

Newton-type Methods and Newton Differentiability

Titus Pinta

September 24, YAMC
2025, Padua

Problem

We want to solve:

$$F(x) = 0$$

$F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ F non-linear

Classic Newton

Algorithm

Step 1: Pick x^0

Step 2: Pick x^{k+1} as the solution of

$$\mathcal{L}_{x^k} F(y) := F(x^k) + \nabla F(x^k)(y - x^k) = 0$$

Step 2: Equivalently

$$x^{k+1} = x^k - \nabla F(x^k)^{-1} F(x^k)$$

Convergence Proof 1

Theorem

$F \in \mathcal{C}^{1+}$ and $F(\bar{x}) = 0$ and $\det \nabla F(x) \neq 0$ (this implies that $\|\nabla F(x)^{-1}\|$ is bounded on a neighborhood of \bar{x}), then $x^k \rightarrow \bar{x}$ quadratically

Proof.

$$0 = F(\bar{x}) = F(x^k) + \nabla F(x^k)(x^k - \bar{x}) + \mathcal{O}(\|x^k - \bar{x}\|^2)$$

$$x^k - \nabla F(x^k)^{-1}F(x^k) - \bar{x} = x^{k+1} - \bar{x} = \nabla F(x^k)^{-1}\mathcal{O}(\|x^k - \bar{x}\|^2)$$

□

Convergence Proof 2

Theorem

$F \in \mathcal{C}^1$ and $F(\bar{x}) = 0$ and $\det \nabla F(x) \neq 0$ (this implies that $\|\nabla F(x)^{-1}\|$ is bounded on a neighborhood of \bar{x}), then $x^k \rightarrow \bar{x}$ superlinearly

Proof.

$$0 = F(\bar{x}) = F(x^k) + \nabla F(x^k)(x^k - \bar{x}) + o(\|x^k - \bar{x}\|)$$
$$x^k - \nabla F(x^k)^{-1} F(x^k) - \bar{x} = x^{k+1} - \bar{x} = \nabla F(x^k)^{-1} o(\|x^k - \bar{x}\|)$$

□

Convergence Proof 3

Theorem

$F \in \mathcal{C}^1$ and $F(\bar{x}) = 0$ and $\det \nabla F(x) \neq 0$ (this implies that $\|\nabla F(x)^{-1}\|$ is bounded on a neighborhood of \bar{x}), then $x^k \rightarrow \bar{x}$ superlinearly

Proof.

$$\|F(x) - F(\bar{x}) - \nabla F(x)(x - \bar{x})\| = o(\|x - \bar{x}\|)$$

$$\begin{aligned}\|x^{k+1} - \bar{x}\| &= \|x^k - \nabla F(x^k)^{-1}F(x^k) - \bar{x}\| \\ &= \|x^k - \bar{x} - \nabla F(x^k)^{-1}F(x^k) - \nabla F(x^k)^{-1}F(\bar{x})\| \\ &\leq \|\nabla F(x^k)^{-1}\| \|\nabla F(x^k)(x^k - \bar{x}) - F(x^k) + F(\bar{x})\| \\ &\leq Mo(\|x^k - \bar{x}\|)\end{aligned}$$



Fixed Point Theory

Consider $x^{k+1} = F(x^k)$

$$\exists c < 1, \forall x, y \in \mathbb{R}^n \quad \|F(x) - F(y)\| \leq c\|x - y\|$$

$\Rightarrow \exists \bar{x}, F(\bar{x}) = \bar{x}$ and x^k converges linearly to \bar{x}

$$\forall x, y \in \mathbb{R}^n \quad \|F(x) - F(y)\|^2 \leq \|x - y\|^2 - \alpha \|x - F(x) - y - F(y)\|^2$$

if $\exists \bar{x}, F(\bar{x}) = \bar{x} \Rightarrow x^k$ converges to \bar{x}

$$\exists c < 1, \forall x \in \mathbb{R}^n \quad \|F(x) - F(\bar{x})\| \leq c\|x - \bar{x}\|$$

$\Rightarrow x^k$ converges linearly to \bar{x}

$$\forall x \in \mathbb{R}^n \quad \|F(x) - F(\bar{x})\| = o(\|x - \bar{x}\|)$$

$\Rightarrow x^k$ converges superlinearly to \bar{x}

Convergence Conclusions

The key ingredients of the proof were:

1. $\exists \bar{x}$ with $F(\bar{x}) = 0$
2. $\nabla F(x)^{-1}$ bounded (by M) near \bar{x}
- 3.

$$\|F(x) - F(\bar{x}) - \nabla F(x)(x - \bar{x})\| = o(\|x - \bar{x}\|)$$

Semismoothness

Directional Derivative

Definition

$F : U \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^n$ has a *directional derivative* at $\bar{x} \in U$ if for all $y \in U$, and $d = y - x$

$$\lim_{t \downarrow 0} \frac{F(x + td) - F(x)}{t} = F'(x; d)$$

Definition

If F has directional derivatives at all x it is called *Gâteaux differentiable*.

Generalized Derivatives

Set-Valued Maps

$T : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a map $T : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^m)$, so $\forall x, T(x) \subseteq \mathbb{R}^m$

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ Lipschitz continuous. By Rademacher's theorem

$$\exists D \subseteq \mathbb{R}^n \text{ dense} \quad \forall x \in D, \exists \nabla F(x)$$

Definition

For $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ Lipschitz continuous, *the Bouligand derivative*
 $\partial_B F$

$$\partial_B F(x) = \left\{ H \in \mathbb{R}^{n \times m} \mid \exists \{x^k\}_{k \in \mathbb{N}} \in D, \lim_{k \rightarrow \infty} \nabla F(x^k) = H \right\}$$

Generalized Derivatives (cont.)

Definition

For $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ Lipschitz continuous, the Clarke derivative $\partial_B F$

$$\partial F(x) = \overline{\text{conv}} \partial_B F(x)$$

Definition

For $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ Lipschitz continuous, the Qi derivative $\partial_C F$

$$\partial F(x) = \partial F_1(x) \times \cdots \times \partial F_m(x)$$

To be clear: $\partial_B F, \partial F, \partial_C F : \mathbb{R}^n \rightrightarrows \mathbb{R}^{n \times m}$

$$\partial_B F(x) \subseteq \partial F(x) \subseteq \partial_C F(x)$$

Semismooth Functions

Definition

$F : U \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^n$ Lipschitz continuous is *semismooth*¹ at \bar{x} if for any sequences $\{t_k\}_{k \in \mathbb{N}} > 0$ with $\lim_{k \rightarrow \infty} t_k = 0$, $\{d^k\}_{k \in \mathbb{N}}$ convergent to d^* , and $\{H^k\}_{k \in \mathbb{N}} \in \partial F(\bar{x} + t_k d^k)$ the limit $\lim_{k \rightarrow \infty} H^k(d^*)$ exists

¹Mifflin, "Semismooth and Semiconvex Functions in Constrained Optimization", 1971

Properties

See ² for proofs

Lemma

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}$ Lipschitz continuous and $x, y \in \mathbb{R}^n$ there exist $t \in [0, 1]$

$$\exists H \in \partial F(x + t(y - x)), \quad F(x) - F(y) = \langle H, y - x \rangle$$

Lemma

If F is semismooth at \bar{x} then for all d , $F'(\bar{x}; d)$ exists and for any $\{d^k\}_{k \in \mathbb{N}}$ convergent to d , and $\{H^k\}_{k \in \mathbb{N}} \in \partial F(\bar{x} + t_k d^k)$ the limit

$$\lim_{k \rightarrow \infty} H^k(d^*) = F'(\bar{x}; d)$$

²Ulbricht "Nonsmooth Newton-like Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces", 2003, Habitation Thesis, https://www.researchgate.net/publication/2473025_Nonsmooth_Newton-like_Methods_for_Variational_Inequalities_and_Constrained_Optimization_Problems_in_Function_Spaces

Properties 2

The following are equivalent

1. F is semismooth at \bar{x}
2. F is Lipschitz, $F'(\bar{x}; \cdot)$ exists and

$$\sup_{H \in \partial F(x)} \|H(x - \bar{x}) - F'(\bar{x}; x - \bar{x})\| = o(\|x - \bar{x}\|)$$

3. F is Lipschitz, $F'(\bar{x}; \cdot)$ exists and

$$\sup_{H \in \partial F(x)} \|F(x) - F(\bar{x}) - H(x - \bar{x})\| = o(\|x - \bar{x}\|)$$

Proof Idea 2. \Leftrightarrow 3.

Take v such that

$$\langle v, F(\bar{x}) - F(x) - F'(\bar{x}; \bar{x} - x) \rangle = \|F(\bar{x}) - F(x) - F'(\bar{x}; \bar{x} - x)\|$$

Define:

$$\varphi(t) = \langle v, G(\bar{x}) - G(\bar{x} + t(x - \bar{x})) - tG'(\bar{x}; \bar{x} - x) \rangle$$

Compute $\varphi(1) - \varphi(0)$ using φ'

Semismooth Newton Method

Recall

1. $\exists \bar{x}$ with $F(\bar{x}) = 0$
2. $\nabla F(x)^{-1}$ bounded (by M) near \bar{x}
- 3.

$$\|F(x) - F(\bar{x}) - \nabla F(x)(x - \bar{x})\| = o(\|x - \bar{x}\|)$$

Semismooth \Rightarrow 3. Holds with $\nabla F(x)$ replaced by $\partial F(x)$

Semismooth Newton Method

Algorithm

$$x^{k+1} \in \{x^k - H^{-1}F(x^k) \mid H \in \partial F(x^k)\}$$

Theorem

F Lipschitz with and $F(\bar{x}) = 0$ and semismooth at \bar{x} and for all x near \bar{x} for all $H \in \partial F(x)$, $\|H^{-1}\| \leq M$, then $x^k \rightarrow \bar{x}$ superlinearly

Proof.

Let H such that $x^{k+1} = x^k - H^{-1}F(x^k)$

$$\begin{aligned}\|x^{k+1} - \bar{x}\| &= \|x^k - H^{-1}F(x^k) - \bar{x}\| \\ &= \|x^k - \bar{x} - H^{-1}F(x^k) - H^{-1}F(\bar{x})\| \\ &\leq \|H^{-1}\| \|H(x^k - \bar{x}) - F(x^k) + F(\bar{x})\| \\ &\leq M \sup_{H \in \partial F(x^k)} \|H(x^k - \bar{x}) - F(x^k) + F(\bar{x})\| \\ &\leq Mo(\|x^k - \bar{x}\|)\end{aligned}$$

Examples

1. if $F \in \mathcal{C}^1$ then it is semismooth
2. $\log(1 + |x|)$ is semismooth
3. $x^2 \sin(x^{-1})$ is differentiable, but not semismooth
4. F is called \mathcal{PC}^k if it is continuous and there are $F_1, \dots, F_m \in \mathcal{C}^k$ and for all x , $F(x) \in \{F_1(x), \dots, F_m(x)\}$. If F is \mathcal{PK}^1 then it is semismooth at any x
5. $\|x\|$ is semismooth (even 1-order semismooth)

p -order Semismooth Functions

Definition

If F is p -order semismooth at \bar{x} if it is semismooth at \bar{x} and

$$\sup_{H \in \partial F(x)} \|F(x) - F(\bar{x}) - H(x - \bar{x})\| = \mathcal{O}(\|x - \bar{x}\|^p)$$

Theorem

F Lipschitz with and $F(\bar{x}) = 0$ and p -order semismooth at \bar{x} and for all x near \bar{x} for all $H \in \partial F(x)$, $\|H^{-1}\| \leq M$, then $x^k \rightarrow \bar{x}$ with order p , i.e.

$$\|x^{k+1} - \bar{x}\| = \mathcal{O}(\|x^k - \bar{x}\|^p)$$

Theorem

If F is \mathcal{PC}^2 then F is 1-order semismooth at any point

Newton Differentiability

Definition

Definition

$F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is *Newton differentiable*³ at \bar{x} if there is

$\mathcal{H}F : \mathbb{R}^n \rightrightarrows \mathbb{R}^{n \times n}$ with

$$\sup_{H \in \mathcal{H}F(x)} \|F(x) - F(\bar{x}) - H(x - \bar{x})\| = o(\|x - \bar{x}\|)$$

³Pinta, “Newton-type Methods”, 2024, PhD thesis

Relation to Smoothness and Semismoothness

Theorem

If $F \in \mathcal{C}^1$ near \bar{x} then it is Newton differentiable at \bar{x} with the Newton differential $x \mapsto \{\nabla F(x)\}$

Theorem

If F is semismooth at \bar{x} then it is Newton differentiable at \bar{x} with the Newton differential $x \mapsto \partial F(x)$

Other Examples

Example⁴

$$F(x, y) = \begin{bmatrix} 4(x^2 - y) + 2(x - 1) \\ -2(x^2 - y) \end{bmatrix} \quad \mathcal{H}F(x, y) = \left\{ 2 \begin{bmatrix} 1 + 4x^2 & -2x \\ -2x & 1 \end{bmatrix} \right\}$$

F is Newton differentiable at 0

Example

$$F(x) = \begin{cases} x & x \in \mathbb{Q}, \\ -x & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases} \quad \mathcal{H}F(x) = \begin{cases} 1 & x \in \mathbb{Q}, \\ -1 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

F is Newton differentiable at 0

⁴R. Bergmann et al. "The difference of convex algorithm on Hadamard manifolds"

Superlinear Convergence

Algorithm

$$x^{k+1} \in \{x^k - H^{-1}F(x^k) \mid H \in \mathcal{H}F(x^k)\}$$

Theorem

F with $F(\bar{x}) = 0$ and Newton differentiable at \bar{x} with $\mathcal{H}F$ and for all x near \bar{x} for all $H \in \mathcal{H}F(x)$, $\|H^{-1}\| \leq M$, then $x^k \rightarrow \bar{x}$ superlinearly

Proof.

Let H such that $x^{k+1} = x^k - H^{-1}F(x^k)$

$$\begin{aligned}\|x^{k+1} - \bar{x}\| &= \|x^k - H^{-1}F(x^k) - \bar{x}\| \\ &= \|x^k - \bar{x} - H^{-1}F(x^k) - H^{-1}F(\bar{x})\| \\ &\leq \|H^{-1}\| \|H(x^k - \bar{x}) - F(x^k) + F(\bar{x})\| \\ &\leq M \sup_{H \in \partial F(x^k)} \|H(x^k - \bar{x}) - F(x^k) + F(\bar{x})\| \\ &\leq Mo(\|x^k - \bar{x}\|)\end{aligned}$$

p -order Newton Differentiability

Definition

If F is p -order Newton Differentiable at \bar{x} if it is Newton differentiable at \bar{x} with $\mathcal{H}F$ and

$$\sup_{H \in \mathcal{H}F(x)} \|F(x) - F(\bar{x}) - H(x - \bar{x})\| = \mathcal{O}(\|x - \bar{x}\|^p)$$

Theorem

F with $F(\bar{x}) = 0$ and p -order Newton differentiable at \bar{x} with $\mathcal{H}F$ and for all x near \bar{x} for all $H \in \mathcal{H}F(x)$, $\|H^{-1}\| \leq M$, then $x^k \rightarrow \bar{x}$ with order p

Theorem

If F is p -order semismooth at \bar{x} it is p -order Newton Differentiable at \bar{x} with Newton differential ∂F

Weak Newton Differentiability

Definition

Definition

$F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is *weakly Newton differentiable* at \bar{x} if there is $\mathcal{H}F : \mathbb{R}^n \rightrightarrows \mathbb{R}^{n \times n}$ with

$$\lim_{x \rightarrow \bar{x}} \sup_{H \in \mathcal{H}F(x)} \frac{\|F(x) - F(\bar{x}) - H(x - \bar{x})\|}{\|x - \bar{x}\|} = c$$

Linear Convergence

Theorem

F with $F(\bar{x}) = 0$ and weakly Newton differentiable at \bar{x} with $\mathcal{H}F$ and for all x near \bar{x} for all $H \in \mathcal{H}F(x)$, $\|H^{-1}\| \leq M$ and $cM < 1$ then $x^k \rightarrow \bar{x}$ linearly

Proof.

Let H such that $x^{k+1} = x^k - H^{-1}F(x^k)$

$$\begin{aligned}\|x^{k+1} - \bar{x}\| &= \|x^k - H^{-1}F(x^k) - \bar{x}\| \\ &= \|x^k - \bar{x} - H^{-1}F(x^k) - H^{-1}F(\bar{x})\| \\ &\leq \|H^{-1}\| \|H(x^k - \bar{x}) - F(x^k) + F(\bar{x})\| \\ &\leq M \sup H \in \partial F(x^k) \|H(x^k - \bar{x}) - F(x^k) + F(\bar{x})\| \\ &\leq M(c + \varepsilon) \|x^k - \bar{x}\|\end{aligned}$$



Strongly Convex Functions

Theorem

Let $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ be a bounded from below, m -strongly convex with L -Lipschitz gradient, then ∇f is weakly Newton differentiable at \bar{x} with a Newton differential $x \mapsto \alpha I$

Proof.

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq m \|x - y\|^2$$

$$\begin{aligned} & \|\nabla f(x) - \nabla f(y) - \alpha(x - y)\|^2 \\ &= \|\nabla f(x) - \nabla f(y)\|^2 + \alpha^2 \|x - y\|^2 - 2\alpha \langle \nabla f(x) - \nabla f(y), x - y \rangle \\ &\leq L^2 \|x - y\|^2 + \alpha^2 \|x - y\|^2 - 2\alpha \langle \nabla f(x) - \nabla f(y), x - y \rangle \\ &\leq L^2 \|x - y\|^2 + \alpha^2 \|x - y\|^2 - 2\alpha m \|x - y\|^2 \end{aligned}$$

$$\|\nabla f(x) - \nabla f(\bar{x}) - \alpha(x - \bar{x})\| \leq \sqrt{L^2 + \alpha^2 - 2\alpha m} \|x - \bar{x}\|$$

Relatively Strongly Monotone Operators

Definition

$T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is m -relatively strongly monotone for $H \in \mathbb{R}^{n \times n}$ if

$$\forall x, y \in \mathbb{R}^n, \quad \langle \nabla f(x) - \nabla f(y), H(x - y) \rangle \geq m \|x - y\|^2$$

Theorem

Let $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ be a bounded from below, m -relatively strongly monotone for H and L -Lipschitz gradient, then ∇f is weakly Newton differentiable at \bar{x} with a Newton differential $x \mapsto H$

Proof.

$$\begin{aligned} & \|\nabla f(x) - \nabla f(y) - H(x - y)\|^2 \\ & \leq \|\nabla f(x) - \nabla f(y)\|^2 + \|H^2\| \|x - y\|^2 \\ & \quad - \langle \nabla f(x) - \nabla f(y), H(x - y) \rangle \\ & \leq L^2 \|x - y\|^2 + \|H^2\| \|x - y\|^2 - 2m \|x - y\|^2 \end{aligned}$$

$$\|\nabla f(x) - \nabla f(\bar{x}) - H(x - \bar{x})\| \leq \sqrt{L^2 + \|H^2\| - 2m} \|x - \bar{x}\|$$

□

Corollary

If $\alpha^{-1} \sqrt{L^2 + \|H^2\| - 2m} < 1$ then H preconditioned gradient descent converges linearly

Sufficient Decrease

Lemma

Let $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ Lipschitz with a global minimum at \bar{x} and denote by $\partial_1 f(x) = P_{\partial f(x)} 0$. Assume that $\partial_1 f$ is weakly Newton differentiable on U with the same c . Assume $H \in \mathcal{H}f$ is positive definite and bounded by M with $cM < 1$. Then

$$\exists \rho > 0 \forall k \in \mathbb{N}, \quad \rho \|x^{k+1} - x^k\|^2 \leq f(x^k) - f(x^{k+1})$$

Idea

Proof.

$$Q_k(x) = f(x^k) + \langle \partial_1 f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, H^k(x - x^k) \rangle$$

$$\begin{aligned} Q_k(x^{k+1}) - f(x^{k+1}) &\leq \int_0^1 t(c + \varepsilon) \|(x^{k+1} - x^k)\| \|x^{k+1} - x^k\| dt \\ &= \frac{c + \varepsilon}{2} \|x^{k+1} - x^k\|^2 \end{aligned}$$

$$f(x^k) - Q_k(x^{k+1}) \geq \frac{1}{2} \langle \partial_1 f(x^k), H^{k-1} \partial_1 f(x^k) \rangle \geq 0$$

□

Gradient Error Bound

Lemma

Let $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ Lipschitz with a global minimum at \bar{x} and denote by $\partial_1 f(x) = P_{\partial f(x)} 0$. Assume that $\partial_1 f$ is weakly Newton differentiable on U with the same c . Assume $H \in \mathcal{H}f$ is positive definite and bounded by M with $cM < 1$. Then

$$\exists \rho > 0 \forall k \in \mathbb{N}, \quad \|\partial_1 f(x^k)\| \leq \rho \|x^{k+1} - x^k\|$$

Proof.

$$\|x^{k+1} - x^k\| = \|H^{k-1} \partial_1 f(x^k)\| \leq \|H^{k-1}\| \|\partial_1 f(x^k)\|$$



Error Bounds

Theorem (Complexity under Sufficient Decrease)

Let $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ and assume there are c and d such that

$$\forall k \in \mathbb{N}, \quad f(x^k) - f(x^{k+1}) \geq c \|x^{k+1} - x^k\|^2$$

$$\forall k \in \mathbb{N}, \quad \|x^{k+1} - x^k\| \geq d \|\partial_1 f(x^k)\|.$$

Then for any $\varepsilon > 0$

$$\min_{n \leq k} \|\partial_1 f(x^n)\| < \varepsilon$$

for all

$$k > \frac{f(x^0) - f(\bar{x})}{c(d\varepsilon)^2}.$$

Proof.

Assume by contradiction $\|\partial_1 f(x^i)\| > \varepsilon$

$$\|x^{i+1} - x^i\| \geq d\varepsilon$$

$$f(x^i) - f(x^{i+1}) \geq c(d\varepsilon)^2$$

$$f(x^0) - f(x^k) = \sum_{i=0}^{k-1} f(x^i) - f(x^{i+1}) \geq (k-1)c(d\varepsilon)^2$$

$$f(x^0) - f(\bar{x}) \geq (k-1)c(d\varepsilon)^2$$

□

Error Bounds

Corollary

Let $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ Lipschitz with a global minimum at \bar{x} and denote by $\partial_1 f(x) = P_{\partial f(x)} 0$. Assume that $\partial_1 f$ is weakly Newton differentiable on U with the same c . Assume $H \in \mathcal{H}f$ is positive definite and bounded by M with $cM < 1$.

Then for any $\varepsilon > 0$

$$\min_{n \leq k} \|\partial_1 f(x^n)\| < \varepsilon$$

for all

$$k > \frac{f(x^0) - f(\bar{x})}{c(d\varepsilon)^2}.$$

Uniform Newton Differentiability

Definition

Definition

$F : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is *uniformly (weakly) Newton differentiable at \bar{x}* if there are $c \in \mathbb{R} (c \neq 0)$, $\mathcal{H}F : \mathbb{R}^n \rightrightarrows \mathbb{R}^{n \times n}$ such that $\forall \varepsilon > 0 \exists \delta$, $\forall x \in \mathbb{R}^m$, $\forall y \in V$ with $\|x - y\| \leq \delta$

$$\sup_{H \in \mathcal{H}F(x)} \frac{\|F(x) - F(\bar{x}) - H(x - \bar{x})\|}{\|x - \bar{x}\|} \leq c + \varepsilon$$

Uniform Semismooth Functions

Definition

$F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is Lipschitz *uniformly semi-smooth** on V if $\forall \varepsilon > 0$
 $\exists \delta$ such that $\forall x, \forall y \in V$ with $\|x - y\| \leq \delta$,

$$\frac{\|F'(x; x - y) - F'(y; x - y)\|}{\|x - y\|} \leq \varepsilon$$

Proposition

$F : \mathbb{R}^n \rightarrow \mathbb{R}^n$, F **uniformly semi-smooth*** on V , then F is
uniformly Newton differentiable on V with a Newton differential
 $\mathcal{H}(x) := \overline{\text{conv}} \left\{ H \in \mathbb{R}^{n \times m} \mid \exists \{x^k\}_{k \in \mathbb{N}}, \lim_{k \rightarrow \infty} \nabla F(x^k)^T = H \right\}$

\mathcal{C}^2 Functions

Recall \mathcal{PC}^2 functions are 1-order semismooth on the entire domain, for all y there is ε for all x

$$\|F'(x; x - y) - F'(y; x - y)\| \leq \varepsilon \|x - y\|^2$$

We can move $\|x - y\|^\gamma$ into ε in order to make the inequality uniform in y

Theorem

If $F \in \mathcal{PC}^2$ then it is uniformly Newton differentiable on any compact set

Underdetermined Newton-type Methods

Given $F : \mathbb{R}^m \rightarrow \mathbb{R}^n$, find $\bar{x} \in \mathbb{R}^m$

$$F(\bar{x}) = 0$$

Same idea: linearize and solve

$$\mathcal{A}(x) = \{y \in \mathbb{R}^m \mid F(x) + \nabla F(x)^T(y - x) = 0\}$$

$\mathcal{A}(x^k)$ is not a singleton

How do we pick $x^{k+1} \in \mathcal{A}(x^k)$?

Idea: $x^{k+1} = P_{\mathcal{A}(x^k)} x^k$

$$x^{k+1} = x^k - \nabla F(x^k)^+ F(x^k)$$

Superlinear Convergence

Algorithm

$$x^{k+1} \in \{x^k - H^+ F(x^k) \mid H \in \mathcal{H}F(x^k)\}$$

$$\text{Denote } \mathcal{A}(x, H) = \{y \in \mathbb{R}^m \mid F(x) + H(y - x) = 0\}$$

Theorem

F with $\mathcal{Z} = \{\bar{x} \mid F(\bar{x}) = 0\} \neq \emptyset$ and $P_{\mathcal{Z}}$ Lipschitz and uniformly Newton differentiable on \mathcal{Z} with $\mathcal{H}F$ and for all x near \mathcal{Z} for all $H \in \mathcal{H}F(x)$, $\|H^{-1}\| \leq M$ and there is P

$$\|P_{\mathcal{A}(x,H)} x - P_{\mathcal{Z}} P_{\mathcal{A}(x,H)} x\| \leq P \|P_{\mathcal{Z}} P_{\mathcal{A}(x,H)} x - P_{\mathcal{A}(x,H)} P_{\mathcal{Z}} P_{\mathcal{A}(x,H)} x\|$$

then $x^k \rightarrow \mathcal{Z}$ superlinearly

Proof.

See⁵

Let $H \in \mathcal{HF}(x)$ and $y = P_{\mathcal{A}(x,H)} x$

$$\begin{aligned} & \| P_{\mathcal{Z}} y - P_{\mathcal{A}(x,H)} P_{\mathcal{Z}} y \| \\ & \leq \| H^+ \| \| H(P_{\mathcal{Z}} y - P_{\mathcal{A}(x,H)} P_{\mathcal{Z}} y) + F(x) + H(P_{\mathcal{A}(x,H)} P_{\mathcal{Z}} y - x) \|. \\ & = \| H^+ \| \| H(P_{\mathcal{Z}} y - P_{\mathcal{A}(x,H)} P_{\mathcal{Z}} y + P_{\mathcal{A}(x,H)} P_{\mathcal{Z}} y - x) + F(x) \|. \\ & = \| H^+ \| \| F(x) - H(x - P_{\mathcal{Z}} y) \|. \end{aligned}$$

$$\| P_{\mathcal{Z}} y - P_{\mathcal{A}(x,H)} P_{\mathcal{Z}} y \| \leq c \| H^+ \| \| (x - P_{\mathcal{Z}} y) \|.$$

$$\text{dist}(y, \mathcal{Z}) \leq cPM \| (x - P_{\mathcal{Z}} y) \|.$$

⁵Pinta, "A Newton-type method for non-smooth under-determined systems of equations", 2025, <https://doi.org/10.1007/s11075-025-02212-8>

Proof (cont).

$$\begin{aligned}\|x - P_{\mathcal{Z}} y\| &\leq \|x - P_{\mathcal{Z}} x\| + \|P_{\mathcal{Z}} x - P_{\mathcal{Z}} y\| \\ &= \text{dist}(x, \mathcal{Z}) + \|P_{\mathcal{Z}} x - P_{\mathcal{Z}} y\| \\ &\leq \text{dist}(x, \mathcal{Z}) + L_{\mathcal{Z}} \|x - y\|.\end{aligned}$$

$$L_{\mathcal{Z}} \|x - y\| \leq L_{\mathcal{Z}} LM \text{dist}(x, \mathcal{Z}).$$



Calculus of Newton Differentiability

Sum Rule

Proposition

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$ Newton differentiable at \bar{x} with $\mathcal{H}F$ and $\mathcal{H}G$. Then $F + G$ is Newton differentiable at \bar{x} with $x \mapsto \mathcal{H}F(x) + \mathcal{H}G(x)$

Product Rule

Proposition

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}$ and $G : \mathbb{R}^n \rightarrow \mathbb{R}$ Newton differentiable at \bar{x} with $\mathcal{H}F$ and $\mathcal{H}G$. Then $F \cdot G$ is Newton differentiable at \bar{x} with $x \mapsto \mathcal{H}F(x)G(x) + F(x)\mathcal{H}G(x)$

Chain Rule

Proposition

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $G : \mathbb{R}^m \rightarrow \mathbb{R}^k$ Newton differentiable at \bar{x} with $\mathcal{H}F$ and $\mathcal{H}G$. Then $G \circ F$ is Newton differentiable at \bar{x} with $x \mapsto \{H_G H_F \mid H_F \in \mathcal{H}F(x), H_G \in \mathcal{H}G(x)\}$

Parallel Chain Rule

Proposition

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $G : \mathbb{R}^n \rightarrow \mathbb{R}^k$ Newton differentiable at \bar{x} with $\mathcal{H}F$ and $\mathcal{H}G$. Then $F \oplus G : \mathbb{R}^n \rightarrow \mathbb{R}^{m+k}$ is Newton differentiable at \bar{x} with $x \mapsto \mathcal{H}F(x) \oplus \mathcal{H}G(x) \subseteq \mathbb{R}^{n \times (m+k)}$

Inexact Newton Differentials

Proposition

Let $F : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ Newton differentiable at \bar{x} , with $\mathcal{H}F$. Let $\varepsilon : U \rightarrow [0, \infty)$ and $\varepsilon(x) \rightarrow 0$ as $x \rightarrow \bar{x}$ then

$$\mathcal{G}F(x) = \mathcal{H}F(x) + B_{\varepsilon(x)}[0].$$

is also a Newton differential for F at \bar{x}

Proposition

Let $F : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ weakly Newton differentiable at \bar{x} , with $\mathcal{H}F$. Let $\varepsilon : U \rightarrow [0, \infty)$ then

$$\mathcal{G}F(x) = \mathcal{H}F(x) + B_{\varepsilon(x)}[0].$$

is also a weak Newton differential for F at \bar{x}

Chords and Preconditioners

Proposition

Let $F : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ Newton differentiable at \bar{x} , with $\mathcal{H}F$. Then F is weakly Newton differentiable at \bar{x} with $x \mapsto \mathcal{H}F(y)$ for some y close enough to \bar{x}

Algorithm

$$x^{k+1} = x^k - \nabla F(x^0)^{-1} F(x^k)$$

Dennis-Moré Condition 1

Proposition

Let $F : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ Newton differentiable at \bar{x} , with $\mathcal{H}F$. Then any $\mathcal{G}F$ such that

$$\sup_{H \in \mathcal{H}F(x), G \in \mathcal{G}F(x)} \|H - G\| = o(\|x - \bar{x}\|)$$

is also a Newton differential for F at \bar{x}

Proposition

Let $F : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ weakly Newton differentiable at \bar{x} , with $\mathcal{H}F$. Then any $\mathcal{G}F$ such that

$$\sup_{H \in \mathcal{H}F(x), G \in \mathcal{G}F(x)} \|H - G\| = \mathcal{O}(\|x - \bar{x}\|)$$

is also a weak Newton differential for F at \bar{x}

Dennis-Moré Condition 2

Proposition

Let $F : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ Newton differentiable at \bar{x} , with $\mathcal{H}F$. Then any $\mathcal{G}F$ such that

$$\sup_{H \in \mathcal{H}F(x), G \in \mathcal{G}F(x)} \|(H - G)(x - \bar{x})\| = o(\|x - \bar{x}\|^2)$$

is also a Newton differential for F at \bar{x}

Proposition

Let $F : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ weakly Newton differentiable at \bar{x} , with $\mathcal{H}F$. Then any $\mathcal{G}F$ such that

$$\sup_{H \in \mathcal{H}F(x), G \in \mathcal{G}F(x)} \|(H - G)(x - \bar{x})\| = \mathcal{O}(\|x - \bar{x}\|^2)$$

is also a weak Newton differential for F at \bar{x}

Quasi-Newton Methods

A quasi-Newton method is characterized by an update rule

$$\mathcal{U} : \mathbb{R}^n \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$$

Algorithm

$$\begin{cases} x^{k+1} &= x^k - H^{k-1} F(x^k) \\ H^{k+1} &= \mathcal{U}(x^k, H^k) \end{cases}$$

In matrix form:

$$\begin{bmatrix} x^{k+1} \\ H^{k+1} \end{bmatrix} = \begin{bmatrix} x^k \\ H^k \end{bmatrix} - \begin{bmatrix} H^{k-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} F(x^k) \\ H^k - \mathcal{U}(x^k, H^k) \end{bmatrix}$$

Quasi-Newton Methods

Proposition

Let $F : U \subseteq \mathbb{R}^n$ Newton differentiable at \bar{x} $\bar{H} \in \mathcal{H}F(\bar{x})$. Assume

$$\lim_{x \rightarrow \bar{x}} \sup_{H_F \in \mathcal{H}F(x)} \|H_F - \bar{H}\| = 0$$

Assume

$$\lim_{x, H \rightarrow \bar{x}, \bar{H}} \frac{\|\mathcal{U}(x, H) - \bar{H}\|_F}{\|H - \bar{H}\|_F} = 0$$

Consider $G : U \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^n \times \mathbb{R}^{n \times n}$

$$G(x, H) = [F(x) \quad H - \mathcal{U}(x, H)]^T.$$

Then G is Newton differentiable at (\bar{x}, \bar{H}) with

$$\mathcal{H}G(x, H) = \left\{ \begin{bmatrix} H & 0 \\ 0 & I \end{bmatrix} \right\}$$

Why Modified Newton Methods?

We have seen many modifications:

1. Inexact
2. Chord
3. Denis-Moré
4. Quasi-Newton

We get convergence using Newton differentiability

Tradeoff: Slower convergence, but cheaper iterations

1. Start with smooth problem
2. solve with modified Newton
3. Analyze using non-smooth theory

Kantorovich-type Theorem

Classic Kantorovich Theorem

Theorem

Let $F \in \mathcal{C}^1$ and define $B := \|\nabla F(x^0)^{-1}\|$ and

$\eta := \|\nabla F(x^0)^{-1} F(x^0)\|$.

If

1. F is L -smooth,

$$\|\nabla F(x) - \nabla F(y)\| \leq L\|x - y\|$$

2. $LB\eta \leq .5$

then $\exists \bar{x}$ near x^0 such that $F(\bar{x}) = 0$

Kantorovich-type Theorem

Theorem

Let F is Newton differentiable on large enough set

$$\|F(x) - F(y) - \mathcal{H}F(x)(x - y)\| \leq \frac{1}{2}L\|x - y\|^\gamma$$

and define $B := \|\mathcal{H}F(x^0)^{-1}\|$ and $\eta := \|\mathcal{H}F(x^0)^{-1}F(x^0)\|$.

1. $\mathcal{H}F$ is Hölder continuous,

$$\|\mathcal{H}F(x) - \mathcal{H}F(y)\| \leq L\|x - y\|^{\gamma-1}$$

2.

$$\frac{LB}{2} \left(\frac{f(t)}{f'(t)} \right)^\gamma \leq f \left(t - \frac{f(t)}{f'(t)} \right)$$

$$f(0) = \eta, \quad f(t) > 0, \quad f(\bar{t}) = 0,$$

$$f'(t) < 0, \quad f'(t) \geq -(1 + Lt^{\gamma-1})^{-1}, \quad f''(t) > 0$$

then $\exists \bar{x}$ such that $F(\bar{x}) = 0$

Proof.

Step 1 Show that the Newton method applied to f converges superlinearly

Step 2 Define $\Sigma(t) = \{x \mid \|x - x^0\| < t, \|\nabla F(x^0)F(x)\| \leq f(t)\}$

Step 3 Show with recursion $x^k \in \Sigma(t^k)$

Step 4 Telescope and triangle to get

$$\|x^m - x^n\| \leq t^m - t^n$$

Step 5 Cauchy sequence



Further Ideas

Further Ideas

1. Stochastic Newton
2. Root multiplicity
3. Newton in metric spaces
4. Newton for multiobjective
5. Newton for constrained optimization

Why does this matter?

Let's say you have a complicated function

$$F(x) = \begin{cases} F_1(x) & \text{if } G(x) = 0 \\ F_2(x) & \text{else} \end{cases}$$

You use automated differentiation or numerical approximation to compute

$$\partial_A F(x) = \begin{cases} \nabla F_1(x) & \text{if } G_A(x) = 0 \\ \nabla F_2(x) & \text{else} \end{cases}$$

It is very likely that $\partial_A F$ is a (maybe weak) Newton differential

Keep doing what you are doing!