ON ITERATIVE METHODS FOR SOLVING NONLINEAR LEAST SQUARES PROBLEMS OVER CONVEX SETS

BY
ADI BEN-ISRAEL

ABSTRACT

Nonlinear least squares problems over convex sets in R^n are treated here by iterative methods which extend the classical Newton, gradient and steepest descent methods and the methods studied recently by Pereyra and the author. Applications are given to nonlinear least squares problems under linear constraint, and to linear and nonlinear inequalities.

Introduction. Iterative methods for the solution of nonlinear least squares problems are extended here to problems over convex sets, and are applied in particular to linear and nonlinear inequalities and to nonlinear least squares problems with linear constraints.

Linearization plays a dual role in our approach: Convex sets are linearized in the sense that their proximity maps are considered as perpendicular projections on supporting hyperplanes and nonlinear functions are linearized, i.e. replaced by their linear approximations. Similar approaches were successfully used by Cheney and Goldstein [5], Goldstein [7], Bellman and Kalaba [1], Rosen [18], [19], Poljak [17] and many others.

In the problems considered below a certain closed convex set stands for what the origin stands in the classical problems in the sense that belonging to that set replaces vanishing, and the minimization of the distance from that set replaces solving least squares. Our methods are natural extensions of the classical methods of Newton, gradient and steepest descent, and of the methods recently studied by Pereyra [16] and the author [2], [3].

The paper has 4 sections. Notations and preliminaries are concentrated in Section 0. In Section 1 the problems and methods of solution are introduced and their relations to well-known methods are shown. Convergence theorems are stated in Section 2. Selected applications are given in Section 3.

Applications to Mathematical Programming and numerical experience will be given elsewhere.

Received March 21, 1967.

Part of the research underlying this report was undertaken for the Office of Naval Research, Contract Nonr-1228(10), Project NR047–021, and for the U.S. Army Research Office — Durham, Contract No. DA-31-124-ARO-D-322 at Northwestern University.

§0. Preliminaries and Notations

0.1 We denote by

R^n the *real Euclidean n-space* with *inner product*: $(x,y) = \sum_{i=1}^n x_i y_i$

norm: $\|x\| = (x,x)^{1/2}$,
distance between points x, y: $d(x,y) = \|x - y\|$,
distance between a point x *and a closed set* K: $d(x,K) = \inf\{d(x,y): y \in K\}$,
the closed sphere with center x *and radius* r: $S(x,r) = \{y: d(x,y) \leqq r\}$.

The space of linear transformations from R^n *into* R^m is denoted by $L(R^n, R^m)$.

0.2 For an $m \times n$ real matrix A we denote by

A^T the *transpose*
$\|A\|$ the *spectral norm*, e.g. [10]
$R(A)$ the *range space*
$N(A)$ the *null space*
A^+ the *generalized inverse*, e.g. [15], [4].

0.3 A function $f: R^n \to R^m$ is *differentiable at* $x_0 \in R^n$ if there is a $f'(x_0) \in L(R^n, R^m)$ such that

$$\lim_{k \to 0} \frac{\|f(x+h) - f(x) - f'(x_0)h\|}{\|h\|} = 0$$

$f'(x_0)$ is the *derivative of* f *at* x_0, and is represented by the *Jacobian* matrix

$$f'(x_0) = \left(\frac{\partial f_i}{\partial x_j}(x_0)\right), \qquad \begin{matrix} i = 1, \cdots, m \\ j = 1, \cdots, n\,. \end{matrix}$$

The *gradient*, $\nabla f(x_0)$, of a function $f: R^n \to R$ at a point $x_0 \in R^n$ is $f'(x_0)^T$. If $\mho$ is an open set in R^n and the mapping: $x \to f'(x)$ of R^n into $L(R^n, R^m)$ is defined and continuous for all $x \in \mho$ we say that f is *continuously differentiable in* $\mho$ and write $f \in C'(\mho)$.

0.4 With a closed convex set K in R^n we associate the *proximity map* $P_K: R^n \to R^n$ defined by: $P_K(x) \in K$ and $d(x, P_K(x)) = d(x,K)$ for all $x \in R^n$.
For properties, applications and further references on proximity maps see [6], [13] and [20].

We write $P_{K^\perp}(x)$ for $x - P_K(x)$, and recall the following properties:

(i) $P_{K^\perp}(x) = 0$ if, and only if $x \in K$,

(ii) If K is a linear subspace in R^n then P_K is the perpendicular projection on K, and $P_{K^\perp}$ is the perpendicular projection on $K^\perp$, the orthogonal complement of K. In this case we write $P_K x$ for $P_K(x)$.

(iii) If K is a closed convex cone in R^n (with vertex at the origin) then $P_{K^\perp}$ is the proximity map of the cone $K^\perp = \{y: x \in K \Rightarrow (y,x) \leqq 0\}$, e.g. [13].

(iv) If $x \notin K$ then $P_K(x)$ is a boundary point of K and $P_{K^\perp}(x)$ is the (outward pointing) normal to a supporting hyperplane of K at $P_K(x)$, e.g. [6] p. 448 lemma.

§1. Problems and Methods

1.1. *Problems*

Let $f: R^n \rightarrow R^m$, K a closed convex set in R^m, $\mho$ an open subset of R^n and let $d(f(x), K)$ be differentiable in $\mho$.

We consider the following:

Problem 1: *Find an* $x \in \mho$ *for which*

$$\nabla d^2(f(x), K) = 0 \tag{1}$$

The significance of problem 1 is that if solvable, its solutions may include the solutions (provided they exist) of the following problems:

Problem 2: *Find an* $x \in R^n$ *for which*

$$d(f(x), K) \text{ is minimized} \tag{2}$$

Problem 3: *Find an* $x \in R^n$ *such that*

$$f(x) \in K \tag{3}$$

We note that problem 3 includes:

Problem 4: *Find an* $x \in R^n$ *which satisfies*
(3) *and in addition*

$$x \in L \tag{4}$$

where L is a closed convex set in R^n.

Indeed problem 4 is rewritten as:

Find an $x \in R^n$ *which satisfies*

$$g(x) \in M \tag{5}$$

where $g: R^n \rightarrow R^m \times R^n$ *is*

$$g(x) = \begin{pmatrix} f(x) \\ x \end{pmatrix} \tag{6}$$

and

$$M = K \times L, \text{ a convex set in } R^m \times R^n. \tag{7}$$

1.2. *Methods.* For the solution of Problem 1 and the treatment of Problems 2 and 3 we propose the iterative methods

$$x_{\nu+1} = x_\nu - f'(x_\nu)^+ P_{K^\perp}(f(x_\nu)), \qquad \nu = 0, 1, \cdots \tag{8}$$

and

$$x_{\nu+1} = x_\nu - (T_\nu^{-1}) f'(x_\nu)^T P_{K^\perp}(f(x_\nu)), \qquad \nu = 0, 1, \cdots \tag{9}$$

where x_0 is an approximate solution, and $\{T_\nu\}$ is a suitably chosen sequence of nonsingular operators in $L(R^n, R^n)$. In treating Problem 3 it is always possible, and often desirable to replace method (8) by method (9) with positive definite T_ν. This is done by considering an equivalent (artificial) Problem 4 in which the convex set L is taken sufficiently large to include all points of interest, i.e., all relevant x satisfy

$$\alpha x \in L \tag{10}$$

where α is a fixed positive number, or equivalently

$$P_{L^\perp}(\alpha x) = 0. \tag{11}$$

For this artificial problem method (8) becomes

$$x_{\nu+1} = x_\nu - g'(x_\nu)^+ P_{M^\perp}(g(x_\nu)), \qquad \nu = 0, 1, \cdots \tag{12}$$

where

$$g(x) = \begin{pmatrix} f(x) \\ \alpha x \end{pmatrix} \tag{13}$$

$$g'(x) = \begin{pmatrix} f'(x) \\ \alpha I \end{pmatrix} \tag{14}$$

$$g'(x)^+ = (f'(x)^T f'(x) + \alpha^2 I)^{-1}(f'(x)^T, \alpha I) \tag{15}$$

(7) $M = K \times L$ and accordingly

$$P_{M^\perp}(g(x)) = \begin{pmatrix} P_{K^\perp}(f(x)) \\ P_{L^\perp}(\alpha x) \end{pmatrix} \tag{16}$$

Substituting (15), (16) and (11) in (12) we get

$$x_{\nu+1} = x_\nu - (f'(x_\nu)^T f'(x_\nu) + \alpha^2 I)^{-1} f'(x_\nu)^T P_{K^\perp}(f(x_\nu)), \qquad \nu = 0, 1, \cdots \tag{17}$$

which is method (9) with

$$T_\nu = f'(x_\nu)^T f'(x_\nu) + \alpha^2 I, \qquad \nu = 0, 1, \cdots \tag{18}$$

from which the (fictitious) set L has justly disappeared.

Comparing (8) with (17) we see that the latter uses the inverse of the positive definite matrix (18), whereas the former uses the generalized inverse of the arbitrary matrix $f'(x_\nu)$. Another advantage of (17) is that it combines characteristics of both Newton and Gradient methods, e.g. the discussion in Marquardt [11] for the special case $K = \{0\}$.

1.3 *Relations with known methods*. In the special case $K = \{0\}$ we have

$$P_{K^\perp}(y) = y, \qquad \text{for all } y. \tag{19}$$

For $K = \{0\}$ method (8) reduces therefore to the author's *variant of Newton's method*, [2]

(20) $$x_{\nu+1} = x_\nu - f'(x_\nu)^+ f(x_\nu), \qquad \nu = 0, 1, \cdots$$

which for nonsingular $f'(x_\nu)$ is the *Newton method* [9], [1]

(21) $$x_{\nu+1} = x_\nu - f'(x_\nu)^{-1} f(x_\nu), \qquad \nu = 0, 1, \cdots.$$

Method (9) reduces for $K = \{0\}$ to *Pereyra's method* [16]

(22) $$x_{\nu+1} = x_\nu - (T_\nu^{-1}) f'(x_\nu)^T f(x_\nu), \quad \nu = 0, 1, \cdots$$

which includes:

(i) The *Gauss-Newton method* with

(23) $$T_\nu = f'(x_\nu)^T f'(x_\nu), \qquad \nu = 0, 1, \cdots$$

(ii) The *modified Newton method* with

(24) $$T_\nu = f'(x_\nu)^T f'(x_0), \qquad \nu = 0, 1, \cdots$$

(iii) *Gradient methods* since

(25) $$f'(x)^T f(x) = \frac{1}{2}\nabla(f(x), f(x))$$ and in particular the

(iv) *Steepest descent method* with scalar T_ν.

Method (17) reduces for $K = \{0\}$ to *Marquardt's method* [11]

(26) $$x_{\nu+1} = x_\nu - (f'(x_\nu)^T f'(x_\nu) + \alpha^2 I)^{-1} f'(x_\nu)^T f(x_\nu), \qquad \nu = 0, 1, \cdots,$$

see also Morrison [14], and Meeter [12].

The *least-squares method* of [3]

(27) $$x_{\nu+1} = x_\nu - (f'(x_\nu)^T f'(x_\nu) + \alpha^2 I)^{-1}(f'(x_\nu)^T f(x_\nu) + \alpha^2(x_\nu - u)), \qquad \nu = 0, 1, \cdots$$

is obtained for $K = \{0\}$, $L = \{0\}$, from (12) by taking

(28) $$g(x) = \begin{pmatrix} f(x) \\ \alpha(x - u) \end{pmatrix}.$$

§2. Convergence Theorems

2.1 *Method* (8). Sufficient conditions for the convergence of method (8), analogous to those of [2] in the special case $K = \{0\}$, are given in:

THEOREM 1. *Assumptions*: $f: R^n \to R^m$, x_0 *a point in* R^n, K *a closed convex set in* R^m *and* r, M, N *positive constants such that*:

(29) $f \in C'(\mathcal{V})$ *where* $\mathcal{V}$ is an open set *containing* $S(x_0, r)$

For all $u, v \in S(x_0, r)$ *with* $u - v \in R(f'(v)^T)$:

$$\|f'(v)(u-v) - P_{K^\perp}(f(u)) + P_{K^\perp}(f(v))\| \leqq M\|u-v\| \tag{30}$$

and

$$\|(f'(v)^+ - f'(u)^+)P_{K^\perp}(f(u))\| \leqq N\|u-v\| \tag{31}$$

For all $x \in S(x_0, r)$

$$M\|f'(x)^+\| + N = k < 1 \tag{32}$$

$$\|f'(x_0)^+ P_{K^\perp}(f(x_0))\| < (1-k)r \tag{33}$$

Conclusions: The sequence (8) *converges to a solution* x^* *of* (1) *which lies in* $S(x_0, r)$ *and is unique in* $S(x_0, r) \cap \{x^* + R(f'(x^*)^T)\}$.

Proof. The proof follows closely that of theorems 1, 2 of [2].

We prove first that:

$$x_\nu \in S(x_0, r) \qquad \nu = 1, 2, \cdots \tag{34}$$

For $\nu = 0, 1, \cdots$ we write

$$\begin{aligned} (35)\ x_{\nu+1} - x_\nu &= -f'(x_\nu)^+ P_{K^\perp}(f(x_\nu)) \\ &= x_\nu - x_{\nu-1} - f'(x_\nu)^+ P_{K^\perp}(f(x_\nu)) + f'(x_{\nu-1})^+ P_{K^\perp}(f(x_{\nu-1})) \\ &= f'(x_{\nu-1})^+[f'(x_{\nu-1})(x_\nu - x_{\nu-1}) - P_{K^\perp}(f(x_\nu)) + P_{K^\perp}(f(x_{\nu-1}))] \\ &\quad + [f'(x_{\nu-1})^+ - f'(x_\nu)^+]P_{K^\perp}(f(x_\nu)). \end{aligned}$$

where we used

$$x_\nu - x_{\nu-1} = f'(x_{\nu-1})^+ f'(x_{\nu-1})(x_\nu - x_{\nu-1}) \tag{36}$$

which follows from

$$x_\nu - x_{\nu-1} \in R(f'(x_{\nu-1})^+) = R(f'(x_{\nu-1})^T), \text{ e.g. [4].} \tag{37}$$

From (35), (30), (31) and (32) we get

$$\begin{aligned} \|x_{\nu+1} - x_\nu\| &\leqq (M\|f'(x_{\nu-1})^+\| + N)\|x_\nu - x_{\nu-1}\| \\ &= k\|x_\nu - x_{\nu-1}\| < \|x_\nu - x_{\nu-1}\| \end{aligned} \tag{38}$$

Now, by (33) we get

$$\|x_1 - x_0\| < (1-k)r \tag{39}$$

so that (34) holds for $\nu = 1$ and from (38)

$$(40)\qquad \| x_{\nu+1} - x_0 \| \leqq \sum_{j=1}^{\nu} k^j \| x_1 - x_0 \| = \frac{k(1-k^\nu)}{(1-k)} \| x_1 - x_0 \|$$

which by (39) proves (34).

The convergence of (8) to a point $x^* \in S(x_0, r)$ follows now from (38).

Using (8) we note that the limit x^* must satisfy

$$(41)\qquad f'(x^*)^+ P_{K^\perp}(f(x^*)) = 0$$

and from

$$(42)\qquad N(A^+) = N(A^T),$$

e.g. [4], we get

$$(43)\qquad f'(x^*)^T P_{K^\perp}(f(x^*)) = 0$$

Let S be a hyperplane whose translate $P_K(f(x^*)) + S$ supports K at $P_K(f(x^*))$ where its normal is $P_{K^\perp} f(x^*)$. Since $P_{K^\perp}(f(x^*))$ and $P_{S^\perp} f(x^*)$ are collinear

$$(44)\qquad P_{K^\perp}(f(x^*)) = t P_{S^\perp} f(x^*), \text{ for some real } t$$

and since $P_{S^\perp}$ is idempotent and symmetric, we rewrite (43) as

$$(45)\quad f'(x^*)^T P_{S^\perp} f(x^*) = f'(x^*)^T P_{S^\perp} P_{S^\perp} f(x^*) = [(P_{S^\perp} f(x^*))']^T (P_{S^\perp} f(x^*)) = 0$$

As in (25) we note that

$$(46)\qquad [(P_{S^\perp} f(x))']^T (P_{S^\perp} f(x)) = \frac{1}{2} \nabla (P_{S^\perp} f(x), P_{S^\perp} f(x))$$

Combining now (43), (44), (45) and (46) we verify that the limit x^* of (8) is a solution of (1).

The claimed uniqueness of x^* holds because for any other limit x^{**} of (8) in $S(x_0, r) \cap \{x^* + R(f'(x^*)^T\}$ we have

$$
\begin{aligned}
(47)\quad \| x^{**} - x^* \| &= \| x^{**} - x^* - f'(x^{**})^+ P_{K^\perp}(f(x^{**})) + f'(x^*)^+ P_{K^\perp}(f(x^*)) \| \\
&\leqq \| f'(x^*)^+ f'(x^*)(x^{**} - x^*) - f'(x^*)^+ P_{K^\perp}(f(x^{**})) + f'(x^*)^+ P_{K^\perp}(f(x^*)) \| \\
&\quad + \| (f'(x^*)^+ - f'(x^{**})^+) P_{K^\perp}(f(x^{**})) \| \\
&\leqq (M \| f'(x^*)^+ \| + N) \| x^{**} - x^* \|,
\end{aligned}
$$

by (30), (31)

$$< \| x^{**} - x^* \|,$$

by (32), a contradiction. *Q.E.D.*

2.2 *Method* (9). We use the notations

$$(48)\qquad \phi(x) = f'(x)^T f(x)$$

$$(49)\qquad \Psi(x) = f'(x)^T P_{K^\perp}(f(x))$$

in adopting Pereyra's theorem 2.1 of [16] to give sufficient conditions for the convergence of (9).

THEOREM 2. *Assumptions*: $f\colon R^n \to R^m$, x_0 *a point in* R^n, K *a closed convex set in* R^m, $\{T_\nu : \nu = 0, 1, \cdots\}$ *a sequence of nonsingular operators in* $L(R^n, R^n)$, $\rho, \lambda, \varepsilon, \beta, \eta, k, r$ *positive numbers such that*

(50) *f is twice differentiable in an open subset of R^n which contains $S(x_0, \rho)$*

(51) $$\| T_\nu^{-1} \| \leqq \lambda \qquad , \qquad \nu = 0, 1, \cdots$$

(52) $$\| T_\nu - \phi'(x_0) \| \leqq \varepsilon, \qquad , \qquad \nu = 0, 1, \cdots$$

(53) $$\| \Psi(u) - \Psi(v) - \phi'(x_0)(u - v) \| \leqq \beta \| u - v \| \quad \textit{for all} \quad u, \; v \in S(x_0, r)$$

(54) $$\| \Psi(x_0) \| \leqq \eta$$

(55) $$k = \lambda(\beta + \varepsilon) < 1$$

(56) $$r = \frac{\lambda \eta}{1 - k} \leqq \rho$$

Conclusions: *The sequence* (9) *converges to a solution x^* of* (1) *which lies in $S(x_0, r)$ and is unique in $S(x_0, r)$*,

Proof. The method (9) is rewritten as

(57) $$x_{\nu+1} = x_\nu - T_\nu^{-1} \Psi(x_\nu), \qquad \nu = 0, 1, \cdots$$

From (51), (54), (55) and (56) we see that

(58) $$\| x_1 - x_0 \| \leqq \lambda \| \Psi(x_0) \| \leqq \lambda \eta < r \leqq \rho$$

and using (57), (53) and (52) we verify that

(59) $$\begin{aligned} \| \Psi(x_1) \| &= \| \Psi(x_1) - \Psi(x_0) - T_0(x_1 - x_0) \| \\ &\leqq \| \Psi(x_1) - \Psi(x_0) - \phi'(x_0)(x_1 - x_0) \| + \| (\phi'(x_0) - T_0)(x_1 - x_0) \| \\ &\leqq (\beta + \varepsilon) \| x_1 - x_0 \| \end{aligned}$$

We prove now, by induction, that the relations

(60. ν) $$\| x_\nu - x_0 \| \leqq r$$

(61. ν) $$\| x_\nu - x_{\nu-1} \| \leqq \lambda \| \Psi(x_{\nu-1}) \|$$

(62. ν) $$\| \Psi(x_\nu) \| \leqq (\beta + \varepsilon) \| x_\nu - x_{\nu-1} \|$$

hold for all $\nu = 1, 2, \cdots$. Suppose indeed that (60. ν), (61. ν) and (62. ν) hold for $\nu = 1, 2, \cdots, p$ and we will show them to hold for $\nu = p + 1$:

(61. $p+1$) $$\| x_{p+1} - x_p \| \leqq \lambda \| \Psi(x_p) \|, \qquad \text{by (51)}$$

Using (61. $p+1$) and (62. ν), $\nu = 1, \cdots, p$, we get

$$\| x_{p+1} - x_p \| \leqq \lambda(\beta + \varepsilon) \| x_p - x_{p-1} \| \leqq \cdots \leqq k^p \| x_1 - x_0 \| \tag{63}$$

so that

$$\| x_{p+1} - x_0 \| \leqq \sum_{\nu=0}^{p} \| x_{\nu+1} - x_\nu \| \leqq \sum_{\nu=0}^{p} k^\nu \| x_1 - x_0 \| \leqq \frac{\lambda\eta}{1-k} = r \tag{60. $p+1$}$$

And finally, using (57), (53) and (52):

$$\begin{aligned} \| \Psi(x_{p+1}) \| &= \| \Psi(x_{p+1}) - \Psi(x_p) - T_p(x_{p+1} - x_p) \| \\ &\leqq \| \Psi(x_{p+1}) - \Psi(x_p) - \phi'(x_0)(x_{p+1} - x_p) \| \\ &\quad + \| (\phi'(x_0) - T_p)(x_{p+1} - x_p) \| \\ &\leqq (\beta + \varepsilon) \| x_{p+1} - x_p \| \end{aligned} \tag{62. $p+1$}$$

The convergence of (57) to a point x^* in $S(x_0, r)$ is now guaranteed. At the point x^* we have

$$\Psi(x^*) = f'(x^*)^T P_K^{\perp}(f(x^*)) = 0 \tag{64}$$

and reasoning as in the proof of Theorem 1 we verify that x^* is a solution of (1).

The uniqueness of x^* follows from the fact that for any other solution x^{**} of $\Psi(x) = 0$ in $S(x_0, r)$ we have

$$\| x^{**} - x^* \| = \| T_0^{-1} T_0(x^{**} - x^*) \| \tag{65}$$

by (51)

$$\leqq \lambda \| \Psi(x^{**}) - \Psi(x^*) - \phi'(x_0)(x^{**} - x^*) \| + \lambda \| (T_0 - \phi'(x_0))(x^{**} - x^*) \|$$

by (52) and (53)

$$\leqq \lambda(\beta + \varepsilon) \| x^{**} - x^* \|,$$

by (55),

$$< \| x^{**} - x^* \|,$$

a contradiction. Q.E.D.

2.3 *Error bounds.* If the conditions of Theorem 1 are satisfied, then an error bound for (8) is given by

$$\| x_\nu - x^* \| \leqq k^\nu r \tag{66}$$

where k is given by (32).

Indeed, from (38) and (39) we get

(67)
$$\|x_{\nu+p}-x_\nu\| \leqq \sum_{i=1}^{p} \|x_{\nu+i}-x_{\nu+i-1}\|$$
$$\leqq k^\nu \sum_{i=0}^{p-1} k^i \|x_1-x_0\| < \frac{k^\nu}{1-k}\|x_1-x_0\| < k^\nu r.$$

Similarly, (66) is an error bound for the method (9) with k given by (55), provided the conditions of Theorem 2 are satisfied.

§3. Applications

3.1 *Proximity maps: special cases.* The promise of methods (8) and (9) lies in problems where the proximity maps are readily available. Two such cases are when the convex set considered is (i) a *linear manifold*, or (ii) an *orthant* in R^n. The proximity maps in these cases are given below.

(i) *Linear manifold.* If L is a linear manifold in R^n it can be represented as

(68)
$$L = \{x : Ax = b\}$$

for some m, $A \in L(R^n, R^m)$, $b \in R^m$.

Equivalently we write

(69)
$$L = A^+b + N(A)$$

and the proximity map is

(70)
$$P_L(x) = A^+b + P_{N(A)}x, \text{ for all } x \in R^n, \text{ from which}$$

(71)
$$P_{L^\perp}(x) = P_{R(A^T)}x - A^+b.$$

(ii) *Orthant.* If K is the nonpositive orthant R^n_- of R^n, i.e.

(72)
$$K = \{x : x \leqq 0 \text{ i.e. } x_i \leqq 0 \text{ for } i = 1, \cdots, n\}$$

then for any $x \in R^n$, written after a rearrangement as

(73)
$$x = \begin{pmatrix} x_+ \\ \hline x_- \end{pmatrix}$$

where the subvector x_+ is positive and the subvector x_- is nonpositive, we have

(74)
$$P_K(x) = \begin{pmatrix} 0 \\ \hline x_- \end{pmatrix}$$

(75)
$$P_{K^\perp}(x) = \begin{pmatrix} x_+ \\ \hline 0 \end{pmatrix}$$

3.2 *Nonlinear Least-squares problems with linear constraints.*

Let $f: R^n \to R^m$, $A \in L(R^n, R^m)$, $b \in R^m$.

Consider:

Problem 5: *Find an* $x \in R^n$ *which minimizes*

$$(f(x), f(x)) = \sum_{i=1}^{m} f_i^2(x) \tag{76}$$

subject to

$$Ax = b. \tag{77}$$

We will instead consider problem 4 of §1 rewritten as:

Problem 6: *Find an* $x \in R^n$ *for which*

$$g(x) = \begin{pmatrix} f(x) \\ x \end{pmatrix} \in M = K \times L \tag{78}$$

where

$$K = \{0\} \tag{79}$$

and L is given by

$$L = \{x : Ax = b\} \tag{68}$$

This problem is stricter than Problem 5 because in addition to (77) we require here that

$$f(x) = 0. \tag{80}$$

Applying method (12) to problem 6 we get, using (13)–(16) with $\alpha = 1$ together with (19) and (71):

$$x_{\nu+1} = x_\nu - (f'(x_\nu)^T f'(x_\nu) + I)^{-1}[f'(x_\nu)\, f(x_\nu) + P_{R(A^T)} x_\nu - A^+ b], \quad \nu = 0, 1, \cdots. \tag{81}$$

Indeed, if (81) converges to a point x^* then at x^*

$$-f'(x^*)^T f(x^*) = P_{R(A^T)} x^* - A^+ b \tag{82}$$

and by (25), (71) we conclude that the gradient $\nabla(f(x^*), f(x^*))$ is perpendicular to the linear manifold L, which if $x^* \in L$ is the classical necessary condition for the minimization of (76) subject to (77). Conversely, we conclude from (82) that if the limit x^* is a stationary point for (76) then, from the vanishing of the right side of (82), x^* satisfies (77).

Finally, in order to keep the sequence (81) "closer" to L, we can replace a subsequence of (81) by

(83) $$x_{\nu_k+1} = P_L(x_{\nu_k}) = A^+b + P_{N(A)}x_{\nu_k}, \qquad k = 1\ 2, \cdots.$$

3.3 *Linear inequalities.*

Let $C \in L(R^n, R^k)$, $d \in R^k$ and consider

Problem 7: *Find an* $x \in R^n$ *such that*

(84) $$Cx - d \leqq 0$$

This is Problem 3 with

(85) $$f(x) = Cx - d$$

and K, the nonpositive orthant in R^k, given as in (72).

In applying our methods to this problem we encounter terms like

(86) $$BP_{K^\perp}(y) \quad \text{where } B \in L(R^k, R^p),\ y \in R^k.$$

Rearranging y as in (73) and likewise the columns of B

(87) $$B = (B_+, B_-)$$
we get from (75)

(88) $$BP_{K^\perp}(y) = (B_+, B_-)P_{K^\perp}\begin{pmatrix} y_+ \\ y_- \end{pmatrix} = B_+y_+$$

Method (8) now becomes

(89) $$x_{\nu+1} = x_\nu - (C^+)_+(Cx_\nu - d)_+ \qquad \nu = 0, 1, \cdots$$

where the subvector $(Cx_\nu - d)_+$ consists of the positive components of $Cx_\nu - d$ and the matrix $(C^+)_+$ consists of the corresponding columns of C^+.

Similarly, method (9) gives

(90) $$x_{\nu+1} = x_\nu - T_\nu^{-1}(C^T)_+(Cx_\nu - d)_+, \qquad \nu = 0, 1, \cdots$$

and, in particular, from (17)

(91) $$x_{\nu+1} = x_\nu - (C^TC + \alpha^2 I)^{-1}(C^T)_+(Cx_\nu - d)_+, \qquad \nu = 0, 1, \cdots$$

3.4 *Linear equations and inequalities.* Let $A \in L(R^n, R^m)$, $b \in R^m$ $C \in L(R^n, R^k)$, $d \in R^k$

and consider:

Problem 8: *Find an* $x \in R^n$ *such that* (77) $Ax - b = 0$, and (84) $CX - d \leqq 0$.

This is problem 4 with f as in (85), K as in (72), and, if (77) is solvable

(92) $$L = \{x : Ax = b\} = A^+b + N(A)$$

Using (12)–(16) with $\alpha = 1$ we get from (71) and (88):

$$x_{\nu+1} = x_\nu - (C^TC + I)^{-1}[(C^T)_+(Cx_\nu - d)_+ + P_{R(A^T)}x_\nu - A^+b], \quad \nu = 0, 2 \cdots. \tag{93}$$

Alternatively, one could treat (77) and (84) separately, by alternating the iterations (83) and (91).

In the case where $b = 0$, $d = 0$, $A = (B, -I)$, $C = (0, -I)$ and strict inequality in (84) problem 8 was similarly solved by Ho and Kashyap, [8].

3.5 *Nonlinear inequalities.* Let $f\colon R^n \to R^m$ and consider:

Problem 9: *Find an* $x \in R^n$ *for which*

$$f(x) \leqq 0 \tag{94}$$

This is problem 3 with K given by (72). Methods (8) (9) and (17) give respectively:

$$x_{\nu+1} = x_\nu - (f'(x_\nu)^+)_+ f(x_\nu)_+ \qquad \nu = 0, 1, \cdots \tag{95}$$

$$x_{\nu+1} = x_\nu - T_\nu^{-1}(f'(x_\nu)^T)_+ f(x_\nu)_+ \qquad \nu = 0, 1, \cdots \tag{96}$$

$$x_{\nu+1} = x_\nu - (f'(x_\nu)^T f'(x_\nu) + \alpha^2 I)^{-1}(f'(x_\nu)^T)_+ f(x_\nu)_+, \quad \nu = 0, 1, \cdots. \tag{97}$$

If any of the above methods converges then its limit x^* satisfies

$$(f'(x^*)^T)_+ f(x^*)_+ = 0 \tag{98}$$

so that x^* is a stationary point for the partial sum of squares

$$\sum_{\text{over } i} f_i^2(x) \quad \text{for which } f_i(x^*) > 0$$

which thus depends on x^*.

References

1. R. Bellman and R. Kalaba, *Quasilinearization and Nonlinear Boundary-Value Problems*, American Elsevier, New York, 1965.

2. A. Ben-Israel, A Newton Raphson Method for the Solution of Systems of Equations, *J. Math. Anal. Appl.* **15** (1966), 243–252.

3. A. Ben-Israel, *On the Newton-Raphson Method*, Systems Research Memorandum No. 162, Northwestern University, Evanston, September, 1966.

4. A. Ben-Israel and A. Charnes, Contributions to the Theory of Generalized Inverses, *J. Soc. Indust. Appl. Math.* **11** (1963), 667–699.

5. E. W. Cheney and A. A. Goldstein, Newton's Method for Convex Programming and Tchebycheff Approximation, *Numerische Mathematik*, **1** (1959), 253–268.

6. E. W. Cheney and A. A. Goldstein, Proximity Maps for Convex Sets, *Proc. Amer. Math. Soc.*, **10** (1959), 448–450.

7. A. A. Goldstein, *Minimizing Functionals on Hilbert Space*, pp. 159–165 in A. V. Balakrishnan and L. W. Neustadt (ed.) Computing Methods in Optimization Problems, *Academic Press*, New York, 1964.

8. Y-C. Ho and R. L. Kashyap, A Class of Iterative Procedures for Linear Inequalities, *J. SIAM Control* **4** (1966), 112–115.

9. A. S. Householder, *Principles of Numerical Analysis*, McGraw-Will, New York, 1953.

10. A. S. Householder, *Theory of Matrices in Numerical Analysis*, Blaisdell, New York, 1964.

11. D. W. Marquardt, An Algorithm for Least Squares Estimation of Nonlinear Parameters, *J. Soc. Indust. Appl. Math.* **11** (1963), 431–441.

12. D. A. Meeter, On a Theorem Used in Nonlinear Least Squares, *J. SIAM Appl. Math.* **14** (1966), 1176–1179.

13. J. J. Moreau, Proximité et Dualité dans un Espace Hilbertien, *Bull. Soc. Math. France*, **93** (1965), 273–299.

14. D. D. Morrison, *Methods for Nonlinear Least Squares Problems and Convergence Proofs*, Proc. Jet Propulsion Lab. Seminar: Tracking Problems and Orbit Determination, 1–9, 1960.

15. R. Penrose, A Generalized Inverse for Matrices, *Proc. Cambridge Philos. Soc.* **51** (1955), 406–413.

16. V. Pereyra, Iterative Methods for Solving Nonlinear Least Squares Problems, *SIAM J. Numer. Anal.* **4** (1967), 27–36.

17. B. T. Poljak, Gradient Methods for Solving Equations and Inequalities, *Z. Vycisl. Mat i Mat. Fiz.* **4** (1964), 995–1005.

18. J. B. Rosen, *Optimal Control and Convex Programming*, pp. 223–237 in Proceedings of the IBM Sc entific Computing Symposium on Control Theory and Applications, T. J. Watson Research Center, Yorktown Heights, N. Y., October, 1964.

19. J. B. Rosen, Iterative Solution of Nonlinear Optimal Control Problems, *J. SIAM Control* **4** (1966), pp. 223–544.

20. W. J. Stiles, Closest-Point Maps and Their Products, *Niew Archief voor Wiskunde* (3) **XIII** (1965), pp. 19–29 and 212–225.

NORTHWESTERN UNIVERSITY,
EVANSTON, ILLINOIS