iteration.

Proof. If both conditions (3.1) are satisfied at a given iteration, then $\alpha = 1$ and (4.66) holds with α trivially bounded away from zero.

Assume that (3.1) does not hold (i.e., α is computed by safeguarded cubic interpolation). The existence of a step length α that satisfies conditions (3.2) is guaranteed from standard analysis (see, for example, Moré and Sorensen [22]). We need to show that α is uniformly bounded away from zero. There are two cases to consider.

First, assume that (3.1a) does not hold, i.e., $\phi(1) - \phi(0) > \sigma\phi'(0)$. Since $\phi'(0) < 0$, this implies the existence of at least one positive zero of the function

$$\psi(\alpha) = \phi(\alpha) - \phi(0) - \sigma\alpha\phi'(0).$$

Let α^* denote the smallest such zero. Since ψ vanishes at zero and α^*, and $\psi'(0) < 0$, the mean-value theorem implies the existence of a point $\hat\alpha$ ($0 < \hat\alpha < \alpha^*$) such that $\psi'(\hat\alpha) = 0$, i.e., for which

$$\phi'(\hat\alpha) = \sigma\phi'(0).$$

Because $\sigma \le \eta$, it follows that

$$\phi'(\hat\alpha) - \eta\phi'(0) = (\sigma - \eta)\phi'(0) \ge 0.$$

Therefore, since the function $\phi'(\alpha) - \eta\phi'(0)$ is negative at $\alpha = 0$, and non-negative at $\hat{\alpha}$, the mean-value theorem again implies the existence of a smallest value $\bar{\alpha}$ ($0 < \bar{\alpha} \le \hat{\alpha}$) such that

$$\phi'(\bar{\alpha}) = \eta\phi'(0). \tag{4.67}$$

The point $\bar{\alpha}$ is the required lower bound on the step length because (4.67) implies that (3.2b) will not be satisfied for any $\alpha \in [0, \bar{\alpha})$.

Expanding ϕ' in a Taylor series gives

$$\phi'(\bar{\alpha}) = \phi'(0) + \bar{\alpha}\phi''(\theta),$$

where $0 < \theta < \bar{\alpha}$. Therefore, using (4.67) and noting that $\eta < 1$ and $\phi'(0) < 0$, we obtain

$$\bar{\alpha} = \frac{\phi'(\bar{\alpha}) - \phi'(0)}{\phi''(\theta)} = (1-\eta)\frac{|\phi'(0)|}{\phi''(\theta)}. \tag{4.68}$$

(Since $\bar{\alpha} > 0$, θ must be such that $\phi''(\theta) > 0$.) We seek a lower bound on $\bar{\alpha}$, and hence an upper bound on the denominator of (4.68). We know from Lemma 4.8 that

$$\phi''(\theta) = -\phi'(0) + q^T\mu + N\|p\|^2,$$

and from (2.10) that $q^T\mu \le 0$. Therefore,

$$\phi''(\theta) \le |\phi'(0)| + |N|\|p\|^2,$$

and hence

$$\bar{\alpha} \ge \frac{(1-\eta)|\phi'(0)|}{|\phi'(0)| + |N|\|p\|^2}.$$

Dividing by $|\phi'(0)|$ gives

$$\bar{\alpha} \ge \frac{(1-\eta)}{1 + \dfrac{|N|\|p\|^2}{|\phi'(0)|}}. \tag{4.69}$$

Since the algorithm guarantees that $\phi'(0) \le -\frac{1}{2}p^T Hp$, it follows that

$$|\phi'(0)| \ge \tfrac{1}{2}p^T Hp \ge \tfrac{1}{2}\gamma\|p\|^2, \tag{4.70}$$

where γ is bounded below. Thus, the denominator of (4.69) may be bounded above as follows:

$$1 + \frac{|N|\|p\|^2}{|\phi'(0)|} \le 1 + \frac{|N|\|p\|^2}{\frac{1}{2}\gamma\|p\|^2} = 1 + \frac{2|N|}{\gamma}.$$

A uniform lower bound on $\bar{\alpha}$ is accordingly given by

$$\bar{\alpha} \ge \frac{\gamma(1-\eta)}{\gamma + 2|N|}. \tag{4.71}$$

In the second case, we assume that (3.1a) is satisfied, but (3.1b) is not. In this case, it must hold that $\phi'(1) \ge 0$, and hence $\phi'(\alpha)$ must have at least one zero in $(0, 1]$. If $\bar{\alpha}$ denotes the smallest of these zeros, $\bar{\alpha}$ satisfies (4.67), and (4.71) is again a uniform lower bound on the step length. ∎

The proof of a global convergence theorem is now straightforward.

Theorem 4.1. *Under assumptions* (i)–(v)*, the algorithm defined by* (1.2)*,* (1.3)*,* (4.7)*, and* (4.66) *has the property that*

$$\lim_{k\to\infty} \|p_k\| = 0. \tag{4.72}$$

Proof. If $\|p_k\| = 0$ for any finite k, the algorithm terminates and the theorem is true. Hence we assume that $\|p_k\| \neq 0$ for any k.

When there is no upper bound on the penalty parameter, the uniform lower bound on α of Lemma 4.9 and (4.46) imply that, for any $\delta > 0$, we can find an iteration index K such that

$$\|p_k\| \leq \delta \text{ for } k \geq K,$$

which implies that $\|p_k\| \to 0$, as required.

In the bounded case, we know that there exists a value $\tilde{\rho}$ and an iteration index $\tilde{K}$ such that $\rho = \tilde{\rho}$ for all $k \geq \tilde{K}$. We consider henceforth only such values of k.

The proof is by contradiction. We assume that there exists $\epsilon > 0$ and $K \geq \tilde{K}$ such that $\|p_k\| \geq \epsilon$ for $k \geq K$. Every subsequent iteration must therefore yield a strict decrease in the merit function (2.7) with $\rho = \tilde{\rho}$, because, using (4.66),

$$\phi(\alpha) - \phi(0) \leq \eta\alpha\phi'(0) \leq \tfrac{1}{2}\eta\bar{\alpha}\gamma\epsilon^2 < 0.$$

The two final inequalities are derived from Lemma 4.9, since $\alpha \geq \bar{\alpha}$, which is uniformly bounded away from zero. The adjustment of the slack variables s in Step 2 of the algorithm can only lead to a further reduction in the merit function. Therefore, since the merit function with $\rho = \tilde{\rho}$ decreases by at least a fixed quantity at every iteration, it must be unbounded below. But this is impossible, from assumptions (i)–(ii) and Lemma 4.2. Therefore, (4.72) must hold. ∎

Corollary 4.1.

$$\lim_{k\to\infty} \|x_k - x^*\| = 0.$$

Proof. The result follows immediately from Theorem 4.1 and Lemma 4.1. ∎

The second theorem shows that Algorithm NPSQP also exhibits convergence of the multiplier estimates $\{\lambda_k\}$.

Theorem 4.2.

$$\lim_{k\to\infty} \|\lambda_k - \lambda^*\| = 0.$$

Proof. If $\|p_k\| = 0$, then $\mu_k = \lambda^*$ (using (1.1) and (1.4)); in Step 1 of Algorithm NPSQP, λ_{k+1} is set to λ^*, and the algorithm terminates. Thus, the theorem is true if $p_k = 0$ for any k. We therefore assume that $\|p_k\| \neq 0$ for any k.

The definition (4.6) gives

$$\lambda_{k+1} = \sum_{j=0}^{k} \gamma_{jk}\mu_j, \tag{4.73}$$

where

$$\gamma_{kk} = \alpha'_k \text{ and } \gamma_{jk} = \alpha'_j \prod_{r=j+1}^{k} (1 - \alpha'_r), \quad j < k, \tag{4.74}$$

with $\alpha'_0 = 1$ and $\alpha'_j = \alpha_j$, $j \geq 1$. (This convention is used because of the special initial condition that $\lambda_0 = \mu_0$.) From Lemma 4.9 and (4.74), we observe that

$$0 < \bar{\alpha} \leq \alpha'_j \leq 1 \quad \text{for all } j, \tag{4.75a}$$

$$\sum_{j=0}^{k} \gamma_{jk} = 1 \tag{4.75b}$$

$$\text{and} \ \ \gamma_{jk} \leq (1-\bar{\alpha})^{k-j}, \quad j < k. \tag{4.75c}$$

Since we know from Theorem 4.1 that $x_k \to x^*$, the iterates will eventually reach a neighborhood of x^* in which the QP subproblem identifies the correct active set (Lemma 4.1) and $\hat{A}(x_k)$ has full rank (assumption (iv)). Assume that these properties hold for $k \geq K_1$. From the definition (1.4) of μ and assumptions (ii)–(iii), we have for $k \geq K_1$ that there exists a bounded scalar M such that

$$\mu_k = \lambda^* + M_k d_k t_k, \tag{4.76}$$

with $|M_k| \leq M$, $d_k = \max(\|p_k\|, \|x^* - x_k\|)$ and $\|t_k\| = 1$. Given any $\epsilon > 0$, Theorem 4.1 and Corollary 4.1 also imply that K_1 can be chosen so that, for $k \geq K_1$,

$$|M_k d_k| \leq \tfrac{1}{2}\epsilon. \tag{4.77}$$

We can also define an iteration index K_2 with the following property:

$$(1-\bar{\alpha})^k \leq \frac{\epsilon}{2(k+1)(1+\bar{\mu}+\|\lambda^*\|)} \tag{4.78}$$

for $k \geq K_2 + 1$, where $\bar{\mu}$ is an upper bound on $\|\mu\|$ for all k. Let $K = \max(K_1, K_2)$. Then, from (4.73) and (4.76), we have for $k \geq 2K$,

$$\lambda_{k+1} = \sum_{j=0}^{K} \gamma_{jk}\mu_j + \sum_{j=K+1}^{k} \gamma_{jk}(\lambda^* + M_j d_j t_j).$$

Hence it follows from (4.75b) that:

$$\lambda_{k+1} - \lambda^* = \sum_{j=0}^{K} \gamma_{jk}(\mu_j - \lambda^*) + \sum_{i=K+1}^{k} \gamma_{jk} M_j d_j t_j.$$

From the bounds on $\|\mu_j\|$ and $\|t_j\|$ we then obtain

$$\|\lambda_{k+1} - \lambda^*\| \leq (\bar{\mu} + \|\lambda^*\|) \sum_{j=0}^{K} \gamma_{jk} + \sum_{j=K+1}^{k} \gamma_{jk} |M_j d_j|. \tag{4.79}$$

Since $k \geq 2K$, it follows from (4.75a) and (4.75c) that

$$\sum_{j=0}^{K} \gamma_{jk} \leq \sum_{j=0}^{K} (1-\bar{\alpha})^{k-j} \leq \sum_{j=0}^{K} (1-\bar{\alpha})^{2K-j} \leq (K+1)(1-\bar{\alpha})^K.$$

Using (4.78), we thus obtain the following bound for the first term on the right-hand side of (4.79):

$$(\bar{\mu} + \|\lambda^*\|) \sum_{j=0}^{K} \gamma_{jk} \leq \tfrac{1}{2}\epsilon. \tag{4.80}$$

To bound the second term in (4.79), we use (4.75b) and (4.77):

$$\sum_{j=K+1}^{k} \gamma_{jk} |M_j d_j| \le \tfrac{1}{2}\epsilon \sum_{j=K+1}^{k} \gamma_{jk} \le \tfrac{1}{2}\epsilon. \tag{4.81}$$

Combining (4.79)–(4.81), we obtain the following result: given any $\epsilon > 0$, we can find K such that

$$\|\lambda_k - \lambda^*\| \le \epsilon \ \text{ for } \ k \ge 2K + 1,$$

which implies that

$$\lim_{k\to\infty} \|\lambda_k - \lambda^*\| = 0.$$

∎

5. Use within a Superlinearly Convergent Algorithm

As mentioned in Section 2.1, a point of interest is whether superlinear convergence may be impeded by the requirement of a sufficient decrease in the merit function at every iteration. In this section we show that a unit step ($\alpha = 1$) will satisfy conditions (3.1) when the iterates are sufficiently close to the solution and x and λ are converging superlinearly at the same rate. For further discussion of superlinear convergence, see Dennis and Moré [8] and Boggs, Tolle and Wang [6].

In addition to the conditions assumed in the previous section, we assume that for all sufficiently large k:

$$x_k + p_k - x^* = o(\|x_k - x^*\|), \tag{5.1a}$$

$$\lambda_k + \xi_k - \lambda^* = o(\|\lambda_k - \lambda^*\|), \tag{5.1b}$$

$$\frac{\|p_k\|}{\|\xi_k\|} > M > 0, \tag{5.1c}$$

where M is independent of k. Note that (5.1a) implies

$$\|p_k\| \sim \|x_k - x^*\| \ \text{ and } \ \|\xi_k\| \sim \|\lambda_k - \lambda^*\|, \tag{5.2}$$

where "$\sim$" means that the two quantities are of similar order. (See also Dennis and Moré [8].)

First we show that these assumptions imply that the penalty parameter ρ in the merit function (2.7) must remain bounded for all k.

Lemma 5.1. *Under assumptions* (i)–(v) *of Section 3, and conditions* (5.1a)*, there exists a finite $\bar\rho$ such that for all k,*

$$\rho \le \bar\rho. \tag{5.3}$$

Proof. The proof is by contradiction. Assume that (5.3) does not hold, in which case the discussion following Lemma 4.3 shows that (4.13) must hold at an infinite subsequence of iterations. Condition (4.13) states that at any iteration in this subsequence, we have $-2(c-s)^T\xi \ge \frac{1}{2}p^T Hp$. Using assumption (iii) and norm inequalities, (4.13) thus implies

$$\|c - s\| \ge \frac{\frac{1}{4}\gamma\|p\|^2}{\|\xi\|}.$$

Using (5.1c), we obtain

$$\frac{\|c - s\|}{\|\xi\|} \ge \frac{\frac{1}{4}\gamma\|p\|^2}{\|\xi\|^2} = \bar M > 0.$$

Consequently, from Lemma 4.4,

$$\hat{\rho} < \frac{2}{\bar{M}}$$

and must remain bounded over the infinite subsequence of iterations. This contradicts the unboundedness of the penalty parameter, and thereby proves the lemma. ∎

The next two lemmas show that conditions (3.1) are eventually satisfied for all k sufficiently large.

Lemma 5.2. *Under assumptions* (i)–(v) *of Section 3 and conditions* (5.1a)*, the sufficient-decrease condition* (3.1a) *holds for sufficiently large* k*, i.e.,*

$$\phi(1) - \phi(0) \le \sigma\phi'(0),$$

where $0 < \sigma < \frac{1}{2}$.

Proof. As in Powell and Yuan [29], observe that the continuity of second derivatives gives the following relationships:

$$\begin{aligned} f(x+p) &= f(x) + \tfrac{1}{2}\big(g(x) + g(x+p)\big)^T p + o(\|p\|^2),\\ c(x+p) &= c(x) + \tfrac{1}{2}\big(A(x) + A(x+p)\big)p + o(\|p\|^2). \end{aligned}$$

Conditions (5.1a) and (5.2) then imply:

$$f(x+p) = f(x) + \tfrac{1}{2}\big(g(x) + g(x^*)\big)^T p + o(\|p\|^2), \tag{5.4a}$$

$$c(x+p) = c(x) + \tfrac{1}{2}\big(A(x) + A(x^*)\big)p + o(\|p\|^2). \tag{5.4b}$$

We shall henceforth use g to denote $g(x)$ and g^* to denote $g(x^*)$, and similarly for f, c and A.

By definition,

$$\phi(0) = f - \lambda^T(c-s) + \tfrac{1}{2}\rho(c-s)^T(c-s), \tag{5.5a}$$

$$\begin{aligned} \phi(1) = {}& f(x+p) - \mu^T(c(x+p) - c - Ap)\\ &+\tfrac{1}{2}\rho\big(c(x+p) - c - Ap\big)^T\big(c(x+p) - c - Ap\big). \end{aligned} \tag{5.5b}$$

Using Taylor series, we have

$$c(x+p) = c + Ap + O(\|p\|^2). \tag{5.6}$$

Substituting (5.4) and (5.6) into (5.5b), we obtain

$$\phi(1) = f + \tfrac{1}{2}(g+g^*)^T p + \tfrac{1}{2}\mu^T(A - A^*)p + o(\|p\|^2). \tag{5.7}$$

Combining (5.5a) and (5.7) gives

$$\begin{aligned} \phi(1) - \phi(0) = {}& \tfrac{1}{2}p^T g + \tfrac{1}{2}p^T g^* + \tfrac{1}{2}\mu^T Ap - \tfrac{1}{2}\mu^T A^* p + \lambda^T(c-s)\\ &-\tfrac{1}{2}\rho(c-s)^T(c-s) + o(\|p\|^2). \end{aligned} \tag{5.8}$$

Using (2.6), (2.9) and (4.9), we obtain the following expression:

$$\phi'(0) = p^T g + 2\lambda^T(c-s) - \rho(c-s)^T(c-s) + \mu^T Ap - \mu^T q. \tag{5.9}$$

Substituting (5.9) into (5.8), we have

$$\phi(1) - \phi(0) = \tfrac{1}{2}\phi'(0) + \tfrac{1}{2}\mu^T q + \tfrac{1}{2}p^T(g^* - {A^*}^T\mu) + o(\|p\|^2). \tag{5.10}$$

It follows from (5.1b)–(5.1c) that the expression $g^* - {A^*}^T\mu$ is $o(\|p\|)$, which gives

$$\begin{aligned}\phi(1) - \phi(0) &= \tfrac{1}{2}\phi'(0) + \tfrac{1}{2}\mu^T q + o(\|p\|^2)\\ &\le \tfrac{1}{2}\phi'(0) + o(\|p\|^2),\end{aligned}$$

and hence

$$\phi(1) - \phi(0) - \sigma\phi'(0) \le (\tfrac{1}{2} - \sigma)\phi'(0) + o(\|p\|^2). \tag{5.11}$$

Since $\sigma < \frac{1}{2}$ and $\phi'(0)$ satisfies (4.7), (5.11) implies that (3.1a) holds for sufficiently large k. ∎

Lemma 5.3. *Under assumptions* (i)–(v) *and conditions* (5.1a)*, the second line search condition* (3.1b) *holds for sufficiently large k, i.e.,*

$$\phi'(1) \le \eta\phi'(0) \;\; \textit{or} \;\; |\phi'(1)| \le \eta|\phi'(0)|,$$

where $\eta < \frac{1}{2}$.

Proof. In this proof, we use the notation $g(1)$ to denote $g(x+p)$, and similarly for c and A. Using (4.10), $\phi'(1)$ is given by

$$\begin{aligned}\phi'(1) = p^T g(1) - p^T A(1)\mu + \rho p^T\big(c(1) - c - Ap\big)&\\ -(\mu - \lambda)^T\big(c(1) - c - Ap\big) + q^T\mu - \rho q^T\big(c(1) - c - Ap\big)&.\end{aligned} \tag{5.12}$$

From conditions (5.2)–(5.3), we have

$$g(1) = g^* + o(\|p\|^2), \quad A(1) = A^* + o(\|p\|^2) \;\text{ and }\; \mu = \lambda^* + o(\|p\|). \tag{5.13}$$

Substituting from (5.6) and (5.13) into (5.12) then gives

$$\phi'(1) = p^T g^* - p^T A^*\mu + q^T\mu - \rho(c - s)^T\big(c(1) - c - Ap\big) + o(\|p\|^2). \tag{5.14}$$

Consider now the vectors $(c-s)$, q and μ. From (2.8) and (5.2), $\|c-s\| = O(\|p\|)$. From Lemma 4.1, we know that the QP subproblem (1.3) will eventually predict the correct active set, and hence that $\mu_i = 0$ if $c_i(x^*) > 0$. For an active constraint c_i, it follows from (2.8) that s_i will eventually be set to zero at the beginning of every iteration if $\rho > 0$, and hence q_i must also be zero. If $\rho = 0$ and c_i is an active constraint, then $|q_i| = o(\|p\|^2)$. Therefore, in either case we have that

$$|q^T\mu| = o(\|p\|^2) \;\text{ and }\; |\rho(c - s)^T\big(c(1) - c - Ap\big)| = o(\|p\|^2). \tag{5.15}$$

Recalling that $\|g^* - {A^*}^T\mu\| = o(\|p\|)$ and using (5.15) in (5.14), we obtain

$$\phi'(1) = o(\|p\|^2). \tag{5.16}$$

Since $|\phi'(0)| \ge \frac{1}{2}\gamma\|p\|^2$, (5.16) implies that (3.1b) will eventually be satisfied at every iteration. ∎

6. Numerical Results

In order to indicate the reliability and efficiency of a practical SQP algorithm based on the merit function (2.7), we present a selection of numerical results obtained from the Fortran code NPSOL (Gill *et al.* [15]). Table 1 contains the results of solving a subset of problems 70–119 from Hock and Schittkowski [20]. (We have omitted problems that are non-differentiable or that contain only linear constraints. Since NPSOL treats linear constraints separately, they do not affect the merit function.)

The problems were run on an IBM 3081K in double precision (i.e., machine precision ϵ_M is approximately 2.22×10^{-16}). The default parameters for NPSOL were used in all cases (for details, see Gill *et al.* [15]). In particular, the default value for `ftol`, the final accuracy requested in f, is 5.4×10^{-12}, and `ctol`, the feasibility tolerance, is $\sqrt{\epsilon_M}$. Analytic gradients were provided for all functions, and a gradient linesearch was used.

For successful termination of NPSOL, the iterative sequence of x-values must have converged and the final point must satisfy the first-order Kuhn-Tucker conditions (*cf.* (1.1)). The sequence of iterates is considered to have converged at x if

$$\alpha\|p\| \le \sqrt{\texttt{ftol}}\,(1 + \|x\|), \tag{6.1}$$

where p is the search direction and α the step length in (1.2). The iterate x is considered to satisfy the first-order conditions for a minimum if the following conditions hold: the inactive constraints are satisfied to within `ctol`, the magnitude of each active constraint residual is less than `ctol`, and

$$\|Z(x)^T g(x)\| \le \sqrt{\texttt{ftol}}\,\bigl(1 + \max(1 + |f(x)|, \|g(x)\|)\bigr), \tag{6.2}$$

where $Z(x)$ is an orthogonal basis for the null space of the gradients of the active constraints. (Thus, $Z(x)^T g(x)$ is the usual *reduced gradient*.)

Table 1 gives the following information: the problem number in Hock and Schittkowski; the number of variables (n); the number of simple bounds (m_B); the number of general linear constraints (m_L); the number of nonlinear constraints (m_N); the number of iterations; and the number of function evaluations. (Each constraint is assumed to include a lower and an upper bound.)

NPSOL terminated successfully on all the problems except the badly scaled problem 85, for which (6.1) could not be satisfied. However, the optimality conditions were satisfied at the final iterate, which gave an improved objective value compared to that in Hock and Schittkowski [20].

7. Acknowledgements

We are grateful to Professor M. J. D. Powell for his careful reading of the previous version of this paper, and especially for pointing out the need for Lemma 4.7.

References

[1] M. C. Bartholomew-Biggs, *A recursive quadratic programming algorithm based on the augmented Lagrangian function*, Report 139, Numerical Optimisation Centre, Hatfield Polytechnic, Hatfield, England, 1983.

[2] ———, *The development of recursive quadratic programming methods based on the augmented Lagrangian*, Report 160, Numerical Optimisation Centre, Hatfield Polytechnic, Hatfield, England, 1985.

[3] M. C. Biggs, *Constrained minimization using recursive equality quadratic programming*, in Numerical Methods for Nonlinear Optimization, F. A. Lootsma, ed., Academic Press, London and New York, 1972, pp. 411–428.

[4] P. T. Boggs and J. W. Tolle, *A family of descent functions for constrained optimization*, SIAM J. Numer. Anal., 21 (1984), pp. 1146–1161.

[5] ———, *An efficient strategy for utilizing a merit function in nonlinear programming algorithms*, Technical Report 85-5, Department of Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, 1985.

[6] P. T. Boggs, J. W. Tolle, and P. Wang, *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM J. Control Optim., 20 (1982), pp. 161–171.

[7] R. M. Chamberlain, M. J. D. Powell, C. Lemarechal, and H. C. Pedersen, *The watchdog technique for forcing convergence in algorithms for constrained optimization*, Math. Programming Stud., (1982), pp. 1–17. Algorithms for constrained minimization of smooth nonlinear functions.

[8] J. E. Dennis Jr. and J. J. Moré, *Quasi-Newton methods, motivation and theory*, SIAM Review, 19 (1977), pp. 46–89.

[9] G. DiPillo and L. Grippo, *A new class of augmented Lagrangians in nonlinear programming*, SIAM J. Control Optim., 17 (1979), pp. 618–628.

[10] L. C. W. Dixon, *Exact penalty functions in nonlinear programming*, Report 103, Numerical Optimisation Centre, Hatfield Polytechnic, Hatfield, England, 1979.

[11] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, Inc., New York-London-Sydney, 1968.

[12] R. Fletcher, *A class of methods for nonlinear programming with termination and convergence properties*, in Integer and Nonlinear Programming, J. Abadie, ed., North-Holland, The Netherlands, 1970, pp. 157–175.

[13] ———, *Methods related to Lagrangian functions*, in Numerical Methods for Constrained Optimization, P. E. Gill and W. Murray, eds., London and New York, 1974, Academic Press, pp. 219–240.

[14] P. E. Gill and W. Murray, *Safeguarded steplength algorithms for optimization using descent methods*, Report DNAC 37, Division of Numerical Analysis and Computing, 1973.

[15] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, *User's guide for NPSOL (Version 4.0): a Fortran package for nonlinear programming*, Report SOL 86-2, Department of Operations Research, Stanford University, Stanford, CA, 1986.

[16] J. Goodman, *Newton's method for constrained optimization*, Math. Program., 33 (1985), pp. 162–171.

[17] S.-P. Han, *Superlinearly convergent variable metric algorithms for general nonlinear programming problems*, Math. Program., 11 (1976), pp. 263–282.

[18] S. P. Han, *A globally convergent method for nonlinear programming*, J. Optim. Theory Appl., 22 (1977), pp. 297–309.

[19] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320.

[20] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems 187, Springer Verlag, Berlin, Heidelberg and New York, 1981.

[21] N. Maratos, *Exact Penalty Function Algorithms for Finite-Dimensional and Control Optimization Problems*, PhD thesis, Department of Computing and Control, University of London, 1978.

[22] J. J. Moré and D. C. Sorensen, *Newton's method*, in Studies in Mathematics, Volume 24. Studies in Numerical Analysis, Math. Assoc. America, Washington, DC, 1984, pp. 29–82.

[23] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.

[24] T. Pietrzykowski, *An exact potential method for constrained maxima*, SIAM J. Numer. Anal., 6 (1969), pp. 299–304.

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

[26] ———, *Introduction to constrained optimization*, in Numerical Methods for Constrained Optimization, P. E. Gill and W. Murray, eds., London and New York, 1974, Academic Press, pp. 1–28.

[27] ———, *A fast algorithm for nonlinearly constrained optimization calculations*, Tech. Report 77/NA 2, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England, 1977.

[28] ———, *Variable metric methods for constrained optimization*, in Mathematical Programming: the State of the Art (Bonn, 1982), A. Bachem, M. Grötschel, and B. Korte, eds., Springer, Berlin, 1983, pp. 288–311.

[29] M. J. D. Powell and Y.-X. Yuan, *A recursive quadratic programming algorithm that uses differentiable exact penalty functions*, Math. Program., 35 (1986), pp. 265–278.

[30] S. M. Robinson, *Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms*, Math. Program., 7 (1974), pp. 1–16.

[31] R. T. Rockafellar, *The multiplier method of Hestenes and Powell applied to convex programming*, J. Optim. Theory Appl., 12 (1973), pp. 555–562.

[32] K. Schittkowski, *The nonlinear programming method of Wilson, Han, and Powell with an augmented Lagrangian type line search function*, Numer. Math., 38 (1981), pp. 83–127.

[33] ———, *On the convergence of a sequential quadratic programming method with an augmented Lagrangian line search function*, Math. Operationsforschung u. Statistik, Ser. Optimization, 14 (1983), pp. 197–216.

[34] R. A. Tapia, *Diagonalized multiplier methods and quasi-Newton methods for constrained optimization*, J. Optim. Theory Appl., 22 (1977), pp. 135–194.

[35] M. H. Wright, *Numerical methods for nonlinearly constrained optimization*, PhD thesis, Department of Computer Science, Stanford University, Stanford, CA, 1976.

Table 1: Results on the Hock and Schittkowski problems.

Prob	n	m_B	m_L	m_N	Itns	Evals
70	4	4	0	1	35	38
71	4	4	0	2	5	6
72	4	4	0	2	6	7
73	4	4	2	1	3	4
74	4	4	2	3	9	12
75	4	4	2	3	6	7
77	5	0	0	2	14	20
78	5	0	0	3	10	15
79	5	0	0	3	9	12
80	5	5	0	3	8	10
81	5	5	0	3	14	20
83	5	5	0	3	4	6
84	5	5	0	3	2	3
85	5	5	0	38	17	18
93	6	6	0	2	11	14
95	6	6	0	4	1	2
96	6	6	0	4	1	2
97	6	6	0	4	3	6
98	6	6	0	4	3	6
99	7	7	0	2	18	32
100	7	0	0	4	15	34
101	7	7	0	5	17	19
102	7	7	0	5	29	69
103	7	7	0	5	25	60
104	8	8	0	5	17	19
106	8	8	3	3	17	21
107	9	7	0	6	11	18
108	9	1	0	13	24	45
109	9	9	1	8	10	11
111	10	10	0	3	48	64
113	10	0	3	5	14	19
114	10	10	5	6	18	19
116	13	13	5	10	31	64
117	15	15	0	5	17	21