Adaptive Variable Metric Methods for Nondifferentiable Optimization Problems

S.P. URYAS'EV
Research Scholar
International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

Summary. This paper deals with new variable metric algorithms for nonsmooth optimization problems, so-called "adaptive algorithms." The essence of these are as follows: there are two simultaneously working gradient algorithms, the first in the main space, the second with respect to the matrices that modify the space variables. The convergence theorems for these algorithms are given for different cases.

1 Introduction

Many efficient algorithms have been developed for nonsmooth optimization problems (see, for example, Refs. 1–8). Most of these algorithms require the solution of some complicated linear or quadratic programming subproblem at each algorithm iteration. The first subgradient algorithm proposed by N.Z. Shor (Ref. 9) requires only subgradient calculations and projection operations at each iteration. It appears that this algorithm has a low convergence rate for ill-conditioned problems. Variable metric algorithms lie somewhere in between: they require not too many calculations at each iteration, but nevertheless have a good practical rate of convergence for ill-conditioned functions. Variable metric algorithms are widely used for smooth optimization problems (see, for example, Ref. 10). As a rule, these algorithms cannot be generalized to nonsmooth optimization problems. The difficulties are connected with the fact that even if the first and second derivatives exist at some point, they do not give the full local description of the function. Because the function is nonsmooth, a point of nonsmoothness can be arbitrarily close to the point where derivatives exist.

These difficulties have lead to the appearance of new ideas in the construction of variable metric algorithms. In the works of N. Shor and his coworkers (see, for example, Ref. 11), so-called "space-dilation algorithms" were developed. Such an approach offers the opportunity to construct practical and effective algorithms, but the most effective algorithm from this family, the r-algorithm, is not sufficiently understood from the theoretical point of view. References to this and related topics can be found in Ref. 12.

This author proposes an alternative "adaptive" approach that is applicable to optimization and game theoretic problems. The first variable metric algorithms in the framework of this approach are discussed in Ref. 13 for stochastic quasigradient algorithms. (Different optimization algorithms for stochastic optimization problems can be found in Ref. 14 and others.) The results of this paper are based upon the working paper (Ref. 15).

2 Essence of the Approach

Let us consider a convex optimization problem

$$f(x) \to \min_{x \in R^n}, \tag{1}$$

where the function $f(x)$ is convex on the Euclidean space R^n. We use the following recurrent algorithm to solve this problem:

$$x^{s+1} = x^s - \rho_s H^s g^s, \quad s = 0, 1, \dots \tag{2}$$

where s is the iteration number; $\rho_s > 0$ is a step size (scalar value); H^s is an $n \times n$ matrix; and g^s is a subgradient from the subdifferential $\partial f(x)$ of the function $f(x)$ at the point x^s, i.e. $g^s \in \partial f(x^s)$. We recall that the subdifferential of the function $f(x)$ at the point $y \in R^n$ is given by the formula (see, for example, Ref. 16)

$$\partial f(y) = \{g \in R^n : f(x) - f(y) \geq \langle g, x - y \rangle \quad \forall x \in R^n\}.$$

At the s^{th} iteration, the natural criterion defining the best choice of the matrix H^s is via the function

$$\varphi_s(H) = f(x^s - \rho_s H g^s).$$

The best matrix is a solution of the problem

$$\varphi_s(H) \to \min_{H \in R^{n \times n}}. \tag{3}$$

It is easy to see that problem (3) is a reformulation of the source problem (1), since if H^* is a solution of (3) then the point $x^s - \rho_s H^* g^s$ is a solution of (1). Moreover, problem (3) is more complex than (1) because the dimension of problem (3) is n times higher than that of (1). However, at the s^{th} iteration of algorithm (2) we do not need the optimal matrix, it is enough to correct (update) the matrix H^s. If we already have some matrix H_0^s, then the direction of adaptation can be defined by differentiating, in the general sense, the function $\varphi_s(H)$ at the point H_0^s. If the function $f(x)$ is a convex function, then the function $\varphi_s(H)$ is also convex. We can use the following formula (Ref. 17) for the differentiation of the complex function φ_s :

$$\partial \varphi_s(H_0^s) = -\rho_s \left\{ g g^{s^T} : g \in \partial f(x^s - \rho_s H_0^s g^s) \right\};$$

here and below, the superscript T means transposition. If $g_0^s \in \partial f(x^s - \rho_s H_0^s g^s)$, then $-\rho_s g_0^s g^{s^T} \in \partial \varphi(H_0^s)$. With respect to the matrix H, in the direction $g_0^s g^{s^T}$, one can do a step of the generalized gradient method:

$$H_1^s = H_0^s + \lambda_0^s g_0^s g^{s^T}, \quad \lambda_0^s > 0.$$

It is possible either to take $H^s = H_1^s$ or to continue the iterations of the generalized gradient algorithm with respect to H :

$$H_{i+1}^s = H_i^s + \lambda_i^s g_i^s g^{s^T}, \quad \lambda_i^s > 0, \tag{4}$$