

GENERALIZED NEWTON METHOD BASED ON GRAPHICAL DERIVATIVES

BORIS MORDUKHOVICH

Wayne State University

Talk given at the

Fields Industrial Optimization Seminar

Toronto, ON, Canada

April 2011

Supported by [NSF grant](#) DMS-1007132

NONSMOOTH EQUATIONS

are those defined by

$$H(x) = 0$$

where $H: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a continuous vector-valued mapping. Vast majority of applications in optimization, variational inequalities, complementarity and equilibrium problems reduce to nonlinear equations with nonsmooth mappings H . This also includes Robinson's formalism of generalized equations, which can be written in the standard equation form (via, e.g., the projection operator) while with intrinsically nonsmooth mappings H

SMOOTH NEWTON'S METHOD

applies to nonlinear equations with $H \in C^1$.

Newton's iterations

$$x^{k+1} := x^k + d^k \quad \text{for all } k = 0, 1, 2, \dots$$

where $x^0 \in \mathbb{R}^n$ is a given starting point and where $d^k \in \mathbb{R}^n$ is a solution to the linear system of equations

$$H'(x^k)d = -H(x^k)$$

A lot is known about the classical Newton's method: fast local convergence, etc

B-DIFFERENTIABLE NEWTON'S METHOD

developed by Pang et al. Assuming that H is **directionally differentiable**, the direction d^k in the iteration scheme

$$x^{k+1} := x^k + d^k, \quad k = 0, 1, 2, \dots$$

is a solution to the **subproblem**

$$H'(x^k; d) = -H(x^k)$$

via the directional derivative $H'(x^k; d)$. Known results for the B -differentiable Newton's method are established under pretty restrictive assumptions; in particular, under the **strong differentiability** at the solution point

SEMISMOOTH NEWTON'S METHODS

developed by Kummer, Qi and Sun, et al. for the Lipschitz continuous and directionally differentiable mappings H under the so-called semismoothness. There are two basic versions following the scheme

$$A_k d = -H(x^k)$$

for determining the iterative direction, where either $A_k \in \partial_C H(x^k)$ (element of the Clarke's generalized Jacobian) or $A_k \in \partial_B H(x^k)$ (element of the B -subdifferential). These constructions will be discussed below. The B -subdifferential method is better but it has troubles with solvability of the generalized Newton equation.

We develop another approach by using advanced tools of variational analysis involving graphical derivatives

TOOLS OF VARIATIONAL ANALYSIS

Given a set-valued map $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, its **outer limit** at \bar{x} is

$$\text{Lim sup}_{x \rightarrow \bar{x}} F(x) := \left\{ y \in \mathbb{R}^m \mid \exists x_k \rightarrow \bar{x}, y_k \rightarrow y \text{ s.t. } y_k \in F(x_k), \forall k \right\}$$

The **graphical derivative** of F at $(\bar{x}, \bar{y}) \in \text{gph } F$ is

$$DF(\bar{x}, \bar{y})(u) := \text{Lim sup}_{t \downarrow 0} \frac{F(\bar{x} + tu) - \bar{y}}{t}, \quad u \in \mathbb{R}^n$$

The **coderivative** of F at $(\bar{x}, \bar{y}) \in \text{gph } F$ is

$$D^*F(\bar{x}, \bar{y})(v) := \left\{ w \in \mathbb{R}^n \mid (w, -v) \in N((\bar{x}, \bar{y}); \text{gph } F) \right\}, \quad v \in \mathbb{R}^m$$

defined via the **normal cone**

$$N(\bar{z}; \Omega) = \text{Lim sup}_{z \rightarrow \bar{z}} \left[\text{cone}(z - \Pi(z; \Omega)) \right]$$

for $\Omega = \text{gph } F$ and $\bar{z} = (\bar{x}, \bar{y})$, where $\Pi(xz; \Omega)$ stands for the **Euclidean projection**

PROPERTIES OF DERIVATIVES AND CODERIVATIVES

- If F is smooth, then

$$DF(\bar{x})(u) = \nabla F(\bar{x})u, \quad D^*F(\bar{x})(v) = \nabla F(\bar{x})^T v$$

- If F is directionally differentiable, then

$$DF(\bar{x})(u) = \left\{ F'(\bar{x}; u) \right\}, \quad u \in \mathbb{R}^n$$

- Both derivatives and coderivatives satisfy calculus rules while full calculus has been developed for coderivatives in general non-smooth and nonconvex settings

THE GENERALIZED NEWTON ALGORITHM

We propose to the following generalized Newton equation (GNE)

$$-H(x^k) \in DH(x^k)(d^k), \quad k = 0, 1, 2, \dots$$

to find directions d^k . Then

THE ALGORITHM

- **STEP 1:** Choose a starting point $x^0 \in \mathbb{R}^n$
- **STEP 2:** Check a suitable termination criterion
- **STEP 3:** Compute $d^k \in \mathbb{R}^n$ such that the GNE holds
- **STEP 4:** Set $x^{k+1} := x^k + d^k$, $k \leftarrow k + 1$, and go to Step 1

METRIC REGULARITY

A set-valued mapping $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is **metrically regular** around $(\bar{x}, \bar{y}) \in \text{gph } F$ if there are neighborhoods U of \bar{x} and V of \bar{y} as well as a number $\mu > 0$ such that

$$\text{dist}(x; F^{-1}(y)) \leq \mu \text{dist}(y; F(x)), \quad \forall x \in U, y \in V$$

This property (related to **error bounds**, **Hoffman estimates**, is crucial for many aspects of variational analysis and optimization.

There is the **coderivative criterion** [Mor84] for it:

F is metrically regular around (\bar{x}, \bar{y}) if and only if

$$\ker D^*F(\bar{x}, \bar{y})(z) = 0 \quad \text{i.e.} \quad 0 \in D^*F(\bar{x}, \bar{y})(z) \implies z = 0$$

SOLVABILITY OF THE GENERALIZED NEWTON EQUATION

THEOREM. Assume that $H: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is **metrically regular** around \bar{x} , i.e., $\ker D^*H(\bar{x}) = \{0\}$. Then there is a constant $\varepsilon > 0$ such that for all $x \in B_\varepsilon(\bar{x})$ the equation

$$-H(x) \in DH(x)(d)$$

admits a solution $d \in \mathbb{R}^n$. Furthermore, the set $S(x)$ of solutions to it is **computed by**

$$S(x) = \limsup_{t \downarrow 0, h \rightarrow -H(x)} \frac{H^{-1}(H(x) + th) - x}{t} \neq \emptyset$$

BASIC CONDITIONS FOR THE ALGORITHM

(H1) There exist a constant $C > 0$, a neighborhood U of \bar{x} , and a neighborhood V of the origin in \mathbb{R}^n such that the following holds:

For all $x \in U$, $z \in V$, and for any $d \in \mathbb{R}^n$ with $-H(x) \in DH(x)(d)$ there is a vector $w \in DH(x)(z)$ such that

$$C\|d - z\| \leq \|w + H(x)\| + o(\|x - \bar{x}\|)$$

(H2) There exists a neighborhood U of \bar{x} such that for all $v \in DH(x)(\bar{x} - x)$ we have

$$\|H(x) - H(\bar{x}) + v\| = o(\|x - \bar{x}\|)$$

SUPERLINEAR LOCAL CONVERGENCE

THEOREM. Let $\bar{x} \in \mathbb{R}^n$ be a solution to the nonsmooth equation $H(x) = 0$ for which the underlying mapping $H: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is **metrically regular** around \bar{x} and assumptions **(H1)** and **(H2)** are satisfied. Then there is a number $\varepsilon > 0$ such that for all $x^0 \in B_\varepsilon(\bar{x})$ the following assertions hold:

- (i)** The Algorithm is **well defined** and generates a sequence $\{x^k\}$ **converging to \bar{x}**
- (ii)** The rate of convergence $x^k \rightarrow \bar{x}$ is at least **superlinear**

SUFFICIENT CONDITIONS FOR (H1)

We say that $H: \mathbb{R}^m \rightarrow \mathbb{R}^m$ is **directionally bounded** around \bar{z} if

$$\limsup_{t \downarrow 0} \left\| \frac{H(x + tz) - H(x)}{t} \right\| < \infty, \quad \forall z \in \mathbb{R}^n \text{ and } x \text{ near } \bar{x}$$

This holds when H is either **directionally differentiable** or **locally Lipschitzian** but **not vice versa**

PROPOSITION. Condition (H1) is satisfied if H is **one-to-one**, **directionally bounded**, and **metrically regular**

There are many further sufficient conditions for the fulfillment of all the above assumptions. Consider those related to the previous developments on the generalized Newton method

***B*-SUBDIFFERENTIAL AND GENERALIZED JACOBIAN**

Given $H: \mathbb{R}^n \rightarrow \mathbb{R}^n$ **locally Lipschitzian** around \bar{x} , define

$$S_H := \{x \in \mathbb{R}^n \mid H \text{ is differentiable at } x\},$$

the ***B*-subdifferential** of H at \bar{x}

$$\partial_B H(\bar{x}) := \left\{ \lim_{k \rightarrow \infty} H'(x^k) \mid \exists \{x^k\} \subset S_H, x^k \rightarrow \bar{x} \right\}$$

and the **generalized Jacobian**

$$\partial_C H(\bar{x}) := \text{co}\{\partial_B H(\bar{x})\}$$

We have the **strict inclusion**

$$DH(\bar{x})(z) \subset \partial_C H(\bar{x})z, \quad \forall z \in \mathbb{R}^n.$$

COROLLARY. Let H be **locally Lipschitzian** around \bar{x} . Then condition **(H1)** is satisfied if **$\det A \neq 0$ for all $A \in \partial_C H(\bar{x})$**

SEMISMOOTHNESS AND CONDITION (H2)

A mapping $H: \mathbb{R}^n \rightarrow \mathbb{R}^m$, locally Lipschitzian and directionally differentiable around \bar{x} , is **semismooth** at this point if the limit

$$\lim_{\substack{h \rightarrow z, t \downarrow 0 \\ A \in \partial_C H(\bar{x} + th)}} \{Ah\}$$

exists for all $z \in \mathbb{R}^n$

Semismoothness always implies (H2) but not vice versa.

Both conditions (H1) and (H2) hold for non-semismooth, Lipschitzian, metrically regular, and one-to-one mappings

A number of examples are available in both **Lipschitzian** and **non-Lipschitzian** cases

APPLICATION TO B-DIFFERENTIABLE NEWTON METHOD

The subdifferential of $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is

$$\partial f := \operatorname{Lim\,sup}_{x \rightarrow \bar{x}} \hat{\partial} f(x)$$

where

$$\hat{\partial} f(x) := \left\{ p \in \mathbb{R}^n \mid \liminf_{u \rightarrow x} \frac{f(u) - f(x) - \langle p, u - x \rangle}{\|u - x\|} \geq 0 \right\}$$

THEOREM. Let $H: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be semismooth, one-to-one, and metrically regular around a reference solution \bar{x} to $H(x) = 0$ satisfying

$$0 \in \partial \langle z, H \rangle(\bar{x}) \implies z = 0$$

Then the B -differentiable Newton method is well defined and converges at least superlinearly to the solution \bar{x}

GLOBAL CONVERGENCE

THEOREM. Let x^0 be a starting point of the Algorithm, and

$$\Omega := \{x \in \mathbb{R}^n \mid \|x - x^0\| \leq r\}, \quad r > 0$$

Assume that:

- The mapping $H: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is **metrically regular** on Ω with modulus $\mu > 0$
- The set-valued map $DH(x)(z)$ uniformly on Ω converges to $\{0\}$ as $z \rightarrow 0$ in the sense that: for all $\varepsilon > 0$ there is $\delta > 0$ with

$$\|w\| \leq \varepsilon, \quad \forall w \in DH(x)(z), \quad \|z\| \leq \delta, \quad x \in \Omega$$

- There is $\alpha \in (0, 1/\mu)$ such that

$$\mu \|H(x^0)\| \leq r(1 - \alpha\mu)$$

and for all $x, y \in \Omega$ we have the estimate

$$\|H(x) - H(y) - v\| \leq \alpha \|x - y\|, \quad \forall v \in DH(x)(y - x)$$

Then the Algorithm is well defined, the sequence of iterates $\{x^k\}$ remains in Ω and converges to a solution $\bar{x} \in \Omega$. Moreover, we have the error estimate

$$\|x^k - \bar{x}\| \leq \frac{\alpha\mu}{1 - \alpha\mu} \|x^k - x^{k-1}\|$$

REFERENCES

1. T. HOHEISEL, C. KANZOW, B. S. MORDUKHOVICH and H. PHAN, Generalized Newton's methods based on graphical derivatives, **SIAM J. CONTROL OPTIM.**, to appear.
2. B. S. MORDUKHOVICH, **VARIATIONAL ANALYSIS AND GENERALIZED DIFFERENTIATION, I: BASIC THEORY, II: APPLICATIONS**, Grundlehren Series (Fundamental Principles of Mathematical Sciences), **330, 331**, Springer, 2006.
3. R. T. ROCKAFELLAR and R. J-B. WETS, **VARIATIONAL ANALYSIS**, Grundlehren Series (Fundamental Principles of Mathematical Sciences), **317**, Springer, 1998.