SMOOTHING ITERATIVE METHOD
FOR A CLASS OF NONSMOOTH EQUATIONS*

H. Wang

Nanjing University of Aeronautics and Astronautics, China
College of Sciences, China University of Mining and Technology, Xuzhou
e-mail: `hjwcumt@163.com`

D. Cao

College of Sciences, China University of Mining and Technology, Xuzhou

S. Li

College of Information and Electrical Engineering,
China University of Mining and Technology, Xuzhou

Статья посвящена решению одного класса негладких уравнений с использованием функции-максимума. На основе принципа максимума энтропии строится гладкая функция, аппроксимирующая исходную негладкую. Предложена новая итерационная формула и дан алгоритм решения аппроксимирующей задачи. Указаны условия, при которых обеспечивается сходимость. Теоретический анализ и результаты расчетов показали эффективность и надежность алгоритма.

Introduction

We consider the following nonsmooth equations

$$F(x) = 0 \tag{0.1}$$

where $F : D \subset R^n \to R^n$, $F(x) = (f_1(x), f_2(x), \cdots, f_n(x))^T$ and $f_i(x) = \max_{j \in J_i} f_{ij}(x)(i = 1, \ldots, n)$, $f_{ij}(x)(j \in J_i)$ are continuously differentiable on D, J_i for $i = 1, \ldots, n$ are finite index sets. Obviously, $f_i(x)(i = 1, \ldots, n)$ are nondifferentiable functions on D.

Recently, the problem (0.1) has been discussed by several authors. In papers [1–3], Newton method and quasi-Newton method for solving nonsmooth equations have been described. Based on the same rate of convergence in [1–3], an improved algorithm with less computing cost was suggested in [4] and [6]. But the above algorithm is hard to accomplish in actual numerical computation. In this paper, we propose a new idea of solving nonsmooth equations (0.1) and our method can improve these defects. The following two steps are required.

First, we use the maximum entropy method to transform nonsmooth functions $f_i(x)(i = 1, \ldots, n)$ to smooth functions.

Second, we apply known techniques and new iterative formula to solve the resulting smooth equations.

As a particular case of this idea, we propose a new algorithm for solving problem (0.1). The structure of the paper is as follows: In Section 1, we will describe maximum entropy function and design smooth approximating function. In Section 2, the proposed algorithm and its convergence property are described. In Section 3, we give numerical tests to illustrate our theoretical results. Finally, the last section contains the conclusions and an outline of the directions for future research.

1. Design Smooth Approximating Function

For nondifferentiable problem this is hard to solve it by traditional methods including derivative. Usually we handle it with a differentiable function approaching to objective function. A maximum entropy function of uniform approximation to objective function is given in [5] and [9, 10].

Maximum Entropy Function. The maximum entropy function originated in the work of L. Boltzmann [10] who studied the relation between entropy and probability in physical systems. Subsequently, entropy concepts and ideas have been introduced in physics, chemistry, meteorology, information science, etc. The maximum entropy function, which was introduced by Jayned [9] has proved to be useful for obtaining convergent upper bounds on nonsmooth objective functions.

Definition 1.1. Let p be a positive integer and let $\varphi_i : D \to R^1 (i = 1, \ldots, k)$ be given continuously differentiable functions and let $\varphi(x) = \max_{1 \le i \le k}\{\varphi_i(x)\}$. The maximum entropy function $\varphi_p : D \to R^1$ of $\varphi_1, \ldots, \varphi_k$ is defined by

$$\varphi_p(x) = \frac{1}{p} \ln \left\{ \sum_{i=1}^{k} \exp\left(p\varphi_i(x)\right) \right\}. \tag{1.1}$$

Obviously, $\varphi_p(x)$ is also continuously differentiable on D.

Lemma 1.1 (Jaynes [9]). *Given $\varepsilon > 0, \exists \bar{p} > 0$ such that $\forall p > \bar{p}$ and $\forall x \in D$*

$$|\varphi_p(x) - \varphi(x)| < \varepsilon. \tag{1.2}$$

Also, if $1 \le p \le q$, then

$$\varphi_q(x) \le \varphi_p(x). \tag{1.3}$$

Proof. Let $v_i(x) = \exp(\varphi_i(x))$ $(i = 1, \ldots, k)$ and let $v(x) = (v_i(x))_{k \times 1}$. For each $x \in D$ $(\forall p \ge 1)$

$$\|v(x)\|_p = \left[\sum_{i=1}^{k} v_i(x)^p \right]^{1/p} = \left[\sum_{i=1}^{k} \exp(p\varphi_i(x)) \right]^{1/p}.$$

So

$$\varphi_p(x) = \ln(\|v(x)\|_p).$$

Now

$$\|v(x)\|_\infty = \lim_{p \to \infty} \|v(x)\|_p = \max_{1 \le i \le k}\{\exp(\varphi_i(x))\}.$$

So

$$\ln(\|v(x)\|_\infty) = \max_{1 \le i \le k}\{\varphi_i(x)\},$$
$$\lim_{p \to \infty} \varphi_p(x) = \max_{1 \le i \le k} \varphi_i(x).$$

This proves (1.2). Finally, according to Jensen's inequality, $(1 \le p \le q) \Rightarrow (\|v(x)\|_q \le \|v(x)\|_p)$, so $\ln(\|v(x)\|_q \le \ln\|v(x)\|_p)$ whence $\varphi_q(x) \le \varphi_p(x)$. So (1.3) is proved. $\square$

Lemma 1.2. $\forall x \in D$, *we have*

$$\varphi(x) \le \varphi_p(x) \le \varphi(x) + \frac{\ln k}{p}. \tag{1.4}$$

Proof.

$$\varphi_p(x) - \varphi(x) = \frac{1}{p}\ln\left\{\sum_{i=1}^{k}\exp(p\varphi_i(x))\right\} - \varphi(x) =$$
$$= \frac{1}{p}\ln\left\{\sum_{i=1}^{k}\exp(p(\varphi_i(x) - \varphi(x)))\right\}.$$

By definition of $\varphi(x)$, $\varphi_i(x) - \varphi(x) \le 0$ $(i = 1, \ldots, k)$ and there is at least one $j \in \{1, \ldots, k\}$ such that $\varphi_j(x) - \varphi(x) = 0$. Therefore $1 \le \sum_{i=1}^{k}\exp(p(\varphi_i(x) - \varphi(x))) \le k$. Therefore we have

$$0 \le \frac{1}{p}\ln\left\{\sum_{i=1}^{k}\exp(p(\varphi_i(x) - \varphi(x)))\right\} \le \frac{1}{p}\ln(k).$$

So we complete the proof.

Because of differentiability and other good properties of maximum entropy function, it is extensively applied to the field of nonlinear programming, stochastic programming, nondifferentiable programming, multiobjective programming, etc.

Smooth Approximating Function. Let m_i be the cardinality of set J_i $(i = 1, \ldots, n)$, so maximum entropy function of $f_{ij}(x)$ $(j \in J_i)$ is defined as follows

$$f_{ip}(x) = \frac{1}{p}\ln\left\{\sum_{j \in J_i}\exp\left(pf_{ij}(x)\right)\right\} \tag{1.5}$$

where $p(\ge 1)$ is entropy factor. And $\forall x \in D$, we also have $f_i(x) \le f_{ip}(x) \le f_i(x) + \dfrac{\ln m_i}{p}$, $f_{iq}(x) < f_{ip}(x) \ \forall \ q > p > 1$, $i = 1, \ldots, n$. And the function $f_{ip}(x)$ is continuously differentiable on D.

From Lemma 1.1 and Lemma 1.2, we know that

$$\lim_{p \to +\infty} f_{ip}(x) = f_i(x), \quad i = 1, \ldots, n.$$

Namely, $f_{ip}(x)$ is uniform approximation of $f_i(x)$. Obviously, it is reasonable and feasible that we substitute a differentiable function $f_{ip}(x)$ for $f_i(x)$. With the idea of maximum entropy function, we can transform (0.1) to the approximation differentiable equations:

$$F_p(x) = 0, \tag{1.6}$$

where $F_p(x) = (f_{1p}(x), f_{2p}(x), \cdots, f_{np}(x))^T$.

Theorem 1.2. *Given* $\varepsilon > 0, \exists p_k > 1$, *such that* $\left(p = p_k > \dfrac{\ln m}{\varepsilon}, \ m = \max\limits_{1 \leq i \leq 1} \{m_i\} \right)$,

$$\|F(\hat{x}) - F(x^*)\|_\infty \leq \varepsilon \quad or \quad \|F_p(\hat{x}) - F_p(x^*)\|_\infty \leq \varepsilon \tag{1.7}$$

where $\hat{x}$ *is a solution of problem (1.6),* x^* *is a solution of problem (0.1).*

Proof. Because x^* is a solution of problem (0.1), we have

$$f_i(x^*) = \max_{i \in J_i}\{f_{ij}(x)\} = 0, \quad i = 1, \ldots, n,$$

$$|f_{ip}(x^*) - f_i(x^*)| = |f_{ip}(x^*)| = \left| \frac{1}{p} \ln \left\{ \sum_{j \in J_i} \exp\left(p f_{ij}(x^*)\right) \right\} \right| \leq$$

$$\leq \frac{\ln m_i}{p} \leq \frac{\ln m}{p}, \quad i = 1, \ldots, n. \tag{1.8}$$

And $\hat{x}$ is a solution of problem (1.6), so $f_{ip}(\hat{x}) = 0 \ (i = 1, \ldots, n)$. Then we can prove

$$f_i(\hat{x}) \leq f_{ip}(\hat{x}) \leq f_i(\hat{x}) + \frac{\ln m_i}{p}, \quad i = 1, \ldots, n,$$

$$|f_i(\hat{x})| \leq \frac{\ln m_i}{p} \leq \frac{\ln m}{p}, \quad i = 1, \ldots, n. \tag{1.9}$$

Let $p = p_k > \dfrac{\ln m}{\varepsilon}$, from (1.8) and (1.9), we have

$$\|F(\hat{x}) - F(x^*)\|_\infty = \|F(\hat{x})\|_\infty \leq \varepsilon,$$
$$\|F_p(\hat{x}) - F_p(x^*)\|_\infty = \|F_p(x^*)\|_\infty \leq \varepsilon.$$

So we complete the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

2. Algorithm and Convergence

Convergence. Let $\gamma \in R(\gamma \neq 0)$. We consider fixed point problem as follows

$$x_i - \gamma f_{ip}(x) = x_i, \quad i = 1, \ldots, n. \tag{2.1}$$

Obviously, fixed point problem (2.1) and problem (1.6) are equivalent. So we have

$$x_i^{(k+1)} = x_i^{(k)} - \gamma^{(k)} f_{ip}(x^{(k)}), \quad i = 1, \ldots, n, \quad k = 0, 1, \ldots \tag{2.2}$$

According to properties of maximum entropy function, it is clear that the solution of problem (2.2) approaches the solution of problem (0.1) under the condition of iteration convergence with $p \to +\infty$.

Let

$$G(x^{(k)}) = \begin{pmatrix} x_1^{(k)} - \gamma^{(k)} f_{1p}(x^{(k)}) \\ \cdots \qquad \cdots \qquad \cdots \\ x_n^{(k)} - \gamma^{(k)} f_{np}(x^{(k)}) \end{pmatrix}, \quad \text{then} \quad G'(x^{(k)}) = I - \gamma^{(k)} U^{(k)} M^{(k)},$$

where $U^{(k)} = \text{diag}(u_1^{(k)}, \ldots, u_n^{(k)})$, $u_i^{(k)} = \dfrac{1}{\sum\limits_{j \in J_i} \exp(pf_{ij}(x^{(k)}))}$,

$$M^{(k)} = \begin{pmatrix} \sum\limits_{j \in J_1} \left\{ \exp(pf_{1j}(x^{(k)})) \dfrac{\partial f_{1j}}{\partial x_1^{(k)}} \right\} & \cdots & \sum\limits_{j \in J_1} \left\{ \exp(pf_{1j}(x^{(k)})) \dfrac{\partial f_{1j}}{\partial x_n^{(k)}} \right\} \\ \cdots & \cdots & \cdots \\ \sum\limits_{j \in J_n} \left\{ \exp(pf_{nj}(x^{(k)})) \dfrac{\partial f_{nj}}{\partial x_1^{(k)}} \right\} & \cdots & \sum\limits_{j \in J_n} \left\{ \exp(pf_{nj}(x^{(k)})) \dfrac{\partial f_{nj}}{\partial x_n^{(k)}} \right\} \end{pmatrix}.$$

Then we have

$$x^{(k+1)} = G(x^{(k)}), \quad k = 0, 1, \ldots \tag{2.3}$$

Theorem 2.1 [8]. *The sufficient condition of iteration convergence of (2.3) is $\rho(G'(x^{(k)})) < 1$, where $\rho(G'(x^{(k)})) = \sqrt{\max \lambda((G'(x^{(k)})^T(G'(x^{(k)})))}$ is spectral radius of $G'(x^{(k)})$.*

Theorem 2.2. *Suppose $x^* \subset int(D)$ is a fixed point of the map $G : D \subset R^n \to R^n$. If $\gamma^{(k)} \leq \dfrac{1}{\rho(M^{(k)})}$ $(k = 0, 1, \ldots)$, then $\forall x^{(0)} \in D$, x^* is an attracting point of iterative sequence (2.3).*

Proof. Suppose λ is an arbitrary eigenvalue of the matrix

$$(U^{(k)} M^{(k)})^T U^{(k)} M^{(k)} = (M^{(k)})^T (U^{(k)})^2 M^{(k)}$$

and $z \in R^n$ is an eigenvector corresponding to λ, then $(M^{(k)})^T(U^{(k)})^2 M^{(k)} z = \lambda z$ and

$$\lambda = \frac{z^T((M^{(k)})^T(U^{(k)})^2 M^{(k)})z}{z^T z} = \frac{(M^{(k)}z)^T(U^{(k)})^2(M^{(k)}z)}{z^T z} =$$

$$= \frac{\sum\limits_{i=1}^{n} u_i^2 (M^{(k)}z)_i^2}{z^T z} \leq \frac{\sum\limits_{i=1}^{n} \max_{1 \leq i \leq n}\{u_i\}^2 (M^{(k)}z)_i^2}{z^T z} < \frac{(M^{(k)}z)^T(M^{(k)}z)}{z^T z} = \frac{z^T((M^{(k)})^T M^{(k)})z}{z^T z}.$$

Obviously, we can deduce that $\rho^2(U^{(k)} M^{(k)}) < \rho^2(M^{(k)})$ (namely $\rho(U^{(k)} M^{(k)}) < \rho(M^{(k)})$), and $\gamma^{(k)} \leq \dfrac{1}{\rho(M^{(k)})}$, such that $\rho(\gamma^{(k)} U^{(k)} M^{(k)}) = \gamma^{(k)} \rho(U^{(k)} M^{(k)}) < \gamma^{(k)} \rho(M^{(k)}) \leq 1$. Therefore $\rho(G'(x^{(k)})) = \rho(I - \gamma^{(k)} U^{(k)} M^{(k)}) < 1$.

From Theorem 2.1, we know that $\forall x^{(0)} \in D$, x^* is an attracting point of iterative sequence (2.3).

Algorithm. With the preceding ideas, we suggest the following algorithm for solving problem (0.1).The detailed steps of the algorithm are as follows.

Step 1. Construct corresponding maximum entropy function and transform the nonsmooth equations (0.1) to the following smooth equations

$$f_{\text{ip}}(x) = \frac{1}{p} \ln \left\{ \sum_{j \in J_i} \exp\left(pf_{ij}(x)\right) \right\}, \quad i = 1, \ldots, n.$$

Step 2. With given initial point $x^{(0)} \in D$ and entropy factor $p(> 1)$, solve the above smooth equations by following iteration $\left(\text{let} \gamma^{(k)} \leq \dfrac{1}{\rho(M^{(k)})} \right)$

$$x_i^{(k+1)} = x_i^{(k)} - \frac{\gamma^{(k)}}{p} \ln \left\{ \sum_{j \in J_i} \exp\left(pf_{ij}(x^{(k)})\right) \right\}, \quad i = 1, \ldots, n, \quad k = 0, 1, \ldots$$

Step 3. If $\|F(x^{(k)})\|_\infty \geq \varepsilon$ and $\|x^{(k)} - x^{(k-1)}\| \geq \varepsilon$, then go to Step 2.

Step 4. Output $x^{(k)}$, end.

In the process of actual computing, we ought to notice the following three points:

— according to Theorem 1.2, let $p > \dfrac{\ln m}{\varepsilon}$;

— in numerical computation, we can select $\gamma^{(k)} = \dfrac{1}{\|M^{(k)}\|}$;

— in order to accelerate the convergence speed of iteration, we can also improve the algorithm by applying Gauss — Seidel iterative technique:

$$x_i^{(k+1)} = x_i^{(k)} - \frac{\gamma^{(k)}}{p} \ln \left\{ \sum_{j \in J_i} \exp\left(p f_{ij}(x_1^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, x_i^{(k)}, \ldots, x_n^{(k)})\right) \right\},$$
$$i = 1, \ldots, n, \quad k = 0, 1, \ldots$$

3. Numerical Results

The algorithm has been implemented using Matlab 5.2 on Pentium-IV computer. Numerical results for the following three examples are reported in this section, in each of which the actual minimum is explicitly known. The symbols p, ε, x^*, $x^{(0)}$, $\hat{x}$ denote the entropy factor, the precision of minimum value, the solution of equations (0.1),the initial iteration point, the approximate solution of equations (0.1).

Example 1.

$$\max_{x \in D}\{f_1(x), \ f_2(x), \ f_3(x)\} = 0$$

where $f_1(x) = x(x-2)(x-8)(x-4)(x-6)$, $f_2(x) = x(2-x)(x-8)(x-4)(x-6)$, $f_3(x) = (x-2)(x-7)$, $D = [0, 6]$. The solutions of above equation $x_1^* = 2.0$, $x_2^* = 4.0$, $x_3^* = 6.0$. Let $\varepsilon = 1e - 4$, $p = 10^4$, $f(x) = \max_{x \in D}\{f_1(x), \ f_2(x), \ f_3(x)\}$, with the above algorithm, the result is as follows:

let $x^{(0)} = 1.90$,   then $\hat{x} = 2.000251$, $f(\hat{x}) = -0.001255$,

let $x^{(0)} = 4.20$,   then $\hat{x} = 3.998995$, $f(\hat{x}) = 0.064320$,

let $x^{(0)} = 5.70$,   then $\hat{x} = 5.999987$, $f(\hat{x}) = 0.001248$.

Example 2.

$$\max_{x \in D}\{f_1(x), \ f_2(x), \ f_3(x)\} = 0$$

where $f_1(x) = x + 3$, $f_2(x) = \dfrac{5}{2}x + 1$, $f_3(x) = -\dfrac{1}{2}(x+4)$, $D = [-6, 0]$. The solutions of equation $x_1^* = -4.0$, $x_2^* = -\dfrac{5}{6}$. Let $\varepsilon = 1e - 4$, $p = 10^4$, $f(x) = \max_{x \in D}\{f_1(x), \ f_2(x), \ f_3(x)\}$, with the above algorithm, the result is as follows:

let $x^{(0)} = -1.90$,   then $\hat{x} = -1.2.000$, $f(\hat{x}) = 0.0000$,

let $x^{(0)} = -4.20$,   then $\hat{x} = -4.0000$, $f(\hat{x}) = 0.0000$.

Example 3.

$$\begin{pmatrix} \max\limits_{x \in D}\{f_{11}(x), \ f_{12}(x)\} \\ \max\limits_{x \in D}\{f_{21}(x), f_{22}(x)\} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

where $f_{11}(x) = x_1^2 + (x_2-1)^2 + x_2 - 1$, $f_{12}(x) = -x_1^2 - (x_2-1)^2 + x_2 + 1$, $f_{21}(x) = \sin(x_1 + x_2)$, $f_{22}(x) = x_1(x_1 + 3x_2) + x_2(x_2 - x_1) - 2$, $D = \left(\left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right)^T$. The solution of equations $x^* = (0,0)^T$. With the above algorithm ($\varepsilon = 1e-4$, $p = 10^4$) and let initial point $x^0 = (0.1, 0.1)^T$, the approximate solution is $\hat{x} = (-0.001861, 0.00027)^T$.

The examples above and more numerical results indicate that our algorithm works reliably and efficiently.

Conclusions

With the preceding study, to eliminate the defect of nondifferentiable problem with proper smoothing method is a kind of feasible method. Though the convergence speed of algorithm only reach linear convergence, the format of algorithm is simple, not need to compute derivative of nonsmooth functions $F(x)$, require low memory capacity and be easy to implement in computer. Another point in the paper is that, $\forall x^{(0)} \in D$, the iterative sequence (2.3) is convergent with suitable γ^k ($k = 0, 1, \ldots$), and so on. In articles [1, 2], algorithm requires to compute set of $\partial_B F(x^{(k)})$ and $\partial_b F(x^{(k)})$ which is hard to do in general (where the definition of $\partial_B F(x^{(k)})$, $\partial_b F(x^{(k)})$ is given in [1]. In articles [4, 6], one also requires to compute $\nabla f_{1j_1}(x^{(k)}), \ldots, f_{nj_n}(x^{(k)})$, where $j_i \in J_i(x)$, defined in [4, 6]. In the paper, we improve the above defects.

A direction for future research is to investigate the existence and uniqueness theorems for solution of nonsmooth equations (0.1) within a given region of n-space, and give sufficient condition which guarantees that a given multidimensional interval contains exact solution of the problem (0.1). Another area for future research is to find additional theoretical convergence results for the new iteration formula and to find a way to produce the $\{x^{(k)}\}$ sequence converging to x^* without a prior knowledge.

References

[1] QI L. Convergence analysis of some algorithms for solving nonsmooth equations // Mathmatics of Operations Research. 1993. Vol. 18. P. 227–224.

[2] SUN D., HAN J. Newton and quasi-Newton for a class of nonsmooth and related problems // SIAM. J. Optimization. 1997. Vol. 7. P. 463–480.

[3] NASHED M.Z., CHEN X. Convergence for Newton-like methods for singular operator equations using outer inverses // Numer. Math. 1993. Vol. 66. P. 235–257.

[4] GAO Y. Newton and quasi-Newton methods for nonsmooth equations of max-type function // Nature Magazine. 2001. Vol. 22. P. 61–63.

[5] SHEN Z., HUANG Z., WOLFE M.A. An interval maximum entropy method for a discrete minimax problem // Appl. Math. and Comput. 1997. Vol. 87. P. 49–68.

[6] GAO Y. Newton method and inexact-Newton method for solving singular nonsmooth equations // Math. Appl. 2002. Vol. 15. P. 57–61.

[7] Gao Y. Newton methods for solving nonsmooth equations via a new subdifferential // Mathmatics of Operations Research. 2001. Vol. 54.

[8] Ortega J., Rheinboldt W. Iterative solution of nonlinear equation in several variables. N.Y.: Acad. Press, 1970.

[9] Jaynes E.T. Information theory and statistical mechanics // Phys. Rev. 1957. Vol. 106. P. 620–630.

[10] Boltzmann L. Weitere studien über das warmegleichgewicht unter gasmolekulen // Wien. Akad. Sitz. 1972. Bd. 66. P. 275–370.