Convergence Properties of the Modified Subgradient Method of Camerini et al.

Sehun Kim and Hyunsil Ahn
Department of Management Science, Korea Advanced Institute of Science and Technology, P.O. Box 150, Chungyang Seoul, Korea

In this article we provide a lower bound on the improvement of the Euclidean distance to an optimal solution in the modified subgradient method of Camerini et al. This is a stronger convergence property than that originally derived by those authors. Furthermore, this lower bound is shown to be strictly better than that of the standard subgradient method. This result may partially explain the successful computational improvement of the modified subgradient method.

1. INTRODUCTION

Let us consider the problem of minimizing a convex, not necessarily differentiable, function $f: R^n \rightarrow R^1$ with a "known" objective function value. The subgradient method based on Poljak's scheme [5] is one of the most widely used tools in the field of nondifferentiable optimization (Held et al. [3]). This method is very simple and generates a sequence which eventually converges to an optimal solution. However, many researchers have experienced an erratic behavior of this method due to its slow convergence, and several improvements [1,2,4,5,8,10] have been suggested to overcome this problem.

Camerini, Fratta, and Maffioli [1] proposed a modification of the subgradient method which uses the information about the direction applied at the previous step. Although their article deals with a piecewise-linear convex function, the algorithm suggested may be applied with no essential changes to any convex function. The main idea of the method consists in avoiding some troublesome effects due to the "subgradient's alternating components" by providing the direction which forms a smaller angle with the direction towards the minimum than does the direction negative to the subgradient, thus enhancing the speed of convergence.

They applied this modification to Lagrangian dual of the traveling salesman problem. Their computational results [1] indicate that this modification is superior to the standard subgradient method (Shor [9,p. 42]). Sarin, Karwan, and Rardin [7] suggested a new strategy for computing surrogate-dual multipliers. Their method involves solving several Lagrangian duals of a specially defined problem. The Lagrangian search was conducted using the modified subgradients. The advantage of modified subgradients over the standard subgradients in this application was demonstrated in Sarin and Karwan [6].

Theoretically, Camerini et al. showed that the search direction by their method forms a smaller angle with the direction towards the minimum points than does the direction by the standard subgradient method. They also showed that the Euclidean distance to an optimal solution is strictly decreasing.

Naval Research Logistics, Vol. 37, pp. 961–966 (1990)
 CCC 0028-1441/90/060961-06$04.00

In this article we provide a lower bound on the decrement of the Euclidean distance to an optimal solution in the modified subgradient method of Camerini et al. This is a stronger convergence property than that derived by those authors. Furthermore, this lower bound is strictly better than that of the standard subgradient method. This result may partially explain the successful computational improvment of the modified subgradient method.

2. PRELIMINARIES

For a convex function f, g is a subgradient of f at x, if

$$f(y) \geq f(x) + g(y - x), \qquad \text{for all } y \in R^n.$$

The subdifferential $\partial f(x)$ is the set of subgradients of f at x. Since f is finite and convex, $\partial f(x)$ is nonempty for all x.

Assuming that $f^* = \min f$ is known, the subgradient method based on Poljak's scheme [5] works as follows:

$$x^{k+1} = x^k - \lambda_k \frac{f(x^k) - f^*}{|g^k|^2} g^k, \qquad 0 < \epsilon_1 \leq \lambda_k \leq 2 - \epsilon_2 < 2, \tag{1}$$

where g^k is a subgradient of f at x^k. It is known that if $\lambda_k = 1$,

$$|x^{k+1} - x^*|^2 \leq |x^k - x^*|^2 - (f(x^k) - f^*)^2/|g^k|^2, \tag{2}$$

where x^* is an optimal solution.

Camerini et al. [1] suggested a modification of the subgradient method as follows:

$$x^{k+1} = x^k - \lambda_k \frac{f(x^k) - f^*}{|s^k|^2} s^k, \qquad 0 < \epsilon_1 \leq \lambda_k \leq 1. \tag{3}$$

$$s^k = g^k + \beta_k s^{k-1}; \qquad s^0 = 0. \tag{4}$$

$$\beta_k = \begin{cases} -\gamma_k(s^{k-1}g^k)/|s^{k-1}|^2, & \text{where } 0 \leq \gamma_k \leq 2 \quad \text{if } s^{k-1}g^k < 0; \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Note that s^k in (4) is in fact equivalent to a weighted sum of all preceding subgradients, which has been successfully used by Crowder [2] in order to avoid some possible troublesome effects due to the "subgradient alternating components."

Camerini et al. [1] provided the following lemma and two theorems as convergence properties of their algorithm. Let x^* be an optimal solution and suppose that the algorithm has not terminated at iteration k.

LEMMA 2.1: $(x^* - x^k)s^k \leq (x^* - x^k)g^k$.

THEOREM 2.1: $(x^* - x^k)(-s^k)/|s^k| \geq (x^* - x^k)(-g^k)/|g^k|$.

THEOREM 2.2: $|x^* - x^k| > |x^* - x^{k+1}|$.

Theorem 2.1 implies that the direction by this method forms a smaller angle with the direction towards the minimum than does the direction by the standard subgradient method. Theorem 2.2 implies the monotone decrease of distance to an optimal solution.

3. STRONGER CONVERGENCE PROPERTIES

In this section we consider the case $\lambda_k = 1$ in (3). Our results can be extended to other cases.

Let

$$\bar{x}^{k+1} = x^k - (f(x^k) - f^*)g^k/|g^k|^2 \tag{6}$$

which is obtained from x^k through the standard subgradient method (1) with $\lambda_k = 1$.

Let

$$s_c^k = g^k - (s^{k-1}g^k)s^{k-1}/|s^{k-1}|^2 \tag{7}$$

and

$$x_c^{k+1} = x^k - (f(x^k) - f^*)s_c^k/|s_c^k|^2, \tag{8}$$

where s_c^k and x_c^{k+1} are obtained from s^{k-1} and x^k through (3) and (4) with $\gamma_k = 1$.

THEOREM 3.1: Suppose that the algorithm (3) has not terminated at iteration k and $s^{k-1}g^k < 0$. Then for an optimal solution x^*,

$$|x^* - x^{k+1}|^2 \le |x^* - \bar{x}^{k+1}|^2 - \gamma_k(2 - \gamma_k)(|s_c^k|^2/|s^k|^2)|x_c^{k+1} - \bar{x}^{k+1}|^2,$$

where $0 \le \gamma_k \le 2$.

PROOF: Assume that the algorithm has not terminated at iteration k and $s^{k-1}g^k < 0$. Let x^* be an optimal solution.

$$\begin{aligned}
&|x^* - x^{k+1}|^2 - |x^* - \bar{x}^{k+1}|^2 \\
&\quad = |\bar{x}^{k+1} - x^{k+1}|^2 + 2(x^* - \bar{x}^{k+1})(\bar{x}^{k+1} - x^{k+1}) \\
&\quad = |\bar{x}^{k+1} - x^k + x^k - x^{k+1}|^2 + 2(x^* - x^k + x^k - \bar{x}^{k+1})(\bar{x}^{k+1} - x^k + x^k - x^{k+1}) \\
&\quad = |x^{k+1} - x^k|^2 - |\bar{x}^{k+1} - x^k|^2 \\
&\qquad + 2\{(x^* - x^k)(\bar{x}^{k+1} - x^k) - (x^* - x^k)(x^{k+1} - x^k)\}.
\end{aligned} \tag{9}$$

From (3) and (6),

$$|x^{k+1} - x^k|^2 - |\bar{x}^{k+1} - x^k|^2 = (f(x^k) - f^*)^2(|s^k|^{-2} - |g^k|^{-2}). \tag{10}$$

From (3) and (6) and Lemma 2.1,

$$\begin{aligned}&(x^* - x^k)(\bar{x}^{k+1} - x^k) - (x^* - x^k)(x^{k+1} - x^k)\\ &\quad = (f(x^k) - f^*)\{(x^* - x^k)s^k|s^k|^{-2} - (x^* - x^k)g^k|g^k|^{-2}\}\\ &\quad \leq (f(x^k) - f^*)\{(x^* - x^k)g^k(|s^k|^{-2} - |g^k|^{-2})\}.\end{aligned} \tag{11}$$

From (4) and (5),

$$\begin{aligned}|s^k|^2 &= |g^k + \beta_k s^{k-1}|^2\\ &= |g^k|^2 - \gamma_k(2 - \gamma_k)(s^{k-1}g^k)^2/|s^{k-1}|^2 \leq |g^k|^2.\end{aligned} \tag{12}$$

From the definition of subgradient,

$$g^k(x^* - x^k) \leq f^* - f(x^k). \tag{13}$$

Combining (12) and (13) with (11), we have

$$\begin{aligned}&(x^* - x^k)(\bar{x}^{k+1} - x^k) - (x^* - x^k)(x^{k+1} - x^k)\\ &\quad \leq -(f(x^k) - f^*)^2(|s^k|^{-2} - |g^k|^{-2}).\end{aligned} \tag{14}$$

Combining (9), (10), and (14)

$$|x^* - x^{k+1}|^2 - |x^* - \bar{x}^{k+1}|^2 \leq -(f(x^k) - f^*)^2(|s^k|^{-2} - |g^k|^{-2}). \tag{15}$$

Since $s^k = \gamma_k s_c^k + (1 - \gamma_k)g^k$ and $s_c^k g^k = |s_c^k|^2$ from (4) and (7),

$$\begin{aligned}|s^k|^{-2} - |g^k|^{-2} &= (|g^k|^2 - |s^k|^2)/(|s^k|^2|g^k|^2)\\ &= (|g^k|^2 - \gamma_k^2|s_c^k|^2 - (1 - \gamma_k)^2|g^k|^2 - 2\gamma_k(1 - \gamma_k)s_c^k g^k)/(|s^k|^2|g^k|^2)\\ &= \gamma_k(2 - \gamma_k)(|g^k|^2 - |s_c^k|^2)/(|s^k|^2|g^k|^2)\\ &= \gamma_k(2 - \gamma_k)(|s_c^k|^{-2} - |g^k|^{-2})|s_c^k|^2/|s^k|^2.\end{aligned} \tag{16}$$

From (6) and (8),

$$|x_c^{k+1} - \bar{x}^{k+1}|^2 = (f(x^k) - f^*)^2(|s_c^k|^{-2} - |g^k|^{-2}). \tag{17}$$

Combining (15)–(17), we obtain the asserted result. ■

COROLLARY 3.1: Under the same condition as that in Theorem 3.1, for an optimal solution x^*,

$$|x^* - x^{k+1}|^2 \le |x^* - x^k|^2 - (f(x^k) - f^*)^2/|g^k|^2 - \gamma_k(2 - \gamma_k)(|s_c^k|^2/|s^k|^2)|x_c^{k+1} - \bar{x}^{k+1}|^2,$$

where $0 \le \gamma_k \le 2$.

PROOF: This is a direct result of Theorem 3.1 and (2). ■

This is a stronger property than Theorem 2.2. Furthermore, this result compared with (2) shows an additional negative term on the right-hand side, which is an additional improvement of the modified method on the Euclidean distance to an optimal solution.

Especially, if we choose $\gamma_k = 1$, then we obtain the following corollary.

COROLLARY 3.2: Assume the same condition as that in Theorem 3.1. Let x^* be an optimal solution and $\gamma_k = 1$. Then

$$|x^* - x^{k+1}|^2 \le |x^* - x^k|^2 - (f(x^k) - f^*)^2/|g^k|^2 - |x^{k+1} - \bar{x}^{k+1}|^2.$$

The additional improvement of the modified method is nonzero unless the modified algorithm happens to generate an identical point as the standard subgradient method.

4. CONCLUSIONS

We provide a lower bound on the improvement of the Euclidean distance to an optimal solution. We believe that this result partially explains the successful computational improvement of this modified method over the standard subgradient method.

REFERENCES

[1] Camerini, P. M., Fratta, L., and Maffioli, F., "On Improving Relaxation Methods by Modified Gradient Technique," *Mathematical Programming Study*, **3**, 26–34 (1975).
[2] Crowder, H., "Computational Improvements for Subgradient Optimization," IBM Research Report No. RC 4907 (No. 21841). Thomas J. Watson Research Center, June, 1974.
[3] Held, M., Wolfe, P., and Crowder, H., "Validation of Subgradient Optimization," *Mathematical Programming*, **6**, 62–88 (1974).
[4] Kim, S., Koh, S., and Ahn, H., "Two-Direction Subgradient Method for Non-Differentiable Optimization Problems," *Operations Research Letters*, **6**(1), 43–46 (1987).
[5] Poljak, B. T., "Minimization of Nonsmooth Functionals," *USSR Computational Mathematics and Mathematical Physics*, **9**, 14–29 (1969).
[6] Sarin, S., and Karwan, M. H., "Computational Evaluation of Two Subgradient Search Methods," *Computers and Operations Research*, **14**(3), 241–247 (1987).

[7] Sarin, S., Karwan, M. H., and Rardin, R. L., "A New Surrogate-Dual Multiplier Search Procedure," *Naval Research Logistics*, **34**, 431–450 (1987).
[8] Shor, N. Z., "Utilization of the Operation of Space Dilation in the Minimization of Convex Functions," *Cybernetics*, **6**(1), 7–15 (1970).
[9] Shor, N. Z., *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, 1985.
[10] Shor, N. Z., and Zhurbenko, N. G., "A Minimization Method Using Space Dilation in the Direction of Difference of Two Successive Gradients," *Cybernetics*, **7**(3), 450–459 (1971).