Journal of Mathematical Research & Exposition,
Vol.12, No.2, May 1992

Toward Newton Method in Nondifferentiable Optimization*

E.A. Nurminski
(Inst. of Appl. Math., Far Eastern Branch Ac. Sci. USSR, Vladivostok, USSR)

Abstract. For optimization problems

$$\min_{x \in E} f(x)$$

where $f(x)$ is a convex but possibly nondifferentiable function, we propose an analogue of Newton Method with theoretical superlinear rate of convergence and discuss its implementation.

One of the main subjects of current research in nondifferentiable optimization (NDO) is the development of an analogue to Newton Method (NM) which is known for its remarkable computational properties, but applicable only to smooth strongly convex twice differentiable function. The studies [1–3] were based on different generalizations of the Hessian contributed to this aim, but there were no detailed investigations of computational properties of the proposed algorithms.

In this paper the conceptual version of the algorithm following the general ideas of NM and possessing an attractive rate of convergence is suggested and its implementation is discussed.

1. Algorithm

What follows is based on the relationship [4,5]

$$-\min_{x \in E} f(x) = \inf_{0 \in \partial_\varepsilon f(0)} \varepsilon = \varepsilon_*. \tag{1}$$

This establishes the correspondence between minima of convex function $f(x)$ on a finite dimensional Euclidean space $\mathbf{E}$ and the minimal root of an ε−subdifferential mapping $\partial_\varepsilon f(0)$. The inner product of vectors x, y from $\mathbf{E}$ will be denoted by xy.

The ε-subdifferential multivalued mapping is defined as usual [6]:

$$\partial_\varepsilon f(0) = \{g : f(x) \geq gx - \varepsilon,\ x \in \mathbf{E}\}.$$

Without loss of generality it is assumed that $f(0) = 0$ and, hence, $\partial_\varepsilon f(0)$ is not empty. Technical considerations require, however, the stronger assumption $0 \in$ int dom f, i.e.,

*Received May 12, 1990.

boundness of f in the neighbourhood of the origin, this will be supposed further on. One may expect that htis assumption be relaxed.

Due to monotonicity with respect ot the set-theoretical inclusion the right- hand part of (1) is equivalent to solving

$$\inf_{\|p\|\leq 1} \sup_{g\in\partial_\varepsilon f(0)} pg = \inf_{\|p\|\leq 1} \phi_p(\varepsilon) = \phi(\varepsilon) = 0. \tag{2}$$

It is easy to demonstrate from the definitions that $\phi_p(\varepsilon)$ is concave with respect to ε for any p and consequently $\phi(\varepsilon)$ is concave as well. By the known rules of sub(super)differential calculus its superdifferential can be determined as

$$\partial\phi(\varepsilon) = \ \mathrm{co}\{\partial\phi_p(\varepsilon), p \in p(\varepsilon)\},$$

where

$$P(\varepsilon) = \{p: \quad \phi_p(\varepsilon) = \phi(\varepsilon), \| p \| \leq 1\}$$

and $\partial\phi_p(\varepsilon)$ denotes the superdifferential of $\phi_p(\varepsilon)$ for a fixed p.

In its turn the set $\partial\phi_p(\varepsilon)$ can be found as

$$\partial\phi(\varepsilon) = \bigcap\nolimits_{\delta\geq 0} \mathrm{co}\ \{u: \ u(f(p/u)+\varepsilon) \geq \phi_p(\varepsilon) - \delta, u > 0\}. \tag{3}$$

This can be derived from [7] or obtained form the first principles taking into account that

$$\phi_p(\varepsilon) = \inf_{u>0} u\{f(p/u)+\varepsilon\} \tag{4}$$

Consider the following algorithm for solving (2). Let ε_0 be such that $\phi(\varepsilon_0) < 0$. Define in a recursive manner the sequence $\{\varepsilon_k\}$:

$$\varepsilon_{k+1} = \varepsilon_k - \phi(\varepsilon_k)u_k^{-1}, \quad k = 0, 1, \cdots, \tag{5}$$

where

$$u_k \in \partial\phi_{p^k}(\varepsilon_k, \quad p^k \in P(\varepsilon_k). \tag{6}$$

If for some $k, u_k = 0$, then it means that (1) is unbounded.

As u_k is in a certain sense the directional derivative of $\phi(\varepsilon_k)$, the algorithm (5)–(6) is the NM for solving (2), which is equivalent to (1) with respect to the optimal value.

Convergence properties of the algorithm are described by the following theorem.

Theorem *Let (1) have a k attainable and finite solution ε_* with respect to x. Then*

i. *The algorithm (5)–(6) converges, i.e.,*

$$\lim_{k\to\infty} \varepsilon_k = \varepsilon_*.$$

ii. *The rate of convergence is superlinear:*

$$| \varepsilon_{k+1} - \varepsilon_* | \leq \lambda_k | \varepsilon_k - \varepsilon_* |, \quad \lambda_k \to 0.$$

Proof We show first that $\phi(\varepsilon) \leq 0$ for any k. Indeed by the definition

$$\phi(\varepsilon) \leq \phi(\varepsilon_k) + u_k(\varepsilon - \varepsilon_k)$$

and taking $\varepsilon = \varepsilon_{k+1}$ yields

$$\phi(\varepsilon_{k+1}) \leq \phi(\varepsilon_k) - u_k \phi(\varepsilon_k) u_k^{-1} = 0.$$

It immediately follows that the sequence $\{\varepsilon_k\}$ is monotone and bounded:

$$\varepsilon_k \leq \varepsilon_{k+1} \leq \varepsilon_*.$$

From the convergence of $\{\varepsilon_k\}$

$$\lim_{k\to\infty} (\varepsilon_k - \varepsilon_{k+1}) = \lim_{k\to\infty} \phi(\varepsilon_k) u_k^{-1} = 0.$$

Superdifferential $\partial\phi$ is monotone:

$$(\partial\phi(\varepsilon') - \partial\phi(\varepsilon))(\varepsilon' - \varepsilon) \leq 0$$

for any ε', $\varepsilon < 0$ or $\partial\phi(\varepsilon') \leq \partial\phi(\varepsilon)$ for for $\varepsilon' > \varepsilon$. Consequently $u_k \geq u_* > 0$ for any $u_* \in \partial\phi(\varepsilon_*)$ and $\lim_{k\to\infty} \phi(\varepsilon_k) = 0$. Therefore $\lim_{k\to\infty} \varepsilon_k = \varepsilon_*$, now we estimate the rate of convergence:

$$\begin{aligned} \varepsilon_* - \varepsilon_{k+1} &\leq \varepsilon_k - \phi(\varepsilon_k) u_k^{-1} - \varepsilon_{k+1} = \varepsilon_k - \phi(\varepsilon_k) u_*^{-1} - \varepsilon_k + \phi(\varepsilon_k) u_k^{-1} \\ &= -\phi(\varepsilon_k)(u_*^{-1} - u_k^{-1}). \end{aligned}$$

Taking inf with respect to $u_* \in \partial\phi(\varepsilon_*)$ yields

$$\varepsilon_* - \varepsilon_{k+1} \leq -\phi(\varepsilon_k)\{\inf_{u_* \in \partial\phi(\varepsilon_*)} u_*^{-1} - u_k^{-1}\}. \tag{7}$$

It follows from the upper semicontinuity of $\partial\phi(\varepsilon)$ that when $\varepsilon_k \to \varepsilon_*$

$$0 \leq \lim_{k\to\infty} \mu_k = \lim_{k\to\infty} \{\inf_{u_* \in \partial\phi(\varepsilon_*)} u_*^{-1} - u_k^{-1}\} \leq \inf_{u_* \in \partial\phi(\varepsilon_*)} u_*^{-1} - \inf_{u_* \in \delta\phi(\varepsilon_*)} u_*^{-1} = 0.$$

From the Lipschitz property of $\partial\phi(\varepsilon)$[8, 9, etc.]

$$0 \leq -\phi(\varepsilon_k) = \phi(\varepsilon_*) - \phi(\varepsilon_k) \leq L(\varepsilon_* - \varepsilon_k),$$

combining this with (7), we obtain

$$\varepsilon_* - \varepsilon_{k+1} \leq L\mu_k(\varepsilon_* - \varepsilon_k) = \lambda_k(\varepsilon_* - \varepsilon_k)$$

with $\lambda_k \to 0$. The proof is completed. □

2. Implementation of the algorithm

The implementation of the algorithm (5)–(6) requires that the practical solution of two problems: finding $u \in \partial\phi_p(\varepsilon)$ and calculating $\phi(\varepsilon)$. Since $h(u) = uf(p/u) + u\varepsilon$ is convex [10, p.69] the problem (4) of finding $\phi_p(\varepsilon)$ for a fixed p is well-defined and for its solution within an arbitrary precision the number of line-search alfgorithms with good computational records can be applied. Solution u_* of this problem provides a supergradient of $\phi(\varepsilon_p)$.

The simplest case is $h(u)-$differentiable at minimal point u_*. For notational simplicity we drop the dependence on p and ε. Then

$$pg^*/u_* = f(p/u_*) + \varepsilon,$$

where $g^* = f'(p/u_*)$ is the gradient of f at the point p/u_*. It is easy to show that $g^* \in \partial_\varepsilon f(0)$. Moreover it follows from (8) that

$$pg^* = u_*(f(p/u_*)) + \varepsilon) = \inf_{u>0} u(f(p/u_*) + \varepsilon) = \sup_{g \in \partial_\varepsilon f(0)} pg,$$

i.e., g^* yields a maxima of the linear form pg on $\partial_\varepsilon f(0)$.

In nonsmooth case (8) may not hold for arbitrary $g^* \in \partial f(p/u_*)$. However, such vector can be easily obtained by the following means: solution (4) is commonly constructed as a sequence of nested intervals $[u_k^l,\ u_k^r]$ containing u_* and converging towards this point. Directional dervatives at the ends of the intervals with respect to converging interval directions have opposite signs:

$$f(p/u_k^l) + \varepsilon - pg_k^l/u_k^l < 0,\ g_k^l \in \partial f(p/u_k^l), \tag{8}$$

$$f(p/u_k^l) = \varepsilon - pg_k^r/u_k^r > 0,\ g_k^r \in \partial f(p/u_k^l). \tag{9}$$

If so, there exists $\alpha_k \in [0,\ 1]$ such that the weighted sum of (9), (10) equals to zero:

$$\varepsilon + f(p/u_k^r) + \alpha_k(f(p/u_k^l) - f(p/u_k^r) - \alpha_k g_k^l p/u_k^l(1-\alpha_k)g_k^r p/u_k^r = 0,$$

and for any limit point g^* of the sequence $\{\alpha_k g_k^l + (1-\alpha_k)g_k^r\}$,

$$g^* \in \partial f(p/u_*),$$

due to the continuity of f and uppersemicontinuity of ∂f

$$\varepsilon + f(p/u_*) - pg^*/u_* = 0$$

and the same conclusions hold for the smooth case.

Calculation of $\phi(\varepsilon)$ amounts to finding the shortest vector in the set $\partial_\varepsilon f(0)$. For this purpose the iterative process based on supremas of linear forms on $\partial_\varepsilon f(0)$ can be applied. When such supremas are known the shortest vector can be obtained, for instance, by the algorithms of [11].

References

[1] A. Auslender, *On the differential properties of the support function of the ε−subdifferential of a convex function*, Math. Program., **24:3**(1982), 257–268.

[2] C. Lemarechal and J. Zowe, *Some remarks on the construction of higher order algorithms in convex optimization*, Appl. Math. Optimizat., **10:1**(1983), 51–68.

[3] M. Gaudioso, *An algorithm for convex NDO based on properties of the contour lines of convex quadratic function*, in *Nondifferentiable Optimization: Motivations and Applications* , V.F. Demyanov and D. Pallaschke eds., Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1985, 190-196.

[4] E. A. Nurminski, *E−subdifferanciallI'nye otobrazheniya ivypuklaya optimizaciya,* Kibernetika, **85:6**(1985), 63– 65. (in Russian).

[5] E. A. Nurminskii, *Globalnye svoistva ε−subgradientnyh otobrazhenii,* Kibernetika, **1**(1986), 120–122.(in Russian).

[6] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1972.

[7] A. A. Skokov Obobschennaya Differenciuemost Funkcii Maximuma, Kibernetika, 5(1973) 47-62 (in Russian).

[8] E. A. Nurminskii, *O nepreryvnosti ε-subgradientnyh otobrazhenii* , Kibernetika, **5**(1977), 148–149. (in Russian).

[9] J-B. Hiriart-Urruty,*Lipschitz r−continuity of the approximate subdifferential of a convex function*, Math. Scand., **47**(1980), 123–134.

[10] N. Bourbaki, *Funkcii Deistvitelnogo Peremennogo*, Nauka,Moskva, 1965.

[11] V.F. Demyanov, *V. N. Malozemov, Vvedenie v Minimaks*,Nauka, Moskva, 1972.