Lecture 11

C-regular Maps

Newton-Type Methods

October 8, 2008

Outline

- C-regular maps
- Newton Method
    - for Differentiable Maps
    - for Nondifferentiable Maps
        - □ Exact

Class of C-regular Functions

Def. (C-**regularity**)
Let $f : \Omega \to \mathbb{R}$ be continuous and locally Lipschitz on an open set Ω. We say that **f is C-regular** at $\bar{x} \in \Omega$ when f is directionally differentiable at $\bar{x}$ and

$$f^\circ(\bar{x}; d) = f'(\bar{x}; d) \qquad \text{for all } d \in \mathbb{R}^n$$

- The class of C-regular functions includes
  - Convex functions
  - Continuously differentiable functions
- This is a class of functions for which the generalized Jacobians are fruitful in algorithmic development

- A C-regular function at a point $\bar{x}$ admits accurate first order approximation in any given direction d, i.e.,

$$f(\bar{x}+d) = f(\bar{x}) + f'(\bar{x}; d) + o(\|d\|) \qquad \text{with} \quad \lim_{\|d\|\to 0} \frac{o(\|d\|)}{\|d\|} = 0$$

- This property is crucial for the family of Newton methods for solving equations $G(x) = 0$ with $x \in X$

- Subdifferential Operations
  - Most of the properties on composition, mean value and optimality conditions have "equivalent" expressions in terms of subdifferentials
  - Few are different (see Section 7.1 FP-II)

Generalized Jacobian for Mapping

Proposition 7.1.14.
Let $\Omega \subseteq \mathbb{R}^n$ be an open set, and let $\bar{x} \in \Omega$. Let $G : \Omega \to \mathbb{R}^m$ be a locally Lipschitz continuous map. Then, we have

$$\partial G(\bar{x}) \subseteq (\partial G_1(\bar{x}) \times \cdots \times \partial G_m(\bar{x}))^T$$

- If the nondifferentiabilities of components G_i are "unrelated" the inclusion is an equality

- Otherwise, the inclusion is usually "strict"

Examples: Let $(\bar{x}_1, \bar{x}_2) = 0$

- Consider the function

$$F(x_1, x_2) = \begin{bmatrix} |x_1| - x_2 \\ x_1 + |x_2| \end{bmatrix}$$

- Consider now the following function

$$F(x_1, x_2) = \begin{bmatrix} \sin x_1 - |x_2| \\ x_1^2 + |x_2| \end{bmatrix}$$

Directional Derivative and Generalized Jacobian

Proposition 7.1.17 FP-II.
Let $\Omega \subseteq \mathbb{R}^n$ be an open set, and let $\bar{x} \in \Omega$. Let $G : \Omega \to \mathbb{R}^m$ be a **locally Lipschitz continuous map on Ω and directionally differentiable at $\bar{x}$**. Then, for every vector $d \in \mathbb{R}^n$, there is a matrix $H \in \partial G(\bar{x})$ such that

$$G'(\bar{x}; d) \; = \; Hd$$

- Different directions d's have different H's

Proof: See the book.

Newton's Method for System of Equations

- We are interested in numerical methods for solving a system of equations
$$G(x) = 0, \qquad G : \mathbb{R}^n \to \mathbb{R}^m$$
- Algorithms of the form $x^{k+1} = x^k + d^k$ starting with some initial x^0
- When G is continuously differentiable, the classical Newton method is based on a natural (local) approximation of G: **linearization**
  - Given an iterate x^k, the map G is approximated at x^k by the following linear map
$$L(x; x^k) = G(x_k) + JG(x^k)(x - x^k)$$
  - We have $L(x; x^k) \approx G(x)$
  - System $G(x) = 0$ is "replaced" by the system $L(x; x^k) = 0$
  - The resulting solution is defining a new iterate x^{k+1},
$$x^{k+1} = x^k - JG(x^k)^{-1}G(x^k)$$

- We study methods motivated by Newton's method

- This method serves as a prototype for development of local, fast algorithms solving (differentiable) system of equations

- Our focus is on algorithms' extensions for potentially non-differentiable equations

- We study local convergence and convergence rate of such methods; local in the sense that initial guess x^0 is in a neighborhood of the solution x^*

- When studying the convergence properties, we use different concepts of asymptotic "rate"

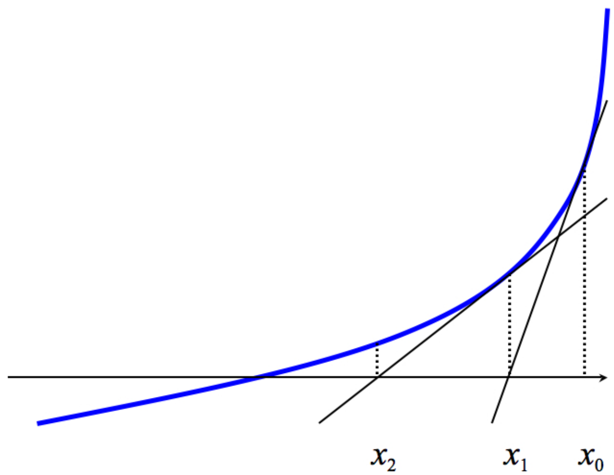
Convergence Rate Terminology

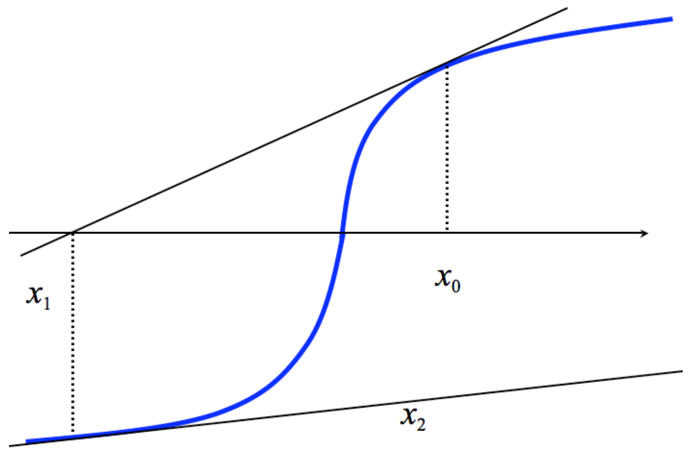
Definition 7.2.1 Let $\{x^k\} \subseteq \mathbb{R}^n$ be a sequence converging to some $x^* \in \mathbb{R}^n$. The convergence rate is said to be

- *Q-linear* if
$$\limsup_{k\to\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} < \infty$$
- *Q-superlinear* if
$$\limsup_{k\to\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = 0$$
- *Q-quadratic* if
$$\limsup_{k\to\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^2} < \infty$$
- *R-linear* if
$$\limsup_{k\to\infty} \left(\|x^{k+1} - x^*\|\right)^{1/k} < 1$$

Properties of Newton's Method

$$x^{k+1} = x^k - JG(x^k)^{-1}G(x^k)$$





- Fast local convergence, but globally the method can fail
  - When started far from a solution

- Numerical instabilities occur when $JG(x^*)$ is singular (or nearly singular)

- Two main properties that make the process work
  - A linear model $L(x; x^k)$ that provides good approximation of G near x^k, when x^k is near a solution
  - Solvability of linear equation $L(x; x^k) = 0$ when $JG(x^k)$ is invertible

- These two properties are "guaranteed" when
  - G is continuously differentiable and
  - $JG(x^*)$ is invertible (nonsingular)

- When G is nondifferentiable, the requirements on the local linear model are not that visible or simple

Newton Iteration

- Newton iteration can be equivalently written as

$$x^{k+1} = x^k + d^k$$

with d^k being a solution to the following linear system

$$G(x^k) + JG(x^k)d = 0$$

- The term $JG(x^k)d$ can be viewed as approximation of

$$G(x^k + d) - G(x^k)$$

- The quality of approximation is good when d is sufficiently "small"
- When G is nondifferentiable the Jacobian need not exist
- When "generalized Jacobian" exists, not all the matrices in the generalized Jacobian set provide good local approximations

Nondifferentiable G

When G is nondifferentiable, the d is selected as a solution to

$$G(x^k) + A(x^k, d) = 0$$

where $A(x^k, d)$ is an approximation of $G(x^k + d) - G(x^k)$ around $d = 0$

- There could be more than such approximations
- Let $\mathcal{A}(x)$ denote a family of such approximations

For example, we will often consider an approximation of the form

$$G(x^k) + Hd, \qquad H \in \partial G(x)$$

In this case, $\mathcal{A}(x) = \partial G(x)$

- Introducing the family of approximations allows us to consider methods using more than one local approximation model
- Not all approximations result a convergent algorithm, so we specify some essential features that are needed for convergence

Newton Approximation Scheme

Definition Let $\Omega \subset \mathbb{R}^n$ be an open set and $\bar{x} \in \Omega$. Let $G : \Omega \to \mathbb{R}^m$ be locally Lipschitz. We say that G has a **Newton approximation at $\bar{x}$**, if there exists a neighborhood $U \subset \Omega$ containing $\bar{x}$ and a family $\mathcal{A}(x)$ of approximations for $x \in U$ such that

(a) $A(x, 0) = 0$ for every $A(x, \cdot) \in \mathcal{A}(x)$ and all $x \in U$

(b) For some function $\Delta : \mathbb{R}_{++} \to \mathbb{R}_{+}$ with $\lim_{t \downarrow 0} \Delta(t) = 0$, we have

$$\frac{\|G(x) + A(x, \bar{x} - x) - G(\bar{x})\|}{\|x - \bar{x}\|} \le \Delta(\|x - \bar{x}\|)$$

for all $A(x, \cdot) \in \mathcal{A}(x)$ and all $x \in U$ with $x \neq \bar{x}$

We refer to $\mathcal{A}$ as a **Newton approximation scheme** for G at $\bar{x}$.

- Condition (b) plays the same role as first-order Taylor expansion at $\bar{x}$
  - When G is differentiable $\mathcal{A}(x) = \{-JG(\bar{x})(x - \bar{x})\}$
  - $\mathcal{A}$ is a Newton approximation when G is continuously differentiable
  - For nondifferentiable mapping G, it ensures that the first-order Taylor expansion at $\bar{x}$ is "accurate"
  - It is equivalent to requirement

$$\lim_{\substack{x \to \bar{x} \\ A(x,\cdot) \in \mathcal{A}(x)}} \frac{\|G(x) + A(x, \bar{x} - x) - G(\bar{x})\|}{\|x - \bar{x}\|} = 0$$

- Existence of Newton approximation scheme is crucial for **convergence**

- A stronger property is needed for "fast" convergence

Strong Approximation

Definition We say that G has a **strong Newton approximation at $\bar{x}$**, if G has a Newton approximation $\mathcal{A}$ at $\bar{x}$ such that for some scalar L,

$$\frac{\|G(x) + A(x, \bar{x} - x) - G(\bar{x})\|}{\|x - \bar{x}\|^2} \le L$$

for all $A(x, \cdot) \in \mathcal{A}(x)$ and all $x \in U$ with $x \neq \bar{x}$.
We refer to $\mathcal{A}$ as a **strong Newton approximation scheme** for G at $\bar{x}$.

- For a differentiable G, having a strong approximation scheme is "equivalent" to having Lipschitz Jacobians
- For a nondifferentiable G, a strong approximation scheme plays the role of Lipschitz Jacobians "Lipschitz Jacobians". The condition is equivalent to

$$\limsup_{\substack{x \to \bar{x} \\ A(x,\cdot) \in \mathcal{A}(x)}} \frac{\|G(x) + A(x, \bar{x} - x) - G(\bar{x})\|}{\|x - \bar{x}\|^2} < \infty$$

Homeomorphism

- Let $X, Y \subseteq \mathbb{R}^n$. A map $\Phi : X \to Y$ is a **homeomorphism** if Φ is continuous and bijective, and its inverse $\Phi^{-1} : X \to Y$ is also continuous

- A map Φ is **Lipschitz homeomorphism on** X if there exist open sets $V_X \subset X$ and $V_Y \subset Y$ such that the restricted map $\Phi : V_X \to V_Y$ and its inverse $\Phi^{-1} : V_Y \to V_X$ are both Lipschitz continuous

Homeomorphism

- Let $\mathcal{F}$ be a family of homeomorphisms on X. We say that $\mathcal{F}$ is a family of **uniformly Lipschitz homeomorphisms** when there exist scalars $\epsilon > 0$, L and L' such that for every $\Phi \in \mathcal{F}$, we can find open sets $V_X \subset X$ and $V_Y \subset Y$ satisfying the following conditions
  - The open ball $B(0, \epsilon)$ is contained in both V_X and V_Y,

$$B(0, \epsilon) \subseteq X, \qquad B(0, \epsilon) \subseteq Y.$$

  - The restricted map $\Phi : V_X \to V_Y$ is Lipschitz on V_X with constant L
  - Its restricted inverse $\Phi^{-1} : V_Y \to V_X$ is Lipschitz on V_Y with constant L'

Nonsingular Approximation

This is pertinent to the case when $m = n$.

Let $\Omega \subset \mathbb{R}^n$ be an open set and $\bar{x} \in \Omega$. Let $G : \Omega \to \mathbb{R}^n$ be locally Lipschitz.

Definition We say that Newton **approximation scheme $\mathcal{A}$ is nonsingular** when

- $\mathcal{A}$ is Newton approximation and
- $\mathcal{A}$ is a family of uniformly Lipschitz homeomorphisms on U

The latter means:

- There exist $\epsilon_{\mathcal{A}}$, $L_{\mathcal{A}}$, and $L'_{\mathcal{A}}$ such that for any $x \in U$ and any $A(x, \cdot) \in \mathcal{A}(x)$,
  - The map $A(x, \cdot)$ is Lipschitz with $L_{\mathcal{A}}$ and its inverse is Lipschitz with $L'_{\mathcal{A}}$ on some open sets, both containing $B(0, \epsilon_{\mathcal{A}})$

The role of nonsingular approximations is the same as the role of nonsingular Jacobians $JG(x)$ in a neighborhood of x^*.

Newton's Method

Step 0 Select initial vector x^0 and $\epsilon > 0$. Set $k = 0$.

Step 1 If $G(x^k) = 0$, then stop.

Step 2 Select an approximation $A(x^k, \cdot)$ in $\mathcal{A}(x^k)$ and **find a vector** $d^k \in B(0, \epsilon)$ such that

$$G(x^k) + A(x^k, d^k) = 0 \qquad \textbf{Newton equation}$$

Step 3 Set $x^{k+1} = x^k + d^k$ and $k := k + 1$, and go to Step 2.

- The stopping criteria in practice is:

$$\|G(x^k)\| \textit{ not exceeding a desired tolerance level}$$

- We use a perfect tolerance to study the asymptotic convergence properties, namely, rate of convergence

- The only difference from the classical method is the computation of d^k
  - In the classical method, G is assumed to be continuously differentiable and therefore, Newton equation has one and only one solution
  - The direction is not restricted to belong to a small ball

- Due to nondifferentiability of G, here,
  - Multiple directions may exist solving the equation
  - Only the directions with small length are safe

- The choice of ϵ is crucial for convergence
  - A choice of "good ϵ" is related to the neighborhood of x^* where the approximation family $\mathcal{A}$ for G at x^* is "good enough"

Convergence Result

Theorem 7.2.5.
Let $\Omega \subseteq \mathbb{R}^n$ be an open set containing x^*. Let $G : \Omega \to \mathbb{R}^n$ be locally Lipschitz on a neighborhood U of x^* satisfying $G(x^*) = 0$. Assume that G has a **nonsingular Newton approximation $\mathcal{A}$** at x^*. Then, we have

- For every ϵ with $0 < \epsilon \le \epsilon_{\mathcal{A}}$, there exists $\delta > 0$ such that
  - When $x^0 \in B(0, \delta)$, the method generates a sequence $\{x^k\}$ converging **Q-superlinearly to x^***
  - If the **Newton approximation $\mathcal{A}$ is strong**, then the sequence $\{x^k\}$ converges **Q-quadratically**

Proof

Let ϵ with $0 < \epsilon \leq \epsilon_{\mathcal{A}}$ be arbitrary but fixed.
Since $\mathcal{A}$ is nonsingular Newton approximation, we can choose δ small enough so that

1. $A(x, \cdot)$ are uniformly Lipschitz homeomorphisms for all $x \in B(x^*, \delta)$.

2. $A(x, \cdot)$ are homeomorphisms on some open sets containing $B(0, \epsilon)$

- Solution of Newton equation exists:
  For such a δ and any $x \in B(x^*, \delta)$, we have

$$\|G(x)\| = \|G(x) - G(x^*)\| \leq L_{\mathcal{A}}\|x - x^*\| \leq \epsilon.$$

  Thus, if $x_k \in B(x^*, \delta)$, then $A(x^k, d) = -G(x^k)$ has a unique solution, i.e.,

$$d^k = A^{-1}(x^k, -G(x^k)).$$

- When $x^0 \in B(0, \delta)$, the Newton method generates the sequence $\{x^k\}$ that stays in $B(0, \delta)$.

  - Suppose that $x^k \in B(x^*, \delta)$. Consider $x^{k+1} - x^*$. We have

$$
\begin{aligned}
\|x^{k+1} - x^*\| &= \|x^k + d^k - x^*\| \\
&= \|A^{-1}(x^k, -G(x^k)) - (x^* - x^k)\| \\
&= \|A^{-1}(x^k, -G(x^k)) - A^{-1}(x^k, A(x^k, x^* - x^k))\| \\
&\le L'_{\mathcal{A}} \, \|G(x^k) + A(x^k, x^* - x^k)\| \qquad (1)
\end{aligned}
$$

- Since $\mathcal{A}$ is an N-approximation of G at x^*, there is a neighborhood U containing x^* such that for all $x \in U$ with $x \neq x^*$ and any $A(x, \cdot) \in \mathcal{A}(x)$,

$$\|G(x) + A(x, x^* - x) - G(x^*)\| \leq \Delta(\|x - x^*\|)\ \|x - x^*\| \quad (2)$$

- Choose δ small so that $B(x^*, \delta) \subseteq U$ and

$$\Delta(t) \leq \frac{1}{2L'_{\mathcal{A}}} \qquad \text{whenever } t \in (0, \delta)$$

Then from Eq. (2) we see that for all $x \in B(x^*, \delta)$ with $x \neq x^*$ and $A(x, \cdot) \in \mathcal{A}(x)$,

$$\|G(x) + A(x, x^* - x) - G(x^*)\| \leq \frac{1}{2L'_{\mathcal{A}}}\ \|x - x^*\| \qquad (3)$$

- By assuming that x^0 and x^k are in $B(x^*, \delta)$, from Eqs. (1)–(3) we have

$$\|x^{k+1} - x^*\| \leq \frac{1}{2}\|x^k - x^*\|$$

  Thus, when $x^0 \in B(x^*, \delta)$, the whole sequence is in $B(x^*, \delta)$

- Hence Eq. (2) holds with $x = x^k$ for any k

- Furthermore, from Eqs. (1)–(2) we obtain

$$\|x^{k+1} - x^*\| \leq L'_{\mathcal{A}} \Delta(\|x^k - x^*\|)\ \|x^k - x^*\|$$

  implying Q-superlinear convergence

- When $\mathcal{A}$ is strong N-approximation, then it can be seen that

$$\|x^{k+1} - x^*\| \leq L'_{\mathcal{A}}\ L_{\mathcal{A}}\ \|x^k - x^*\|^2$$

  implying Q-quadratic convergence

Inexact Newton's Method

Step 0 Select vector x^0, $\epsilon > 0$, and a sequence $\{\eta^k\}$ of nonnegative scalars. Set $k = 0$.

Step 1 If $G(x^k) = 0$, then stop.

Step 2 Select an approximation $A(x^k, \cdot)$ in $\mathcal{A}(x^k)$ and **find a vector** $d^k \in B(0, \epsilon)$ such that

$$G(x^k) + A(x^k, d^k) = r^k \qquad \textbf{inexact Newton equation}$$

where r^k is such that

$$\|r^k\| \leq \eta^k \, \|G(x^k)\|$$

Step 3 Set $x^{k+1} = x^k + d^k$ and $k := k + 1$, and go to Step 2.

- The inexactness of computing d^k is proportional to the residual $\|G(x^k)\|$

- Such a direction exists when G has a nonsingular Newton approximation at x^* (see Lemma 7.2.7)

Inexact Method Convergence

Theorem 7.2.5. Let Ω be an open set containing x^*. Let $G : \Omega \to \mathbb{R}^n$ be locally Lipschitz on a neighborhood U of x^* satisfying $G(x^*) = 0$. Assume that G has a **nonsingular Newton approximation $\mathcal{A}$** at x^*. Then, there exists $\bar{\eta}$ such that when $\eta^k \le \bar{\eta}$ for all k, we have

- For every ϵ with $0 < \epsilon \le \epsilon_{\mathcal{A}}$, there exists $\delta > 0$ such that
  - When $x^0 \in B(0, \delta)$, the method generates a sequence $\{x^k\}$ converging **Q-linearly to x^***
  - If $\eta^k \to 0$, then the sequence $\{x^k\}$ converges **Q-superlinearly**
  - If the **Newton approximation $\mathcal{A}$ is strong** and

$$\eta^k \le \tilde{\eta}\, \|G(x^k)\| \qquad \text{for some } \tilde{\eta} \text{ and all } k$$

then the sequence $\{x^k\}$ converges **Q-quadratically**