

ON THE LOCATION OF DIRECTIONS OF INFINITE DESCENT FOR NONLINEAR PROGRAMMING ALGORITHMS*

ANDREW R. CONN† AND NICHOLAS I. M. GOULD‡

Abstract. There is much current interest in general equality constrained quadratic programming problems, both for their own sake and for their applicability to active set methods for nonlinear programming. In the former case, typically, the issues are existence of solutions and their determination. In the latter instance, nonexistence of solutions gives rise to directions of infinite descent. Such directions may subsequently be used to determine a more desirable active set.

The generalised Cholesky decomposition of relevant matrices enables us to answer the question of existence and to determine directions of infinite descent (when applicable) in an efficient and stable manner.

The methods examined are related to implementations that are suitable for null-space, range-space and Lagrangian methods.

1.1. Introduction. The ability to solve equality constrained quadratic programs is of fundamental importance in the theory of nonlinear programming. Firstly, it is the simplest nonlinear programming problem and secondly, much of the basis for current algorithms for nonlinear programming depends upon solving quadratic programming problems as subproblems. Well-known examples include successive quadratic programming (see, for example Chamberlain, Lemaréchal, Pedersen and Powell (1982) and Powell (1978)), active set strategies (see, for example Murray and Wright (1978) and Biggs (1975)) and the method of Coleman and Conn (Coleman and Conn (1982a) and (1982b)).

Consequently, there is much interest in the design of robust and efficient methods for handling general quadratic programs stably.

Our own particular interest in nonlinear programming concerns those methods which attempt to minimize some quadratic modelling function in a particular subspace that represents a linearization of the constraints that it is *supposed* are active at the solution. Such methods give rise to equality constrained quadratic programs.

In this paper we consider the Equality Constrained Quadratic Program.

$$\begin{aligned} EQP: \quad & \underset{p \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2} p^T H p + g^T p \equiv Q(p) \\ & \text{subject to} \quad A p = 0, \end{aligned}$$

where H is $n \times n$ symmetric, A is $t \times n$, rank t ($t \leq n$), and g is an n -vector.

We take the point of view that it is important that our method of determining p is not disjoint from our method of determining existence. Furthermore, if possible, we would like to exploit any structure available, in the implementation.

In the special case when a finite solution to EQP exists, a number of different procedures have been developed for locating it (see, for example, Fletcher (1981)).

*Received by the editors July 5, 1983, and in revised form March 1, 1984. This research was partially supported by Natural Sciences and Engineering Research Council of Canada grants A-8639 and A-8442. This paper was typeset using software developed at Bell Laboratories and the University of California at Berkeley. Final copy was produced on Autologics APS- μ 5 typesetter.

† Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

‡ Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

Recently, Gould (1983) has given a complete characterization of the existence of such a solution appropriate for many of these procedures. In this paper, we shall be principally concerned with the construction of a vector p along which it is possible to decrease $Q(p)$ whilst maintaining $Ap = 0$ for the cases for which EQP has no finite solution.

We say that the vector p , satisfying $Ap = 0$, is a *direction of infinite feasible descent* if $Q(\alpha p) \rightarrow -\infty$ monotonically as the scalar $\alpha \rightarrow +\infty$: Our goal is to investigate the existence of directions of infinite descent for $Q(p)$ and to describe practical algorithms for locating such directions. The existence of a direction of infinite descent for a given EQP often indicates that the current estimate of the solution to the nonlinear program is far from its correct value or that the prediction of the constraints active at the solution (signified by their linearized form $Ap = 0$) is incorrect. Fortunately, a direction of infinite descent may then be used to improve the estimate of the solution (by decreasing an appropriate merit function) and to improve the prediction of the active set (by moving so that an inactive constraint becomes active). In this paper, we consider how methods normally used to find a finite solution to EQP can be adapted to find suitable directions of infinite descent whenever this is appropriate.

We will consider three methods, termed, null, range and Lagrangian methods. The null-space method is basically that of Bunch and Kaufman (1980).

The extensions to Lagrangian and range space methods are new. A discussion of the usefulness of these extensions is given in §5, below.

1.2. A preliminary discussion on the existence of a finite solution to EQP . In this section, we use the well-known characterization of the existence of a finite minimizer of a quadratic function (see, e.g. Gill, Murray and Wright (1981, pp. 65-67)) to obtain similar results for EQP .

Let Z be any $n \times (n-t)$ matrix such that $AZ = 0$, $\text{rank}(A^T:Z) = n$ (see, e.g., Gill and Murray (1974, pp. 57-62)). Then any vector p which satisfies $Ap = 0$ may be expressed as $p = Zp_Z$ for some $n-t$ vector p_Z . Consequently EQP is equivalent to the problem

$$NS: \quad \underset{p_Z \in \mathbb{R}^{n-t}}{\text{minimize}} \quad \frac{1}{2} p_Z^T (Z^T H Z) p_Z + p_Z^T (Z^T g).$$

It is well known that NS has

- (i) a unique global minimizer if and only if $Z^T H Z$ is positive definite;
- (ii) weak global minimizers if and only if $Z^T H Z$ is positive semi-definite and $Z^T g$ lies in the range of $Z^T H Z$;
- (iii) no finite minimizer if $Z^T H Z$ is positive semi-definite and $Z^T g$ does not lie in the range of $Z^T H Z$; and
- (iv) no finite minimizer if $Z^T H Z$ is indefinite.

In order to distinguish between cases (iii) and (iv) we make the following definitions.

Any vector p such that $Ap = 0$, $p^T H p = 0$ and $g^T p < 0$ is known as a *feasible direction of linear infinite descent*, henceforth referred to as a *dolid*, for EQP . Any vector p such that $Ap = 0$ and $p^T H p < 0$ is known as a quadratic feasible direction of infinite descent or a *direction of negative curvature*, henceforth referred to as a *donc*. (Note it is often convenient to introduce scaling in the definition and thus we will sometimes require $p^T H p = -1$ or $g^T p = -1$ rather than $p^T H p < 0$ or $g^T p < 0$.) The importance of such definitions is that for any dolid, p is feasible and $Q(\alpha p) = \alpha(g^T p)$ (and hence approaches minus infinity linearly as $\alpha \rightarrow \infty$) and for any donc, p is feasible and $Q(\alpha p) = \alpha^2 p^T H p + \alpha g^T p$ (which approaches minus

infinity quadratically as $\alpha \rightarrow \pm \infty$).

We are interested in the calculation of doncs and dolid. We have a preference (whenever possible) for determining doncs rather than dolid because of the faster asymptotic behaviour of the former.

LEMMA 1.1. (a) *There exists a direction of negative curvature if and only if $Z^T HZ$ is indefinite.* (b) *There exists a direction of linear infinite descent if $Z^T HZ$ is positive semi-definite and $Z^T g$ does not lie in the range of $Z^T HZ$.*

Proof. (a) follows from the definition. (b) Suppose $Z^T HZ$ is positive semi-definite, $Z^T g$ does not lie in the range of $Z^T HZ$ and that any vector p which satisfies $Ap = 0$, $p^T Hp = 0$ also satisfies $g^T p = 0$. In particular, all vectors $p = Zp_Z$ which satisfy $Z^T HZp_Z = 0$ satisfy $p_Z^T (Z^T g) = 0$ also. As $Z^T HZ$ is singular there is at least one p_Z which satisfies $Z^T HZp_Z = 0$. Then $Z^T g$ is orthogonal to the null-space of $Z^T HZ$ and therefore lies in its range. This contradicts the assumptions made and hence there is at least one dolid.

Lemma 1.1 motivates the following: conceptually we should like to obtain the unique global minimizer of EQP . However, if $Z^T HZ$ is indefinite, such a minimizer does not exist, and we then wish to locate directions of negative curvature. If $Z^T HZ$ is positive semi-definite, neither a unique global minimizer nor a direction of negative curvature exists. In this case we wish to determine a weak global minimizer, if possible, or otherwise determine a suitable dolid.

In the section which follows we will describe three approaches to finding stationary points for EQP , when such points exist. Each gives rise to particular matrix decompositions. Whenever minimizers do not exist, a significant feature of the approach in this paper is the method used to determine a suitable donc or dolid.

In the calculation of doncs and dolid, we shall explicitly exploit the matrices that arise in the particular underlying method for finding stationary points for EQP .

1.3. Methods for finding a stationary point for EQP . We briefly survey the existing procedures for finding a stationary point for EQP . Such methods normally assume that second order sufficiency conditions will be satisfied at the stationary point and hence that the stationary point found will be a strong global minimizer for the problem.

(a) *Null-space methods.* Find an appropriate $n \times (n-t)$, matrix Z such that $AZ = 0$ and $\text{rank}(A^T : Z) = n$. Then solve the *null-space equations*

$$\begin{aligned} Z^T HZp_Z &= -Z^T g, \\ p &= Zp_Z. \end{aligned}$$

(b) *Lagrangian or Kuhn-Tucker methods.* Find the vector p and the t -vector λ (a vector of first order estimates of the Lagrange multipliers at the solution) by solving the *Kuhn-Tucker equations*

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix}.$$

(c) *Decomposition or range-space methods.* These methods make use of the structure in the Lagrangian matrix,

$$K = \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}$$

to decompose the Kuhn-Tucker equations to obtain p and λ separately. For example, if H is nonsingular, p and λ may be found from the *range-space equations*

$$\begin{aligned} AH^{-1}A^T\lambda &= AH^{-1}g, \\ Hp &= A^T\lambda - g. \end{aligned}$$

The matrix equations defining each of the methods a), b) and c), are either solved directly, (by appropriate matrix factorizations) or iteratively. The decision as to which method to use and how to solve the resulting linear equations depends on the number of variables, n , and the number of constraints, t . In what follows, we shall assume that n is sufficiently small that we are able to solve the relevant equations directly. Normally it is most efficient to use null-space methods when t is large relative to n , and Lagrangian or range-space methods otherwise.

For a more detailed discussion of the linear equation solving techniques appropriate for these methods, and the reasons for choosing a particular method, see, for example Fletcher (1981), Gill, Murray and Wright (1981), Gill et al. (1982) and Gould (1983).

1.4. The nature of solutions to EQP. In § 1.2, the existence of finite solutions to *EQP* was discussed in terms of the definiteness of the matrix Z^THZ normally associated with null-space methods for *EQP*. In this section the question of existence of finite solutions to *EQP* is extended to the other methods introduced in § 1.3.

For the remainder of this paper we shall use the following notation. We define the $(n + t) \times (n + t)$ Kuhn-Tucker matrix K by

$$K = \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}.$$

We shall denote the inertia of any $m \times m$ matrix M by

$$In(M) = (m_+, m_-, m_0),$$

where m_+ , m_- and m_0 are respectively the number of positive, negative and zero eigenvalues of M (counted with appropriate multiplicities) such that

$$m = m_+ + m_- + m_0.$$

Define

$$In(H) = (h_+, h_-, h_0),$$

$$In(K) = (k_+, k_-, k_0)$$

and

$$In(Z^THZ) = (z_+, z_-, z_0)$$

for any appropriate Z . Furthermore if H is nonsingular (ie. $h_0 = 0$) define

$$In(AH^{-1}A^T) = (a_+, a_-, a_0).$$

The following lemma, a special case of Lemma 3.4 (Gould (1983)), is crucial to our discussion.

LEMMA 1.2. $In(K) = In(Z^THZ) + (t, t, 0)$. Furthermore if $h_0 = 0$

$$In(K) = In(H) + In(-AH^{-1}A^T).$$

We may thus derive the following results (see Gould (1983)).

THEOREM 1.3. *Let EQP be as given. Then the statements (i), (ii) and (iii), below, are equivalent.*

- (i) *EQP has a unique global minimizer,*
- (ii) *$z_- = z_0 = 0$ (or $z_+ = n - t$) and*
- (iii) *$k_- = t, k_0 = 0$ (or $k_+ = n$).*

Furthermore if $h_0 = 0$, (i), (ii), (iii) and

- (iv) *$h_- + a_+ = t$ (or $h_- = a_-$), $a_0 = 0$,*
- are equivalent.*

Proof. Follows directly from case (i) in § 1.2 and Lemma 1.2.

THEOREM 1.4. *Let EQP be as given. Then the statements (i), (ii) and (iii), below, are equivalent.*

- (i) *EQP has weak global minimizers,*
- (ii) *$z_- = 0, z_0 > 0$ and $Z^T g$ lies in the range of $Z^T HZ$ and*
- (iii) *$k_- = t, k_0 > 0$ and $\begin{pmatrix} -g \\ 0 \end{pmatrix}$ lies in the range of K .*

Furthermore, if $h_0 = 0$, (i), (ii), (iii) and

- (iv) *$h_- + a_+ = t, a_0 = 0$ and $AH^{-1}g$ lies in the range of $AH^{-1}A^T$,*
- are equivalent.*

Proof. Follows directly from case (ii) § 1.2 and Lemmas 1.1 and 1.2.

THEOREM 1.5. *Let EQP be as given. Then the statements (i) and (ii), below, are equivalent.*

- (i) *$z_- = 0, z_0 > 0$ and $Z^T g$ does not lie in the range of $Z^T HZ$ and*
- (ii) *$k_- = t, k_0 > 0$ and $\begin{pmatrix} -g \\ 0 \end{pmatrix}$ does not lie in the range of K .*

Furthermore, if $h_0 = 0$, (i), (ii) and

- (iii) *$h_- + a_+ = t, a_0 = 0$ and $AH^{-1}g$ does not lie in the range of $AH^{-1}A^T$,*
- are equivalent.*

The existence of a direction of linear infinite descent is implied by any of (i), (ii) and (iii).

Proof. Follows directly from cases (iii), § 1.2 and Lemmas 1.1 and 1.2.

THEOREM 1.6. *Let EQP be as given. Then the statements (i), (ii) and (iii), below, are equivalent.*

- (i) *There exists a direction of negative curvature for EQP,*
- (ii) *$z_- > 0$ and*
- (iii) *$k_- > t$.*

Furthermore, if $h_0 = 0$, (i), (ii), (iii) and

- (iv) *$h_- + a_+ > t$*
- are equivalent.*

Proof. The proof follows directly from case (iv) in § 1.2 and Lemmas 1.1 and 1.2.

The remainder of this paper will be taken up with a discussion of appropriate techniques for obtaining doncs and dolids.

2. The calculation of feasible directions of infinite descent.

2.1. Using the generalized Cholesky factorization. The numerical implementation of any method for the solution of EQP's depends essentially upon the technique used to solve the resulting system(s) of linear equations. All the systems in § 1.3 have symmetric coefficient matrices whose inertias are required to determine the existence of optimal solutions.

It is this latter requirement that predisposes us to consider an approach based

upon the Bunch-Parlett-Fletcher-Kaufman generalized Cholesky factorization (Bunch and Parlett (1971), Fletcher (1976), Bunch and Kaufman (1977)).

The basis of the generalised Cholesky factorization is a matrix formulation of the classical approach to diagonalizing a quadratic form by completing the square, with the additional observation that $xy = u^2 - v^2$ whenever $x = u + v$ and $y = u - v$. Essentially, a permutation matrix P is found such that a given real symmetric matrix G is factored as $P^TGP = MDM^T$ where M is unit lower triangular and D is block diagonal, with 1×1 and 2×2 diagonal blocks. The matrix D has the same inertia as G and this inertia is easily recoverable. For instance the number of negative eigenvalues of G is the number of 2×2 blocks plus the number of negative elements which occur in 1×1 blocks. Such a factorization may be achieved in about $n^3/6$ multiplications and n^2 comparisons.

The generalized Cholesky factorization has been used to calculate directions of negative curvature in the context of unconstrained optimization problems (see eg. Fletcher and Freeman (1977), Sorensen (1977), Moré and Sorensen (1979), Goldfarb (1980) and in a particular null-space method for *EQP* (Bunch and Kaufman (1980)).

Bunch (1971), Bunch and Kaufman (1977) and Fletcher (1976), have developed "stable" implementations in the sense that Bunch (1971) is as stable as Gaussian elimination with total pivoting and the other two methods are as stable as Gaussian elimination with partial pivoting.

2.2. Calculating dolidis. We have already remarked that, whenever possible, we prefer to calculate directions of negative curvature rather than dolidis. Consequently, as a dolid exists if and only if Z^THZ is indefinite, for this section we shall assume that Z^THZ is positive semi-definite.

The methods we propose to use to determine dolidis all depend fundamentally on the following well-known elementary result.

LEMMA 2.1. *Let N be any real symmetric, $m \times m$ matrix and b be any m -vector. Then either a) $\exists x : Nx = b$ or b) $\exists y : Ny = 0, b^Ty = 1$.*

Proof. Let P denote the orthogonal projector into the null-space of N . $Pb = 0 \Rightarrow b = Nx$. Otherwise define $y = Pb / \|Pb\|^2$.

Considering the separate methods of § 1.3 in turn, we then obtain as corollaries to Lemma 2.1, constructive means of obtaining dolidis whenever they exist.

(a) *Null-space methods.* Identifying Z^THZ with N and $-Z^Tg$ with b in Lemma 2.1, it is apparent that, either a) $\exists p_Z$ such that $Z^THZp_Z = -Z^Tg$ (in which case there is a weak global minimizer) or b) $\exists p_Z$ such that $Z^THZp_Z = 0$ and $p_Z^T Z^Tg = -1$ (in which case $p = Zp_Z$ is a dolid).

(b) *Lagrangian methods.* Identifying K with N , b with $-\begin{pmatrix} g \\ 0 \end{pmatrix}$ and writing $x = \begin{pmatrix} p \\ \lambda \end{pmatrix}$, we obtain that either a) $\exists x$ such that $Kx = b$ or b) $\exists x$ such that $Kx = 0, x^Tb = 1$. Case a) implies that p is a weak minimizer of *EQP*. b) implies that $Hp - A^T\lambda = 0$ (and thus $Z^THp = 0$) and $Ap = 0$ (from $Kx = 0$) and that $p^Tg = -1$ (from $x^Tb = 1$). Thus p is a dolid.

(c) *Range-space methods.* In this case we identify $AH^{-1}A^T$ with N and b with $AH^{-1}g$. Case a) implies p is a weak minimizer of *EQP*, case b) implies p is a dolid where $p = -H^{-1}A^T u$, and u is any vector such that $AH^{-1}A^T u = 0$ and $u^T AH^{-1}g = 1$.

Thus in each of the three cases considered above we have that, either a weak minimizer is determined, or a dolid is given by the orthogonal projection of the right-hand side into the null-space of the corresponding coefficient matrix.

Assuming we have a generalised Cholesky factorization of N , it is a straightforward task to determine the existence or otherwise of a solution to $Nx = b$, and thus establish optimality or find a dolid. We now give the necessary construction.

Suppose we have a generalized Cholesky factorization of N given by

$$P^TNP = MDM^T$$

where P is a permutation matrix. Let I denote the index set for the zero 1×1 blocks of D . Let $\bar{b} = P^Tb, \bar{x} = P^Tx$. Define r such that $Mr = \bar{b}$. The existence or otherwise of a solution to $Nx = b$ depends upon whether the equation $Ds = r$ has a solution. This equation has a solution if and only if $r_i = 0$ for all $i \in I$. If $r_i = 0$ for all $i \in I$, then the nonzeros of D determine $s_i, i \notin I$ and the $s_i, i \in I$ may be chosen arbitrarily. The equation $M^T\bar{x} = s$ may be solved to give \bar{x} and $x = P\bar{x}$ then satisfies $Nx = b$. Conversely suppose $r_i \neq 0$ for some $i \in I$. Let s be any nonzero solution of $Ds = 0$. Then $s_i = 0$ for all $i \notin I$ necessarily. Let \bar{x} be such that $M^T\bar{x} = s$ and $x = P\bar{x}$. Then, using $P^TNP = MDM^T, x = P\bar{x}$ and $Ds = 0$ we have that $Nx = PMDs = 0$. Similarly, using $Mr = \bar{b}$ and $b = P\bar{b}$, we have that $b^Tx = r^Ts$. Now, suppose we chose our s_i such that $s_i = 0, i \notin I$ and $s_i = \alpha r_i, i \in I$ for some $\alpha \neq 0$, then $r^Ts \neq 0$. The latter result follows from

$$r^Ts = \sum_{i \in I} r_i s_i + \sum_{i \notin I} r_i s_i = \alpha \sum_{i \in I} r_i^2,$$

which is nonzero, since at least one $r_i \neq 0$. Finally, upon letting $\alpha = 1 / \sum_{i \in I} r_i^2$, and using $b^Tx = r^Ts$, we have that $b^Tx = 1$.

2.3. Locating directions of negative curvature. In this section we show how directions of negative curvature may be calculated for null-space, range-space and Lagrangian methods.

DEFINITION. We say that x is a *negative vector* for N if and only if $x^TNx < 0$. We say that x is a *positive vector* for N if and only if $x^TNx > 0$. We say that vectors u and v are *N -conjugate* if $u^TNv = 0$.

(a) *Null-space methods.* Suppose $z_- > 0$. Then there is a vector v_Z such that $v_Z^TZ^THZv_Z < 0$. On letting $v = Zv_Z, v^THv < 0$ and $Av = AZv_Z = 0$. Hence v is a donc.

In general, suppose that $z_- = s > 0$. Then there are s vectors v_{Z_1}, \dots, v_{Z_s} such that

$$\begin{aligned} v_{Z_i}^T(Z^THZ)v_{Z_i} &< 0, \quad 1 \leq i \leq s, \\ v_{Z_i}^T(Z^THZ)v_{Z_j} &= 0, \quad 1 \leq i \neq j \leq s. \end{aligned}$$

For instance, the eigenvectors of Z^THZ corresponding to the negative eigenvalues are such a set. (The two conditions above imply that on the span of the set of vectors $v_{Z_i}, i = 1, \dots, s$ the matrix Z^THZ is negative definite, and on this subspace the v_{Z_i} satisfy the usual definition of conjugacy.) Then the vector

$$(2.1) \quad v = \sum_{i=1}^s \alpha_i Z v_{Z_i}$$

is a donc. We note that different choices of α_i result in different doncs. Little is known concerning the best choices for the α_i 's (see Bunch and Kaufman (1980) for a discussion of this issue). Ideally, one would like to chose the α_i 's so that the cost, measured in terms of number of iterations and work per iteration, is minimized.

(b) *Lagrangian methods* . Let

$$K = \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}.$$

Let $k = k_- > t$. Then there are k K -conjugate negative vectors v_1, \dots, v_k (for instance v_1, \dots, v_k might be the eigenvectors of K corresponding to the negative eigenvalues). Then $v = \sum_{i=1}^k \alpha_i v_i$ is a negative vector for K for any scalars $\alpha_1, \dots, \alpha_k$ not all zero. Choose the scalars $\alpha_1, \dots, \alpha_k$ so that

$$(2.2) \quad (A : 0)v = \sum_{i=1}^k \alpha_i (A : 0)v_i = 0.$$

(Since $(A : 0)v_i$ is a t -vector and $k > t$, it is always possible to find such a combination.) Then if $v = \begin{pmatrix} p \\ q \end{pmatrix}$, p is a direction of negative curvature.

(c) *Range-space methods* . We shall only be concerned with the case that H is nonsingular. For other cases, range-space methods are difficult to implement and Lagrangian methods may be preferred. It is easy to see that

$$(2.3) \quad K = \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} = \begin{pmatrix} I & 0 \\ AH^{-1} & I \end{pmatrix} \begin{pmatrix} H & 0 \\ 0 & -AH^{-1}A^T \end{pmatrix} \begin{pmatrix} I & H^{-1}A^T \\ 0 & I \end{pmatrix}.$$

Let $h_- = k$, $a_+ = l$ and $k + l > t$, (thus ensuring the existence of a donc, from Theorem. 1.6). Let h_1, \dots, h_k be a set of H -conjugate negative vectors, a_1, \dots, a_l be a set of $AH^{-1}A^T$ -conjugate positive vectors. Then from (2.3), it follows that

$$\begin{pmatrix} h_i \\ 0 \end{pmatrix} \quad (1 \leq i \leq k) \quad \text{and} \quad \begin{pmatrix} -H^{-1}A^T a_i \\ a_i \end{pmatrix} \quad (1 \leq i \leq l)$$

are K -conjugate negative vectors. Hence

$$u = \sum_{i=1}^k \alpha_i \begin{pmatrix} h_i \\ 0 \end{pmatrix} + \sum_{i=1}^l \beta_i \begin{pmatrix} -H^{-1}A^T a_i \\ a_i \end{pmatrix}$$

is a negative vector for any scalars α_i, β_i (not all zero). Furthermore, since $k + l > t$, it is always possible to choose the scalars α_i, β_i so that

$$(A \ 0)u = 0.$$

Then, writing $u = \begin{pmatrix} p \\ q \end{pmatrix}$, p is a direction of negative curvature for EQP .

3. Calculating directions of negative curvature. In this section we present methods for calculating doncs for the three classes of algorithm discussed in § 1.3. All three methods require the calculation of negative and/or positive vectors of appropriate coefficient matrices as indicated in § 2.3. Such vectors may be obtained directly from the generalised Cholesky factorization of the given matrices. However, modifications of the direct procedure results in simplifications and significant savings.

3.1. Calculating negative and positive vectors. Let the real symmetric $m \times m$ matrix, N have a generalized Cholesky factorization

$$N = PMDM^T P^T$$

where P is a permutation matrix, M is a unit lower triangular matrix and D is block diagonal with 1×1 and 2×2 diagonal blocks.

Define the ordered index sets

$$\begin{aligned}
 I_{-1}(D) &= \{i \mid d_{ii} < 0 \text{ and } d_{ii+1} = d_{ii-1} = 0\}, \\
 I_{+1}(D) &= \{i \mid d_{ii} > 0 \text{ and } d_{ii+1} = d_{ii-1} = 0\}, \\
 I_{-2}(D) &= \{i \mid d_{ii+1} \neq 0\}, \\
 I_{+2}(D) &= \{i \mid d_{i-1i} \neq 0\},
 \end{aligned}$$

where D has elements d_{ij} and, by convention, $d_{i0} = d_{im+1} = 0$. These sets identify the indices of 1×1 blocks with negative and positive eigenvalues and the first and second index of each 2×2 block (which by construction has one positive and one negative eigenvalue) respectively. Let $I_{-}(D) = I_{-1}(D) \cup I_{-2}(D)$ and $I_{+}(D) = I_{+1}(D) \cup I_{+2}(D)$. Then $|I_{-}(D)| = n_{-}$ and $|I_{+}(D)| = n_{+}$ where $In(N) = (n_{+}, n_{-}, n_0)$. In order to construct an N -conjugate set of n_{-} negative and n_{+} positive vectors, we proceed as follows. Let e_i be the i th column of the identity matrix.

Define the vectors $u_i (i \in I_{-}(D))$ and $u_i (i \in I_{+}(D))$ such that

$$\begin{aligned}
 u_i &= \begin{cases} \gamma_i e_i, & i \in I_{-1}(D), \\ \gamma_i (e_i + \beta_i e_{i+1}), & i \in I_{-2}(D), \end{cases} \\
 u_i &= \begin{cases} \eta_i e_i, & i \in I_{+1}(D), \\ \eta_i (e_{i-1} + \theta_{i-1} e_i), & i \in I_{+2}(D), \end{cases}
 \end{aligned}$$

where $\beta_i = (\lambda_{-} - d_{ii})/d_{ii+1}$, $\theta_i = (\lambda_{+} - d_{ii})/d_{ii+1}$, λ_{-} and λ_{+} are the negative and positive eigenvalues of

$$\begin{pmatrix} d_{ii} & d_{ii+1} \\ d_{ii+1} & d_{i+i+1} \end{pmatrix}$$

and the scalars γ_i and η_i are arbitrary. Then it is easy to see that the vectors $v_i (i \in I_{-}(D))$, obtained from

$$(3.1) \quad M^T P^T v_i = u_i,$$

are N -conjugate negative vectors. Similarly the vectors $v_i, i \in I_{+}(D)$, where $M^T P^T v_i = u_i$, are N -conjugate positive vectors. Hence any vector

$$v_{-} = \sum_{i \in I_{-}(D)} \alpha_i v_i$$

(with the scalars α_i not all zero) is a negative vector and any vector

$$v_{+} = \sum_{i \in I_{+}(D)} \delta_i v_i$$

(for scalars δ_i not all zero) is a positive vector.

3.2. Practical aspects. Null-space methods. Let $Z^T H Z$ have a generalised Cholesky factorization $P_Z M_Z D_Z M_Z^T P_Z^T$. From (2.1), the vector

$$v = Z \left(\sum_{i \in I_{-}(D_Z)} \alpha_i v_{Zi} \right)$$

is a direction of negative curvature for any (nonzero) set of scalars α_i and any set of

$Z^T HZ$ -conjugate negative vectors v_{Z_i} . Suitable negative vectors may be obtained as described above.

However as

$$v = \sum_{i \in I_-(D_Z)} \alpha_i v_{Z_i} = \sum_{i \in I_-(D_Z)} \alpha_i P M^{-T} u_{Z_i} = P M^{-T} \sum_{i \in I_-(D_Z)} \alpha_i u_{Z_i},$$

it is more efficient to obtain

$$\sum_{i \in I_-(D_Z)} \alpha_i u_{Z_i}$$

and then perform *one* backsolve to find v . A particularly appealing direction of negative curvature is obtained by picking $\alpha_i = 0$ for all $i (\neq j) \in I_-(D_Z)$ and $\alpha_j = 1$ where j is the smallest index in $I_-(D_Z)$. With this choice the “backsolve” $M^{-T} u_{Z_j}$ may be performed relatively efficiently since the last $n - t - j$ components of $M^{-T} u_{Z_j}$ are then zero.

3.3. Practical aspects. Lagrangian methods. Let K have a generalised Cholesky factorization $P_K M_K D_K M_K^T P_K^T$. Let $v_i (i \in I_-(D_K))$ be K -conjugate negative vectors obtained from the vectors u_i by (3.1) as described in § 3.1. According to (2.2), we require scalars α_i such that

$$\sum_{i \in I_-(D_K)} \alpha_i Q(A : 0) v_i = 0,$$

where the permutation matrix Q is introduced for convenience. As $K = P_K M_K D_K M_K^T P_K^T$, we may write $(A : 0) = (0 : I_t) P_K M_K D_K M_K^T P_K^T$, where I_t is the $t \times t$ identity matrix. Therefore we must find scalars α_i such that

$$(3.2) \quad \sum_{i \in I_-(D_K)} \alpha_i \bar{M} D_K u_i = 0,$$

where $\bar{M} = Q(0 : I_t) P_K M_K$. The matrix $(0 : I_t) P_K M_K$ is made up of rows of the unit lower triangular matrix M_K and Q is now chosen so that the rows of \bar{M} occur in the same order as in M_K (see Fig. 1).

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ x & 1 & 0 & 0 & 0 & 0 & 0 \\ u & u & 1 & 0 & 0 & 0 & 0 \\ x & x & x & 1 & 0 & 0 & 0 \\ u & u & u & u & 1 & 0 & 0 \\ u & u & u & u & u & 1 & 0 \\ x & x & x & x & x & x & 1 \end{pmatrix} \quad \begin{pmatrix} u & u & 1 & 0 & 0 & 0 & 0 \\ u & u & u & u & 1 & 0 & 0 \\ u & u & u & u & u & 1 & 0 \end{pmatrix}$$

M_K
 \bar{M}

FIG. 1. An example of how \bar{M} is obtained from M_K by deleting those rows indicated with entries x . {Thus $t = 3$ and P_K points to rows 3, 5 and 6.}

Now $D_K u_i = \gamma_i d_{ii} e_i$ if $i \in I_{-1}(D_K)$ and $D_K u_i = \lambda_- \gamma_i (e_i + \beta_i e_{i+1})$ if $i \in I_{-2}(D_K)$. Suppose $\gamma_i = 1/d_{ii}$ if $i \in I_{-1}(D_K)$ and $\gamma_i = 1/\lambda_- \sqrt{1 + \beta_i^2}$ if $i \in I_{-2}(D_K)$. Hence $D_K u_i = e_i$ if $i \in I_{-1}(D_K)$ and $D_K u_i = \Phi_i e_i + \psi_i e_{i+1}$, if $i \in I_{-2}(D_K)$, where $\Phi_i = 1/\sqrt{1 + \beta_i^2}$, and $\psi_i = \beta_i/\sqrt{1 + \beta_i^2}$. These choices of γ_i , ϕ_i and ψ_i are made so that $D_K u_i$ is a unit vector in all cases.

Using this construction, (3.2) gives

$$(3.3) \quad \sum_{i \in I_-(D_K)} \alpha_i \bar{M} D_K u_i = \sum_{i \in I_{-1}(D_K)} \alpha_i \bar{M} e_i + \sum_{i \in I_{-2}(D_K)} \alpha_i \bar{M} (\Phi e_i + \psi e_{i+1}) = 0.$$

As $\bar{M} e_i$ is the i th column of \bar{M} and $M(\Phi e_i + \psi e_{i+1})$ is a linear combination of the i th and $(i + 1)$ st columns of \bar{M} , (3.3) is equivalent to finding a suitable linear combination of columns of \bar{M} which is zero.

Let N be the $t \times k$ matrix whose columns are $\bar{M} e_i$ if $i \in I_{-1}(D_K)$ and $\bar{M}(\Phi e_i + \psi e_{i+1})$ if $i \in I_{-2}(D_K)$ and let these columns be arranged in increasing order according to the set $I_-(D_K)$ (see Fig. 2).

$$\begin{array}{ccc} \begin{pmatrix} u & u & 1 & 0 & 0 & 0 & 0 \\ u & u & u & u & 1 & 0 & 0 \\ u & u & u & u & u & 1 & 0 \end{pmatrix} & & \begin{pmatrix} v & 0 & 0 & 0 \\ v & u & 1 & 0 \\ v & u & u & 0 \end{pmatrix} \\ \bar{M} & & N \end{array}$$

FIG. 2. An example of how N is obtained from \bar{M} . In this case $I_{-1}(D_K) = \{4,5,7\}$ and $I_{-2}(D_K) = \{1\}$. The entries v in column 1 of N are obtained from an appropriate linear combination of columns 1 and 2 of \bar{M} .

Let α be the vector make up of the unknown scalars α_i ordered according to $I_-(D_K)$. Then (3.3) is the same as finding a nontrivial solution to

$$(3.4) \quad N\alpha = 0.$$

There may be many independent solutions to (3.4). Our intention is to obtain such a solution in an efficient manner. By the construction of N from M_K , any column of N which has a structural zero as its j th entry will also have structural zeros in its i th entry for all $i \leq j$. (A structural zero of N is an element of N which is obtained as a linear combination of elements of the upper-triangular (zero) part of M_K .) Furthermore the introduction of the permutation matrix Q in the definition of N ensures that any row which has a structural zero in its j th entry will also have a structural zero in its i th entry for all $i \geq j$. Hence the matrix N may be thought of as being “lower triangular-like”. In practice the matrix N may be anything from completely dense to the zero matrix. Our hope is that, for a given problem, N is closer to the latter than the former (see § 4).

We propose to find a nontrivial solution to (3.4) by (lower) triangulating the last $t \times t$ block of N using elementary row operations to eliminate super-diagonal elements and row interchanges to limit growth of the off-diagonal elements. No structural fill-in occurs as a result of the partial pivoting because of the structure of N . Recall that N is a $t \times k_-$ ($k_- > t$) matrix. The process starts in the k_- th column of N and proceeds one column backwards through N until either there are no nonzero pivots in the current column or t columns have been triangulated. If the former case is reached at column $k_- - k$, the last $k + 1$ columns of N are linearly dependent. Furthermore the last $k + 1$ columns of N have then been reduced to the form

$$\begin{pmatrix} 0 & 0 \\ \bar{n} & \bar{N} \end{pmatrix}$$

where \bar{N} is a $k \times k$ lower triangular matrix and \bar{n} a k -vector.

A suitable solution to (3.4) is then found as $\alpha = \begin{pmatrix} 0 \\ \alpha' \end{pmatrix}$ where α' is a $k + 1$ vector whose first entry is 1 and whose remaining entries $\bar{\alpha}$ are found by forward substitution in the system $\bar{N}\bar{\alpha} = -\bar{r}$. In the latter case (3.4) has been reduced to the equivalent system

$$(N_1 : \bar{N})\alpha = 0$$

where \bar{N} is $t \times t$ triangular. If we form any nonzero linear combination $N_1\alpha_1$, a suitable solution α to (3.4) is given by $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$ where α_2 is found by forward substitution in the system $\bar{N}\alpha_2 = -N_1\alpha_1$.

Obviously the efficiency of this method depends upon how triangular-like N is and how many columns of N must be processed before a suitable dependency may be found. In the worst case, when N is completely dense and no dependencies are found until the final stage, $t^3/3 + O(t^2)$ operations will be carried out to find α . This cost may be acceptable if t is relatively small. This is an assumption which is normally made whenever Lagrangian methods are to be used. In practice we should hope that the cost will be substantially smaller than $t^3/3$. (See § 4 for numerical results.)

Once the scalars α_i have been obtained, the calculation of a direction of negative curvature proceeds exactly as described in § 2.3. Namely the vector $\sum_{i \in I_-(D_k)} \alpha_i u_i$ is formed and a single back-solve is used to obtain the direction of negative curvature as the first n components of the vector v where

$$M_k^T P_k^T v = \sum_{i \in I_-(D_k)} \alpha_i u_i.$$

3.4. Practical aspects. Range-space methods. Recall, that we assume $h_0 = 0$.

Suppose we have a generalised Cholesky factorization $P_1 M_1 D_1 M_1^T P_1^T$ for H and $P_2 M_2 D_2 M_2^T P_2^T$ for $AH^{-1}A^T$. Thus we may determine vectors $h_i (i \in I_-(D_1))$ and $a_i (i \in I_+(D_2) \equiv I_-(D_2))$ such that the h_i 's are a set of h_- negative vectors for H and the a_i 's are a set of a_+ positive vectors for $AH^{-1}A^T$, where $a_i^T A H^{-1} A^T a_j = 0, i \neq j$. Following § 2.3c, the vector $\begin{pmatrix} h_i \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -H^{-1}A^T a_i \\ a_i \end{pmatrix}$ are $\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}$ conjugate vectors. However, we need to satisfy $(A \ 0)u = 0$ where

$$u = \sum_{i \in I_-(D_1)} \alpha_i \begin{pmatrix} h_i \\ 0 \end{pmatrix} + \sum_{i \in I_-(D_2)} \beta_i \begin{pmatrix} -H^{-1}A^T a_i \\ a_i \end{pmatrix}$$

and the α_i 's and β_i 's are suitably chosen.

As before, it is computationally more attractive to work with the negative vectors of the block diagonal matrices D_1 and D_2 rather than the a_i 's and h_i 's. Thus we consider the following development:

It is not difficult to see that one may write, using $H = P_1 M_1 D_1 M_1^T P_1^T$ and $A = P_2 N D_1 M_1^T P_1$,

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} \begin{pmatrix} M_1 & 0 \\ N & M_2 \end{pmatrix} \begin{pmatrix} D_1 & 0 \\ 0 & -D_2 \end{pmatrix} \begin{pmatrix} M_1^T & N^T \\ 0 & M_2^T \end{pmatrix} \begin{pmatrix} P_1^T & 0 \\ 0 & P_2^T \end{pmatrix},$$

where $N = P_2^T A P_1^T M_1^{-T} D_1^{-1}$ (Recall that $h_0 = 0 \Rightarrow D_1^{-1}$ exists.)

In addition, we note that

$$\begin{pmatrix} M_1^T & N^T \\ 0 & M_2 \end{pmatrix} \begin{pmatrix} P_1^T & 0 \\ 0 & P_2^T \end{pmatrix} \begin{pmatrix} h_i \\ 0 \end{pmatrix} = \begin{pmatrix} M_1^T P_1^T h_i \\ 0 \end{pmatrix}$$

and

$$\begin{pmatrix} M_1^T & N^T \\ 0 & M_2 \end{pmatrix} \begin{pmatrix} P_1^T & 0 \\ 0 & P_2^T \end{pmatrix} \begin{pmatrix} -H^{-1}A^T a_i \\ a_i \end{pmatrix} = \begin{pmatrix} -M_1^T P_1^T H^{-1}A^T a_i + N^T P_2^T a_i \\ M_2 P_2^T a_i \end{pmatrix} = \begin{pmatrix} 0 \\ M_2 P_2^T a_i \end{pmatrix}.$$

Thus, writing $M_1^T P_1^T h_i = u_i^{(1)}$, $M_2^T P_2^T a_i = u_i^{(2)}$, and

$$\bar{M} = (0 : I_t) \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} \begin{pmatrix} M_1 & 0 \\ N & M_2 \end{pmatrix} (P_2 N \ P_2 M_2)$$

(3.4) becomes

$$(3.5) \quad \sum_{i \in I_-(D_1)} \alpha_i P_2 N D_1 u_i^{(1)} - \sum_{i \in I_-(D_2)} \beta_i P_2 M_2 D_2 u_i^{(2)} = 0.$$

Let the matrix \tilde{N} be made up of columns $ND_1 u_i^{(1)}$ ($i \in I_-(D_1)$). By construction these columns are either single columns or linear combinations of two columns of N . Similarly let T be made up of the columns $M_2 D_2 u_i^{(2)}$ ($i \in I_+(D_2)$). As $M_2 D_2 u_i^{(1)}$ is either a single column of M_2 or a linear combination of two adjacent columns, T preserves the “triangular-like” structure of M_2 . Moreover, given M_2 , T is trivial to obtain. Defining the vectors α and β to have elements $\alpha_i (i \in I_-(D_1))$ and $\beta_i (i \in I_+(D_2))$, (3.5) becomes

$$(3.6) \quad (\tilde{N} : T) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 0.$$

Each column of N , and hence \tilde{N} , must be calculated from the definition $N = P_2^T A P_1^T M_1^{-T} D_1^{-1}$. As this could prove expensive if many columns of \tilde{N} are required, we try to obtain a solution to (3.6) which has as few nonzero components of α as possible. As a dependence amongst the columns of $(\tilde{N} : T)$ may require $t + 1$ columns, we look for such a dependence in the a_+ columns of T and any $t - a_+ + 1$ columns at \tilde{N} . Clearly it is not possible to make statements about a_+ being close to t , but since we are using a range-space method we do at least expect t to be reasonably small. Certain columns of \tilde{N} are easier to obtain than others. Observing that $u_i^{(1)}$ is either $\gamma_i^{(1)} e_i$ or $\gamma_i^{(1)}(e_i + \beta_i^{(1)} e_{i+1})$ for appropriate $\gamma_i^{(1)}$ and $\beta_i^{(1)}$, and that $ND_1 u_i^{(1)} = P_2^T A P_1^T M_1^{-T} u_i^{(1)}$, it is easy to show that $DN_1 u_i^{(1)}$ may be calculated in about $i^2/2 + i \times i$ multiplications. (This may be a significant overestimate if A or M_1 are sparse.) Therefore it is advisable to select those columns of \tilde{N} corresponding to small indices i in preference to larger indices.

Let \bar{N} be made up of any $t - a_+ + 1$ columns of \tilde{N} and let $\bar{\alpha}$ be the corresponding elements of α .

Then (3.6) gives

$$(3.7) \quad (\bar{N} : T) \begin{pmatrix} \bar{\alpha} \\ \beta \end{pmatrix} = 0.$$

Suppose $\bar{N} = (b : B)$. A solution of (3.7) may be found from the solution $\begin{pmatrix} x \\ y \end{pmatrix}$ to the equations $(B : T) \begin{pmatrix} x \\ y \end{pmatrix} = b$ by letting $\bar{\alpha}_1 = -1$, $\bar{\alpha}_{i+1} = x_i$ and $\beta_i = y_i$. This later equation is triangular-like and may be reduced to lower triangular form in a similar

fashion to that described for Lagrangian methods. In this case, however T is of rank a_+ and so unlike the Lagrangian case there is little chance of obtaining "easy" solutions.

3.5. Operation counts. Table 1 provides a summary of the operation counts required to calculate dolids and doncs. In addition, we provide a breakdown of the costs of the necessary factorizations and their totals. Our calculations for doncs are in the worst case. Moreover, it is assumed that factorizations are done from scratch given A and H .

TABLE 1

	Null space	Lagrangian	Range space
factorizations	Z (eg. orthogonal columns) $Z^T H Z$	$\begin{pmatrix} H A^T \\ A \ 0 \end{pmatrix}$	H $AH^{-1}A^T$
calculate	$nt^2 - \frac{1}{3}t^3$: find Z		$\frac{n^3}{6}$: factor H
factorizations	$n^2(n-t) + \frac{n}{2}(n-t)^2$: find $Z^T H Z$ $\frac{(n-t)^3}{6}$: factor $Z^T H Z$	$\frac{1}{6}(n+t)^3$	$nt(n+t/2)$: find $AH^{-1}A^T$ $\frac{t^3}{6}$: factor $AH^{-1}A^T$
total	$\frac{1}{3}(n+t)^2 + \frac{1}{6}(2n-t)^2 - \frac{3n^2t}{2} - \frac{2t^3}{3}$	$\frac{1}{6}(n+t)^3$	$\frac{1}{6}(n+t)^3 + \frac{n^2t}{2}$
calculate DOLID	$(2n-t)^2 - n^2$	$(n+t)^2$	$(n+t)^2 + n^2$
calculate DONC	$O(n-t)^2$	$\frac{t^3}{3} + O(n^2)$	$\frac{t}{12}(3n^2 + 9nt - t^2) + O(n^2)$

4. Numerical experience. There are two unresolved issues connected with the methods for calculating directions of negative curvature suggested in § 3. Firstly, it has been noted that although the matrix \bar{N} , associated with the Lagrangian method, is "triangular-like", it is not clear just how triangular it will be. It is obviously desirable that this matrix be small and as triangular as possible. However, examples may be constructed where it is $t \times t$ (in the worst case) and totally dense. Our hope is that in practice \bar{N} is much more likely to be almost triangular. Secondly, the efficiency of the procedure suggested for range-space methods depends crucially upon the number of columns $t - a_+ + 1$ of \bar{N} needed to determine a dependency among the columns of $(\bar{N} : T)$. We should like $t - a_+ + 1$ to be as small as possible.

In this section, numerical results are presented which indicate experience with both of these issues. The numerical experiments are by no means exhaustive; it is merely intended that they illustrate that the approaches discussed in § 3 are viable.

Directions of negative curvature were obtained for EQP's for which the matrices H and A were generated (pseudo-) randomly. The number of variables, n , for each problem solved was fixed at 30, while the number of constraints, t , was allowed to vary from 1 to 29. As Lagrangian and range-space methods are normally only

TABLE 2

Numerical experiments with Lagrangian and range-space methods.

t	Lagrangian		$t - a_+ + 1$	Range-space	
	Number of columns of N processed	Proportion nonzero super diagonals in \bar{N}		Approximate no. mults. to calculate \bar{N}	Proportion nonzero super-diags. in (B:T)
1	1	0/0	1	2	0/0
2	1	0/0	1	3	0/1
3	1	0/0	1	4	0/3
4	1	0/0	2	47	2/6
5	1	0/0	2	54	2/10
6	1	0/0	2	61	2/15
7	1	0/0	2	68	2/21
8	1	0/0	2	75	2/28
9	1	0/0	3	241	8/36
10	4	0/3	4	547	17/45
11	3	0/1	4	581	16/55
12	4	0/3	4	615	18/66
13	5	0/6	4	649	18/78
14	6	0/10	4	683	19/91
15	7	0/15	4	717	19/105
16	8	0/21	4	751	19/120
17	10	0/36	4	785	18/136
18	16	9/105	5	1417	35/153
19	18	10/136	5	1472	33/171
20	19	10/153	5	1527	33/190
21	21	28/190	6	2466	53/210
22	22	29/210	6	2547	54/231
23	23	22/231	6	2628	53/253
24	24	27/253	6	2709	54/276
25	25	27/276	6	2790	54/300
26	26	20/300	6	2871	54/325
27	27	24/325	6	2952	55/351
28	NO DONC	NO DONC	NO DONC	NO DONC	NO DONC
29	NO DONC	NO DONC	NO DONC	NO DONC	NO DONC

appropriate when t is relatively small (compared to n), it would be unlikely that they would be used when H has many negative eigenvalues. The matrix H was chosen to have 6 negative eigenvalues and for convenience was diagonal. The matrix A in every case was full. The results obtained are indicated in Table 2. In each case a direction of negative curvature was generated whenever possible; the residuals of the constraints Ap for normalized directions of negative curvature tended to have magnitudes of about 10^{-16} (using double precision on the Honeywell 6066 at the University of Waterloo).

For the Lagrangian method, it is seen that for problems for which $1 \leq t \leq 17$,

the matrix \bar{N} is already triangular and therefore no effort need be expended to triangularize it. For $t \geq 18$, a small amount of elimination must be performed to triangularize \bar{N} . Typically between 3 and 15% of the super diagonal elements needed to be eliminated. In no case did a column have more than 2 super diagonal elements. Thus, for our test problems, the procedure described for the Lagrangian methods proved extremely efficient.

In the case of the range-space method, it is seen that the number of columns of \tilde{N} which must be computed increases gradually as t increases. We note the effect that an increase in the number of columns of \tilde{N} has on both the effort to compute \bar{N} and the number of nonzero super diagonals of $(B : T)$. Although the extra amount of work required to find a donc for range-space rather than Lagrangian methods is relatively small for our test problems, it is significantly more work than required for the Lagrangian methods themselves. This suggests that it may be preferable to use Lagrangian methods rather than range-space methods for nonconvex problems for which h_- is at all large.

5. Comments and conclusions. In this paper, different strategies for calculating directions of infinite descent and directions of negative curvature for the problem EQP have been presented. These methods differ according to whether a null-space, Lagrangian or range-space method is being used to solve the EQP .

Problems of the form EQP arise in a variety of ways; two common occurrences are as subproblems in active set methods for quadratic programming and as subproblems in successive quadratic programming methods for nonlinear programming. These two applications are slightly different, since for the first H will remain constant whereas for the second H may change (significantly the inertia of H may change). As a rule, null-space methods should be used if the number of constraints, t , represented in A is large relative to n and Lagrangian or range-space methods when t is small. In general, this is not known before the sequence of subproblems is solved. However if t is less than the number of negative eigenvalues of H , it is straightforward to show (see, e.g. Gould (1983, Theorem 2.1)) that there exists a direction of negative curvature for EQP and hence EQP does not have a finite solution. Therefore, if H has many negative eigenvalues there must be many constraints active for a finite solution to exist. For quadratic programming applications, a null-space method should always be used when H has many negative eigenvalues. When H has few negative eigenvalues the converse does not apply (namely it is not clear which is the best method to use). For successive Q.P. applications, the changing inertia of H will make it more difficult to select the appropriate method a priori; some upper bound on t may be known and this may enable a sensible choice to be made.

It is our opinion that, at least from the point of calculating directions of negative curvature, Lagrangian methods are likely be superior to range-space methods, since, for the former, the only extra overhead incurred while calculating a donc rather than moving to a minimizer of the EQP is just that in solving (3.4). Although the solution may require as many as $1/3t^3$ multiplications (using Gaussian elimination with interchanges), our limited experience is that N is already essentially triangular and a solution may be obtained almost trivially. On the other hand, a direction of negative curvature for range-space methods may require the calculation of a significant number of columns in \bar{N} , each of which is relatively expensive. Furthermore, we require the solution of

$$(B : T) \begin{pmatrix} x \\ y \end{pmatrix} = b$$

using Gaussian elimination. Here $(B : T)$ can be significantly nontriangular and the elimination cost could be substantial. A possible remedy for the first drawback is to calculate N during the factorizations of H and $AH^{-1}A^T$. However this can be shown to be equivalent to finding the generalized Cholesky factorization of K while insisting that row and column interchanges are only permitted in the leading $n \times n$ and remaining $t \times t$ block of K . This is then just a variation on the theme of Lagrangian methods (exemplifying the close connection between the two approaches).

More generally, one might wish to consider the problem

$$\begin{aligned} EQP': \quad & \underset{p \in \mathbb{R}^n}{\text{minimize}} \quad Q(p) \\ & \text{subject to} \quad Ap = d. \end{aligned}$$

This problem is easily solved by expressing $p = p_1 + p_2$, where p_1 is chosen such that $Ap_1 = d$ (for instance by solving $AA^T p_a = d$ and letting $p_1 = A^T p_a$, and then solving

$$\begin{aligned} & \underset{p_2 \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2} p_2^T H p_2 + p_2^T (H p_1 + c) \\ & \text{subject to} \quad A p_2 = 0. \end{aligned}$$

For problems which are either structured or sparse, it may be worthwhile compromising stability of the appropriate factorizations, in order to maintain, to some extent, any available structure. We have in mind a variant of the generalized Cholesky algorithm for which the choice of pivotal elements is made with respect to the fill-in which may result. Of course, there should be some overriding stability restrictions (e.g. threshold pivoting). For range-space methods for quadratic programming, it is probably worth spending extra effort in obtaining a factorization of H which maintains some of the structure in H . For Lagrangian methods, it is possible to insist that row and column interchanges are made so that the first $n \times n$ blocks of the generalized Cholesky factors of K are the factors of H . This is important since, for the range-space methods just discussed, it is possible that H can be factored, so as to maintain any structure present. Furthermore, if H does not change from one subproblem to the next, the factorization of K will only change in its last t rows and its last t columns. It is to be expected that simple changes to the matrix A (as may result from quadratic programming applications) will result in simple changes to the appropriate factorizations. For a discussion of the changes which become necessary, see Sorensen (1977) and Bunch and Kaufman (1980).

Acknowledgments. We thank Helen Warren and Yvonne Fink for their excellent typing and ability to decipher our often cryptic handwriting.

REFERENCES

- M. C. BIGGS (1975), *Constrained minimization using recursive quadratic programming: some alternative subproblem formulations*, in *Towards Global Optimization*, L. C. W. Dixon and G. P. Szego, Eds., North-Holland, (Amsterdam).
- J. R. BUNCH (1971), *Analysis of the diagonal pivoting method* this Journal, 8, pp. 656-680.

- J. R. BUNCH AND L. C. KAUFMAN (1977), *Some stable methods for calculating inertia and solving symmetric linear systems*, Math. Comp. 31, pp. 163-179.
- J. R. BUNCH AND L. C. KAUFMAN (1980), *A computational method for the indefinite quadratic programming problem*, Linear Alg. and Applics. 34, pp. 341-370.
- J. R. BUNCH AND B. N. PARLETT (1971) *Direct methods for solving symmetric indefinite systems of linear equations*, this Journal, 8, pp. 639-655.
- R. M. CHAMBERLAIN, C. LEMARECHAL, H. C. PEDERSEN AND M. J. D. POWELL (1982) *The watchdog technique for forcing convergence in algorithms for constrained optimization*, Math. Programming Study, 16, pp. 1-17.
- T. F. COLEMAN AND A. R. CONN (1982a) *Nonlinear Programming via an exact penalty function: Asymptotic Analysis*, Math. Programming, 24, pp. 123-136.
- T. F. COLEMAN AND A. R. CONN (1982b) *Nonlinear Programming via an exact penalty function: Global Analysis*, Math. Programming, 24, pp. 137-161.
- R. FLETCHER (1976) *Factorizing symmetric indefinite matrices*, Linear Algebra and Appl., 14, pp. 257-272.
- R. FLETCHER (1981) *Practical Methods of Optimization, Vol. 2: Constrained Optimization*, John Wiley, New York.
- R. FLETCHER AND T. L. FREEMAN (1977) *A modified Newton method for minimization*, J.Optim.Theory Appl., 23, pp. 357-372.
- P. E. GILL, N. I. M. GOULD, W. MURRAY, M. A. SAUNDERS AND M. H. WRIGHT (1982) *Range space methods for convex quadratic programming*, Tech. Rept SOL. 82-14 Stanford Univ., Stanford, CA.
- P. E. GILL AND W. MURRAY (1974) *Numerical Methods for Constrained Optimization*, Academic Press, New York.
- P. E. GILL, W. MURRAY AND M. H. WRIGHT (1981) *Practical Optimization*, Academic Press, New York.
- D. GOLDFARB (1980) *Curvilinear path steplength algorithms for minimization which use directions of negative curvature*, Math. Programming, 18, pp. 31-40.
- N. I. M. GOULD (1983) *On practical conditions for the existence and uniqueness of solutions to general equality quadratic programming problems*, Research Report CORR 83-1, Univ. of Waterloo, Waterloo, Ontario.
- J. J. MORE AND D. C. SORENSEN (1979) *On the use of directions of negative curvature in a modified Newton method*, Math. Programming, 16, pp. 1-20.
- W. MURRAY AND M. H. WRIGHT (1978) *Projected Lagrangian methods based on the trajectories of penalty and barrier functions*, Research Report SOL 78-23, Stanford Univ. Stanford, CA.
- M. J. D. POWELL (1978) *A fast algorithm for nonlinearly constrained optimization calculations*, in Numerical Analysis, Dundee 1977, G.A. Watson, ed., Lecture Notes in Mathematics 630, Springer Verlag, Berlin.
- D. C. SORENSEN (1977) *Updating the symmetric indefinite factorization with applications in a modified Newton's method*, report ANL-77-49, Argonne National Lab., Argonne, IL.