Open access • Journal Article • DOI:10.1137/0707032

The Conjugate Residual Method for Constrained Minimization Problems

David G. Luenberger

Institutions: Stanford University

Published on: 01 Sep 1970 - SIAM Journal on Numerical Analysis (Society for Industrial and Applied Mathematics)

Topics: Conjugate residual method, Conjugate gradient method, Positive-definite matrix and Conjugate

The Conjugate Residual Method for Constrained Minimization Problems

David Luenberger

THE CONJUGATE RESIDUAL METHOD FOR CONSTRAINED MINIMIZATION PROBLEMS

DAVID G. LUENBERGER

1. Introduction. The method of conjugate gradients [1]–[5], a particular version of the general class of conjugate direction methods, was originally developed for solving a linear system of equations

$$Ax = b, \tag{1.1}$$

where A is an $n \times n$ symmetric, positive definite matrix, x is an unknown n-vector and b is a fixed n-vector. The conjugate gradient method can be viewed as a descent procedure for the problem:

Minimize

$$(x, Ax) - 2(x, b); \tag{1.2}$$

and through this viewpoint the method can be generalized so as to apply to the solution of general unconstrained minimization problems. For these general problems conjugate gradient methods are among the most effective first order descent procedures.

Minimization problems subject to constraints cannot be treated directly by the conjugate gradient methods although they have often been employed indirectly through the artifice of introducing penalty functions or of suitably projecting the descent directions onto the constraint surface. The major motivation for the research presented in this paper is to develop a conjugate direction procedure that is capable of handling constraints directly. This quickly leads one to consider equations of the form (1.1) with the matrix A symmetric and nonsingular but not necessarily positive definite. For, consider the simple problem:

Minimize

$$(u, Qu) - 2(c, u) \tag{1.3}$$

subject to

$$Bu = d,$$

where Q is an $n \times n$ symmetric, positive definite matrix, B is an $m \times n$ matrix of rank $m \leqq n$. The n-vector u is unknown and c and d are fixed n- and m-vectors respectively. By introducing an m-dimensional Lagrange multiplier this problem is found to be equivalent to that of solving the linear vector equation

$$\begin{bmatrix} Q & B^t \\ B & 0 \end{bmatrix} \begin{bmatrix} u \\ \lambda \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}. \tag{1.4}$$

Denoting the $(n + m) \times (n + m)$ matrix in (1.4) by A we see that A is symmetric, nonsingular, but not positive definite.

Of course, (1.1) can be modified to the equivalent form

$$A^2x = Ab, \tag{1.5}$$

where now A^2 is positive definite, but this stratagem requires roughly twice the computational effort of methods which could be applied directly if A itself were positive definite [3].

In this paper we discuss methods requiring essentially the same computational effort as those employed if A is positive definite. An alternative approach to this problem, which salvages the standard conjugate gradient algorithm, is considered in [6].

2. The method of conjugate residuals.

DEFINITION. Given an $n \times n$ matrix A, a pair of n-vectors x, y is said to be *A-orthogonal* if $(x, Ay) = 0$.

We now develop an iterative technique for solving

$$Ax = b \tag{2.1}$$

which has the same important characteristics as the method of conjugate gradients but which in addition is applicable in the case where A is symmetric but not necessarily positive definite. The procedure is a descent procedure in that a sequence of approximate solutions $x_1, x_2, \cdots, x_n$ is generated with the property that the objective functional

$$E(x) = \|Ax - b\|^2 \tag{2.2}$$

decreases monotonically. Furthermore, as is common with descent procedures, the approximation x_{k+1} is generated from x_k by moving from x_k in a certain direction p_k to a minimum point of E in this direction. In other words, $x_{k+1} = x_k + \alpha_k p_k$, where α_k is chosen to minimize E. The novel feature of the method given below is of course the method for selecting the direction vectors p_k.

In the procedure the direction vectors are chosen to be an A^2-orthogonalized version of the residuals $r_k = b - Ax_k$. This is similar to the method of conjugate gradients but now since the gradient of the underlying functional E at x_k is $-2Ar_k$, and not $-2r_k$, we refer to the scheme as conjugate residuals.

Here is the fundamental iteration scheme.

Conjugate residual scheme. Select x_1 arbitrarily. Set $p_1 = r_1 = b - Ax_1$, and repeat the following steps omitting (2.3a, b) on the first step.

If $\alpha_{k-1} \neq 0$,

$$p_k = r_k - \beta_k p_{k-1}, \qquad \beta_k = \frac{(Ar_k, Ap_{k-1})}{(p_{k-1}, A^2 p_{k-1})}. \tag{2.3a}$$

If $\alpha_{k-1} = 0$,

$$p_k = Ar_k - \gamma_k p_{k-1} - \delta_k p_{k-2}, \tag{2.3b}$$

$$\gamma_k = \frac{(Ar_k, A^2 p_{k-1})}{(p_{k-1}, A^2 p_{k-1})}, \qquad \delta_k = \frac{(Ar_k, A^2 p_{k-2})}{(p_{k-2}, A^2 p_{k-2})}.$$

(2.3c) $$x_{k+1} = x_k + \alpha_k p_k, \qquad \alpha_k = \frac{(r_k, Ap_k)}{(p_k, A^2 p_k)},$$

(2.3d) $$r_{k+1} = b - Ax_{k+1} = r_k - \alpha_k A p_k.$$

We prove below that the process defined by (2.3) converges to the unique solution of (2.1) in n steps or less. Furthermore, the residuals $r_1, r_2, \cdots$ form an A-orthogonal sequence and the direction vectors $p_1, p_2, \cdots$ form an A^2-orthogonal sequence.

If the matrix A is positive definite, then for each k, $\alpha_k > 0$ and hence (2.3b) is never used. In this case the procedure can be regarded as one of the family of general conjugate gradient methods discussed by Hestenes [2]. However, even in the positive definite case this method has not been explicitly pointed out and in terms of total computation it is as efficient as all standard procedures.

DEFINITION. Corresponding to the symmetric matrix A, a vector x is said to be *singular* if $(x, Ax) = 0$.

If A is not positive definite, α_k may be positive, negative or zero. The condition $\alpha_k = 0$ corresponds, as shown below, to singularity of the residual r_k.

THEOREM 1. *The following relations hold for the method of conjugate residuals*:

(a) $(p_i, A^2 p_j) = 0, \qquad i \neq j,$

(b) $(r_i, Ap_j) = 0, \qquad i \geqq j + 1,$

(c) $Ap_i \in [p_1, p_2, \cdots, p_{i+1}], \qquad r_i \in [p_1, p_2, \cdots, p_i],$

(d) $(r_i, Ar_j) = 0, \qquad i \neq j,$

(e) $\alpha_{i-1} \neq 0 \quad \textit{implies} \quad (r_i, Ap_i) = (r_i, Ar_i),$

$\beta_i = -(r_i, Ar_i)/(r_{i-1}, Ar_{i-1}),$

(f) $\alpha_{i-1} = 0 \quad \textit{and} \quad r_i \neq 0 \quad \textit{implies} \quad \alpha_i \neq 0,$

(g) $r_i \neq \theta \quad \textit{implies} \quad p_i \neq \theta \qquad (\theta \textit{ denotes zero vector}).$

Proof. The proof is by induction. The vectors $r_1, p_1 = r_1$ and r_2 satisfy these relations since:

(a) If r_1 is singular,

$$r_2 = r_1 = p_1 \quad \text{and} \quad (r_1, Ar_1) = 0.$$

(b) If r_1 is not singular,

$$(r_1, Ar_2) = (p_1, Ar_2) = (r_1, Ar_1) - \alpha_1 (p_1, A^2 p_1)$$
$$= 0 \qquad \text{by (2.3c)}.$$

Now assume that these relations are true for the first $k - 1$ steps, i.e., through $r_k, p_{k-1}, \alpha_{k-1}, \beta_{k-1}$. There are now two distinct cases of interest: $\alpha_{k-1} \neq 0$ and $\alpha_{k-1} = 0$. We must show that r_{k+1} and p_k can be added to these relations in either case.

Case 1. $\alpha_{k-1} \neq 0$. To prove (e) we have immediately

$$(r_k, Ap_k) = (r_k, Ar_k) - \beta_k (r_k, Ap_{k-1}) = (r_k, Ar_k)$$

by the induction hypothesis (b). Also,

$$(r_k, Ar_k) = (r_k, Ar_{k-1}) - \alpha_{k-1}(r_k, A^2 p_{k-1})$$

whence, by (d),

$$(r_k, A^2 p_{k-1}) = -\frac{1}{\alpha_{k-1}}(r_k, Ar_k)$$

which gives the formula for β_k.

To prove (a) we write

$$(p_k, A^2 p_i) = (r_k, A^2 p_i) - \beta_k(p_{k-1}, A^2 p_i).$$

For $i < k - 1$ the first term on the right vanishes because of the induction hypotheses (c) and (b) and the second term vanishes by (a). For $i = k - 1$ the two terms cancel.

To prove (b) we write

$$(r_{k+1}, Ap_i) = (r_k, Ap_i) - \alpha_k(Ap_k, Ap_i).$$

For $i < k$ the right side vanishes by (b) and (a). For $i = k$ the two terms on the right cancel.

To prove (c) we have obviously, for $j \geqq 1$,

$$r_j = p_1 - \sum_{i=1}^{j-1} \alpha_i A p_i \tag{2.4}$$

and hence setting $j = k$ and using (2.3a), we have

$$p_k = p_1 - \sum_{i=1}^{k-2} \alpha_i A p_i - \alpha_{k-1} A p_{k-1} - \beta_k p_{k-1}.$$

Thus, in view of the induction hypothesis (c) and $\alpha_{k-1} \neq 0$ we obtain $Ap_{k-1} \in [p_1, p_2, \cdots, p_k]$.

To prove (d) we write

$$(r_{k+1}, Ar_i) = (r_k, Ar_i) - \alpha_k(Ap_k, Ar_i).$$

For $i < k$ the first term on the right vanishes by the induction hypothesis. The second term also vanishes for $i < k$ by the second part of (c) and (a). For $i = k$ the right side becomes, by (e),

$$(r_k, Ar_k) - \frac{(r_k, Ar_k)}{(p_k, A^2 p_k)}(Ap_k, Ar_k).$$

But this vanishes because $(p_k, A^2 p_k) = (r_k - \beta_k p_{k-1}, A^2 p_k) = (r_k, A^2 p_k)$.

To prove (g) suppose $p_k = \theta$. Then

$$r_k = \beta_k p_{k-1}.$$

But by (e) and (b) this implies $\beta_k = 0$ and hence $r_k = \theta$.

Case 2. $\alpha_{k-1} = 0$. To prove (f) we note that by (2.3b),

$$(r_k, Ap_k) = (r_k, A^2 r_k) - \gamma_k(r_k, Ap_{k-1}) - \delta_k(r_k, Ap_{k-2}). \tag{2.5}$$

The second two terms on the right are zero by the hypothesis (b), and hence if $r_k \neq \theta$ we have $\alpha_k \neq 0$. This proves that there cannot be generated two consecutive singular points in the procedure.

To prove (a) we write

$$(p_k, A^2 p_i) = (Ar_k, A^2 p_i) - \gamma_k (p_{k-1}, A^2 p_i) - \delta_k (p_{k-2}, A^2 p_i).$$

For $i = k-1$, $i = k-2$ the terms on the right cancel. For $i < k-2$ we have by the induction hypothesis on (c) that $A^2 p_i \in [p_1, p_2, \cdots, p_{k-1}]$ and hence by (b) the first term vanishes. The other two terms vanish by (a).

The proof of (b) is the same as for Case 1.

Now if $\alpha_{k-1} = 0$, then we have $\alpha_{k-2} \neq 0$ and also $(r_{k-1}, Ap_{k-1}) = 0$. Since $(r_{k-1}, Ap_{k-1}) = (r_{k-1}, Ar_{k-1}) - \beta_{k-1}(r_{k-1}, Ap_{k-2})$, it follows, using (b), that $(r_{k-1}, Ar_{k-1}) = 0$, and by (e), we have $\beta_{k-1} = 0$. (Equivalently, $\alpha_{k-1} = 0$ if and only if r_{k-1} is singular.)

To prove (c) we note that since $\beta_{k-1} = 0$ we have $p_{k-1} = r_{k-1}$ and since $\alpha_{k-1} = 0$ we have $r_k = r_{k-1}$. Thus, by (2.3b),

$$p_k = Ap_{k-1} - \gamma_k p_{k-1} - \delta_k p_{k-2}$$

which proves (c).

To prove (d) we write

$$(r_{k+1}, Ar_i) = (r_k, Ar_i) - \alpha_k (Ap_k, Ar_i).$$

For $i < k$, this vanishes by the same argument as Case 1. For $i = k$ we note that $r_k = r_{k-1} = p_{k-1}$ and hence the first term vanishes by the induction hypothesis on (d) and the second term vanishes by (a).

Part (g) follows immediately from (2.5). This completes the proof of Theorem 1.

We now state the fundamental minimization property of the method of conjugate residuals.

THEOREM 2. *For each k the point x_{k+1} minimizes the functional $E(x) = \|Ax - b\|^2$ over the linear variety $x_1 + [p_1, p_2, \cdots, p_k]$.*

Proof. This follows immediately from the projection theorem and the fact that r_{k+1} is A-orthogonal to the subspace $[p_1, p_2, \cdots, p_k]$.

THEOREM 3. *The method of conjugate residuals converges in n steps or less to the unique solution of $Ax = b$.*

Proof. From part (g) of Theorem 1 we see that the sequence $p_1, p_2, \cdots$ does not terminate until a solution is obtained. Furthermore, by the A^2-orthogonality of the p_i's this sequence is linearly independent. The result then follows from Theorem 2.

To summarize the results derived in this section, the method of conjugate residuals progresses in a fashion entirely analogous to that of conjugate gradients by generating A^2-orthogonal direction vectors and minimizing $\|Ax - b\|^2$ over increasing subspaces. If at some stage the residual r_k should be singular, then the next direction vector p_k is equal to r_k and the functional E cannot be reduced by moving in that direction since x_k is already at the minimum along that line. The next direction p_{k+1} is then generated by a special formula and this new direction will not be singular. Thus singular directions occur only as isolated members of the sequence $p_1, p_2, \cdots, p_n$. The process converges within n steps.

3. Computational requirements. An important feature of the method of conjugate residuals, which sets it apart from other schemes for handling the non-positive definite case, is that only a single multiplication of a vector by a matrix is required at each step. To see this, suppose that $p_{k-2}, Ap_{k-2}, p_{k-1}, Ap_{k-1}, r_k$ have been stored. Then we have the following.

Case 1. $\alpha_{k-1} \neq 0$.

Step 1a. Calculate Ar_k. This requires one matrix multiplication.

Step 1b. Calculate β_k, p_k from (2.3a).

Step 1c. Calculate Ap_k from $Ap_k = Ar_k - \beta_k Ap_{k-1}$, and then α_k, x_{k+1} from (2.3c).

Step 1d. Calculate r_{k+1} from $r_{k+1} = r_k - \alpha_k Ap_k$.

Case 2. $\alpha_{k-1} = 0$. In this case, r_{k-1} is singular and it follows from part (e) of Theorem 1 that $\beta_{k-1} = 0$ and hence $p_{k-1} = r_{k-1}$. Also since $\alpha_{k-1} = 0$ we have $r_k = r_{k-1}$. Thus $Ar_k = Ap_{k-1}$ is on hand.

Step 2a. Calculate $A^2 r_k = A(Ap_{k-1})$. This requires one multiplication of a vector by a matrix.

Step 2b. Calculate γ_k, δ_k and p_k from (2.3b).

Step 2c. Calculate Ap_k from $Ap_k = A^2 r_k - \gamma_k Ap_{k-1} - \delta_k Ap_{k-2}$ and then α_k, x_{k+1} from (2.3c).

Step 2d. Calculate $r_{k+1} = r_k - \alpha_k Ap_k$.

Note that the storage requirements are really not as large as indicated since p_{k-2}, Ap_{k-2} are only needed in the singular case and then $p_{k-1} = r_{k-1} = r_k$, so extra storage is available.

4. A triangulation method. In this section we show that the nonsingular part of the scheme of conjugate residuals can be implemented by performing only a sequence of line searches (i.e., one-dimensional minimization problems) and residual evaluations, without explicitly computing inner products. Although this procedure is of little value for solving a linear equation, its obvious generalization to nonlinear equations is of substantial importance.

The method is more or less a straightforward extension of the method of parallel tangents (PARTAN) which implements the method of conjugate gradients when A is positive definite [7], [8]. The short derivation of the method given here is of some independent interest in view of the lengthy standard derivations. Our derivation exploits the results of Theorem 1.

Triangulation procedure. Given an arbitrary x_1 let $p_1 = r_1$ and determine x_2 as the point minimizing $E(x) = \|Ax - b\|^2$ along the line $x_1 + \alpha p_1$, $-\infty < \alpha < \infty$. After that the process is defined by induction. Having determined $x_1, x_2, \cdots, x_k$, put

$$y_k = x_k + \mu_k r_k,$$

where μ_k is determined to minimize $E(y_k)$. Then let

$$x_{k+1} = x_{k-1} + w_k[y_k - x_{k-1}],$$

where w_k is determined to minimize $E(x_{k+1})$.

We shall show that the triangulation procedure described above is equivalent to the method of conjugate residuals in the nonsingular case. If during the process

a residual r_k should become singular, the process breaks down and recovery must be made in some other fashion.

THEOREM 4. *The method of triangulation, in the nonsingular case, is equivalent to the method of conjugate residuals.*

Proof. The proof is by induction. The statement is obviously true for the first step. Suppose it is true for $x_1, x_2, \cdots, x_k$. For simplicity and without loss of generality we assume that $x_1 = \theta$. According to the method of conjugate residuals the point x_{k+1} is chosen so as to minimize the functional E over the subspace $[p_1, p_2, \cdots, p_{k-1}, r_k]$. Furthermore, by the orthogonality relations of the method this is equivalent to minimizing E over the subspace determined by p_{k-1} and r_k. This in turn is equivalent to finding a point x_{k+1} in that plane such that the residual there, $r_{k+1} = b - Ax_{k+1}$, is A-orthogonal to both r_k and p_{k-1}.

Consider the points x_{k-1} and y_k. The residuals at these points, $r_{k-1}, b - Ay_k$, are both A-orthogonal to r_k, the first by the induction hypothesis and part (d) of Theorem 1, the second because y_k is the minimizing point of E in the direction r_k. Therefore, by linearity, the residual at any point along the line through x_{k-1} and y_k is A-orthogonal to r_k.

It is therefore only necessary to find a point on this line where the residual is A-orthogonal to p_{k-1}. This clearly can be accomplished by minimizing E along the line.

5. Application to constrained minimization problems and nonlinear equations. Consider the nonlinear equation

$$F(x) = \theta, \tag{5.1}$$

where F is a mapping from E^n into E^n and θ is the zero vector in E^n. We restrict attention to the case where the matrix of first partial derivatives $F'(x)$ is symmetric.

As a primary motivating example consider the constrained minimization problem:

Minimize

$$f(u) \tag{5.2}$$

subject to

$$H(u) = \theta,$$

where f is a nonlinear functional on E^m and H is a nonlinear mapping from E^m to E^p, where $p \leqq m$. If u_0 is the solution of (5.2) and the rank of $H'(u_0)$ is equal to p, then, as is well known, there is a $\lambda_0 \in E^p$ such that (u_0, λ_0) is a solution of the simultaneous equations

$$\begin{aligned} f'(u) + \lambda^t H'(u) &= \theta, \\ H(u) &= \theta. \end{aligned} \tag{5.3}$$

The system (5.3) is of the type (5.1) with $n = m + p$.

The procedures developed here also have application to minimization problems involving inequality constraints, but this topic is deferred to another paper.

Solution of (5.1) is, of course, equivalent to minimizing the form $\psi(x) = \|F(x)\|^2$ and indeed this is the approach taken here. Standard methods of minimization,

however, require the direction vectors to be selected on the basis of the gradient. In the case of problem (5.2) this in turn would require knowledge of second derivatives of f and H. By employing the philosophy of the conjugate residual method, the descent directions are based on the residuals $F(x)$.

We shall briefly discuss three ways in which the conjugate residual scheme can be generalized to nonlinear problems. These are identical in spirit to the standard methods for adopting the conjugate gradient procedure to nonlinear problems and hence we shall not pursue the details.

A. If one is willing to assume the availability of $F'(x)$, then the local approximation $A \simeq F'(x)$ can be used in the formulas (2.3). The conjugate gradient analogue of this procedure has been developed by Daniel [9], [10].

B. Multiplication of a vector by the matrix A can be approximated by the difference of two residuals by the formula

$$F(x_k + y) - F(x_k) \simeq Ay$$

which is exact in the linear case (the step size can of course be adjusted for best results). Since, in the linear case, the conjugate residual scheme can be implemented with a single A multiplication, so in the nonlinear case it can be approximated by taking a single extra step. We give here the procedure corresponding to Case 1 of § 3, the other case being a direct analogue. Assume x_k, p_{k-1}, q_{k-1}, r_k are at hand, where p_{k+1} is the actual direction vector used to get x_k, q_{k-1} is an artificial vector used to approximate Ap_{k-1}, and $r_k = F(x_k)$. Then:

Calculate an approximation s_k to Ar_k by

$$s_k = F(x_k + r_k) - F(x_k). \tag{5.4a}$$

Calculate

$$\beta_k = \frac{(s_k, q_{k-1})}{(q_{k-1}, q_{k-1})}, \qquad p_k = r_k - \beta_k p_{k-1}. \tag{5.4b}$$

Calculate

$$q_k = s_k - \beta_k p_{k-1}, \qquad \alpha_k = \frac{(r_k, q_k)}{(q_k, q_k)}, \tag{5.4c}$$

$$x_{k+1} = x_k + \alpha_k p_k.$$

Calculate

$$r_{k+1} = F(x_{k+1}). \tag{5.4d}$$

A more refined procedure would be to calculate p_k from (5.4a) and (5.4b) and then search along the direction p_k for a point minimizing $\|F(x)\|^2$ initiating the search at the point predicted by (5.4c). For details on such line searching techniques and the analogue of this procedure for conjugate gradients, see [11].

C. The method of triangulation can be employed directly since it simply entails evaluation of the residual and line searching. There is a slight difficulty if a singular point is encountered. In this case a good procedure is probably to take a gradient step and then start again. As in B above the gradient can be approximated by $F(x_k + r_k) - F(x_k)$.

There are a few general remarks that apply to the application of any of the above schemes when applied to nonlinear equations. The first consideration is the necessity for occasionally restarting the process. The methods of conjugate gradients and conjugate residuals that we have discussed are not self-correcting. Specifically, if the initial direction vector should for some reason be in error, the usual conjugate properties of the subsequently generated direction vectors will not hold and the process may not terminate in a finite number of steps. This implies, in particular, that even if a nonlinear equation is exactly linear in a region containing the solution, any of the nonlinear versions of conjugate residuals (or conjugate gradients) may not converge in a finite number of steps. For this reason, it is good practice to restart these procedures every n steps or so. In view of this comment, the ploy of taking a gradient step and restarting when encountering a singular point does not seem unattractive.

The second comment relates to the detection of singular points. In a nonlinear problem, a singular direction (or a nearly singular direction) is characterized by a halt or near stoppage of the descent process. In other words, if a singular point is approached, the values of the objective functional $\|F(x)\|^2$ will converge to some positive value. However, since the desired value of this objective functional is known to be zero, such a singularity is easily identified and the singularity procedure or a gradient step can be employed to move away and continue the descent.

REFERENCES

[1] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 409–436.

[2] M. R. Hestenes, *The conjugate gradient method for solving linear systems*, Proc. Symposia in Appl. Math., vol. VI, Numerical Analysis, McGraw-Hill, New York, 1956, pp. 83–102.

[3] F. S. Beckman, *The solution of linear equations by the conjugate gradient method*, Mathematical Methods for Digital Computers, A. Ralston and H. S. Wilf, eds., John Wiley, New York, 1960.

[4] H. A. Antosiewicz and W. C. Rheinboldt, *Numerical analysis and functional analysis*, Survey of Numerical Analysis, J. Todd, ed., McGraw-Hill, New York, 1962, Chap. 14.

[5] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley, New York, 1969.

[6] ———, *Hyperbolic pairs in the method of conjugate gradients*, SIAM J. Appl. Math., 17 (1969), pp. 1263–1267.

[7] B. V. Shah, R. J. Buehler and O. Kempthorne, *Some algorithms for minimizing a function of several variables*, Ibid., 12 (1964), pp. 74–92.

[8] Phillip Wolfe, *Methods of nonlinear programming*, Nonlinear Programming, J. Abadie, ed., John Wiley, New York, 1967, pp. 97–131.

[9] J. W. Daniel, *The conjugate gradient method for linear and nonlinear operator equations*, this Journal, 4 (1967), pp. 10–26.

[10] ———, *Convergence of the conjugate gradient method with computationally convenient modifications*, Numer. Math., 10 (1967), pp. 125–131.

[11] R. Fletcher and C. M. Reeves, *Functional minimization by conjugate gradients*, Comput. J., 7 (1964), pp. 149–154.