

THE USE OF NEGATIVE CURVATURE
IN MINIMIZATION ALGORITHMS

by

Donald Goldfarb*

TR 80-412

Department of Computer Science
Cornell University
Ithaca, New York 14853

* Research supported in part by the U.S. Army Research Office
under Grant DAAG 29-77-G-0114.

THE USE OF NEGATIVE CURVATURE
IN MINIMIZATION ALGORITHMS

by
Donald Goldfarb

Abstract:

In this paper we examine existing algorithms for minimizing a nonlinear function of many variables which make use of negative curvature. These algorithms can all be viewed as modified versions of Newton's method and their merits and drawbacks are discussed to help identify new and more promising methods. The algorithms considered include ones which compute and search along nonascent directions of negative curvature and ones which search along curvilinear paths generated by these directions and descent directions. Versions of the Goldfeld-Quandt-Trotter method, or equivalently, methods based upon a trust region strategy, and gradient path methods are also considered. When combined with the numerically stable Bunch-Parlett factorization of a symmetric indefinite matrix the latter two approaches give rise to new, and what appears to be, efficient and robust minimization methods which can take advantage of negative curvature when it is encountered. Several suggestions are made for further research in this area.

The Use of Negative Curvature in Minimization Algorithms

1. Introduction

We consider here the problem of how to best take advantage of negative curvature when minimizing a nonlinear function $f(x)$ of a vector $x \in R^n$. This problem has been ignored for the most part because the Hessian matrix, $G(x)$, is usually positive definite in the neighborhood of a local minimum of $f(x)$ and because using a positive definite approximation B to $G(x)$ ensures that the direction $s = -B^{-1}g$, where $g = \nabla f(x)$ is the gradient of $f(x)$, is a descent direction. However, far from a local minimum it is often the case that the true Hessian is indefinite. Moreover, in such regions, directions of negative curvature which are also descent directions are excellent choices for decreasing $f(x)$ as much as possible.

Taking advantage of directions of negative curvature should also prove useful to algorithms for solving the general nonlinear programming problem:

$$(1) \quad \begin{array}{ll} \text{Minimize } f(x), & \text{subject to } c_i(x) \geq 0 \quad i \in I \\ x & c_i(x) = 0 \quad i \in E. \end{array}$$

Currently, the most successful algorithms for solving (1) are the so-called projected Lagrangian or recursive quadratic programming methods, [3], [14], [22], [23], [39]. These are based upon applying Newton or quasi-Newton methods to the Lagrangian function $L(x, \lambda) = f(x) - \sum_1^r \lambda_i c_i(x)$ while linearizing the constraints. Although the second order necessary conditions require that the projected Hessian of $L(x, \lambda)$ with respect to the active constraint manifold must be positive semi-definite at the optimum, away from this point, the Hessian of $L(x, \lambda)$ and various projected Hessians can be, and commonly are, indefinite. Augmented Lagrangian methods [25], [35], [40] which search along Newton or quasi-Newton directions should also benefit from research on negative curvature directions.

In the next three sections we describe three different approaches for using directions of negative curvature in minimization algorithms. In addition to discussing methods that have appeared in the literature, we propose several new methods. Finally, in section 5 we state what we believe to be the most promising directions for future research in this area.

2. Methods which make direct use of Negative Curvature Directions.

In this section we consider methods which either search along "nonascent" directions of negative curvature, d , (i.e. $d^T g \leq 0$ and $d^T G d \leq 0$), when they exist or which search along a curvilinear path

$$(2) \quad x(\alpha) = x + a_1(\alpha)s + a_2(\alpha)d, \alpha \geq 0$$

where s is a "descent" direction, (i.e. $s^T g < 0$), d is as above, and $a_1(\alpha)$ and $a_2(\alpha)$ are increasing functions of α , with $a_1(0) = a_2(0) = 0$. Essentially all proposals to explicitly take advantage of negative curvature in minimization algorithms that have appeared in print

have used one of these approaches [9], [13], [15], [17], [29], [30], [33], [43]. However, before we describe these methods we first shall discuss how "nonascent" directions of negative curvature can be obtained.

Let the Bunch-Parlett [6],[7] factorization of G be

$$(3) \quad P^T G P = L D L^T$$

where D is block diagonal with 1×1 or symmetric 2×2 blocks, L is unit lower triangular and P is a permutation matrix. Moreover, L has zeros wherever D has nonzero off-diagonal elements. It can be shown that $(1/6)n^3 + O(n^2)$ operations and $O(n^2)$ comparisons are required to compute this factorization.

Let $V = [v_1, \dots, v_k]$ be an $(n \times k)$ matrix of rank k where $DV = \Lambda V$. $V^T V = I$,

$$\Lambda = \text{diag} \{ \lambda_1, \dots, \lambda_k \}, \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k \leq 0,$$

and k is as large as possible. That is, V is an orthogonal matrix whose columns are normalized eigenvectors of D corresponding to zero and negative eigenvalues. These eigenvectors have either one or two nonzeros and they, i.e. V , and Λ , are trivial to compute.

If we let the k -component vector $e^T = (1, 1, \dots, 1)$, and define d by

$$(4) \quad d = \begin{cases} Pt, & \text{if } g^T Pt \leq 0 \\ -Pt, & \text{if } g^T Pt > 0 \end{cases}$$

where

$$(5) \quad L^T t = Ve,$$

then d is clearly a descent direction by (4). From (5) it follows

that

$$\begin{aligned} d^T G d &= t^T L D L^T t = e^T V^T D V e \\ &= e^T \Lambda e = \sum \lambda_i \leq 0, \\ &\quad \lambda_i \leq 0 \end{aligned}$$

and thus d is a direction of zero or negative curvature.

In [13], Fletcher and Freeman describe a method which, when G is indefinite, alternatively searches along the direction d above and the Newton direction restricted to a subspace of positive curvature, given by

$$(6) \quad p = -(L\bar{D}L^T)^{-1}g$$

where \bar{D} is D in (3) with any negative eigenvalues made zero.

Fiacco and McCormick [9] were the first to propose a modified Newton method which used directions of negative curvature. Their method was based upon the Cholesky factorization of G , which, unfortunately, is not numerically stable and may not even exist when G is indefinite.

Although Fletcher and Freeman's method corrects this defect, it still has several unsatisfactory aspects. First, if we consider some region where $f(x)$ is approximated well by a quadratic with Hessian G , then the direction d given by (4) will be (approximately) the same no matter where we are in that region. That is except for its use of negative curvature information the direction d is, in a certain sense, arbitrary.

Second, the search direction is generally dependent upon how the factorization (3) is carried out. For example, different symmetric pivoting strategies give different matrices L and D and, hence, different search directions d .

Third, the strategy of alternatively searching in negative and positive curvature directions is ad hoc. Unfortunately, non-convergence may occur in Fletcher and Freeman's method if only negative curvature directions are followed.

When the gradient g is not orthogonal to the column space of

$$(7) \quad N = PL^{-T}V$$

we can overcome the first of the above unsatisfactory features in Fletcher and Freeman's method by choosing d to be the direction of steepest descent in this space. That is

$$(8) \quad d = -QQ^Tg$$

where

$$N = QR$$

and Q is an $n \times k$ matrix with orthonormal columns and R is a $k \times k$ upper triangular matrix. This QR factorization can be computed in a numerically stable fashion using either Householder or Givens transformations [44].

Although the column space of N is one of nonpositive curvature, it is dependent upon the pivoting strategy used in factorizing G . On the other hand, the nonpositive eigenspace of G is unique, as is the direction of steepest descent in it. To compute this direction we could use a block Lanczos method [8] to compute the k algebraically smallest eigenvalues and an orthonormal basis Q for the associated eigenspace of G and then orthogonally project $-g$ onto this subspace using equation (8) above. From Sylvester's

Theorem of Inertia we know that the number of zero and negative eigenvalues of G is equal to the number of 2×2 and negative 1×1 diagonal blocks in D . Also N is an excellent choice for starting the block Lanczos method since all of its columns have nonzero projections on the nonpositive eigenspace of G .

We can, also, use the Lanczos method to find the direction of most negative curvature. If this direction d is used alone, then it suffers from the first of the problems mentioned before for Fletcher and Freeman's method. Combining it, or any other direction of negative curvature which is independent of g , with a descent direction s based upon the local gradient g in the manner of (2) overcomes this drawback. Moreover, by searching along the curvilinear path (2), we no longer must resort to some ad hoc strategy such as alternating between negative and positive curvature search directions as used by Fletcher and Freeman.

McCormick was the first to suggest such an approach and in [30] he showed how to modify the Armijo step-size rule for searching along curves of the form

$$(9) \quad x(\alpha) = x + \alpha^2 s + \alpha d, \quad \alpha \geq 0$$

where $s^T g < 0$ and $d^T g < 0$ and $d^T G d < 0$, when G has at least one negative eigenvalue. Subsequently, Moré and Sorensen [33] obtained similar results for a more general line search based upon (9).

In [17], we proposed instead searching along curves of the form

$$(10) \quad x(\alpha) = x + \alpha s + \alpha^2 d, \quad \alpha \geq 0.$$

Our reason for this choice is the following. At a point \bar{x} , where $g(\bar{x}) \neq 0$, if one takes infinitesimally small steps, then the greatest decrease in $f(x)$ is obtained by moving in the direction of the negative gradient, $-g(\bar{x})$. If $G(\bar{x})$ has at least one negative eigenvalue and $G(x)$ varies slowly, then for very large steps, the best direction in which to move is $\pm u$, the direction of most negative curvature, with sign chosen so that $u^T g(\bar{x}) \leq 0$. This is precisely what our search procedure, with $s = -g$ and $d = u$, will do, while one based upon (9) will result in just the opposite strategy.

In [17], we showed how the Armijo [2] and Goldstein [20] step-size rules could be modified for searching along the curve (10). Moreover, we were able to prove, under fairly mild assumptions, that the sequences of iterates produced by these algorithms converge to stationary points at which the Hessian is positive semi-definite. Similar results were obtained by McCormick [30] and Moré and Sorensen [33] for their step-length algorithms. Sorensen [43] implemented a modified Newton algorithm based upon (9) and reported promising numerical results. In his algorithm, Sorensen chooses s as the solution of

$$B s = -g$$

where $PBP^T = L\bar{D}L^T$, and \bar{D} and L are obtained from the Bunch-Parlett factorization $PGP^T = LDL^T$ of G and the eigenvector-eigenvalue decomposition of D , i.e.

$$D = U \Lambda U^T,$$

by replacing the diagonal elements λ_j of \mathcal{A} by

$$\bar{\lambda}_j = \max \{ |\lambda_j|, \epsilon \frac{1}{n} \max_{1 \leq i \leq n} |\lambda_i|, \epsilon \}.$$

Here, ϵ is the relative machine precision, and

$$\bar{D} = U \bar{\mathcal{A}} U^T = U \text{diag}(\bar{\lambda}_1, \dots, \bar{\lambda}_n) U^T.$$

d is chosen as the solution of

$$L^T P d = \pm |\min \{ \lambda(D), 0 \}|^{1/2} u$$

where $\lambda(D)$ is the smallest eigenvalue of D and u is the associated normalized eigenvector. It would be interesting to compare results obtained using the curvilinear steplength algorithms described in [17] which are based upon (10) with those obtained using the algorithms described in [30] and [33] which are based upon (9) in otherwise identical modified-Newton codes.

Another method which makes direct use of directions of negative curvature has been proposed by Gill and Murray [15],[34]. Their method is based upon forming the Cholesky factorization

$$LDL^T = G + E$$

where E is a diagonal matrix whose diagonal elements are chosen so as to make the matrix $G + E$ positive definite. The matrix E is determined during a modified Cholesky factorization algorithm, and when G has at least one negative eigenvalue, it is possible to identify a column of L^{-T} which is a direction of negative curvature. Gill and Murray search in this direction only when $\|g\| \leq \epsilon$ and $E \neq 0$, i.e. when one is in the neighborhood of a

saddle point. Clearly when $E \neq 0$, the direction $-(LDL^T)^{-1}g$ is biased towards the eigenvectors of $G + E$ corresponding to the smallest eigenvalues. An analysis of this bias should prove illuminating when G has negative eigenvalues.

3. Goldfeld-Quandt-Trotter Type Methods.

In the Goldfeld-Quandt-Trotter (GQT) algorithm [18] one computes a step $\Delta x = \Delta x(\alpha) = x(\alpha) - x$ by solving

$$(11) \quad (\alpha I + G) \Delta x = -g$$

where the parameter α is chosen large enough to ensure that the matrix $(\alpha I + G)$ is positive definite and that there is a sufficient decrease in the objective function. Clearly as $\alpha \rightarrow \infty$, $\Delta x(\alpha) \rightarrow -g/\alpha$, and an infinitesimal step is taken in the direction of steepest descent. If G is nonsingular, then $\alpha = 0$ gives the Newton step.

If $(\alpha I + G)$ is positive definite and $g \neq 0$, then the step $\Delta x(\alpha)$ determined by (11) has the property that it solves the problem

$$\text{Minimize } q(x + s) = 1/2 s^T G s + g^T s + f$$

$$\text{subject to } \|s\| \leq \|\Delta x(\alpha)\|$$

where q is a local quadratic approximation to $f(x)$ at the point x . The method as stated has two computational drawbacks: (i) it requires the solution of the linear systems (11) for possibly several values of α at each iteration and (ii) it requires the

knowledge of a lower bound for the smallest eigenvalue of G at each iteration. As Fletcher [12] has pointed out the use of a QTQ^T decomposition where Q is orthogonal and T tridiagonal makes the solution of (11) for several values of α not much more time consuming than for just one value of α . (Of course, a full eigenvalue-eigenvector decomposition of G is also possible.) Having the QTQ^T decomposition of G also makes the computation of the smallest eigenvalue of G easier. An efficient $2/3 n^3 + 0(n^2)$ algorithm exists for the computation of $G = QTQ^T$, (e.g., see [44; pages 335-336]). However, as far as this author knows, this factorization has not been used in optimization algorithms.

Several other effective implementations of the GQT approach, (or for nonlinear least squares, the Levenberg-Marquardt [26],[27] approach), have been tried, [11],[24],[32],[36],[37],[38]. These algorithms are based upon considering $q(x+s)$ to be an accurate approximation to $f(x)$ in some trust region $\|s\| \leq \rho$ about x , which is varied from one iteration to the next. How to choose ρ , or equivalently, how to choose α , has been studied for the case of G positive semi-definite by Fletcher [11], Hebden [24], and More [32]. Powell suggested replacing the curvilinear path $x(\alpha)$ determined by (11) by a piecewise linear two-segment path, which is usually referred to as a dog-leg approximation [36],[37], [38]. This approximation makes the computation for several values of α , (or equivalently ρ), easy to carry out.

We now describe a new approach to these problems which we believe to be quite promising. Consider the GQT algorithm applied

to the objective function as a function of a new set of variables z , where z is related to x by the linear transformation

$$(12) \quad z = L^T P^T x.$$

and L and P are determined by performing the Bunch-Parlett factorization (3) of G . In the z variables the gradient $\nabla_z f = L^{-1} P^T g$ and the Hessian $\nabla_z^2 f = D$. Consequently the GQT algorithm in the z space determines a $\Delta z = \Delta z(\alpha)$ by solving

$$(13) \quad (\alpha I + D) \Delta z = -\nabla_z f = -L^{-1} P^T g.$$

for a given α . Transforming back to the x variables we have

$$(14) \quad (\alpha I + D) L^T P^T \Delta x = -L^{-1} P^T g.$$

Thus $\Delta x(\alpha)$ is determined by backsolving

$$L^T P^T \Delta x = -(\alpha I + D)^{-1} w$$

where $w = L^{-1} P^T g$. Because D is diagonal, $1/2 n^2 + 0(n)$ operations are required for each additional choice of α , and it is trivial to determine a strict lower bound for α so that $(\alpha I + D)$ is positive definite. It is also possible to use the factorization $P^T G P = L T L^T$ computed by Aasen's method [1] in $1/6 n^3 + 0(n^2)$ operations, where P , L , and T are permutation, unit lower triangular and tridiagonal matrices, respectively. The variant of the GQT algorithm described above remains the same except that D is replaced by T .

Premultiplying (14) by PL we obtain

$$(15) \quad (\alpha M + G) \Delta x = -g$$

where $M = PLL^T P^T$. Hence, our algorithm is of the modified GQT type described in [19] and has the property that for $(\alpha M + G)$

positive definite and $g \neq 0$, a point is found which minimizes the local quadratic approximation $q(x+s)$ to $f(x)$ over an ellipsoid. Assuming G is nonsingular, $\Delta x(0)$ is the Newton direction as in (11). However, $\Delta x(\infty)$ is no longer a direction of steepest descent in the x space, (although it is a steepest descent direction in the metric M). This may not be very crucial since we would like to choose α as close to $\hat{\alpha} = -\min \{0, \lambda(D)\}$ as possible where $\lambda(D)$ is the smallest eigenvalue of D . When G , and hence D , has at least one negative eigenvalue, if we solve (13) for $\Delta z(\alpha)$, and hence (14) or (15) for $\Delta x(\alpha)$, then in the limit as $\alpha \rightarrow \lambda(D)_+$, $\Delta z(\alpha) \rightarrow$ the corresponding eigenvector, u , of D as long as $L^{-1}P^Tg$ is not orthogonal to u . The problem with using (13) and (14) is not the determination of a suitable initial search direction--that is just $\pm P L^{-T}u$ --but, rather, how large a step to take. Analogous remarks apply to the GQT equation (11).

One interpretation of (11), (Jane Cullum--private communication), that may be of help is the following. Consider the system of differential equations

$$(16) \quad \dot{x}(t) = -g(x(t))$$

with initial conditions

$$(17) \quad x(0) = x_0.$$

A solution $x(t)$ satisfying (16) and (17) defines the "gradient path," i.e., a path that is normal to the contours of $f(x)$, that passes through x_0 . If one applies the backward Euler method to (16) one obtains

$$(18) \quad \Delta x(h) = x(h) - x(0) = -hg(x(h))$$

which for small h should give an $x(h)$ with a lower function value than $x(0)$. Solving the set of nonlinear equations (18) for $x(h)$ is, in general, a difficult task. However, if we linearize $g(x(h))$ about $x(0)$, i.e. set

$$g(x(h)) = g(x(0)) + G(x(0)) \cdot \Delta x(h) ,$$

or equivalently, apply one step of Newton's method to the solution of

$$F(x) \equiv x - x(0) + h g(x) = 0$$

we obtain

$$(19) \quad (I + h G) \Delta x = -h g .$$

This, of course, is just the GQT equation (11) with $\alpha = 1/h$. As the backward Euler method is the most elementary method used for solving stiff differential equations, there may be results in that field that could be of help in choosing h in (19), (i.e. α in (11)). Moreover, the above interpretation of the GQT method seems to indicate that it should perform well on ill-conditioned problems. Minimization methods derived from other stiff schemes may also prove to be of interest.

Before leaving the GQT method we observe that there are many possibilities for the choice of M in (15). One possibility is to use a positive definite quasi-Newton matrix which is updated by, say the BFGS formula [5],[10],[16],[41]. If a quasi-Newton updating algorithm is used which results in an indefinite matrix

approximation B to G, then the GQT method (11) with G replaced by B is a natural choice for determining the next iterate. This may prove to be very useful in implementing quasi-Newton updates which preserve known sparsity in the Hessian. These methods [28],[42],[45] do not produce matrices B which are positive definite.

4. Newton and quasi-Newton Based Gradient Path Methods.

In the last section we defined the gradient path $x(t)$ passing through a point x_k as the solution of the system of differential equations (16) subject to the initial conditions (17). If $g(x)$ is a nonlinear function, it is, in general, not possible to obtain a closed form expression for $x(t)$. In fact, in the last section we were able to obtain such an expression only after replacing (16) by its backward Euler approximation (18) and replacing $g(x(t))$ by its first order Taylor series approximation about $x(0)$. Recently, several new minimization methods have been proposed which are based upon solving (16) and (17) using just the latter approximation [4],[46],[47].

This is equivalent to solving

$$(20) \quad \dot{x}(t) = -\nabla f(x(t))$$

given that

$$(21) \quad x(0) = x_k$$

assuming that $f(x)$ is a quadratic form. If

$$(22) \quad G = Q\Lambda Q^T, \quad Q^T Q = I$$

is the eigenvalue-eigenvector decomposition of G, i.e. Λ is a diagonal matrix of eigenvalues and Q is a matrix of normalized eigenvectors, then it is well known that the solution of (20) and (21) is

$$(23) \quad x(t) = x_k - Q(I - e^{-\Lambda t})\Lambda^{-1} Q^T \nabla_x f(x_k).$$

Botsaris and Jacobson [4] show that the matrix $(I - e^{-\Lambda t})\Lambda^{-1}$ is well-defined and positive definite for all $t > 0$, even if some $\lambda_i = 0$. Their method uses (23), choosing $t = \infty$ as a first estimate of x_{k+1} given x_k . As this gives an infinite step if any $\lambda_i < 0$, they replace Λ in this case by $|\Lambda|$. If this does not result in a decrease in $f(x)$, they perform a search along the curve determined by (23) using the golden section method to decrease the size of the interval to (1/10)-th of its original size.

A major drawback of this method is the need to compute the full eigenvector-eigenvalue decomposition of G on each iteration. This problem is avoided by Zang [47] who works with a Cholesky factorization LDL^T of a positive definite matrix B which is updated by a quasi-Newton method after each iteration. Zang determines the gradient path of the quadratic approximation $\phi(x)$ to $f(x)$, where $\nabla^2 \phi(x) = B$, in a transformed space whose variables z are related to x by

$$(24) \quad z = L^T x.$$

Since in this space $\nabla^2 \phi(z) = D$, (23) becomes

$$(25) \quad z(t) = z_k - (I - e^{-Dt})D^{-1} \nabla_z f(z_k).$$

In order to take advantage of directions of negative curvature, we propose a somewhat different approach. At each iteration compute the Bunch-Parlett factorization (3) of G, and determine the eigensystem of D, i.e.,

$$(26) \quad D Q = Q \Lambda.$$

The total work involved in (3) and (26) is only $1/6 n^3 + O(n^2)$ operations. Using the same transformation of variables (12) as we used in the previous section, we obtain (23) with x replaced wherever it appears by z. Transforming back into our original x variables yields

$$(27) \quad x(t) = x_k - PL^{-T} Q(I - e^{-\Lambda t}) \Lambda^{-1} Q^T L^{-1} P^T \nabla f(x_k).$$

As in the modified GQT method described in the previous section $\Delta x(t) = x(t) - x_k$ ranges from the Newton step for $t = \infty$, ($\alpha = 0$), assuming G, and hence D, is positive definite to an infinitesimal steepest descent step in the metric $M = PLL^T P^T$ for $t \rightarrow 0$, ($\alpha \rightarrow \infty$). If D has negative eigenvalues, then in general, $\Delta x(\infty)$ is an infinite step in the direction of the eigenvector corresponding to the most negative eigenvalue of D, transformed back into the x-space. This method shares with Zang's methods [47], and the method of the last section, the property that computing $x(t)$ for several values of t requires only a relatively small increase in work over that required for one value of t. Moreover, it is possible to directly use directions of negative curvature by searching

along the ray $x(t) - x_k$, $t > 0$, when G is indefinite, assuming that the component of $Q^T L^{-1} P^T \nabla f(x_k)$ corresponding to the most negative eigenvalue of D is not equal to zero.

Another possibility is to orthogonally reduce G to tridiagonal form, i.e.

$$G = \bar{Q} T \bar{Q}^T,$$

which allows us to express the gradient path as

$$(28) \quad x(t) = x_k - \bar{Q}(I - e^{-Tt})T^{-1}\bar{Q}^T \nabla f(x_k).$$

Moreover, one can use a Padé approximation for e^{-Tt} [31], or better still, for $(I - e^{-Tt})T^{-1}$, such as

$$(29) \quad (I - e^{-Tt})T^{-1} \approx t(I + tT/3)^{-1} (I - tT/6).$$

Since T is tridiagonal, approximate values of $x(t)$ for various values of t along the curve (28) can be computed in a reasonable amount of time using approximations like (29).

It is important to note that replacing Λ by $|\Lambda|$ and computing a (positive-definitized) Newton direction

$$- Q |\Lambda|^{-1} Q^T \nabla f(x_k)$$

as proposed, for example, by Greenstadt [21] is biased towards directions which have the smallest curvature independent of whether that curvature is positive or negative while methods (14) and (27) produce directions that are biased towards the direction of most negative curvature. A similar remark applies to the gradient path method of Botsaris and Jacobsen [4].

5. Suggestions for Research:

Before some of the new methods proposed in the previous sections can be implemented, several questions remain to be answered. For example, for the generalized Goldfeld-Quandt-Trotter method based upon the Bunch-Parlett factorization, how large an initial step should one take in the direction of negative curvature obtained when G has negative eigenvalues and α is chosen equal to $\hat{\alpha} = -(\text{minimum eigenvalue of } D)$? Alternatively, how should one choose $\alpha > \hat{\alpha}$ initially at each iteration? The latter approach seems to be a better one to follow, as some of the tactics employed for determining the appropriate size to use for a "trust region" in [11], [24], and [32] are probably applicable to the indefinite case. The convergence and rate of convergence of this type of method should also be studied.

The generalized GQT method that we proposed searches along a curvilinear path which is very similar to the path (10) when $f(x)$ is locally nonconvex at the current point x . With proper choice of "descent pairs" s and d , the algorithms given in [17] guarantee, under fairly mild conditions on $f(x)$, that convergence to a stationary point where the Hessian is positive definite or positive semi-definite occurs. It would be interesting to develop specific versions of the generalized GQT method proposed in section 3 that have this property. When $g = 0$ this can be achieved only if the choice of α and hence $\Delta z = \Delta z(\alpha)$ in (13) is determined from (13) itself. That is when $g = 0$ and G is not positive semi-definite, one should choose $\alpha = \hat{\alpha}$, the most negative eigenvalue of D, and

Δz equal to the corresponding eigenvector. If some other arbitrary value of α is chosen, then (13) will in general only have the solution $\Delta z = 0$ and the method will halt.

Another aspect of GQT methods that deserves study is their relationship to methods for solving stiff ordinary differential equations. Such research might lead to better ways of solving ill-conditioned minimization problems.

Almost all of the questions raised regarding the generalized GQT method also apply to the "transformed" gradient path method based upon the Bunch-Parlett factorization that was described in the last section. For example, how should the initial step, i.e., t , be chosen, and can an algorithm be devised which guarantees convergence to a positive semi-definite point? The relationship between this method and the generalized GQT method should also be explored further.

An analysis of how modified Newton methods, such as the Gill-Murray algorithm [15], and quasi-Newton methods model negative curvature might prove very useful. Both those methods that force the approximation B to the Hessian to be positive definite and those that do not should be studied. Using the results of this analysis, it might be possible to derive methods based upon the methods proposed in the preceding sections by replacing G by B.

Specializing the methods described here to the solution of nonlinear least squares problems and generalizing them for the solution of constrained nonlinear programming problems are areas ripe for research. In particular, recursive quadratic programming

type methods might be developed which could effectively use negative curvature information.

Finally, in order to computationally compare different algorithms, a collection of locally nonconvex minimization problems needs to be created and/or assembled.

III. References

- [1] Aasen, J.O., "On the reduction of a symmetric matrix to tridiagonal form," BIT 11 (1971), 233-242.
- [2] Armijo, L., "Minimization of functions having Lipschitz continuous first partial derivatives," Pacific J. Math 16 (1966), 1-3.
- [3] Biggs, M.C., "Constrained minimization using recursive equality quadratic programming," Numerical Methods for Nonlinear Optimization, ed. F.A. Lootsma, Academic Press (London), (1972).
- [4] Botsaris, C.A. and D.H. Jacobsen, "A Newton-type curvilinear method for optimization," J. Math. Anal. Appl. 54, (1976), 217-229.
- [5] Broyden, C.G., "The convergence of a class of double rank minimization algorithms 2, the new algorithm," JIMA 6 (1970) 222-231.
- [6] Bunch, J.R. and Parlett, B.N., "Direct methods for solving symmetric indefinite systems of linear equations," SIAM J. Numer. Anal. 8 (1971), 639-655.
- [7] Bunch, J.R. and Kaufman, L., "Some stable methods for calculating inertia and solving symmetric linear equations," Math. Comp. 31 (1977), 163-179.
- [8] Cullum, J. and R.A. Willoughby, "Computing eigenvectors and eigenvalues of large, symmetric matrices using Lanczos tridiagonalization," Proc. Dundee Biennial Conference on Numerical Analysis (1979), ed. G.A. Watson, University of Dundee, to be published by Springer-Verlag (1980).
- [9] Fiacco, A.V. and G.P. McCormick, Nonlinear Programming, Sequential Unconstrained Minimization Techniques, John Wiley & Sons, New York (1968), 170-172.
- [10] Fletcher, R., "A new approach to variable metric algorithms," Computer J. 13 (1970), 317-322.
- [11] Fletcher, R., "A modified Marquardt subroutine for non-linear least squares," A.E.R.E. Harwell Report R6799 (1971)

- [12] Fletcher, R., "On Newton's method for minimization," Dundee Numerical Analysis Report No. 19, University of Dundee (1976).
- [13] Fletcher, R., and T.L. Freeman, "A modified Newton method for minimization," J. Opt. Theory Applics. 23, (1977), 357-372.
- [14] Garcia Palomares, U.M. and O.L. Mangasarian, "Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems," Math. Prog. 7, Vol. 11, (1976), 1-13.
- [15] Gill, P.E. and W. Murray, "Newton type methods for unconstrained and linearly constrained optimization," Math. Prog. 7 (1974), 311-350.
- [16] Goldfarb, D., "A family of variable-metric methods derived by variational means," Math. Comp 24 (1970), 23-26.
- [17] Goldfarb, D., "Curvilinear path steplength algorithms for minimization which use directions of negative curvature," Math. Prog. 17 (1979), in press.
- [18] Goldfeld, S.M., R.E. Quandt, H.F. Trotter, "Maximization by quadratic hill-climbing," Econometrica, Vol. 34 (1966), 541-551.
- [19] Goldfeld, S.M., R.E. Quandt, H.F. Trotter, "Maximization by improved quadratic hill-climbing and other methods," Econometric Res. Prog. Memo. No. 95 (1968), Princeton University.
- [20] Goldstein, A.A., "On steepest descent," SIAM J. on Control, 3 (1965) 147-151.
- [21] Greenstadt, J., "On the relative efficiencies of gradient methods," Math. Comp. 21 (1967), 360-367.
- [22] Han, S-P., "A globally convergent method for nonlinear programming," Department of Computer Science, TR 75-257, Cornell University (1975).
- [23] Han, S-P., "Superlinearly convergent variable metric algorithms for general nonlinear programming problems," Math. Prog., Vol. 11 (1976), 263-282.
- [24] Hebden, M.D., "An algorithm for minimisation using exact second derivatives," A.E.R.E. Harwell Report TP515 (1973).

- [25] Hestenes, M.R., "Multiplier and gradient methods," J. Opt. Theory Applics. 4, (1969) 303-320.
- [26] Levenberg, K., "A method for the solution of certain nonlinear problems in least squares," Quart. Appl. Math 2 (1944), 164-168.
- [27] Marquardt, D.W., "An algorithm for least squares estimation of nonlinear parameters," SIAM J. Appl. Math. 11 (1963), 431-441.
- [28] Marwil, E.S., "Exploiting sparsity in Newton-like methods," Ph.D. thesis, Cornell University, (1978).
- [29] McCormick, G.P., "On computing the positive part and a direction of negative curvature for a symmetric matrix given in dyadic form," Tech. Paper Serial T-337, Inst. for Management Sci. and Engineering, The George Washington University (1976).
- [30] McCormick, G.P., "A modification of Armijo's step-size rule for negative curvature," Math. Prog. 13 (1977), 111-115.
- [31] Moler, C. and C. VanLoan, "Nineteen dubious ways to compute the exponential of a matrix," SIAM Review 20 (1978), 801-836.
- [32] Moré, J.J., "The Levenberg-Marquardt algorithm: implementation and theory," Proc. 1977 Dundee Conf. on Numer. Anal. (1977).
- [33] Moré, J.J. and Sorensen, D.C., "On the use of directions of negative curvature in a modified Newton method," Math. Prog. 16 (1979), 1-20.
- [34] Murray, W., "Second derivative methods," Numerical Methods for Unconstrained Optimization, ed. W. Murray, Academic Press (New York) (1972), 57-71.
- [35] Powell, M.J.D., "A method for nonlinear constraints in minimization problems," Optimization, ed. R. Fletcher, Academic Press (London and New York), (1969), 283-298.
- [36] Powell, M.J.D., "A hybrid method for nonlinear equations," Numerical Methods for Nonlinear Algebraic Equations, ed. P. Rabinowitz, Gordon and Breach (London) (1970).
- [37] Powell, M.J.D., "A FORTRAN subroutine for solving systems of nonlinear algebraic equations," Numerical Methods for Nonlinear Algebraic Equations, ed. P. Rabinowitz, Gordon and Breach (London) (1970).
- [38] Powell, M.J.D., "A new algorithm for unconstrained optimization," Nonlinear Programming, eds. J.B. Rosen, O.L. Mangasarian, K. Ritter, Academic Press (New York), (1970).

- [39] Powell, M.J.D., A fast algorithm for nonlinearly constrained optimization calculations," Proceedings Dundee Biennial Conference on Numerical Analysis (1977), ed., G.S. Watson, Springer-Verlag Lecture Notes in Mathematics, No. 630, 144-157.
- [40] Rockafellar, R.T., "Augmented Lagrange multiplier functions and duality in non-convex programming," SIAM J. Control 12, (1974), 268-285.
- [41] Shanno, D.F., "Conditioning of quasi-Newton methods for function minimization," Math. Comp. 24 (1970), 647-656.
- [42] Shanno, D.F., "On variable-metric methods for sparse Hessians," Math. of Comp. (1979), in press.
- [43] Sorensen, D.C., "Updating the symmetric indefinite factorization with applications in a modified Newton's method," Ph.D. thesis, University of California at San Diego, Argonne National Laboratory Report ANL-77-49, (1977).
- [44] Stewart, G.W. Introduction to Matrix Computations, Academic Press, New York (1973).
- [45] Toint, Ph.L., "On sparse and symmetric updating subject to a linear equation," Math. Comp. 31 (1977), 954-961.
- [46] Vial, J. and Zang, I., "Unconstrained optimization by approximation of the gradient path," Math. of Operations Research 2 (1977), 253-265.
- [47] Zang, I., "A new arc algorithm for unconstrained optimization," Math. Programming 15 (1978), 36-52.

