Mathematical Programming 7 (1974) 384–387. North-Holland Publishing Company

SHORT COMMUNICATION

NOTE ON AN EXTENSION OF "DAVIDON" METHODS TO NONDIFFERENTIABLE FUNCTIONS

Claude LEMARECHAL
IRIA–Laboria, Le Chesnay, France

Received 3 June 1974
Revised manuscript received 2 August 1974

Some properties of "Davidon", or variable metric, methods are studied from the viewpoint of convex analysis; they depend on the convexity of the function to be minimized rather than on its being approximately quadratic. An algorithm is presented which generalizes the variable metric method, and its convergence is shown for a large class of convex functions.

1. This note summarizes a paper [4] to appear in full elsewhere. It presents an algorithm for the minimization of a general (not necessarily differentiable) convex function. Its central idea is the construction of descent directions as projections of the origin onto the convex hull of previously calculated subgradients as long as satisfactory progress can be made.

Using projection to obtain a direction of movement for the conjugate gradient method is not new [5,8], but previous methods have projected onto the affine manifold spanned by the gradients, rather than onto the convex hull. Alternatively, Demjanov [2] has employed a projection onto the convex hull of a *local* set of subgradients to obtain a direction of steepest descent. The present algorithm may be said to combine these ideas. It is also connected with proposals of Bertsekas and Mitter [1] (by Theorem 2 below), and has many points in common with a method due to Wolfe [7].

2. Let H denote Euclidean n-space, A be an n-by-n matrix, and $b \in H$. Let $f(x) = \frac{1}{2}(Ax, x) - (b, x)$ and $g(x) = Ax - b$. We study minimizing sequences of the form $x_{n+1} = x_n + \rho_n s_n$, where

$$\begin{aligned} (s_n, g_{n+1}) &= 0 \quad \text{for } n \geqslant 0\,, \\ (As_i, s_j) &= 0 \quad \text{for } i \neq j\,. \end{aligned} \tag{1}$$

It is well known that (1) implies $(g_n, s_j) = 0$ for $j < n$. We will suppose further that

$$s_n = \sum_{i=0}^{n} \lambda_i^n g_i \qquad \text{with } \lambda_i^n \in \quad , \quad \lambda_n^n \neq 0 \, . \tag{2}$$

Theorem 1. *If a sequence satisfies* (1), (2), *then for each n, s_n is parallel to the projection of the origin onto the convex hull of $g_0, ..., g_n$.*

Since the projection is unique and f is minimized on the ray $x_n + \rho s_n$, the sequence of directions s_n depends only on x_0, and is the same for any procedure for which (1), (2) hold.

3. When f is not quadratic, the definition of s_n as a projection onto the convex hull of the preceding gradients offers itself as a natural replacement for the usual formulas of the variable metric method. For a convex function f on H, we recall the ϵ-subdifferential of Rockafellar [6, p. 219],

$$\partial_\epsilon f(x) = \{g \in H: f(y) \geqslant f(x) + (g, y-x) - \epsilon \text{ for all } y \in H\},$$

and the formula

$$\sup \{(s, g): g \in \partial_\epsilon f(x)\} = \inf_{\rho > 0} \, [f(x+\rho s) - f(x) + \epsilon]/\rho \, . \tag{3}$$

Theorem 2. *If f is convex and quadratic, then for a sequence in which* (1.2) *hold, $g_n \in \partial_{\epsilon_n}(x_0)$ for all n, where $\epsilon_n = f(x_0) - f(x_n)$.*

(The proof is almost immediate, using the fact that $(g_n, x_n - x_0) \leqslant 0$ for all n.)

Now if the conclusion of Theorem 2 held for some nonquadratic convex function f, we could show that $f(x_n)$ tended to the minimum value of f: since ϵ_n is an increasing sequence, $g_m \in \partial_{\epsilon_n} f(x_0)$ for $m \leqslant n$, so $s_n \in \partial_{\epsilon_n} f(x_0)$. The limiting values $f(x_n) \to f^*$, $\epsilon_n \to \epsilon^* = f(x_0) - f^*$ exist, and we can show that $s_n \to 0$, whence $0 \in \partial_{\epsilon^*} f(x_0)$. The left-hand side of (3) is thus nonnegative for any s when $x = x_0$; we conclude for any $y \in H$ that $f(y) \geqslant f(x_0) - \epsilon^* = f^*$, which is then the minimum value of f. This observation motivates our algorithm.

4. Let f be a convex function on H, let $x_0 \in H$ and $g_0 \in \partial f(x_0)$ be given, and let $\alpha, \beta, \epsilon, \eta$ be positive numbers with $\alpha, \beta < 1$. Set $n = p = 0$.

Step 1. Determine s_n as the negative of the projection of 0 onto the convex hull of $g_p, g_{p+1}, \ldots, g_n$. If $|s_n| \leqslant \eta$, stop.

Step 2. Find (by dichotomous search) $x_{n+1} = x_n + \rho_n s_n$ $(\rho_n \geqslant 0)$ and $g_{n+1} \in \partial_{\beta\epsilon} f(x_{n+1})$ such that $(g_{n+1}, s_n) \geqslant -\alpha |s_n|^2$.

Step 3. Increase n by 1. If $(g_n, x_n - x_p) \leqslant (1-\beta)\epsilon$, go to step 1.

Step 4. Set $p = n$ and go to step 1.

(Step 1 is a simple quadratic programming problem, and it can be shown that ρ_n in step 2 is an approximation of the optimal step in the direction s_n.)

5. As long as step 4 is not taken, the algorithm approximates a conjugate gradient algorithm, for if $g_n \in \partial_{\beta\epsilon} f(x_n)$ and $(g_n, x_n - x_p) \leqslant (1-\beta)\epsilon$, then $g_n \in \partial_{\epsilon'} f(x_p)$ with $\epsilon' = f(x_p) - f(x_n) + \epsilon$. When the algorithm terminates in step 1, we have $s_n \in \partial_{\epsilon'} f(x_p)$ and,,for all $y \in H$,

$$f(y) \geqslant f(x_p) + (s_n, y - x_p) - \epsilon' = (s_n, y - x_p) + f(x_n) - \epsilon\,.$$

It follows that $f(x_n) \leqslant f(y) + \eta |y - x_p| + \epsilon$ for all y, so that x_n approximately minimizes f over H, to the degree measured by η and ϵ.

The class of functions f for which termination can be shown includes those which are strongly convex, that is, for which there exists $\delta > 0$ such that for all $x, y \in H$ and $0 \leqslant \lambda \leqslant 1$.

$$f[\lambda x + (1-\lambda) y] \leqslant \lambda f(x) + (1-\lambda) f(y) - \delta\lambda(1-\lambda) |y - x|\,.$$

Theorem 3. *If f is strongly convex, the algorithm will terminate after some finite number of steps in the stop of step* 1.

References

[1] D.P. Bertsekas and S.K. Mitter, "A descent numerical method for optimization problems with nondifferentiable cost functional", *SIAM Journal on Control* 11 (1973) 637–652.

[2] V.F. Demjanov, "Algorithms for some minimax problems", *Journal of Computer and Systems Science* 2 (1968) 342–380.

[3] C. Lemarechal, "An algorithm for minimizing convex functions", in: J.L. Rosenfeld, ed., *Information processing '74* (North-Holland, Amsterdam, 1974) pp. 552–556.

[4] C. Lemarechal, "An extension of "Davidon" methods to nondifferentiable functions", in: M.L. Balinski and P. Wolfe, eds., *Nondifferentiable optimization,* Mathematical Programming Study 3 (North-Holland, Amsterdam), to appear.

[5] G.P. McCormick and K. Ritter, "Projection method for unconstrained optimization", *Journal of optimization Theory and Applications* 10 (1972) 57–66.
[6] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, N.J., 1970).
[7] P. Wolfe, "A method of conjugate subgradients for minimizing nondifferentiable functions", in: M.L. Balinski and P. Wolfe, eds., *Nondifferentiable optimization,* Mathematical Programming Study 3 (North-Holland, Amsterdam), to appear.
[8] G. Zoutendijk, "Nonlinear programming, computational methods". in: J. Abadie, ed., *Integer and nonlinear programming* (North-Holland, Amsterdam, 1970) pp. 37–86.