GENERALIZED NEWTON ALGORITHMS FOR TILT-STABLE MINIMIZERS

BORIS MORDUKHOVICH

Wayne State University

talk given in Optimization Seminar, CMM

based on joint work with Ebrahim Sarabi (Miami Univ., OH)

Santiago, Chile, September 2020

Supported by NSF and Air Force grants

CLASSICAL NEWTON METHOD

to solve equations $f(x) = 0$ for $\mathcal{C}^1$-smooth mappings $f\colon I\!R^n \to I\!R^m$ is written in the case of unconstrained optimization

$$\text{minimize } \ \varphi(x) \ \text{ subject to } \ x \in I\!R^n \ \text{ for } \ \varphi \in \mathcal{C}^2$$

as $f = \nabla\varphi$ via the following iterations: $x^0 \in I\!R^n$ is given, then

$$x^{k+1} := x^k + d^k \text{ with } \ -\nabla\varphi(x^k) = \nabla^2\varphi(x^k)d^k, \quad k = 0, 1, \ldots$$

This algorithm is locally well-defined and superlinearly converges to a local minimizer $\bar{x}$ if $\nabla^2\varphi(\bar{x}) > 0$.

The most popular extension of Newton's algorithm to solve nonsmooth equations $f(x) = 0$ is the semismooth Newton method based on Clarke's generalized Jacobian. In the case of optimization, it addresses local minimizers of $C^{1,1}$ functions φ, i.e., smooth with Lipschitzian gradients.

MAJOR GOALS

To develop two versions of the generalized Newton method of computing not just arbitrary local minimizers for nonsmooth problems of unconstrained and constrained optimization, but those which exhibit certain stability behavior known as tilt stability. We first proceed with $\mathcal{C}^{1,1}$ unconstrained optimization, and then propagate the algorithms to minimizing extended-real-valued prox-regular functions which cover, in particular, broad classes in constrained optimization. The main results verify well-posedness/solvability of subproblems and Q-superlinear convergence of iterates. To proceed, we are based on advanced results of second-order variational analysis.

GENERALIZED DIFFERENTIATION

Tangent cone to a set $\Omega \subset I\!R^n$ at $\bar{x} \in \Omega$

$$T_\Omega(\bar{x}) := \Big\{ w \in I\!R^n \Big| \, \exists\, t_k \downarrow 0, \ \ w_k \to w \ \text{ as } \ k \to \infty \ \text{ with } \ \bar{x} + t_k w_k \in \Omega \Big\}$$

Normal cone to $\Omega \subset I\!R^n$ at $\bar{x} \in \Omega$

$$N_\Omega(\bar{x}) := \Big\{ v \in I\!R^n \Big| \, \exists\, x_k \xrightarrow{\Omega} \bar{x}, \ v_k \to v \ \text{ with } \ v_k \in T^*_\Omega(x_k) \Big\}$$

Graphical derivative of a mapping $F\colon I\!R^n \rightrightarrows I\!R^m$ at $(\bar{x}, \bar{y}) \in \text{gph}\, F$

$$DF(\bar{x}, \bar{y})(u) := \Big\{ v \in I\!R^m \Big| \, (w, v) \in T_{\text{gph}\, F}(\bar{x}, \bar{y}) \Big\}, \ u \in I\!R^n$$

Coderivative of $F\colon I\!R^n \rightrightarrows I\!R^m$ at $(\bar{x}, \bar{y}) \in \text{gph}\, F$

$$D^*F(\bar{x}, \bar{y})(v) := \Big\{ u \in I\!R^n \Big| (u, -v) \in N_{\text{gph}\, F}(\bar{x}, \bar{y}) \Big\}, \ v \in I\!R^m$$

Subderivative of $\varphi\colon I\!R^n \to \overline{I\!R} := (-\infty, \infty]$ at $\bar{x} \in \operatorname{dom} \varphi$

$$\mathrm{d}\varphi(\bar{x})(\bar{w}) := \liminf_{t \downarrow 0, w \to \bar{w}} \frac{\varphi(\bar{x} + tw) - \varphi(\bar{x})}{t}, \ \bar{w} \in I\!R^n$$

Subdifferential of $\varphi\colon I\!R^n \to \overline{I\!R}$ at $\bar{x} \in \operatorname{dom} \varphi$

$$\partial\varphi(\bar{x}) := \left\{ v \in I\!R^n \middle| \ (v, -1) \in N_{\operatorname{epi} \varphi}\big(\bar{x}, \varphi(\bar{x})\big) \right\}$$

TILT-STABLE MINIMIZERS

DEFINITION (Poliquin and Rockafellar, 1998) Given $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, a point $\bar{x} \in \operatorname{dom}\varphi$ is said to be a tilt-stable local minimizer of φ if for some $\gamma > 0$ the argminimum mapping

$$M_\gamma\colon v \mapsto \text{argmin}\Big\{\varphi(x) - \langle v, x\rangle \Big|\, x \in \mathbb{B}_\gamma(\bar{x})\Big\}$$

is single-valued and Lipschitz continuous on a neighborhood of $\bar{v} = 0$ with $M_\gamma(\bar{v}) = \{\bar{x}\}$.

This notion is very well investigated and comprehensively *characterized* in second-order variational analysis with many applications to various classes of problems in constrained optimization. In particular, tilt-stable local minimizers of prox-regular functions $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ are characterized via second-order subdifferential/generalized Hessians by

$$\partial^2\varphi(\bar{x}, 0) > 0$$

where $\partial^2\varphi(\bar{x}, \bar{v}) := (D^*\partial\varphi)(\bar{x}, \bar{v})$ for $\bar{v} \in \partial\varphi(\bar{x})$ from [Mor92].

CODERIVATIVE-BASED ALGORITHM FOR $\mathcal{C}^{1,1}$

Consider the set

$$Q^*(x) := \Big\{ y \in I\!R^n \Big| -\nabla\varphi(x) \in (D^*\nabla\varphi)(x)(-y) \Big\}$$

where we have the representation

$$D^*\nabla\varphi(x)(-y) = \partial\langle -y, \nabla\varphi\rangle(x) \text{ for } x \text{ near } \bar{x},$$

Algorithm 1 Pick $x_0 \in I\!R^n$ and set $k := 0$
Step 1: If $\nabla\varphi(x_k) = 0$, then stop
Step 2: Otherwise, select a direction $d_k \in Q^*(x_k)$ and set $x_{k+1} := x_k - d_k$
Step 3: Let $k \leftarrow k+1$ and then go to Step 1

PERFORMANCE OF CODERIVATIVE-BASED ALGORITHM

THEOREM Let φ be a $\mathcal{C}^{1,1}$ function on a neighborhood of $\bar{x}$, which is a tilt-stable local minimizer φ. Then there exists a neighborhood O of $\bar{x}$ such that the set-valued mapping $Q^*(x)$ is nonempty and compact-valued for all x in O.

DEFINITION (Gfrerer and Outrata, 2019) A mapping $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is semismooth* at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ if whenever $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$ we have the condition

$$\langle u^*, u\rangle = \langle v^*, v\rangle \quad \text{for all} \quad (v^*, u^*) \in \operatorname{gph} D^*F\Big((\bar{x}, \bar{y}); (u, v)\Big)$$

THEOREM Let φ be a $\mathcal{C}^{1,1}$ function on a neighborhood of its tilt-stable local minimizer $\bar{x}$, and let the gradient mapping $\nabla\varphi$ be semismooth* at $\bar{x}$. Then there is $\delta > 0$ such that for any starting point $x_0 \in \mathbb{B}_\delta(\bar{x})$ we get that every sequence $\{x_k\}$ constructed by Algorithm 1 converges to $\bar{x}$ and the rate of convergence is superlinear.

$\mathcal{C}^{1,1}$ ALGORITHM BASED ON GRAPHICAL DERIVATIVES

Consider the set

$$Q(x) := \left\{ y \in I\!R^n \,\middle|\, -\nabla\varphi(x) \in (D\nabla\varphi)(x)(y) \right\}$$

Algorithm 2 Pick $x_0 \in I\!R^n$ and set $k := 0$
Step 1: If $\nabla\varphi(x_k) = 0$, then stop
Step 2: Otherwise, select a direction $d_k \in Q(x_k)$ and set $x_{k+1} := x_k - d_k$
Step 3: Let $k \leftarrow k + 1$ and then go to Step 1

THEOREM Let φ be a $\mathcal{C}^{1,1}$ function on a neighborhood of $\bar{x}$, which is a tilt-stable local minimizer φ. Then there exists a neighborhood O of $\bar{x}$ such that the set-valued mapping $Q(x)$ is nonempty and compact-valued for all x in O.

SECOND SUBDERIVATIVES

The second subderivative of $\varphi\colon I\!R^n \to \overline{I\!R}$ at $\bar{x}$ for $\bar{v} \in I\!R^n$ is

$$\mathrm{d}^2\varphi(\bar{x},\bar{v})(w) := \liminf_{t\downarrow 0,\, w'\to w} \Delta_t^2\varphi(\bar{x},\bar{v})(w')$$

where the second-order finite difference are

$$\Delta_t^2\varphi(\bar{x},\bar{v})(w) := \frac{\varphi(\bar{x}+tw') - \varphi(\bar{x}) - t\langle\bar{v},w'\rangle}{\frac{1}{2}t^2}$$

φ is twice epi-differentiable at $\bar{x}$ for $\bar{v}$ if for every $w \in I\!R^n$ and $t_k \downarrow 0$ there is $w_k \to w$ with $\Delta_{t_k}^2\varphi(\bar{x},\bar{v})(w_k) \to \mathrm{d}^2\varphi(\bar{x},\bar{v})(w)$.

The latter class includes fully amenable functions, parabolically regular functions, etc.

SUBPROBLEMS ASSOCIATED WITH ALGORITHM 2

Subproblems for directions: At each iteration x^k with $v^k := -\nabla\varphi(x^k)$ find $w = D^k$ as a stationary point of

$$\text{min } \varphi(x^k) + \langle v^k, w\rangle + \frac{1}{2}\text{d}^2\varphi(x^k, v^k)(w)$$

Constructive implementations of **subproblems** are given, in particular, for the classes of extended linear-quadratic programs and for minimization of augmented Lagrangians.

THEOREM Let $\varphi\colon I\!R^n \to \overline{I\!R}$ be a $\mathcal{C}^{1,1}$ function around $\bar{x}$, where $\bar{x}$ is its tilt-stable local minimizer, and let φ be twice epi-differentiable at x for $v = \nabla\varphi(x)$. Then for each large $k \in I\!N$ the subproblem admits a unique optimal solution.

SUPERLINEAR CONVERGENCE OF ALGORITHM 2

THEOREM Let $\varphi\colon I\!R^n \to \overline{I\!R}$ be a $\mathcal{C}^{1,1}$ function on a neighborhood of its tilt-stable local minimizer $\bar{x}$, and let $\nabla\varphi$ be semismooth* at $\bar{x}$. Then there exists $\delta > 0$ such that for any starting point $x_0 \in I\!B_\delta(\bar{x})$ we have that every sequence $\{x_k\}$ constructed by Algorithm 2 converges to $\bar{x}$ and the rate of convergence is superlinear.

PROX-REGULAR FUNCTIONS AND MOREAU ENVELOPES

$\varphi\colon I\!R^n \to \overline{I\!R}$ is prox-regular at $\bar{x} \in \operatorname{dom}\varphi$ for $\bar{v} \in \partial\varphi(\bar{x})$ if φ it is l.s.c. and there are $\varepsilon > 0$ and $\rho \geq 0$ such that for all $x \in I\!B_\varepsilon(\bar{x})$ with $\varphi(x) \leq \varphi(\bar{x}) + \varepsilon$ we have

$$\varphi(x) \geq \varphi(u) + \langle \bar{v}, x - u \rangle - \frac{\rho}{2}\|x - u\|^2 \text{ for } (u, v) \in (\operatorname{gph}\partial\varphi) \cap I\!B_\varepsilon(\bar{x}, \bar{v})$$

φ is subdifferentially continuous at $\bar{x}$ for $\bar{v}$ if the convergence $(x_k, v_k) \to (\bar{x}, \bar{v})$ with $v_k \in \partial\varphi(x_k)$ yields $\varphi(x_k) \to \varphi(\bar{x})$. If both properties hold, φ is continuously prox-regular. This is the major class in second-order variational analysis.
The **Moreau envelopes** of φ for $r > 0$ is

$$e_r\varphi(x) := \inf_w \left\{ \varphi(w) + \frac{1}{2r}\|w - x\|^2 \right\}$$

THEOREM (Poliquin-Rockafellar, 1996) If φ is continuously prox-regular at $\bar{x}$ for $\bar{v}$, then its Moreau envelope for small $r > 0$ is a $\mathcal{C}^{1,1}$ function with $\nabla e_r\varphi(\bar{x} + r\bar{v}) = \bar{v}$.

ALGORITHMS FOR PROX-REGULAR FUNCTIONS

Minimizing an extended-real-valued continuously prox-regular function φ is equivalent to the unconstrained $\mathcal{C}^{1,1}$ problem

$$\text{minimize } \; e_r\varphi(x) \; \text{ subject to } \; x \in I\!R^n$$

THEOREM Let $\varphi\colon I\!R^n \to \overline{I\!R}$ be continuously prox-regular at $\bar{x}$ for $\bar{v} = 0$, where $\bar{x}$ is a tilt-stable local minimizer of φ. Assume that the $\partial\varphi$ is semismooth* at $(\bar{x}, \bar{v})$. Then for any small $r > 0$ there exists $\delta > 0$ such that for each starting point $x_0 \in I\!B_\delta(\bar{x})$ both Algorithms 1 and 2 and are well-defined, and every sequence of iterates $\{x_k\}$ superlinearly converges to $\bar{x}$.

APPLICATIONS TO CONSTRAINED OPTIMIZATION

Consider the constrained problem

$$\text{minimize} \quad \psi(x) \quad \text{subject to} \quad f(x) \in \Theta$$

where the functions $\psi \colon I\!R^n \to I\!R$ and $f \colon I\!R^n \to I\!R^m$ are $\mathcal{C}^2$-smooth and the set $\Theta \subset I\!R^m$ is closed and convex. Denote

$$\varphi(x) := \psi(x)_{(}x) \quad \text{with} \quad \Omega := \Big\{ x \in I\!R^n \;\Big|\; f(x) - \Theta \Big\}$$

Algorithm 3 Set $k := 0$, and pick any $r > 0$
Step 1: If $0 \in \partial\varphi(x_k)$, then stop.
Step 2: Otherwise, let $v_k = \nabla(e_r\varphi)(x_k)$, select w_k as a stationary point of the subproblem

$$\min_{w \in I\!R^n} \langle v_k, w \rangle + \frac{1}{2} \mathrm{d}^2 \varphi(x_k - r v_k, v_k)(w)$$

and then set $d_k := w_k - r v_k$, $x_{k+1} := x_k + d_k$
Step 3: Let $k \leftarrow k + 1$ and then go to Step 1

SUPERLINEAR CONVERGENCE OF ALGORITHM 3

THEOREM Let $\bar{x}$ be a tilt-stable local minimizer of the constrained problem, and let Θ be $\mathcal{C}^2$-cone reducible at $f(\bar{x})$ under the constraint qualification condition

$$\text{span}\Big\{N_\Theta\Big(f(\bar{x})\Big)\Big\} \cap \ker \nabla f(\bar{x})^* = \{0\}$$

Assume that the mapping $x \mapsto f(x) - \Theta$ is metrically subregular at $(\bar{x}, 0)$ and that the normal cone mapping N_Θ is semismooth* at $(f(\bar{x}), \bar{\lambda})$, where $\bar{\lambda}$ is the unique Lagrange multiplier. Then for any small $r > 0$ there ia $\delta > 0$ such that for each starting point $x_0 \in I\!B_\delta(\bar{x})$ the sequence $\{x_k\}$ constructed by Algorithm 3 superlinearly converges to $\bar{x}$.

Proof Besides the above reduction and propagation, the proof is based on the recently developed theory of parabolic regularity [MMS20]

REFERENCES

[GO19] H. Gfrerer and J. V. Outrata, On a semismoothness* Newton method for solving generalized equations, arXiv:1904.09167.

[MMS20] A. Mohammadi, B. S. Mordukhovich and M. E. Sarabi, Parabolic regularity in geometric variational analysis, Trans. Amer. Math. Soc., to appear.

[Mor92] B. S. Mordukhovich, Sensitivity analysis in nonsmooth optimization, in Theoretical Aspects of Industrial Design, SIAM Proc. Appl. Math. 58, pp. 32–46, Philadelphia, PA, 1992.

B. S. Mordukhovich, Variational Analysis and Applications, Springer, 2018.

[MS20] B. S. Mordukhovich and M. E. Sarabi, Generalized Newton algorithms for tilt-stable minimizers in nonsmooth optimization, arXiv:1909.00241

[PR96] R. A. Poliquin and R. T. Rockafellar, Prox-regular functions in variational analysis, Trans. Amer. Math. Soc. 348 (1996), 1805–1838.

[PR98] R. A. Poliquin and R. T. Rockafellar, Tilt stability of a local minimum, SIAM J. Optim. 8 (1998), 287–299.

[RW98] R. T. Rockafellar and R. J-B Wets, Variational Analysis, Springer, 1998.