J. Austral. Math. Soc. Ser. B 38(1996), 194–200

A COMPARATIVE STUDY OF SMOOTHING APPROXIMATIONS

X. Q. YANG[1]

(Received 12 January 1995; revised 13 February 1995)

Abstract

It is known that many optimization problems can be reformulated as composite optimization problems. In this paper error analyses are provided for two kinds of smoothing approximation methods of a unconstrained composite nondifferentiable optimization problem. Computational results are presented for nondifferentiable optimization problems by using these smoothing approximation methods. Comparisons are made among these methods.

1. Introduction

Consider the following composite nondifferentiable optimization problem:

$$\min G(x) \qquad \text{subject to } x \in \mathbb{R}^n, \tag{P}$$

where $G(x) := h(x, |g_1(x)|, \cdots, |g_m(x)|) : \mathbb{R}^n \longrightarrow \mathbb{R}$, $h : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$ is a continuously differentiable function and $g_1, \cdots, g_m : \mathbb{R}^n \longrightarrow \mathbb{R}$ are continuously differentiable functions.

Several optimization problems such as exact penalty problems, minimax problems and l_1-norm minimization problems, can be reformulated as problem P. Similar composite models have been considered by Bertsekas [3], Ben-Tal and Teboulle [1] and Yang [9]. Bertsekas [3] considered a composite nondifferentiable optimization problem where the nondifferentiable feature is brought in by $\max\{0, g(x)\}$ (so-called kinks). In Ben-Tal and Teboulle [1] a nondifferentiable convex function is smoothed by a differentiable recession function. In Yang [9] a composite nondifferentiable optimization problem is considered using a two-parameter approximation where nondifferentiable functions can be smoothed by adding a power greater than or equal to 2. Teo and Goh [7] studied constrained optimization problems with nonsmooth objective functions (L_1 functionals) by smoothing the objective function with a single parameter approximation. This method has been extended for solving more general

[1]Department of Mathematics, University of Western Australia, Nedlands, 6009, Australia.

https://doi.org/10.1017/S0334270000000588 Published online by Cambridge University Press

nondifferentiable optimization problems in, for example, Jennings and Teo [6] and Teo et al. [8].

In this paper we restrict our study to two kinds of smoothing approximation methods for P, given in Teo and Goh [7] and in [2,4,9], respectively. We show comparisons between these two methods. Error analyses between P and smoothing approximation problems are presented under certain assumptions. These results are illustrated by giving computational results for nondifferentiable optimization problems via the computer package MATLAB.

2. Error analyses

In this section various error bounds are presented for the function values of the objective function of P at an approximate optimal solution and at the optimal solution.

A smoothing approximation method is given in [7] to approximate absolute value functions. For a given $\epsilon > 0$, each nondifferentiable function $|g_i(x)|$ in P is replaced by a differentiable function $\hat{g}_i^\epsilon(x)$ which is defined by

$$\hat{g}_i^\epsilon(x) = \begin{cases} |g_i(x)|, & \text{if } |g_i(x)| \geq \epsilon/2, \\ \left[g_i(x)^2 + \epsilon^2/4\right]/\epsilon, & \text{if } |g_i(x)| < \epsilon/2. \end{cases}$$

The resulting differentiable optimization problem is denoted by $P_1(\epsilon)$ with the objective function $G_\epsilon(x) := h(x, \hat{g}_1^\epsilon(x), \dots, \hat{g}_m^\epsilon(x))$.

In [2,4,9], a least-square based method (LSBM) is presented. If this method is applied to P, then the nondifferentiable function $|g_i(x)|$ is replaced by

$$\bar{g}_i^\epsilon(x) = \sqrt{g_i(x)^2 + \epsilon^2}.$$

The resulting differentiable optimization problem is denoted by $P_2(\epsilon)$.

Assume that x^* is a minimum of P and that L is the Lipschitz constant of h.

THEOREM 2.1. *Let $x_{1\epsilon}^*$ be a minimum of $P_1(\epsilon)$. Then*

$$0 \leq G(x_{1\epsilon}^*) - G(x^*) \leq \frac{1}{2} L\sqrt{m}\,\epsilon. \tag{2.1}$$

PROOF. It is easy to see that

$$0 \leq \hat{g}_i^\epsilon(x) - |g_i(x)| \leq \frac{\epsilon}{4}, \quad \forall x \in \mathbb{R}^n.$$

From the Lipschitz property of h, we have

$$|G_\epsilon(x) - G(x)| \leq \frac{1}{4} L\sqrt{m}\,\epsilon.$$

Thus

$$G(x_{1\epsilon}^*) - G(x^*) \le G_\epsilon(x_{1\epsilon}^*) - G(x^*) + \frac{1}{4}L\sqrt{m}\,\epsilon.$$

Then

$$\begin{aligned} 0 &\le G(x_{1\epsilon}^*) - G(x^*) \\ &\le G_\epsilon(x_{1\epsilon}^*) - G(x^*) + \frac{1}{4}L\sqrt{m}\,\epsilon \\ &\le G_\epsilon(x^*) - G(x^*) + \frac{1}{4}L\sqrt{m}\,\epsilon \\ &\le \frac{1}{4}L\sqrt{m}\,\epsilon + \frac{1}{4}L\sqrt{m}\,\epsilon = \frac{1}{2}L\sqrt{m}\,\epsilon. \end{aligned}$$

Thus (2.1) holds.

THEOREM 2.2. *Let $x_{2\epsilon}^*$ be a minimum of $P_2(\epsilon)$. Then*

$$0 \le G(x_{2\epsilon}^*) - G(x^*) \le 2L\sqrt{m}\,\epsilon. \tag{2.2}$$

PROOF. We have

$$0 \le \bar{g}_i^\epsilon(x) - |g_i(x)| \le \epsilon.$$

Thus by using a similar argument as in the proof of Theorem 2.1, the error estimate (2.2) follows.

From (2.1) and (2.2), the convergence rate of the method of [7] is faster (4 times) than the least-square based method corresponding to a decreasing of the smoothing parameter ϵ. However, as shown in [9], the following generalized least-square method is more flexible than Teo and Goh's method and can be used to solve optimization problems with higher degree nondifferentiability.

Consider the following problem

$$\min Q(x) \qquad \text{subject to } x \in \mathbb{R}^n, \tag{P$_3$}$$

where $Q(x) := q(x, q_1(x), \ldots, q_m(x)) : \mathbb{R}^n \longrightarrow \mathbb{R}$, $q : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$ is a differentiable function. The functions $q_i : \mathbb{R}^n \longrightarrow \mathbb{R}, i = 1, \ldots, m$ are not, in general, differentiable, but for some $\alpha \ge 2$, the functions $q_i^\alpha, i = 1, \ldots, m$ are differentiable. Thus in P_3, $q_i(x)$ is replaced by

$$\bar{q}_i^\epsilon(x) = (q_i(x)^\alpha + \epsilon^\alpha)^{1/\alpha}, \quad i = 1, \ldots, m.$$

The resulting differentiable optimization problem is denoted by $P_3(\epsilon)$. Assume that x_3^* is a minimum of P_3 and that L_1 is the Lipschitz constant of q.

THEOREM 2.3. *Let $x^*_{3\epsilon}$ be a minimum of $P_3(\epsilon)$. Then*

$$0 \le G(x^*_{3\epsilon}) - G(x^*_3) \le 2L_1\sqrt{m}\,\epsilon. \tag{2.3}$$

PROOF. We have

$$0 \le \tilde{q}_i^{\epsilon}(x) - q_i(x) \le \epsilon.$$

Thus, by using a similar argument as in Theorem 2.1, the error estimate (2.3) follows.

Note the error bound (2.3) does not depend on the parameter α.

3. Numerical examples

In this section we present some numerical results by using the two smoothing approximation methods discussed in Section 2. The computations here are carried out by using the computer package MATLAB with analytical gradients supplied. The FUNCTION *fminu* is used which is based on a quasi-Newton method with the BFGS formula for updating the approximation of the Hessian matrix, see [5].

EXAMPLE 3.1. Consider the problem

$$\min\left(1+\sum_{i=1}^{n} i\,|x_i|\right)^2 \qquad \text{subject to } x \in \mathbb{R}^n.$$

(See reference [3].)

We use f and n_f to denote the value of the function at the last evaluation point and the number of function evaluations made, respectively, when the generalized least-square method is used. We use f' and n'_f to denote the value of the function at the last evaluation point and the number of function evaluations made, respectively, when the method in [7] is used. For a given error $\epsilon > 0$, with the least-square based method this problem is approximated by a smooth problem

$$\min\left(1+\sum_{i=1}^{n} i\sqrt{x_i^2+\epsilon^2}\right)^2 \qquad \text{subject to } x \in \mathbb{R}^n.$$

By using Teo and Goh's methods, each term $|x_i|$ is approximated by

$$\hat{x}_i^{\epsilon} = \begin{cases} |x_i|, & \text{if } |x_i| \ge \epsilon/2, \\ \left[(x_i)^2+\epsilon^2/4\right]/\epsilon, & \text{if } |x_i| < \epsilon/2. \end{cases}$$

Thus we obtain a differentiable problem

$$\min \left(1 + \sum_{i=1}^{n} i \hat{x}_i^{\epsilon}\right)^2 \qquad \text{subject to } x \in \mathbb{R}^n.$$

The initial guess for this problem is a vector x_0 with component $x_i^0 = -1$, $i = 1, \ldots, n$. Computational results for $n = 5$ and $n = 50$ are summarized in Table 1 and Table 2, respectively.

TABLE 1. Iterations with $n = 5$.

ϵ	f	n_f	f'	n'_f
5^{-1}	16.0000	29	1.5625	10
5^{-2}	2.5600	5	1.1025	5
5^{-3}	1.2544	2	1.0201	2
5^{-4}	1.0486	9	1.0040	2
5^{-5}	1.0096	5	1.0000	2
5^{-6}	1.0019	9		
5^{-7}	1.0004	6		
5^{-8}	1.0001	9		
5^{-9}	1.0000	2		

TABLE 2. Iterations with $n = 50$.

ϵ	f	n_f	f'	n'_f
5^{-1}	6.5536e+04	175	12.2500	10
5^{-2}	2.7040e+03	65	2.2500	2
5^{-3}	125.4400	2	1.2100	2
5^{-4}	9.2416	2	1.0404	2
5^{-5}	1.9825	15	1.0080	6
5^{-6}	1.1699	5	1.0016	2
5^{-7}	1.0329	24	1.0003	2
5^{-8}	1.0065	16	1.0001	2
5^{-9}	1.0013	6	1.0000	2

The following example shows that the method given in Yang [9] is flexible and can be used to solve optimization problems with higher degree nondifferentiability.

EXAMPLE 3.2. Consider the problem

$$\min 1 + |x|^{1/3}, \qquad \text{subject to } x \in \mathbb{R}.$$

(See [9].) For $\alpha = 6$ this problem is approximated by

$$\min 1 + (x^2 + \epsilon^6)^{1/6}, \qquad \text{subject to } x \in \mathbb{R},$$

and for $\alpha = 12$ by

$$\min 1 + (x^4 + \epsilon^{12})^{1/12}, \qquad \text{subject to } x \in \mathbb{R},$$

respectively. The starting point $x_0 = 1$. The results for the generalized least-square method with $\alpha = 6$ and $\alpha = 12$ are given in Table 3. It is noted from Table 3 that the function values of each iteration for $\alpha = 6$ and $\alpha = 12$ are the same and that the number of function evaluations of each iteration does not change much with the value of α. Hence it is suggested that α is chosen relatively large when the degree of nondifferentiability of the optimization problem is not known.

TABLE 3. Iterations with $\alpha = 6$ and $\alpha = 12$.

ϵ	f	n_f^6	n_f^{12}
5^{-1}	1.2000	20	16
5^{-2}	1.0400	10	13
5^{-3}	1.0080	22	22
5^{-4}	1.0016	33	38

As a conclusion, we see that the convergence rate of the method introduced by Teo and Goh [7] is faster than that of the least-square based method given in [2,4,9], corresponding to a decreasing of the smoothing parameter ϵ. It is clear that the method of [7] also needs fewer function evaluations. However the method given in [9] is flexible and can be applied for optimization problems whenever higher degree nondifferentiability appears.

Acknowledgement

This research is supported by a grant from the Australian Research Council.

References

[1] A. Ben-Tal and M. Teboulle, "A smoothing technique for nondifferentiable optimization problems", in *Optimization - Fifth French-German Conference* (ed. S. Dolecki), Volume 1405 of *Lecture Notes in Math.*, (Springer, New York, 1989) 1–13.

[2] A. Ben-Tal, M. Teboulle and W. H. Yang, "A least-square method for a class of nonsmooth minimization problems with applications in plasticity", *Applied Math. Optimiz.* **24** (1991) 272–288.

[3] D. P. Bertsekas, "Nondifferentiable optimization via approximation", *Math. Prog. Study* **3** (1975) 1–25.

[4] R. A. El-Attar, M. Vidyasagar and R. K. Dutta, "An algorithm for l_1-norm minimization with application to nonlinear l_1 approximation", *SIAM J. Numer. Anal.* **16** (1979) 70–86.

[5] R. Fletcher, *Practical Methods of Optimization* (John Wiley, New York, 1987).

[6] L. S. Jennings and K. L. Teo, "A computational algorithm for functional inequality constrained optimization problem", *Automatica* **26** (1990) 371–375.

[7] K. L. Teo and C. J. Goh, "On constrained optimization problems with nonsmooth cost functionals", *Applied Math. Optimiz.* **18** (1988) 181–190.

[8] K. L. Teo, V. Rehbock and L. S. Jennings, "A new computational algorithm for functional inequality constrained optimization problems", *Automatica* **29** (1993) 789–792.

[9] X. Q. Yang, "Smoothing approximations to nonsmooth optimization problems", *J. Austral. Math. Soc., Series B* **36** (1994) 1–13.