Title	An algorithm for solving nonlinear equations arising from nonsmooth optimization problems via some generalized Newton method (Mathematical Decision Making under uncertainty and ambiguity)
Author(s)	Saito, Seiji; Ishii, Hiroaki
Citation	数理解析研究所講究録 (2000), 1132: 64-67
Issue Date	2000-02
URL	http://hdl.handle.net/2433/63717
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

数理解析研究所講究録
1132 巻 2000 年 64-67

An algorithm for solving nonlinear equations arising from nonsmooth optimization problems via some generalized Newton method

大阪大学大学院工学研究科応用物理学専攻　齋藤誠慈 (Seiji Saito)
大阪大学大学院工学研究科応用物理学専攻　石井博昭 (Hiroaki Ishii)

E-mail: { saito-se, ishiiha }@ap.eng.osaka-u.ac.jp

Keywords : Nonlinear Optimization Problem, Generalized Newton Method, Mollifier

1 Introduction

Consider the following optimization problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x)=0, x\geq 0 \end{array} \qquad (1)$$

where $f : R^{n}\rightarrow R$ and $g : R^{n}\rightarrow R^{m}$. In this investigation we show a new algorithm for solving nonlinear equations arising from the problem under the conditions of semismoothness for f and g rather than they are both twice differentiable (see [QS]). Here f is called to be semismooth on R^{n} if f is locally Lipshitzian on R^{n} and at any $x\in R^{n}$ there exists the limit

$$\lim_{V\in\partial f(x+t\eta),\eta\rightarrow h,t\rightarrow+0}\{V\eta\}$$

for any $h\in R^{n}$. The set $\partial f(x)$ is the generalized Jacobian at $x\in R^{n}$ defined by Clarke ([C]) as follows:

$$\partial f(x)=co\{\lim_{\xi\rightarrow x,\xi\in D_{f}}Jf(\xi)\}$$

and D_{f} is the set of points at which f is differentiable. $Jf(\xi)$ means the usual Jacobian matrix of partial derivatives whenever ξ is a point at which the necessary partial derivatives exist.

If f is locally Lipshitzian then f is differentiable almost everywhere. f is semismooth if and only if, for any $V\in\partial f(x+h)$ such that $h\rightarrow 0$, it follows that

$$Vh-f^{'}(x;h)=o(\| h\|),$$

where $f^{'}(x;h)$ is the classic directional dervative defined by

$$f^{'}(x;h)=\lim_{t\rightarrow+0}\frac{f(x+th)-f(x)}{t}.$$

In [QS] the authors deal with solving $F(x)=0$, where $F : R^{n}\rightarrow R^{n}$. They assume the existence of the generalized Jacobian matrix V, which is nonsingular, and they get a theorem for the global convergence by some Newton method as follows:

Theorem([QS]). Suppose that F is semismooth on $S=\{x\in R^{n} :\| x-x^{0}\|\leq r\}$, where $x^{0}\in R^{n}, r>0$. Also suppose that for any $V\in\partial F(x)$, x, $y\in S$ such that V is nonsingular and satifies the following three conditions:

$$\| V^{-1}\|\leq\beta,$$
$$\| V(y-x)-F^{'}(x;y-x)\|\leq\gamma\| y-x\|,$$
$$\| F(y)-F(x)-F^{'}(x;y-x)\|\leq\delta\| y-x\|,$$

where

$$\alpha=\beta(\gamma+\delta)<1,\beta\| F(x^{0})\|\leq r(1-\alpha).$$

Then the iterates $x^{k+1}=x^{k}-V_{k}^{-1}F(x^{k}), V_{k}\in\partial F(x^{k})$, remain in S and converge to the unique solution x^{*} of $F(x)=0$ in S. Moreover it follows that

$$\| x^{k+1}-x^{*}\|\leq\frac{\alpha}{1-\alpha}\| x^{k}-x^{k-1}\|$$

for $k=1,2,\cdots$.

In an example we illustrate the convergence of our iterate without the assumption for the existence of the inverse generalized Jacobian matrices as well as we show the nonsingularity of the matrices, which are sufficiently closed to the one at the optimal solution x^{*}. Moreover we show the superlinear convergence of the above iterates via generalized Newton method.

2 Mollifier

In order to apply generalized Newton method with twice differentiability we consider the following function $\rho : R^{n}\rightarrow[0,\infty)$ due to Friedrichs and treat a convolution with $\rho_\eta : R^n \to R$ and a locally summable function, where $\eta > 0$ is sufficiently small. ρ is called to be a mollifier if ρ belongs to C^k-class, where $0 \le k \le \infty$, and has the compact support in R^n. There are many types of weight functions ρ. For example we denote

$$\rho(x) = \begin{cases} ce^{-\frac{1}{1-\|x\|^2}} & \text{for } \| x \| \le 1; \\ 0 & \text{for } \| x \| > 1. \end{cases}$$

Here c is a positive number such that $\int_{\|x\|\le 1} \rho(x)dx = 1$. For $\eta > 0$ and $\| x \| \le \eta$ we denote

$$\rho_\eta(x) = \left(\frac{1}{\eta}\right)^n \rho(\frac{x}{\eta}).$$

It is well-known that if $F : R^n \to R$ is in C^p-class, $p \ge 0$, then the convolution

$$F_\eta(x) = \int_{\|x-\xi\|\le\eta} F(\xi)\rho_\eta(x-\xi)d\xi$$

is in C^p-class and

$$F_\eta \to F \text{ as } \eta \to +0 \text{ in } C^p.$$

In what follows we define

$$\begin{aligned} f_\eta(x) &= (f * \rho_\eta)(x); \\ g_\eta(x) &= (g * \rho_\eta)(x), \end{aligned}$$

where

$$\begin{aligned} (f * \rho_\eta)(x) &= \int_{\|x-\xi\|\le\eta} f(\xi)\rho_\eta(x-\xi)d\xi; \\ (g * \rho_\eta)(x) &= \int_{\|x-\xi\|\le\eta} g(\xi)\rho_\eta(x-\xi)d\xi. \end{aligned}$$

We define

$$\mathcal{L}_\eta(w) = f_\eta(x) + y^T g_\eta(x) - z^T x,$$

where $w = (x^T, y^T, z^T)^T \in R^l, l = 2n + m$. And also we deal with nonlinear equations arising from the optimization problem as follows:

$$r_\eta(w) = \begin{pmatrix} \nabla_x \mathcal{L}_\eta(w)^T \\ g_\eta(x) \\ x_1 z_1 \\ \vdots \\ x_n z_n \end{pmatrix} = 0 \qquad (2)$$

where $r_\eta : R^l \to R^l$. And also we show a new algorithm for solving the above nonlinear equations.

Then it follows that $\mathcal{L}_\eta(\cdot)$ is twice continuously differentiable on R^l for any $\eta > 0$. In this paper we suppose that there exists a unique optimal solution of (1):

Assumption(A1). There exists a unique optimal solution x^* of (1).

3 Generalized Newton Method

We suppose that the following conditions hold in order to show a new algorithm by some generalized Newton method.

Assumption(A2). Suppsoe that there exists an $\eta_0 > 0$ as follows. Let $w^* = (x^{*T}, y^{*T}, z^{*T})^T \in R^l$ be a saddle point of $\mathcal{L}_\eta$ for $0 < \eta \le \eta_0$. Denote

$$D = \{w \in R^l : \| w - w^* \| \le \varepsilon\}$$

for $\varepsilon > 0$. Suppose that the following conditions (i)-(v) hold for $0 < \eta \le \eta_0$:

(i) Both f_η and g_η are twice continuously differentiable and their derivatives satisfy the following Lipschitz conditions:

$$\begin{aligned} &\| \nabla_x^2 f_\eta(x_1) - \nabla_x^2 f_\eta(x_2) \| \le L \| x_1 - x_2 \|, \\ &\| \nabla_x^2(y_1^T g_\eta(x_1)) - \nabla_x^2(y_2^T g_\eta(x_2)) \| \\ &\qquad \le L(\| y_1 - y_2 \| + \| x_1 - x_2 \|) \end{aligned}$$

for $(x_1, y_1, z_1), (x_2, y_2, z_2) \in D$;

(ii) Let $I^* = \{i : x_i^* = 0\}$. When $i \in I^*$, then $z_i^* > 0$, where z_i^* is the i-th element of z^* and x_i^* the i-th element of x^* ;

(iii) The set
$\{\nabla_x g_i^\eta(x^*)^T : i = 1, \cdots, m\} \cup$
$\{e_i = (0, \cdots, 0, \overset{i}{1}, 0, \cdots, 0)^T : i \in I^*\}$
is linearly independent for $w \in D$, where $g_\eta = (g_1^\eta, \cdots, g_m^\eta)^T$;

(iv) $\nabla_x^2 \mathcal{L}_\eta(w^*) > 0$;

(v) Let $0 < \delta < 1$ and $0 < \varepsilon < 1$ satisfy

$$\begin{aligned} &B_0\delta + \varepsilon L_0 < \sqrt{2} - 1, \\ &1 - M\varepsilon R_1 > 0. \end{aligned}$$

Let an integer $N \ge 3$ be

$$\left(\frac{B_0\delta+\varepsilon L_0+1}{(B_0\delta+\varepsilon L_0)^{-1}-1}\right)^{N-2} \leq \frac{\delta-\delta^2}{B_0(\delta^2+1)\varepsilon\left[L+1+\sqrt{(L+1)^2+\frac{8(n+1)L(\delta-\delta^2)}{B_0(\delta^2+1)}}\right]}.$$

Here $R_2 = \sup_{w\in D,\eta\in(0,\eta_0]} \| \nabla_w r_\eta(w) \|$,

$$B_0 = \frac{R_1(M+\varepsilon R_2)}{1-M\varepsilon R_1},$$
$$M = L+1,$$
$$R_1 = \sup_{w\in D,\eta\in(0,\eta_0]} \| (\nabla_w r_\eta(w))^{-1} \|,$$

$$L_0 = \sup_{w\in D,\eta\in(0,\eta_0]} \| \nabla_x^2(y^T g_\eta(x)) \| \frac{R_1(1+\delta^2)}{1-M\varepsilon R_1}.$$

By applying the generalized Newton method, we get the following algorithm of solving $r_\eta(w) = 0$. Denote norms by

$$\| x \| = \sum_i |x_i|,$$
$$\| A \| = \sum_{i,j} |a_{ij}|$$

for $x \in R^n$ and $A = (a_{ij}), 1 \leq i,j \leq l$, respectively.

Algorithm. Find a sequence

$$\{w^{(k)} = (x^{(k)T}, y^{(k)T}, z^{(k)T})^T : k = 1,2,\cdots\}$$

such that

$$w^{(k+1)} = w^{(k)} - V^{(k)}(k) r_\eta(w^{(k)}),$$

where $V^{(k)}(k)$ is closed to the Jacobian matrix $(\nabla_w r_\eta(w^*))^{-1}$ for the sufficiently large k. We construct two $l\times l-$ matrices

$$\{U(k) : k = 1,2,\cdots\};$$
$$\{V^{(p)}(k) : k = 1,2,\cdots; p = 1,2,\cdots,k\}.$$

Let $\eta = \eta_0/k$. Denote $r_\eta = r_k$.

Choose $w^{(1)} \in D$. For $k = 1,2,\cdots$, do the following steps.

Step 1. For $1 \leq k \leq N$, find $U(k)$ such that

$$\| \nabla_w r_k(w^{(k)}) U(k) - I \| \leq \delta,$$

where I is the identity matrix. Put

$$V^{(0)} = U(k)$$

and go to Step 2.

For $k \geq N+1, N+2, \cdots$, put

$$U(k) = V^{(k-1)}(k-1).$$

Go to Step 2.

Step 2. For $p = 1,2,\cdots,k$, compute $\{V^{(p)}(k)\}$ such that

$$V^{(p)}(k) = V^{(p-1)}(k)[2I - \nabla_w r_k(w^{(k)}) V^{(p-1)}(k)].$$

Step 3. Compute

$$w^{(k+1)} = w^{(k)} - V^{(k)}(k) r_k(w^{(k)}).$$

Go to Step 1.

In [NI] the authors show an algorithm by generalized Newton method, which is applied by the idea of C.Neumann expansion. In the algorithm one finds an matrix $U(k)$ satisfying the inequality of Step1 for each k. In the above algorithm we find $U(k)$ with $k = 1,\cdots,N$.

The following theorem can be proved by the mathematical induction.

Theorem 1. We have

$$\| I - \nabla_w r_k(w^{(k)}) V^{(k)}(k) \| \leq \delta^{2^k}$$

for $k = 1,2,\cdots$.

Our iterate by the generalized Newton method gives the superlinear convergence to the optimal solution of (1) as follows:

Theorem 2. It follows that

$$\| w^{(k+1)} - w^{(k)} \| \leq B_0\delta^{2^k} \| w^{(k)} - w^* \| + L_0 \| w^{(k)} - w^* \|^2$$

for $k \geq 1$.

In the following example we show our main results which are important estimates and the superlinear convergence by the generaliezd Newton method.

Example. Consider an optimization problem:

$$\begin{array}{lll} \text{minimize} & f(x) = x^3(|\sin\frac{\pi}{x}| + x^4) & \\ \text{subject to} & 0 \le x \le 1. & (P) \end{array}$$

Denote

$$\begin{aligned} g(x) &= x_1 - 1 + x_2 = 0; \\ x &= (x_1, x_2)^T \ge 0; \\ f_\eta(x) &= \int_{R^2} f(x_1)\rho_\eta(x-\xi)d\xi; \\ \mathcal{L}_\eta(w) &= f_\eta(x) + yg(x) - (z_1x_1 + z_2x_2); \\ \xi &= (\xi_1, \xi_2)^T; \\ w &= (x^T, y, z^T)^T, \quad z = (z_1, z_2)^T \ge 0. \end{aligned}$$

In the same way as in Section 1 we get

$$r_\eta(w) = \begin{pmatrix} \nabla_x \mathcal{L}_\eta(w)^T \\ g(x) \\ x_1 z_1 \\ x_2 z_2 \end{pmatrix} = 0;$$

$$\nabla_x \mathcal{L}_\eta(w)^T = \begin{pmatrix} \frac{\partial f_\eta}{\partial x_1} + y - z_1 \\ \frac{\partial f_\eta}{\partial x_2} + y - z_2 \end{pmatrix}.$$

We have the optimal solution of (P) : $w^* = (0, 1, y^*, z_1^*, z_2^*)$ satisfying

$$\exists y^* \in R: \quad z_i^* = \frac{\partial f_\eta}{\partial x_i}(0,1) - y^* > 0,$$

where $i = 1, 2$. Then (A1) and Condition(ii) in (A2) hold for sufficiently small δ, ε . Since there exists an L such that

$$2 \sup_{w \in D, 0 < \eta \le 1} \| \nabla_x (\nabla_x^2 \rho_\eta(w))_{ij} \| \le L$$

and $\max_{0 \le x \le 1} |f(x)| \le 1$, where the above left-hand term is the gradient of the ij-element of $\nabla_x^2 \rho_\eta(w)$, we have

$$\| \nabla_x^2 f_\eta(x_1) - \nabla_x^2 f_\eta(x_2) \| \le L \| x_1 - x_2 \| .$$

It can be easily seen that (iv) in (A2) is satisfied. Since $\nabla_x g(x) = (1, 1)$ and $\{e_i : i \in I^*\} = \{(1, 0)^T\}$, Condition (iii) in (A2) holds. We get the following estimates:

$$\begin{aligned} \| \nabla_w r_k(w^{(k)}) - \nabla_w r_k(w^*) \| &\le M \| w^{(k)} - w^* \|; \\ \| \nabla_w r_k(w^{(k)}) \| &\le R_2 + M\varepsilon; \\ \| I - \nabla_w r_k(w^{(k)})(\nabla_w r(w^*))^{-1} \| &\le \varepsilon R_2 M; \\ \| (\nabla_w r_k(w^{(k)}))^{-1} \| &\le \frac{R_2}{1 - M\varepsilon R_2}; \\ \| V^{(k)} - (\nabla_w r_k(w^{(k)}))^{-1} \| &\le \frac{R_2 \delta^{2^k}}{1 - M\varepsilon R_2}; \end{aligned}$$

$$\begin{aligned} &\| V^{(k)}(k) - (\nabla_w r_k(w^{(k)}))^{-1} \| \\ &\le R_3 \delta^{2^k} + R_3 R_2 M \| w^{(k)} - w^* \|, \end{aligned}$$

where $R_3 = \frac{R_2}{1 - M\varepsilon R_2}$. Since $r_{+0}(w^*) = 0$, we have, as $k \to \infty$,

$$\begin{aligned} &\| w^{(k+1)} - w^* \| \\ &\le B_0 \delta^{2^k} \| w^{(k)} - w^* \| + L_0 \| w^{(k)} - w^* \|^2 . \end{aligned}$$

Hence we have the superlinear convergence to the optimal solution of (P).

References

[C] F.H.Clarke: " Optimization and Nonsmooth Analysis," *Wiely*, New York, 1983.

[NI]T. Noda and H. Ishii: "A General Newton Method for Systems of Nonlinear Equations II" *Mathematica Japonica*, Vol.48, pp. 447-452,1998.

[SI] S.Saito and H.Ishii:" On solving equations arising from optimizations problems by some generalized Newton method"(to be submitted in *J. Math. Anal. Appl.*)

[QS] L.Qi and J.Sun: "A nonsmooth version of Newton's method," *Math. Prog.*, Vol. 58, pp,353-367, 1993.